Thèse délivrée par
L'Université de Lille

THÈSE

Pour obtenir le grade de Docteur en Sciences de Gestion
Présentée et soutenue publiquement par

PHOOL KUMAR
Le: 19 / 12 / 2025

Étude explicative des facteurs psychologiques à l'origine du comportement de navigation des internautes français en matière de recherche d'informations sur la santé sur les SERP dans le cadre de G-AI

En bref, HISB français sur SERP

JURY

Directeur de thèse:	**Monsieur Dominique Crié -** Professeur, IAE - Université de Lille
Co-directrice de thèse:	**Madame Annabel Martin-Salerno -** Professeure, IAE - Université de Lille
Président du jury:	**Monsieur Christophe Benavent -** Professeur, Université Paris Dauphine
Rapporteurs:	**Monsieur Kristof Coussement** - Professeur, IÉSEG School of Management, Lille **Madame Nadia Steils** - Professeure associée, HEC, Liège
Suffragant:	**Monsieur Mihai Calciu** - Maitre de Conférences-HDR Emérite, IAE, Université de Lille

Thesis awarded by
The University of Lille

THESIS

To obtain the degree of Doctor of Management Sciences
Presented and publicly defended by

PHOOL KUMAR
Date: 19 / 12 / 2025

Explanatory study of the psychological factors behind French Web Searcher's Health Information Search Navigation Behaviour on SERP in purview of G-AI

In short **French HISB on SERP**

JURY

Thesis Director:	**M. Dominique Crié -** Professor, IAE - Université de Lille
Thesis Codirector:	**Mme. Annabel Martin-Salerno -** Professor, IAE - Université de Lille
Jury President:	**M. Christophe Benavent -** Professor, Université Paris Dauphine
Rapporteurs:	**M. Kristof Coussement** - Professor, IÉSEG School of Management, Lille **Mme. Nadia Steils** - Associate Professor, HEC, Liège
Suffragant:	**M. Mihai Calciu** - Maitre de Conférences-HDR Emérite, IAE, Université de Lille

Thesis Layout, Text Formatting and Navigation Guidelines

Thesis Layout:

1. **Thesis Index:** Units - 3, Chapters - 10, Pillar Topics - 48, Topic Cluster - 55; Index Table at page 587 to 590.
2. **Graphics-Visuals** (Charts, Diagrams, Tables, Infographics, Screenshots)**:**
   a. **In-Thesis-built-Originals:** Total - 105; Complete list at page 591 to 594.
   b. **Borrowed/Sourced:** Total - 90 in Appendix-9 at page 464 to 503.
3. **Citations, References and/or Sources:** Bibliography - 1123 (from page 504 to 586), Weblink Footnotes - 274, Abbreviations - 199, Appendices - 9 (from page 412 to 503),

Text Formatting Scheme:

1. **Unit Title:** Bold, Maroon Colour, 28px, Times New Roman
2. **Chapter Title:** Bold, Sky-blue Colour, 16px, Times New Roman
3. **Important Topic Title:** Bold, Sky-blue Colour, 12px, Times New Roman
4. **General Topic Title:** Bold, Black Colour, 12px, Times New Roman
5. **General Manuscript Text:** Normal, Black, 12px, Times New Roman, 1.5 Interline Gap
6. **Graphics-Visuals Naming Convention:** Normal, Maroon Colour, 12px, Times New Roman for serial number, rest of text in Normal, Black Colour, 12px, Times New Roman
7. **Border around all types of Visuals used:** It should help our readers to differentiate between the start and end of a Visual *(Table, Diagram, Screenshot, Chart or Infographic)* area from the general text areas.
8. **Reference Citation pattern in Text:** Author name (italic/normal) and Publication year in [….] or (….), Normal, Maroon Colour, 12px, Times New Roman
9. **Bibliography Format:** References/Citations/Sources are written in one of the following formats wherever applicable:
   a. **Article/Paper/Book/Blog of One Author:** *Author's Surname or Full Name* [Year of publication] - Author; Year; Article/Paper/Book/Blog Title; Chapter & Page No.; Publishing details if any; Source URL. *Where there was conflict between two references due to same surname, name is also used.*
   b. **Article/Paper/Book/Blog of Two Authors:** *Surname of both authors separated by "&"* [Year of publication] - Authors; Year; Article/Paper/Book/Blog Title; Chapter & Page No.; Publishing details if any; Source URL

c. **Article/Paper/Book/Blog of more than two authors:** *One Author Surname et al.* [Year of publication] - Authors; Year; Article/Paper/Book/Blog Title; Chapter & Page No.; Publishing details if any; Source URL. Main author surname with priority, if recognizable in list.

d. **Blog Source without author name:** *Blog Name 1, 2, 3…..* [Year of publication] - Blog Name; Year; Blog Title; Source URL

e. Where finding or differentiating between name and surname was difficult, one or both are written; applicable to b, c and d points above.

Navigating across citations and abbreviations:

1. **Use of "here onward" in abbreviations:** It is used with each such term appearing first time in order of reading this thesis from top to bottom. For example, where 'Hypertext Transfer Protocol Secure' appeared first time, just next to it, its abbreviation is made in this format '*(HTTPs **here onward**)'; collectively both written as* Hypertext Transfer Protocol Secure *(HTTPs here onward)*. Accumulated list of Abbreviations is given in Appendix-7.
2. **Navigating to a Citation:** There are four types of citations used in this thesis, Visuals, Bibliography, Appendix and Chapters or Section inter or within a Chapter; all have one unique colour like this example "*SAS-7".* To search about their details, select that citation and press 'Control F' and click enter to move through all citations including its reference in bibliography.
3. **Footnotes at the end of the page for citations of website or blog articles:** Wherever such references are used, for those alongside adding their publication details in the bibliography list, their direct weblinks are inserted at the bottom of the same page by linking the corresponding citation with it. To open these Footnotes in PDF version, use "Ctrl + click", it shall open you the link in new tab. Be careful, due to different PC environments of treating the weblinks, sometimes some link may not open by simple click, in that case, copy and paste the link in Web Browser or Google Search manually.
4. **Bookmarks Navigation** (Typical Blue Link within thesis text)**:** Seeing large number of Graphics & Visuals; these are divided into two categories. First category is of our **own created original** Graphics & Visuals and Second category is of **borrowed / sources / referenced** Graphics and Visuals. And to maintain the scientific nature of our thesis, the second category *(except some contextually highly important)* are listed under Appendix-9, and a Bookmark Link for each is created to navigate between the location of their mentioning in this thesis and Appendix-9. So, you can go to each of them by either simply clicking or using 'Ctrl + Click' *(depending on your PC environment).*

Acknowledgement

This thesis is the culmination of a journey enriched by guidance, support, and inspiration from many remarkable individuals and institutions. With deep humility and profound gratitude, I take this opportunity to acknowledge those who have contributed to this academic endeavour.

First and foremost, I extend my sincere thanks and heartfelt appreciation to my research guide, **Pr. Dominique Crie,** whose insightful mentorship, unwavering patience, and intellectual rigor shaped the foundation of this work. His critical feedback, academic depth, and encouragement pushed me beyond limits and helped in developing clarity, structure, and purpose in this research. Doing teaching and researching under his directorship of our department of Marketing and Sales; and being a consistent part of **JIMS** Scientific Conference Series Organizing Committee have instilled me with the sophisticated scientific aura, decorum and discipline, I owe all this to him.

I am equally grateful to my research co-guide, **Pr. Annabel-Martin Salerno**, whose complementary guidance, technical wisdom, and steadfast support added valuable dimensions to this thesis. Her thoughtful inputs and dedication to scholarly excellence have been a constant source of motivation for me. Monthly **MTD** research conference series in our **LUMEN Lab** run under her responsibility and being co-director of our Marketing and Sales department, I owe many learnings to her, especially discipline and cooperation within a research lab.

I express my deep gratitude to rapporteurs **Pr. Kristof Coussement** and **Pr. Nadia Steils** and the suffragants **Pr. Mihai Calciu** and **Pr. Christophe Benavent** of our thesis jury for the rigorous, kind and informed attention they paid to our work. Their relevant analyses, constructive comments and encouragement have greatly contributed to enriching the thinking and refining the scientific scope of this research. Their commitment to the evaluation of this thesis demonstrates high academic standards and a strong sense of intellectual sharing, which I acknowledge with respect and gratitude.

A sincere thanks to the doctoral committee **(CSI - Pr. Mihai Calciu, Pr. Christophe Benavent, Pr. Guignard Thomas),** whose annual evaluations, thoughtful feedback, and encouragement sustained the progress of this research. Their commitment to academic rigor and research integrity has been instrumental in shaping the trajectory of our research.

I owe special thanks to my research laboratory, **LUMEN,** its director **Pr. Benoît DEMIL** and my doctoral school **Ecole Doctorale (SJPG)**, for providing a stimulating research environment, access to critical resources, scientific trainings and an inclusive intellectual community; here I would like to extend my gratitude toward **Dr. Sophie Ranchy** and **Mrs. Neve Casene** from this fraternity for organizing all end to end administrative things. With deep love inside, I express my deepest gratitude to **IAE - University School of Management and University of Lille**, my academic home, whose vision of interdisciplinary research made this work realised. Here a special thanks to **Mrs. Sylvie VAN CAENEGHEM** and her administrative team; and **Mr. Bruno Dejaeger** and **Mr. Gregory Vermeersch** for facilitating a pain-free IT infrastructure.

I would also like to warmly thank **my colleagues and peers**, whose collaboration, informal consultations, and shared intellectualism helped refine many ideas presented in this thesis. Here to mention **Pr. Hamdi Dkhil'**s honest analysis and readiness to help made the research process both productive and enjoyable. I am especially indebted to the **video test participants**, whose willingness to contribute their time and experiences provided the invaluable data on which this study is originally based; for this a heartfelt gratitude to responsible of different marketing departments **Pr. Annabel-Martin Salerno, Pr. Thomas Guignard, Pr. Laurent Carpentier, Pr. Marie Beck, Pr. Aurely Lao and Pr. Christell Quero** to motivate participants to take part in research test of this thesis. Without their engagement, this research would not have been this much enthusiastic.

On a deeply personal note, I offer my countless gratitude to **my parents, family and close friends**, whose love, sacrifices, and belief in my journey appeared as my greatest strength. **My cultural roots and values** of both motherland **India** and working-land **France** have constantly guided my decisions, and I dedicate this work to the silent but proactive teachings of these two cultures shaping my worldview.

With reverence, I bow in gratitude to almighty, my **ancestors, culture-kinship**, and the living **wisdom of nature**, whose energies have silently nurtured and grounded me throughout this path.

And finally, to each soul, seen and unseen, who contributed to this journey - I am forever indebted.

Phool Kumar
19 / 12 / 2025

Étude explicative des facteurs psychologiques à l'origine du comportement de navigation des internautes français en matière de recherche d'informations sur la santé sur les SERP dans le cadre de G-AI

Résumé: Cette thèse présente une étude explicative et exploratoire des facteurs psychologiques, cognitifs, de compréhension, de pertinence, d'intention, de personnalisation et émotionnels qui sous-tendent le comportement de recherche d'informations de santé des utilisateurs francophones (HISB) sur les pages de résultats des moteurs de recherche (SERP). L'accent a été mis sur le développement d'un profilage psycho-cognito de nos participants au test, puis sur la classification des intentions de recherche en deux catégories. La recherche adopte une approche mixte, intégrant des tests de suivi du comportement de recherche des utilisateurs par vidéographie et une analyse de la perception basée sur les commentaires. Au total, 160 participants francophones ont créé 2 884 requêtes de recherche uniques liées à la santé, selon 4 scénarios de test avec une moyenne de 4 requêtes par scénario, ce qui a permis de générer 627 enregistrements vidéo valides. Ces enregistrements vidéo ont été transcrits et divisés en 9 330 échantillons de texte valides et logiques, soit l'équivalent de 815 430 jetons de données humaines réelles, ainsi que la création d'un autre ensemble de commentaires d'utilisateurs YouTube. En parallèle, 165 872 échantillons de texte simulés par l'IA (17 773 761 jetons) ont été générés à l'aide d'API (Application Programming Interface) de grands modèles de langage (LLM) ; en utilisant des invites textuelles basées sur des règles sous des capacités et des caractéristiques d'utilisation contrôlées. Les données humaines réelles et les données générées par l'IA ont été analysées à l'aide d'un modèle d'apprentissage automatique supervisé basé sur un réseau neuronal récurrent (RNN) personnalisé et de la méthode BiGRU (Bidirectional Gated Recurrent Unit) développée en Python. Cette étude assimilée du traitement cognitif, de l'engagement émotionnel et de la compréhension sémantique offre également des perspectives cruciales sur les expériences de recherche futuristes médiées par l'IA et la prise de décision marketing dans le contexte des soins de santé français. Cette recherche contribue aux domaines croisés du marketing numérique, de l'informatique de santé grand public et de l'interaction homme-machine en proposant un cadre comportemental raffiné pour la recherche influencée par l'IA. Elle fait également progresser le discours méthodologique sur la manière de comparer systématiquement la cognition humaine à l'intelligence générative dans ce contexte.

Keywords: *Comportement de recherche d'informations sur la santé en français, IA générative dans la recherche d'informations sur la santé sur les moteurs de recherche, classification des intentions de recherche, psychologie du marketing de recherche sur le Web, modélisation RNN-BiGRU, analyse prédictive par apprentissage automatique supervisé, analyse de similarité sémantique, IA explicable dans le marketing de la santé en français.*

Unité de recherche: Lumen Lab. (Axe MTD), IAE Lille, 104 avenue du Peuple Belge - 59043 LILLE
Ecole doctorale: Ecole doctorale des sciences juridiques, politiques et de gestion, n° 74, 1 place Déliot, 59000 Lille, edsjpg@univ-lille.fr, http://edsjpg.univ-lille.fr; Université de Lille, 42 rue Paul Duez, 59000 Lille

Explanatory study of the psychological factors behind French Web Searcher's Health Information Search Navigation Behaviour on SERP in purview of G-AI

Abstract: This thesis presents an **explanatory and exploratory investigation** into the psychological, cognitive, understanding, relevance, intent, personalization and emotional factors underlying French-speaking User's Health Information Search Behaviour (HISB) on Search Engine Result Pages (SERPs) with prime focus on developing Psycho-Cognito Profiling of our test-participants and then doing dual-category Search Intents Classification. The research adopts a mixed-methods design, incorporating testing of user's search behavioural tracking through videography, and commentary-based perception analysis. A total of 160 French speaking participants created 2,884 unique health-related search prompts; under 4-test scenario with average 4 prompts per scenario; using which generating 627 valid video recordings were created. These video recordings were transcribed and divided into 9330 valid and logical text samples, equivalent of 815,430 tokens of real-human data, along with creating another set of YouTube user comments. In parallel, 165, 872 AI-simulated text samples *(17, 773, 761 tokens)* were generated using APIs *(Application Programming Interface)* of Large Language Models (LLMs); using rule-based **text-prompts** under controlled capacities and usage features. Both **real-human data** and **AI-generated-data** were analysed using a custom-built **Recurrent Neural Network (RNN)** based Supervised Machine Learning Model and **BiGRU** *(Bidirectional Gated Recurrent Unit)* method developed in Python to present an assimilated study of cognitive processing, emotional engagement, and semantic understanding alongside offering critical insights into futuristic AI-mediated search experiences and marketing decision making in French healthcare contexts. This research contributes to the intersecting fields of digital marketing, consumer health informatics, and human-computer interaction by proposing a refined behavioural framework for AI-influenced search. It also advances methodological discourse on how to systematically compare human cognition with generative intelligence in the context.

Keywords: *French Health Information Search Behaviour, Generative AI in Health Information Search on Search Engines, Search Intent Classification, Web Search Marketing Psychology, RNN-BiGRU Modelling, Supervised Machine Learning Predictive Analytics, Semantic Similarity Analysis, Explainable AI in French Health Marketing.*

Research unit: Lumen Lab. (Axe MTD), IAE Lille, 104 avenue du Peuple Belge - 59043 LILLE
Doctoral school: Ecole doctorale des sciences juridiques, politiques et de gestion, n° 74, 1 place Déliot, 59000 Lille, edsjpg@univ-lille.fr, http://edsjpg.univ-lille.fr; Université de Lille, 42 rue Paul Duez, 59000 Lille

Introduction

Inspired by the Digital Marketing Psychology, SERP Usage importance in human life globally and G-AI potential, this thesis examines **four interdependent dynamics** that collectively shape the Digital Health Search (DHS) experience, as below:

1. **Universal *(Psychological)* Dynamic:** Search User' Psycho-Cognito Navigation Behaviour on SERPs
2. **Sectorial *(Business)* Dynamic:** Online Health Information Search as a Digital Marketing priority
3. **Market *(Audience)* Dynamic:** French Speaking HIS Audience
4. **State-of-Art *(Technological)* Dynamic:** G-AI transforming Online Search functionality.

The aim is to decode the interplay among these dynamics, probing their individual significance and collective impact both. Search Engine Results Page *(SERP here onward)* is an information providing service to mankind in their day-to-day *(in-fact moment-by-moment)* life which intends to meet five axes of human information needs primarily as **Relevance, Quality, Authenticity, Freshness and Scope**. [1]*Digitalmainstreet* [2019] in reference to Google says that consumers use search to optimize their lives. It lets them feel more confident and less anxious. Whether it is about eating, travelling, shopping etc. everyone feels excited on seeing a good restaurant, jolly tourist place or better product reviews before shopping. This is how SERPs help lives to go confident of their choice thus the additional excitement of creating the best experience. The exponential growth of SERP acceptability in public *{even after entry of Generative Artificial Intelligence (G-AI here onward)}* established that SERP is here to serve the human-life until digital products and services remain the societal needs. Be it **professional search or social search**, SERP proves itself as a digital showcase of both needs by presenting searched information like a show cabinet *(flex, banner, billboard, kiosk etc.)* in any shopping mall. In a way, we call SERPs the most-popular **'INFOMALL'** around the globe. And the owners of these **SERP showcases** on **General Web Search** front are Google, Bing, Yahoo, Qwant, Baidu etc. while on **Social and/or Professional Search** front are Facebook, LinkedIn, Twitter, Instagram, WhatsApp etc. Daily Search Users and Content Providers *(Websites, Blogs, SERP Feature Blocks and Snippets filled with information, news and ads of different Companies and Individuals)* are 'interest-base end user stakeholders' of Search Engines. Now a days the new entrant in new format of content presentation are the **Large Language Model** *(LLM here onward)* **Platforms** such as Chat GPT, Claude, Gemini, Copilot, Llama, Grok, DeepSeek, Mistral etc. The entry of LLMs is so influential that a twin term of Search Engine Optimization *(SEO here onward)* is coined as Generative Engine Optimization *(GEO here onward - twin because race is almost same for both approaches*

[1] https://digitalmainstreet.ca/no-regrets-the-3-things-driving-the-research-obsessed-consumer/

and that is getting ranked or mentioned in Search or Chat Outputs). While before LLMs Content Providers' aim used to be ranking their content on SERPs now LLMs entry has added another ranking-race; but Content Optimization Strategies and Norms remain almost identical in both LLM Chats and SERPs. All of them strive to deliver the best of customer satisfaction in their delivery of searched information based on rigorous research and studies of management philosophies, technical innovations, and user psychologies around the demanded search query non-stop through their iterative and interactive UI and UX platforms. While prime focus of this thesis is SERP, still it shall be touching SEO and GEO contextually. And as an outlier, the prime focus of this thesis is one of the three stakeholders of SERP, i.e. the **Search User** as shown in Table-1 below. However, other two stakeholders shall also be talked contextually, the other two being Content Provider and Search Engine; all three shown in Table-1 below:

Mutual give and take of three stakeholders of Web 3.0 + Web 4.0 SERP		
Stakeholder	**Give**	**Take**
Search User	Search Feedback	Information Needs
Content Provider	Content, Information	Revenue, Recognition
Search Engine	SERP, Information	Search Feedback, Revenue

Table-1: Three Stakeholders of SERP

A distinct philosophical spectrum from above outlier is drawn in Table-2 below. This thesis revolves around the study of "**SERP User Psychology**" influenced by inseparable and associated impact of "**G-AI Technical Innovations**" and "**Management Philosophies**". As said in abstract the domain specificity **French Health Market** is at core of this research, which gives us another important KEYWORD as 'Healthcare Information Search/Seeking' (***HIS*** *here onward*).

SERP Generation Catalysts		
Technical Innovations	User Psychology	Management Philosophies
Data Sciences	Behavioural Traits	E-Marketing Purposes
• SERP Algorithms • Model Distillation • RNN BiGRU Supervised Machine Learning (ML)	• Psycho-Cognito Profiling of SERP Navigators • Dual Category Search Intents Classification through Linguistics	• SERP in Healthcare Managerial Implication in Marketing and Community Management • Search Engine Marketing (SEM)

Modelling in Python • G-AI LLMs • Neural Networks (NNs) • Deep Learning (DL) • Natural Language Processing (NLP)	• Understanding through Art of Phrasing • Cognition through Topicality • Persistency through Search Skills • Personalization through Algorithms • Intuition through Semantics	• Social Media Marketing (SMM) • Social Media Optimization (SMO) • SEO and GEO • Web Content Annotation • Web Content Creation • Search Prompt Engineering

Table-2: SERP Generation Catalysts

[2]*Kai Lemmetty* [2022] finds that "the internet search user behaviour in the French market is characterized by a significant and growing number of **online shoppers**. As of 2022, 41.6 million French people, accounting for 81.4% of internet users, made online purchases, with a strong preference for **Amazon** as their primary shopping destination. Here to further note that **health and beauty products online purchasing stands third with 38.8%** of French web users. 18% of French users discover products through advertising on **Google**. **28% searches first** happen through **Search Engines** in France; ecommerce platforms like Amazon stands first with 47% in this regard. Searching products through social networks stands top among 18-25 age group with 22% of total users. The e-commerce market in France is expected to continue growing, with the number of users projected to rise to 51.5 million and user penetration to increase to 78% by 2025". [3]*Manon Fischer* [2023] states that "French consumers are increasingly using mobile devices for online shopping, with one third of the population making at least one purchase per month on mobile apps. They tend to search for the best bargain but always expect top quality in the products they buy. 52% French consumers are worried about data security when buying on social media and 39% fear scams. Because of these fears, French consumers undergo a lot of research before buying a new product online. In fact, **80% consumers research** before buying and this habit tends them to search more". *Jolivet & Turon* [2019] identifies substantial price dispersion and varying consumer preferences, with some users choosing non-cheapest options. *Muller & Bolger* [1985] highlights differences in search patterns between French and English Canadians, with the former considering fewer alternatives and **spending less time** on the search process. *Bressolles & Durrieu* [2010] categorizes online **wine buyers** into three groups, the 'disappointed', the 'reassurance seeker' and the 'opportunist' based on electronic service quality dimensions. In particular context of Healthcare **55% of French people** have already used a search engine to find out about their symptoms, 34% prefer general information sites *(Doctissimo, PasseportSanté.net, etc.),* 17% institutional sites *(ameli.fr, HAS),* and 11% artificial intelligence *(ChatGPT,*

[2] https://www.gaasly.com/blog/online-shopping-trends-in-france
[3] https://www.hive.app/blog/consumer-habits-france-expansion

Claude, Gemini - here 27% of 18-24-year-old), [4]*Bayle-Iniguez* [2025]; *this study is viable enough about why SERP still matters.* Young adults aged 15-30 years use the internet as a resource for **health information**, with approximately 48.5% of web users in this age group utilizing the internet for health purposes, *Beck et al.* [2014]. Additionally, they found approximately 80% (371/474) of young online health seekers considered the information found online **reliable**. **Women** and people with **higher sociocultural** positions *(for employees and manual workers vs individuals with executive or manager positions)* were more likely to use the Internet for health purposes. As per *Figueiredo & Paiva* [2015] they trust the information found online and consider it reliable. The internet is also extensively used by users of the Mediatheque Simone de Beauvoir **library**, who consider it an important source of information and trust official websites, *Florence* [2005]. *Ducrot et al.* [2021] found in their research of French Health Market that after a rapid growth in the internet use for seeking health information in the 2010 to 2014 period, a **decrease** was recorded in 2017, in parallel with a **decrease in trust** in the quality and reliability of information found online. These findings underline the need for public health authorities to increase citizens' **eHealth literacy** and to provide **alternative trustworthy sources** combining the popularity and accessibility of general health information websites. *Woolford* [2015] said in her research on **Tourism** in France and their internet use that social media is not a valuable communication channel and other channels *(more "traditional")* should not be abandoned in favour of social media. Social media channel should be used as a combination with other channels like emails and display advertising. Purpose of mentioning about it in our context is that healthcare being super sensitive subject, probabilities are even lower than Tourism that people may be satisfying only with the social media content, despite its increasing spread over, so the viability of SERP is still standfast. And that is why out of the two types of search engines *(as talked in beginning)* *the focus of this thesis is the web search engines while maintaining social search engines mentioning contextually.* *Rolland & Freeman* [2010] found in their research on ecommerce users that the French on-line shoppers placed greater emphasis on the **security/privacy** dimension than their English-speaking counterparts. This was found in their "**E-tail SQ**" outcome, a 15-item scale to measure five key user values *(labelled ease of use, information content, fulfilment reliability, security/privacy, and post-purchase customer service).* Although not exactly the SERP research but having on such direct research of this behaviour on SERP; we can relate both being near next to grounds of similar usage, except the money transactional part of ecommerce platforms. So here now we have some explicit evidence of online **user behaviour of French market from health, wine buyers, students, tourism and e-commerce domains**. As per *Renahy et al.* [2007] and *Renahy et al.* [2010] expressed concerns about the **quality and**

[4] https://www.lequotidiendumedecin.fr/sante-societe/e-sante/55-des-francais-ont-deja-utilise-un-moteur-de-recherche-pour-sinformer-sur-leurs-symptomes

reliability of online information, leading to a decrease in trust over time still the trust in the information found online played a significant role in determining how individuals managed their health, with higher trust associated with certain behaviours. The behaviour of healthcare information seekers in the Washington state market indicate that many individuals seek both illness and wellness information online, with **83% using general search engines like Google for health information**, *Weaver et al.* [2010] and *Jia et al.* [2021]. As per a survey done on **28 European Union countries including France**, *Suenaga & Vicente* [2021] found that the demand for physician services is also influenced by Internet-based health information seeking, and there is a strong correlation between individual health status and online health information seeking; henceforth, comes another aspect of viability of **contextual framework** of this thesis i.e. *"Search User's HISB (Health Information Search Behaviour) on **SERP** in French Market"*. Table-3, Apprendix-9 shows the distribution and proportion of the number of articles published in the top 10 countries according to the address of the first author of the article, having France published 9 articles in context of our thesis topic. *It means not much research done around the research framework this thesis addresses.*

As a backdrop of our research, Diagram-1, Apprendix-9 represents an online health seeking information behaviour model. This model comprised of various multiple stages through which HIS is performed and each stage has a challenge of understanding HIS. *Our model aims to address all or some of these crucial stages fully or partially like health demands, demand motivations, accessibility, and decision-making; all are easily relatable with user's intent, cognitive capacity and understanding of their demand types.* The research model in Diagram-2, Apprendix-9 states that the health information consumers with different focuses were found to seek different online health information, which represents the varied behaviour of HIS thus justifies our choice of 10 different intent types divided into two categories. According to research model in Diagram-3, Apprendix-9 online health information seeking behaviour was found to be influenced by perceived usefulness, perceived ease of use and sense of self-worth, Davis F.D. (1989); which can be easily related to reassurance and precautionary intents included in our research context. *Benjelloun et al.* [2020] did research on Google Data set *(2.1M unique datasets, 2.6K domains appeared in the top 100 search results)* topic distribution as shown in Chart-1, Apprendix-9 ; and if to combine Biology and Medicine data sets from it, it makes 25.4% of data storage just related to healthcare and body. *So, from here it is easily catchable why specifically this thesis focuses on health sector.*

Research by [5]*Kyle Byers* [2020] shown in Chart-2, Apprendix-9 **shows the healthcare sector yet relying on and producing highest traffic percentage from Search Engines i.e. 87.85%**. Although it can be due to "Your Money Your Life" ***(YMYL,*** *here onward)* 'search interest' zone but Finance is also a YMYL zone, and it is standing at 67.78%. *So, interpreting what, why and how the people are looking so much on Search Engines in healthcare is the interest of this thesis. It also indicates the viability of Search Engines highest in the healthcare sector.*

The another very important study which frames the ground for this thesis is by *Maynes & Everdell* [2014] shown in Infographic-1, Apprendix-9. They with their agency mediative, did comparative research on **eye tracking behaviour** of the cursor *(mouse)* of the searchers on search engines from 2005 and 2014. Their results have been silent emphasizers for keeping such works continued as a parallel stream of banks of a river to the river's main-water-stream. *And we intend to demonstrate this study psychologically fully or partially through this thesis.* In reference to Infographic-1, while in 2005, this behaviour was triangular concentrated in top-left corner of the SERP screen; in 2014 it dispersed over the left side from top-to-bottom. It shows that with time, people have become more familiar with the SERP parts, and they analyse it with more interest and detailing. *Maynes & Everdell* [2014] reveals significant shifts in user behaviour on SERPs due to the introduction of new elements like the Carousel, Knowledge Graph, and Local Listings, alongside a decline in sponsored ad click-through rates (CTR) from 3.16% in 2005 to 0.7% in 2014. Organic Listings consistently capture the highest click activity (32.8%), despite the presence of new elements, which often provide information without requiring clicks. The Knowledge Graph significantly influences CTR, reducing clicks on other results to 36.5% when relevant, while irrelevant Knowledge Graphs drive 82% CTR to alternative results. Local Listings attract attention when physical location outweighs choice or reviews, and the Carousel, though intent-focused, fails to hold user attention. Traditional search patterns, like the Golden Triangle in Infographic-1, persist with Organic Listings but are disrupted by new elements, leading to quicker, less horizontal page scanning and reduced focus on top Organic Listings. Features like authorship mark-up, peer reviews, star ratings, and thumbnails enhance engagement, drawing clicks regardless of result position, particularly for experience-driven searches. So as per above findings in a mere span of 9 years from 2005 to 2014, the searcher's search behaviour took many remarkable and surprising turns. This induces to our curiosity to dig out current scenarios; about which survey of [6]*Olaf Kopp* [2024] conducted on LinkedIn, Twitter, and his website with over 500 marketing users revealed that **search engines** *(still highest),* social media, YouTube, and

[5] https://growthbadger.com/traffic-study/
[6] https://www.kopp-online-marketing.com/social-media-has-become-one-of-the-most-important-gatekeepers-for-content

newsletters are the most important sources for keeping up with the latest professional developments. He further adds that the share of social media in B2C will be significantly higher than in B2B. From Chart-3, Apprendix-9 of this study, it can easily be related with the research context of this thesis *that why doing research on SERP is still viable at present day.*

Research Aperture and Rationale of this thesis: In order to give a profound base to the knowledge of our research topic, the Literature Review (LR) Methodology of this thesis was divided into five parts viz-a-viz Foreground Literature Review (FLR), Synthesized Literature Review (SLR), Research Methodology, Method, Model, Metrics and Modalities Literature Review (R5MLR), Research Problem Contemporary Literature Review (RPCLR) and Arbitrary Literature Review (ALR) as demonstrated in Table-4 below.

Five Types of Literature Reviews (LR) done in this Thesis

LR Type	Aperture	Description	Purpose	Approach	Sections
Foreground LR - **FLR**	**Domain Level** i.e. SERP	Broader scanning of Literature giving it a universal coverage around the domain of this thesis.	Gaining factual awareness and understanding of knowledge of the research domain in context	Being chronical	Unit-1: Ch. 1, 2, 3, 4
Synthesized LR - **SLR**	**Thematic** i.e. French HISB on SERP	Specific Literature skimmed around the theme question of this thesis but still broader aperture.	Understanding the dedicated Literature developed till date.	Being maestro	Unit-2: Ch. 5
Research Methodology, Method, Model, Metrics and Modalities LR - **R5MLR**	**Technical** i.e. R5Ms relevant to thesis theme	LR of all possible R5Ms relevant to the topic of this thesis.	Getting ideas and updates from all possible **37 research models** relevant to our thesis.	Being architect	Unit-2: Ch. 7, Section 7.2
Research Problem Contemporary LR - **RPCLR**	**Statistical** i.e. Focused to Research Framework of this	Contemporary Literature Review only related to final research framework chosen for this thesis.	Knowing and understanding all possible **methods and metrics** to assess the research	Being critical	**RPCLR-1:** Unit-2: Ch. 7, Section 7.4 **RPCLR-2:**

	thesis		framework only in context		Unit-3: Ch. 9, Section 9.3.2
Arbitrary LR - **ALR**	**Arbitrary** i.e. Facts filling	LR included in arbitrary contexts	Assessing points with facts.	Being detective	Introduction, Ch. 6, Ch. 10

Table-4: Five Types of LR

Further to label our LR on sophisticated scientific norms, the **PRISMA** model of *Moher et al.* [2009] was implemented. Based on PRIMA model *(detailed PRISMA analysis is given in Chapter-8, Section-8.3).* Table-5 below presents the list of inclusion and exclusion of bibliographic document types reviewed for this research:

	Inclusion	**Exclusion**
Language	English in general, French for specific articles	Non-English, Non-French
Article Types	Journal articles, conference articles, conference proceedings, books, book chapters, articles in press, influencers blog articles, research agency blogs, research videos	Preprint, review papers, editorial, commentary, letter to the editor, perspective, opinion, viewpoint, brief communications
For contextual enrichment	Theoretically established coherent conceptual citations, Common ground theories from scientific journals, research agencies & blogs.	Non-scientific general statements, articles, blogs, reviews.
For argument enrichment *(to central research model)*	Logical citations in coherence with the central research model of this thesis which include algorithmic, mathematical, and scientific methods, models and methodologies only from scientific journals.	Research agencies & blogs, non-scientific general statements, articles, reviews.

Table-5: Inclusion and Exclusion criteria for Literature Review

Based on the **Research Rationale** set above, this thesis seeks to find similar or alternate user behaviour factors regarding SERP usage. This thesis aims to cover maximum possible aspects as per the Table-6 given in Chapter-2. Table-2 above presents an amalgamation of influence factors those define the SERP user behaviour distinctively.

Novelty Contribution Aims of our Research: Our thesis aims first to prepare a "**SERP Navigators' Psycho-Cognito Profiling**" of our test participants; then designing a **Supervised Machine Learning model in Python based on RNN BiGRU Dual Category Search Intents Classification both on Real-Human Data and AI Simulated Data combined**. Three analysis metrics of **Accuracy, F1-Score and AUC Scores** are used to identify proportions of Search Intent types in our dual-intent categories. First category is termed as "**Standard Search Intents**" which is comprised of four Search Intents viz-a-viz Informational, Navigational, Transactional and Commercial, whereas the second category is termed as "**Health Search Intents**" consisting of six Health Intents viz-a-viz Precautionary, Reassurance, Diagnostic, Treatment, Urgency and Courtesy. Everything is done in a systematic and structured way, as explained here in next section. **Our Psycho-Cognito Profiling and ML model combined is one a novel contribution in Marketing Psychology**, having huge possibilities of developing further grand **Marketing Strategies** by decodifying the 'Search Behaviours Factors' as promised in the title of this thesis.

Structure of this thesis (Units-3 | Chapters-10 | Pillar-Topics-48 | Topic-Cluster-55)**:**

a. **Unit-1** (Foreground Literature Review - Dynamics of Web Search Marketing Psychology) **covers Chapter 1 to 4:** *Technically this unit is not only about knowing the preface of our research but the grand Dynamics of WEB SEARCH MARKETING PSYCHOLOGY.* So, the purpose of this unit is to learn and explore broader possible literature about the topic of our thesis. **Chapter-1** is dedicated to chronicle history of five different aspects of SERP, which affirm a solid knowledge base in vast purview of G-AI, SEMLAs Evolving, Search Engines Evolving, SERP Interface Evolving and SERP Features Evolving. **Chapter-2** covers the **psychological variables of this research** and **central research model**. **Chapter-3** covers the technical literature around SERP. **Chapter-4** focuses on the literature related to three stakeholders of SERP i.e. the Search, Searchers and Search Producers.

b. **Unit-2** (Research Framework) **covers Chapter 5 to 8:** With 'intense and critical focus' drawn through FLR towards our thesis title, the purpose of this unit is to form the ground for our centric research model sets and building the room for new contribution to academic research. In standard terms it answers about Research AOQ *(aims-objectives-questions)* while standing on an affirmed platform of FLR. Its **Chapter-5** frames possible **'Research Gaps'** and **'Research Hypotheses'** through **Synthesized Literature Review** *(SLR here onward)* around our research theme. **Chapter-6** focuses on developing the **Research Scope** based on critical analysis, finding knowledge gaps and connecting them to variables, hypotheses, and our research motivations. **Chapter-7** prepares the **R5Ms Framework** in which **R5MLR** related to Research Scopes is covered, Five Research Problems are defined, the most feasible out of five problem is adapted to

run our research test, the contemporary LR **(RPCLR-1)** related to this problem is covered and its R5Ms Protocol is figured out. In **Chapter-8** complete **Bibliographic Analysis** is done. Bibliography covers Literature from 1929 to 2025 period.

c. **Unit-3** (Research Test Ontology) **covers Chapter 9 to 10:** Having prepared the 'Research Framework' in Unit-2; this Unit covers all practical work to get the final results of our research. **Chapter-9** is about defining the Research Test Method, Data Collection and Results Validation. Test Data Taxonomy, Psycho-Cognito Profiling of Test Participants, RNN BiGRU Intents Classification Supervised ML Model Testing, RPCLR-2 and Results are presented in this chapter. **Chapter-10** is about discussion, analysis and conclusion of our research findings. sums up of limitations and future scopes emerge from our research.

Unit-1

Foreground Literature Review

(Dynamics of Web Search Marketing Psychology)

Synopsis: This unit aims to skim through the literature in terms of Books, Research Papers, Articles, Blogs, Infographics, and various Marketing Science Databases to first form a foundational and sophisticated foreground of understanding and knowledge of Web Search Marketing Psychology, second to provide a build-up platform to the variables and hypotheses of this thesis. We proceed here by developing an affirm understanding on how the influence and importance of each Search Engine (SE) dwell in user minds along with impressions and believes about corresponding Content Creators and Content Hubs. It also talks about different kinds of Search Algorithms, Intelligences, Machine Learning Programs, Creatives and Contents that keep web search user utilising Search Engines. Addressing to all these aspects, this unit is divided into 4 chapters as follows:

1. Chapter-1: Chronology of Information Creation and Presentation towards SERP: This chapter is a chronological study of evolution of five major dimensions of Information Creation and Presentation in history till date. These dimensions are Generative Artificial Intelligence *(G-AI),* Search Engine Machine Learning *(SEML)* Algorithms, Search Engines, Search Engine Result Pages *(SERP)* and SERP Features.
2. Chapter-2: Psychological Literature Review in the context of SERP: This chapter covers the relevant psychological literary references and discussions of SERP around Web Searcher Behaviour such as Art of Phrasing, Natural Language Ontology, Topicality, Semantic Web, SERP Personalization and Search Skills.
3. Chapter-3: Technical Literature Review in the context of SERP: This chapter covers the relevant technical literary references and discussions of SERP around Web Search Utility such as SERP Data Sciences, Generative Artificial Intelligence, Search Engine Machine Learning, Deep Learning, Large Language Models and Metasearch.
4. Chapter-4: Web Search Dynamics: This chapter spins the focus towards the SERP Model of Google; the three stake holders of SERP *(the Content Creator, the Search Engine, and the Web Searcher),* SERP Interface, Web Search Intelligence, Web Search Market and SERP for Gen-Z.

Chapter-1

Chronology of Information Creation and Presentation towards SERP

Hypothetically, the SERP is a philosophy of representation of knowledge and content falling into same pipeline as asserted by *Alfred North Whitehead* [1929] that, "all philosophy is a series of footnotes to **Plato**, then all theory and practice of knowledge representation are surely a series of footnotes to **Aristotle**". He puts it in his book at page 9 as "*There is no doctrine put forward which cannot cite in its defence some explicit statement of one of this group of thinkers, or of one of two founders of all Western thought, Plato and Aristotle*". So, in terms of philosophy, this thesis being on the field of art of information creation and its presentation is easily connectable to footnotes of Plato and Aristotle with an add-up of most advanced technical manner of representing the information in digitalised form of SERPs. Empirically, Plato and Aristotle provide foundation to Artificial Intelligence *(AI here onward),* the support wall of online information generation through Generative-AI in the form of SERPs and most recently the Large Language Models *(LLMs here onward)*. The Infographic-2, Apprendix-9 represents the historical milestones of AI till 2018.

1.1. Chronology of G-AI Evolving - From AI to G-AI:

Taking a cue from Infographic-2 about historical journey of AI, we shall narrow-down our review towards Generative-AI *(G-AI here onward)* because SERPs are considered as a sub-branch of G-AI in technical terms. The development of G-AI has been a turbulent journey, with significant advancements and challenges. The field of AI, including G-AI, was born in the mid-20th century, with early work at Dartmouth College, *Strickland* [2021]. A little more than half a century ago *(when Norbert Wiener was devising Cybernetics, Wiener N. [1948] and Gray Walter was building robotic tortoises, W.G. Walter [1950])* the English mathematician **Alan Turing** proposed a much-discussed imitation game used as a yardstick for assessing if a machine is intelligent or not, which, since then has been known as the **Turing Test for Artificial Intelligence** *A.M. Turing* [1950] et *Legg & Hutter* [2007]. The advent of the general-purpose computer in the fifties of last century did substantially alter the dreams and ambitions of people and made AI an even more tangible possibility. So, in 1956, the **fathers of "modern" AI,** Marvin Minsky, John McCarthy, Allen Newell, Nathaniel Rochester, Claude Shannon, and Herbert Simon came together for a summer school at Dartmouth College *(Hanover, New Hampshire)* under the premise "*that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it*", *P. McCorduck* [2003]. AI has expanded into a broader study of the interaction between the body, brain, and

environment since the 1980s, *Lungarella et. al.* [2007]. The recent emergence of Chat GPT has sparked a new wave of innovation in this area, *Yu & Guo* [2023]. However, the use of G-AI tools, which are trained on internet data, raises concerns about data quality and potential biases, *Martinez et al.* [2023]. Despite these challenges, the potential of G-AI in education and other fields is significant, *Yu & Guo* [2023]. Here is a chronological overview of some of the significant events in the history of G-AI:

1. **1956 of 1950s, the official birth year of AI:** The term "Artificial Intelligence" was coined, and researchers began exploring the possibility of creating machines that could simulate human intelligence. AI was marked as a field officially at Dartmouth Conference in 1956. John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon organized the event, which aimed to explore the possibilities of creating intelligent machines, *Strickland* [2021].
2. **1960s, the decade of early AI Programs:** Early AI programs focused on symbolic reasoning and problem-solving, such as the General Problem Solver *(**GPS** here onward)* developed by Allen Newell and Herbert A. Simon, *A. Newell* [1963].
3. **1970s brought the Expert Systems:** Expert systems, which used predefined rules to make decisions in specific domains, gained popularity. **MYCIN**, an expert system for diagnosing bacterial infections, was one notable example; *Buchanan et al.* [1984]. Bayesian networks and machine learning were introduced in the 1970s, *R. Evans* [2014].
4. **1980s Neural Networks Resurgence:** Neural networks, inspired by the structure of the human brain, experienced a resurgence. Backpropagation, a key algorithm for training neural networks, was developed during this time. Artificial Neural Networks *(**ANN** here onward)* re-emerged in the early 1980s as an interdisciplinary field, gaining visibility through conferences, professional societies, and journals, and advancing in theory and applications; *McCain* [2006]. Bayesian Neural Networks *(**BNNs** here onward)* were introduced in the 1980s for challenging Machine Learning *(**ML** here onward)* tasks like detection, regression, and classification in computer vision, speech recognition, and Natural Language Processing *(**NLP** here onward)*, *Ethan & Fookes* [2020].
5. **1990s boom of Bayesian Networks and ML:** Bayesian networks and probabilistic graphical models *(**PGMs** here onward)* gained attention. ML techniques, such as support **vector machines and decision trees**, became more prominent. Bayesian networks and ML boomed together in the 1990s, offering advantages for data modelling and avoiding overfitting; *Heckerman* [1997].
6. **2000s entry of Deep Learning and Neural Networks:** Deep Learning *(**DL** here onward)* gained momentum with the development of more sophisticated Neural Network *(**NN** here onward)* architectures. The combined entry of DL and NNs happened in the mid-2000s, with rapid progression since 2012, *Rawat & Wang* [2017]. Breakthroughs in training Deep Neural Networks *(**DNNs** here onward),* including

Unsupervised Learning and Convolutional Neural Networks *(**CNNs** here onward)* contributed to significant advancements.

7. **2010s of Generative Models and GANs and Transfer Learning and Transformer Models:** Generative models, such as Variational AutoEncoders *(**VAEs** here onward)* and Generative Adversarial Networks *(**GANs** here onward),* emerged. GANs, introduced by Ian Goodfellow and his colleagues in 2014 and have been widely applied to different areas since then, including digital image processing and medical imaging analysis, [7]*Bonder* [2023], [8]*GAN Wikipedia, Lan et al.* [2020]. Transfer Learning and Transformer Models *(**TLTMs** here onward)* were introduced around end of 2014 and beginning of 2015. Later on, TLTMs such as OpenAI's GPT series, demonstrated the power of Large Pre-trained Language Models *(**LPLMs** here onward)* in various natural language processing tasks; [9]*TLTM Wikipedia, Wolf et al.* [2019].
8. **2020s for the Continued Advancements:** Ongoing research and development in G-AI continue to evolve rapidly, with ongoing efforts to improve models, address biases, and explore new applications. Elizabeth Reid, the Vice President & GM - Search at Google while talking about "Improving Search with generative AI" states that, *"with new breakthroughs in generative AI, we're again reimagining what a search engine can do. With this powerful new technology, we can unlock entirely new types of questions you never thought Search could answer, and transform the way information is organized, to help you sort through and make sense of what's out there.",* [10]*Reid* [2023]. She further adds that G-AI is transforming searches to contextual conversational searches. G-AI also dominates the online shopping i.e. ecommerce activities tremendously. So, it was necessary to start this literature review by glancing on the G-AI magic on Search methods and grasping a gauze of how crucial could it be of talking about SERP without taking this dig about G-AI. However along with ongoing boom in G-AI, the ethical considerations and responsibilities can be another independent parallel research scope to this thesis. Below Chart-4 depicts the bullet terms invented or introduced in G-AI evolution journey from 1950s to 2020s.

[7] https://reintech.io/blog/understanding-gans-with-r
[8] https://en.wikipedia.org/wiki/Generative_adversarial_network
[9] https://en.wikipedia.org/wiki/Transformer_(deep_learning_acrhitecture)
[10] https://blog.google/products/search/generative-ai-search/

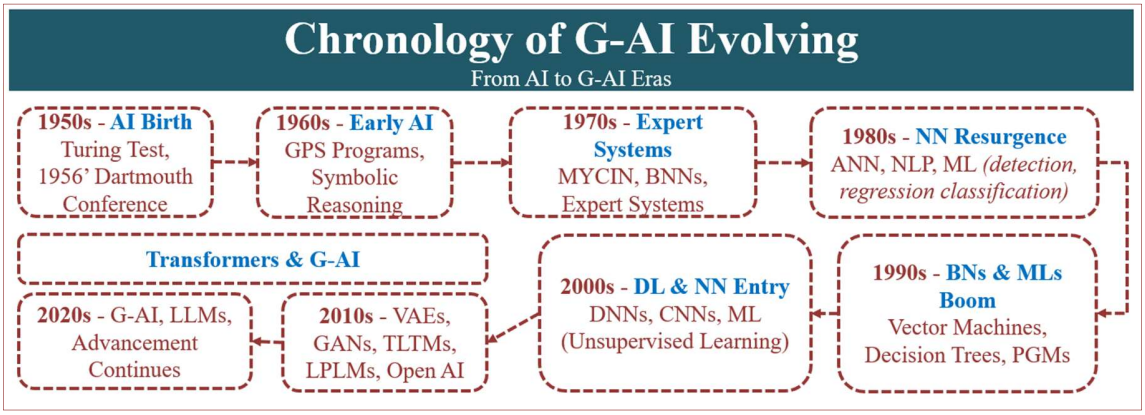


Chart-4: Chronology of G-AI Evolution

1.2. Chronology of SEMLAs Evolving - Teaching & Training SEs:

Search Engine Machine Learning Algorithms *(**SEMLAs** here onward)* can be seen as a superset to G-AI and subset to AI. The evolution of "SERP Algorithms" is a dynamic process for formulating the massive web content through ML**.** This learning is influenced by technological advancements, user behaviour, and the ever-growing volume of web content. Specific algorithms are characterized by a continuous effort to provide more relevant, personalized, and user-friendly search experiences while adapting to the evolving landscape of the internet and user expectations. These ML Algorithms developed for Search Engines to handle the massive online search for producing a high-quality user search satisfaction are termed in combination as SEMLAs. Ongoing research and development in AI, NLP and other fields are expected to shape the future trajectory of Search Engine *(**SE** here onward)* algorithms. And that is why the involvement of SEMLAs has seen significant advancements in recent years. *Real et al.* [2020] and *Kuang & Leung* [2019] both highlight the potential of **AutoML** and **Self-Learning Search Engine Architectures**, respectively, in automating the discovery and evolution of ML Algorithms. *Real et al.* [2020] specifically demonstrates the ability to discover complete Machine Learning Algorithms *(**MLAs** here onward)* using basic mathematical operations, while *Kuang & Leung* [2019] focuses on the use of 'Unsupervised Reinforcement Learning' to continuously improve search performance. These developments have led to the integration of MLAs in various applications, including data collection, preprocessing, visualization, and decision-making, *Kumar et al.* [2023]. Here's a chronological overview of key MLAs related to Search i.e. SEML with their updates:

1. **Early Information Retrieval (Pre-1990s):**

i) **Boolean Models:** Very first algorithm type and very successful till date has been the Boolean Search as [11]*Roger Montti* [2022] emphasizes their importance of Booleans in SEO. Boolean models use simple logics to match keywords in documents and retrieve relevant results, though these systems lack the sophistication of modern algorithms.

2. **Emergence of Web Search Engines (1990s):**

i) **PageRank (1996):** Developed by Larry Page and Sergey Brin at Stanford, **PageRank** was the foundation for Google's search algorithm, which was a breakthrough algorithm that assessed the importance of web pages based on the number and quality of links pointing to them. It assigned a numerical weight to each element of a hyperlinked set of documents, determining the importance of web pages based on link structure. The basic formula of calculating a PageRank is:

$$PR(u) = \sum_{v \epsilon B(u)} \frac{PR(v)}{L(v)}$$

Formula explanation: PageRank value for a page **u** is dependent on the PageRank values for each page **v** contained in the set B(u) *(the set containing all pages linking to page u),* divided by the number L(v) of links from page **v**, [12]*PageRank-Wiki*.

3. **Introduction of ML, SERP Features and Personalization (2000s):** [13]*Danny Sullivan Google* [2023] enlists 25 major SEMLAs traversing from 2000 to 2023, *(point 4 and 5 below also link to this citation)* remarkable changes can be listed as follows:

i) **Introduction of ML:** Search engines began integrating MLAs for tasks such as ranking, indexing, and query understanding. The basic ML models considered factors like keyword relevance and link popularity.

ii) **Specific to Google Algorithm Updates:** Google introduced several major updates, including "Florida" (2003), "Jagger" (2005), and "BigDaddy" (2006), which aimed to refine search results, combat spam, and improve the overall quality of the index.

iii) **Introduction of Local Search (2004):** Search engines integrated algorithms for local search, considering geographic information and providing users with location-specific results.

iv) **Advent of Personalized Search (2005):** Search engines started incorporating personalization features, considering user search history and behaviour to tailor search results. This marked the beginning of more user-centric algorithms. In 2005, Google started incorporating personalized search results based

[11] https://www.searchenginejournal.com/how-to-use-boolean-search-operators-seo-application/7822/
[12] https://en.wikipedia.org/wiki/PageRank
[13] https://blog.google/products/search/25-biggest-google-search-updates/

on a **user's search history, location, and previous interactions** with Google services. This aimed to provide more relevant results tailored to individual preferences.

v) **Universal Search (2007):** Google introduced Universal Search, blending different types of content *(videos, images, news)* into traditional web search results. This provided users with a more diverse and comprehensive set of results, knowns as the expansion of search engine features.

4. **Social Engagement Signals, NN and NLP Era (2010s):**

i) **The rise of Social Engagement Signals and Influence Tracking (2010):** Search engines began incorporating social signals, such as likes and shares, into their algorithms to assess content popularity and relevance.

ii) **Introduction of DL and NNs (2010s):** The use of DL techniques, particularly NNs, became more prevalent in SEML. These models enabled more sophisticated processing of complex data and improved understanding of user intent.

iii) **Google Panda (2011):** Panda targeted low-quality content, excessive keywords, link farms and content farms, penalizing sites with thin or duplicated content. It emphasized the importance of **high-quality, original content** for better rankings.

iv) **Google Penguin (2012):** Penguin focused on penalizing websites that engaged in **manipulative link-building practices**, such as buying links or participating in link schemes and penalized sites with low-quality backlinks and began to value those receiving contextualized links from high-quality and authoritative sites. It aimed to ensure the credibility of the links influencing search rankings.

v) **Hummingbird (2013):** Google introduced the Hummingbird update to better understand the **context and intent** behind search queries. It emphasized **NLP and the semantic meaning** of words, moving beyond keyword matching and improved results classification by focusing on the exact match between the keyword and the search intention.

vi) **Mobile-Friendly Update (2015):** Also known as **Mobilegeddon**, it was brought in response to the increasing use of mobile devices, which was for favouring thus boosting the rankings of mobile-friendly pages in mobile search results.

vii) **RankBrain (2015):** Google's RankBrain is an AI-driven algorithm that uses machine learning to **interpret and understand the meaning** behind search queries. It helps Google deliver more relevant results for ambiguous or new queries.

viii) **Google Fred (2017):** While not officially confirmed by Google, the "Fred" updates were believed to target **low-quality content and aggressive monetization practices**, impacting sites that prioritized revenue over user experience.

ix) **Mobile-First Indexing (2018):** With the rise of mobile usage, search engines introduced algorithms that prioritize mobile-friendly content, ensuring a better user experience for mobile searchers.

x) **BERT and NLP (2019):** Google introduced BERT *(Bidirectional Encoder Representations from Transformers),* an NLP *(Natural Language Processing)* model, to better understand the context, nuances of user queries and improving the accuracy of search results. BERT is an open-source NLP pre-training technique. This marked a significant advancement in understanding conversational search queries.

5. **CWV and G-AI Era (2020s):**

i) **BERT Refinements (2020s and beyond):** Ongoing refinements to BERT and the integration of advanced NLP techniques continue to enhance search engines' ability to understand user intent, context, and complex queries.

ii) **Core Web Vitals (2021):** Google introduced Core Web Vitals ***(CWV** here onward)* as a ranking factor, emphasizing the importance of **page load speed, interactivity, and visual stability** in the user experience. It emphasized the importance of delivering a seamless and fast user experience.

iii) **Page Experience Update (2021):** Building on CWV, Google rolled out the Page Experience Update *(first input delay, largest contentful paint, and cumulative layout shift),* incorporating page experience signals such as CWV, mobile-friendliness, safe browsing, HTTPS, and intrusive interstitial guidelines into ranking considerations.

iv) **Core Algorithm Update (May 2022):** Updated and revamped major aspects of Google's core algorithm and ranking factors for the first time since November 2021.

v) **Core Algorithm and Content Update (Sep. 2022):** The second broad core update, implementing more major changes to Google's ranking system, rewarding sites that produce helpful content, and underscoring the importance of reviews.

vi) **Conversational AI Tool as another new SERP Feature (Nov. 2022):** With the introduction of ChatGPT of Open AI, both Bing and Google started inducing this feature in their Search Engine menu.

vii) **Integration of Chat-GPT in leading SEs (2023):** In early 2023, Bing brought **the Bing Chat** and Google brought **Google Bard.**

viii) **Privacy and Ethical Considerations (Ongoing):** Search engines are increasingly emphasizing privacy and transparency in algorithmic processes. Techniques like **federated learning and differential** privacy are explored to protect user data late 2023 and starting 2024.

ix) **Future Advancements in AI and Quantum Computing (Ongoing):** The future of SEML likely involves further integration of advanced AI techniques and the potential impact of quantum computing

on search algorithms, offering new dimensions in processing and understanding information. Below Chart-5 depicts the bullet terms invented or introduced in decades from pre-1990s to 2020s.

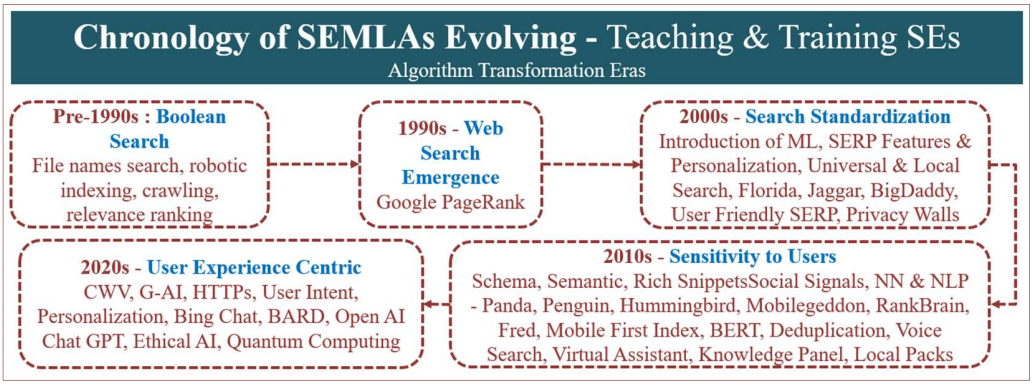


Chart-5: Chronology of SEMLAs Evolving - Teaching & Training SEs

1.3. Chronology of Search Engines Evolving - From Web 1.0 to Web 6.0:

A journey from an Information Gatekeeper to a Marketing & Branding Platform, further to a Personalized Conversational Search: Over the past decades, SEs have become the biggest source of acquiring information, *Croft et al.* [2010]. This is a parallel journey to that of AI to G-AI. SEs have traversed through transformative changes. Saga begins with fulfilling the public need of construed information through computers in 1990s, which along with passage of the time became such a big integral part of human life that in 2012, 91% of online adults found using search engines such as Google, Bing, Yahoo, Baidu, Yandex worldwide to find information on the web, which was 84% in 2004, *Moran & Hunt* [2014] - Page 4. In the early years of Online Search, it was primarily meant for information finding, pertaining more like an **information gatekeeper**. With the passage of time, it enticed the number of purposes and interests of the people, seeing this unprecedented surge in use and importance of SEs, the marketers started using it as a **marketing and branding platform** and by 2020s it transformed to **personalized conversation search platform,** and the sage has no end, at least in decades to come. [14]*Miller & Moses* [2023] work is a corner-stone description of the journey of SE's evolving. Let's visualize this tremendous and tedious journey in three segments viz-a-viz Search Engines, SERP Interface and SERP Features, starting with Search Engines first:

[14] https://medium.com/@efficio/journey-through-the-evolution-of-the-web-from-web-1-0-to-web-6-0-and-beyond-7d87661ddf2a

1. **Pre-web Search Engine Era (Pre-19990s): WHOIS** *(Who Is)* user search dates-back to 1982, and the **Knowbot** Information Service multi-network user search was first implemented in 1989, [15]*Search Engine Wikipedia 4.*
2. **Early Web Search Engine Web 1.0 Era (Early 1990s):** It began with the **Archie** Search Engine created for **indexing and searching file names on FTP sites** in 1990 by Alan Emtage at McGill University Montreal, Canada, [15]*Search Engine Wikipedia 4;* [16]*Beverly Mapes* [2023]. In 1991, **Tim Berners-Lee created a virtual library** to help users find URLs for different websites, [15]*Search Engine Wikipedia 4*. **Gopher** (1991) was an early protocol for **organizing and retrieving documents** on the internet. While not a traditional search engine, it played a role in the early days of web navigation. In 1993, the first **graphical web browser**, **Mosaic**, was introduced, which allowed users to browse the World Wide Web. In 1993, **JumpStation** brought a new leap with their **linear search**, and the first popular search engine on the Web was **Yahoo!,** which started as a Web directory in 1994 and added a **search function** in 1995, [17]*Search Engine Wikipedia 5*. Developed by Jonathan Fletcher at the University of Stirling, JumpStation is considered the first search engine to use a **web robot (or spider) for indexing web pages**. It created an index of titles and headers. **VERONICA** (1993) was developed at the same university as Archie. It indexed and searched file names but focused on Gopher protocol sites. In 1993, **World Wide Web Wanderer** was created by Matthew Gray at MIT, it was **one of the first web crawlers**. It collected/indexed URLs and determined the size and growth of the World Wide Web. In 1993 itself, **ALIWEB** was one of the earliest search engines to **index web pages**. It allowed webmasters to submit their pages for indexing.
3. **Crawler-based Short Web 1.0 Era of Search Engines (Mid-1990s):** In 1994, **WebCrawler**, created by Brian Pinkerton at the University of Washington, was one of the **first full-text web search engines**. It allowed users to search for any word in any web page, [18]*seomechanic.com 1* and [19]*Owen Whitcombe* [2022]. WebCrawler became one of the first search engines to index web pages and **display results in a list format**. In 1994, Michael Mauldin at Carnegie Mellon University developed **Lycos,** the first search engines to incorporate **relevance ranking into its results**. Then in the same year of 1994, Jerry Yang and David Filo founded Yahoo! initially as a directory but later played a significant role in organizing the early web. **Infoseek (1994)** focused on **natural language queries and relevance ranking**. In 1995, Six Stanford students, brought a search engine featured with a combination of human-edited directories and spider-based indexing. In the same year Digital Equipment Corporation built **Altavista**, one of the most powerful and

[15] https://en.wikipedia.org/wiki/Search_engine
[16] https://topofthelist.net/a-history-of-search-engines/
[17] https://en.wikipedia.org/wiki/Timeline_of_web_search_engines
[18] https://www.seomechanic.com/complete-history-search-engines/
[19] https://www.libertymarketing.co.uk/blog/a-history-of-search-engines/

popular search engines of its time. It pioneered many search technologies, including advanced search features. **Excite (1995)** incorporated **NLP** and offered **personalized content**. It expanded its features to include email, news, and other services. *Jose & Lal* [2013] state that SE crawlers are highly automated programs that periodically visit the web site to collect information. The time spent by various crawlers is significant in identifying the server load as major proportion of the server load is constituted by search engine crawlers. As per their findings certain crawlers like **Googlebot, Feedfetcher-Google, Bingbot, Baiduspider** etc. showed consistency in the time spent *(were dynamic and present in almost every hour which contributes a major portion of the server load)* whereas certain other bots like **Gosospider, MSNbot, Discobot** etc. were not consistent in their behaviour.

4. **Modern Web 1.0, Web 2.0, Web 3.0 Era of Search Engines (Late 1990s and 2000s):** In 1997, **Ask Jeeves** *(later Ask.com)* introduced a question-and-answer format, allowing users to pose queries in natural language. It aimed to provide more **human-like interactions** with search. Then enters the largest serving biggest Search Engine till date the **Google** in 1998; [18]*seomechanic.com 1* and [19]*Owen Whitcombe* [2022]. Founded by Larry Page and Sergey Brin at Stanford University, Google revolutionized the search engine landscape with its **PageRank algorithm**. It quickly became the dominant search engine due to its effective and relevant search results. **Microsoft's MSN Search** *(later rebranded as Bing)* entered the SE market in 1998 and became a major competitor to Google. **Bing introduced features like image search and integrated results** from various sources. Yahoo! transitioned from a directory-based approach to a more algorithmic search engine, powered by Inktomi and later Yahoo Search Technology renaming itself as **Yahoo! Search** in 2002. Then around 6 years later in 2008 **DuckDuckGo** brought a **privacy-focused search engine** that gained popularity for not tracking user data. It emphasizes user privacy and delivers unbiased search results. In 2009, Microsoft introduced **Bing** as its search engine with focus on features like a daily changing background image and a unique approach to displaying search results.
5. **Niche Driven Web 3.0 & Web 4.0 Era of Search Engines (2010s):** This era turned to be an era of rigorous developments and updates on technical fronts which catered Search Engines to specific niche-based markets and user preferences. The rise of smartphones led to a shift in search behaviour, with a focus on **mobile-friendly** results. **Voice search**, powered by virtual assistants like **SIRI and Google Assistant**, gained prominence. Google introduced the **BERT** algorithm based on NLP in 2019 which enhanced its ability to understand the context and nuances of user queries, particularly in natural language.
6. **G-AI Web 4.0 & Web 5.0 Era of Search Engines (2020s - Ongoing):** After the innovation of Google and Bing no such big worldwide search engine entered the market except a revolutionary break through that

Google Introduced **CWV** in early 2020, [20]*CWV-Google* [2020] as a ranking factor, emphasizing user experience metrics like **page load speed, interactivity, and visual stability**. Then Open AI in November 2022 brought Chat-GPT. Both are technical enhancements around information processing and presenting instead by Search Engines now. The evolution of search engines showcases a continuous effort to improve information retrieval, user experience, and the relevance of search results. Ongoing technological advancements are likely to shape the future landscape of search engine development. Recent trends include a focus on AI-driven search, federated learning, and evolving privacy considerations. The future may see advancements in quantum computing and further integration of AI technologies. We shall be linking G-AI as the latest innovation in SERP Chronical Journey below.

7. **Predicted Web 5.0 & Web 6.0 Era of SEs (2030s):** Experts assume that the decade of 2030 shall see more vibrant AI applications such as Quasi emotive web, *Eran Fisher* [2014] and Hypothetical transformative web. [14]*Miller & Moses* [2023] Web 6.0 is a vision of the future web that could involve hyper-personalization, advanced AI, quantum computing, and fully integrated web experiences. Below Chart-6 depicts the bullet terms invented or introduced in decades from pre-1990s to 2020s.

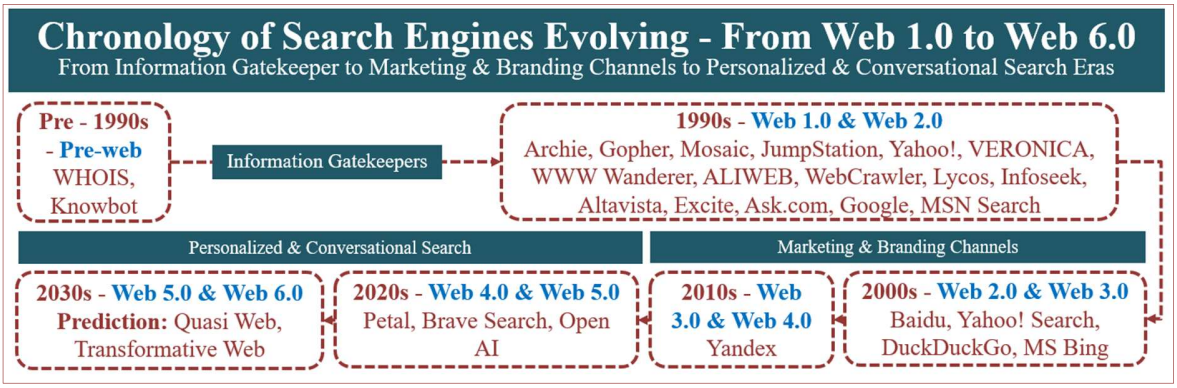


Chart-6: Chronology of Search Engines Evolving

1.4. Chronology of SERP Interface Evolving – From Click-Search to Zero-Click-Search to Conversational Search:

A journey from Click-Search to Zero-Click-Search, further to Conversational Search: The history of SERPs is marked by constant innovation and adaptation to changes in technology and user behaviour, making it a dynamic and evolving aspect of the **digital real estate** and its **digital landscape** the Search Engine. As we

[20] https://developers.google.com/search/blog/2023/05/introducing-inp#

saw in Search Engine chronology above for presenting more structured searched information to searchers there has been a parallel evolution of indexing, crawling, and presenting the information on SEs. [21]*Webdesignmuseum,* [13]*Danny Sullivan Google [2023],* [22]*Google Press [2004],* [23]*Google Search Central,* [24]*similarweb-zero-click,* [25]*Google Timeline Wiki, SERP Features Timeline Brodieclark [2024],* [17]*Search Engine Wikipedia 5,* [15]*Search Engine Wikipedia 4* list out a periodic transformation of SERP Interface of Google and other SEs. This evolution can be visualized as synthesized visualization of above references on timeline as follows:

1. **Pre-Web and Single Column Layout Era (Pre-1990s and 1990s):** Starting with **simple indexing and searching file names** on FTP sites of Archie in 1991, advancing to **web robot for indexing** (spider) web pages algorithm of JumpStation, in 1993 through to one of the first **web crawling technic** by WWW Wanderer in 1993 itself to first indexing based WebCrawler in 1994, to incorporating **relevance ranking** into its results by Lycos in 1994, to **PageRank Algorithm** of Google the display of the search results on SERP was text-based with a clean and simple SERP layout, emphasizing speed and efficiency. In the early days (1995) of search engines like Yahoo! and AltaVista, SERPs primarily consisted of **simple text-based listings**. Basic titles and descriptions were displayed for each result, providing users with links to relevant web pages. Later, these results were categorized as Organic Results to differentiate them from Paid Results of Google AdWord brought in 2000.
2. **2-3 Columns Layout Era (Early 2000s):** As said above Google introduced the AdWords in 2000 to facilitate the online paid advertising thus named as "**Paid Results**". This was the era when these simple Information Gatekeepers *(till now)* started transforming into well echoed marketing and advertising digital landscapes. AdWords facilitated businesses to bid for ad placements on SERPs. The integration of paid advertisements marked a significant shift in the monetization of search engines.
3. **2-3 Columns Layout Era (Late 2000s):** To tackle query ambiguity, search result diversification approaches have recently been proposed to produce rankings aimed to satisfy the multiple possible information needs underlying a query, *Santos et al.* [2015]. The Search Engines like Google, Yahoo, and Bing continued to **diversify** their features on SERPs, including image search, video search, news search, and local search results. In this zest **Universal Search** was introduced by Google in 2007, which integrated various types of content into a single SERP, including images, videos, news, and maps, [25]*Google Timeline Wiki*. This marked the REAL DEPARTURE from traditional **text-only** results, providing users with a more

[21] https://www.webdesignmuseum.org/gallery/google-1998
[22] https://googlepress.blogspot.com/2004/10/google-announces-desktop-search.html
[23] https://developers.google.com/search/blog/2009/05/introducing-rich-snippets
[24] https://support.similarweb.com/hc/en-us/articles/360006488277-Zero-Click-Searches
[25] https://en.wikipedia.org/wiki/Timeline_of_Google_Search

diverse and **multimedia**-rich experience. DuckDuckGo brought **privacy-focused search experience** in 2008, which turned the focus of Search Engines not only towards privacy of searchers but somehow laid the line of user-friendliness and personalization of SERP.

4. **2 Columns Layout Era (Schema & Semantic - Early 2010s):** Now Search Engines started thinking towards a standard and structured **nomenclature** for SERPs which could be coherently understood by Searchers across various Search Engines. It got Bing, Google, Yahoo, and Yandex developed together the Schema.org, a vocabulary for **machine-readable, expressive descriptions** of web content launched on 2nd June 2011, *Ronallo* [2012] & *Karle et al.* [2017]. Due to increased competition among Search Engines, they began thinking by putting themselves in the shoe of the Searcher and it gave birth to the **Intent mapping Semantic Search algorithm** named Hummingbird by Google in 2013. This marked the era of understanding and matching the Search Topics, Latent Semantic Keywords and Searcher Intents. In 2014, Google introduced **featured snippets, displaying concise answers** to user queries at the top of the SERP. Featured snippets aimed to provide quick and direct responses without users having to click through to a specific page. Google introduced its **Mobile-Friendly Update** in 2015, which favoured mobile-optimized websites in mobile search results. Google started displaying **rich snippets**, which include additional information like **ratings, reviews, and event information**, in search results. Featured snippets, which provide **direct answers** to users' queries, also became prominent on SERPs. Google launched **RankBrain** in 2015; similar to Hummingbird, tries to comprehend the client's goal behind questions.
5. **2 Columns Layout Era (Late 2010s):** Around 2015 with the rise of voice-activated **virtual assistants** like SIRI, Google Assistant, and Amazon Alexa, SERPs began to adapt to voice search queries. **Voice search** often provides concise, spoken responses instead of traditional lists of search results. Voice search capabilities have become more prominent, allowing users to perform searches using natural language and voice commands. In 2015, Google announced a shift towards **mobile-first indexing**, prioritizing mobile-friendly websites in SERP rankings. The move responded to the increasing prevalence of mobile devices for internet access. In continuation to Mobile favouring, in 2016, Google rolled out a mobile-friendly update, penalizing websites that were not optimized for mobile devices in mobile search results. This emphasized the importance of responsive web design for improved user experience. In 2019, Google introduced the **BERT** algorithm to better **understand the context and nuances of user queries**. BERT aimed to improve the accuracy of search results by considering the context of words in a search query. The inclusion of various SERP features, such as **knowledge panels, local packs, and rich snippets**, enhances the depth and relevance of search results. Google started **deduplication** on featured snippets from the main organic search results in 2016, ensuring that if a page appeared in the featured snippet, it would not be repeated in the regular search results. Later in 2020 Google rolled back the deduplication policy, allowing

webpages that were featured in snippets to also appear in the main organic search results. This change aimed to provide more visibility to the source pages.

6. **2 Columns Layout Era (Conversational and User Experience Centric 2020s & Ongoing):** Google announced the inclusion of **Core Web Vitals** in 2020, focusing on **user experience** metrics like page load speed, interactivity, and visual stability. Pages optimized for CWVs were given preference in SERP rankings. Search engines, especially Google, continue to integrate AI and ML into their algorithms to refine search results. **Personalization, NLP, and understanding user intent** are areas where ML plays a crucial role. Ongoing developments emphasize the Searcher **privacy and security**. Search engines are placing increased emphasis on user privacy and security, with features like encrypted search (HTTPS) and enhanced privacy settings. Efforts to minimize the storage of personally identifiable information reflect a growing concern for user data protection.

Historical SERP Interface Journey: You can view in below references from *Appendix-1* about how the Google SERP view changes its appearance along with time:

a) Google SERP view 1998: Refer to *SAS-32*.
b) Google SERP view 2000: Refer to *SAS-33*.
c) Google SERP view 2004: Refer to *SAS-38*.
d) Google SERP view 2006: Refer to *SAS-25*.
e) Google SERP view 2008: Refer to *SAS-26*.
f) Google SERP view 2011: Refer to *SAS-27.*
g) Google SERP view 2014: Refer to *SAS-34*.
h) Google SERP view 2017: Refer to *SAS-35*.
i) Google SERP view 2019: Refer to *SAS-28*.
j) Google SERP view 2021: Refer to *SAS-36*.
k) Google SERP view 2024: Refer to *SAS-37*.

Discussion: The chronological evolution of SERP Interface illustrates a progression from simple text-based listings to a multifaceted and dynamic interface that incorporates various content types, personalization, and advanced algorithms. As technology continues to advance, the SERP Interface will likely undergo further refinements to meet the evolving needs and expectations of users. SERPs Interface Structure and its Information Blocks and Snippets *(such as Organic, Paid and Featured Results)* are talked in depth in Chapter-4 of this thesis. Below Chart-7 depicts the bullet terms invented or introduced in decades from pre-1990s to 2020s.

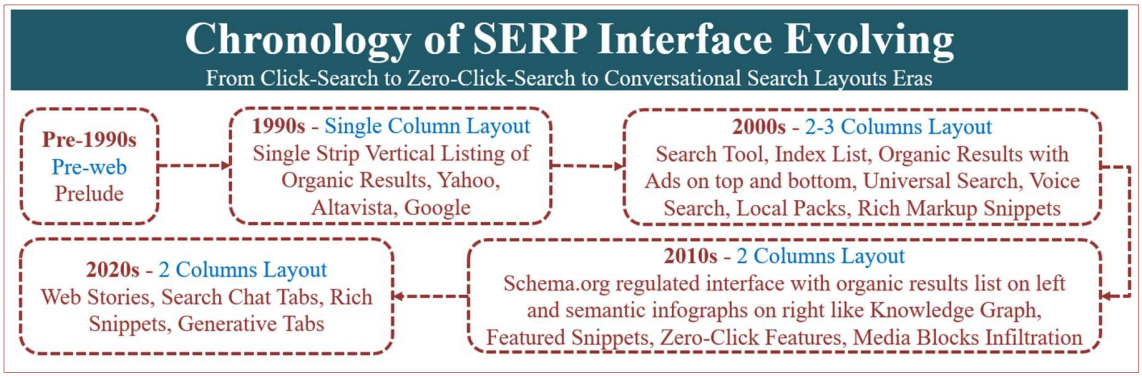


Chart-7: Chronology of SERP Interface Evolving

1.5. Chronology of SERP Features Evolving – From Read-only to Conversational:

A journey from Read-only to Conversational: The evolution of SERP has seen a significant shift from the traditional "10 blue links" to a more diverse range of features, including direct answers to queries, *Oliveira & Teixeira Lopes* [2023]-1 and [26]*Brodieclark.com 1*. *Oliveira & Teixeira Lopes* [2023]-2 depicts the SERP Features chronology in Infographic-3, Appendix-9. This evolution has been driven by the need to provide more comprehensive and relevant information to users. The evolution of SERP features highlights a progression from basic text-based listings to a dynamic and multifaceted interface that incorporates various content types, advanced functionalities, and user-centric features. Featured snippets have become a prominent feature, providing users with concise and direct answers to their queries. Their evolution has been marked by continuous innovation and adaptation to changing user needs. There development journey has been a parallel development to that of AI, ML, Search Engines and SERP Technologies talked in above sections. Having a catchable insight from Infographic-3 and [27]*SERP Bedgarone 1*, let's glance on which feature was developed with what purpose below:

1. **Organic Textual Listing Era (1990s):** Early search engines in 1990s, such as **Yahoo! and AltaVista**, presented **simple text-based listings** on SERPs with titles and brief descriptions. In late 1990s, **Google** introduced a minimalist design with clean and efficient SERP layouts, prioritizing relevance and speed.
2. **Organic and Paid Listing Era (Early 2000s):** Paid search advertising was among the earliest forms of SERP features, with sponsored links appearing at the top and sides of search results. **Google's AdWords**

[26] https://brodieclark.com/notes/
[27] https://bedgarone.github.io/serpevolution/timeline/2020

program, launched in 2000, was a pioneer in this space. Paid Ads got place on SERP along with Text Snippets of 1990s by now started calling as **Organic Results**.

3. **Organic-Paid & Diversified Featured Snippets Era (Late 2000s):** Google introduced Universal Search, integrating different types of content, such as **images, videos, news, and maps**, into the main search results into a single SERP, providing users with a more comprehensive view of information.
4. **Schema and Semantic Snippets Era (Early 2010s): Google Instant** introduced in 2010 first time provided **real-time search suggestions** and **results as users typed their queries** in the search bar, aiming to enhance the **speed of search interactions**. Introduced in 2012, [28]*wordstream-1,* Google's **Knowledge Graph** displays information about people, places, and things directly on the SERP, providing users with **quick answers and context related** to their queries. It contributed to the **understanding of entities and their relationships**, which influenced the display of featured snippets by providing more context. It appears as a box on the right side of the SERP, offering additional context and related entities. 2013 brought two new features. First the **Image packs** display a set of images related to a search query, providing users with visual content directly on the SERP. This feature is commonly seen in searches related to visual topics like **fashion, travel, and design**. Second the **Top Stories** section appears at the top of the SERP, and features **news articles** related to a user's search query, often sourced from authoritative news websites. It provides users with the latest and most relevant news on a given topic. It is a particularly common for current events and trending topics. Then in 2014 Google started **labelling mobile-friendly** websites in search results to encourage mobile optimization. 2014 also saw the entry of **Rich Answers Featured Snippets.** Featured snippets, also known as **answer boxes or position zero**, display concise answers to user queries at the top of the search results providing immediate information without the need to click through to a website. This marked a shift towards providing quick and relevant information to users. Google brought **Local 3-pack** featured block in 2015 that displays three local businesses related to a user's search query, along with their contact information, reviews, and a map snippet. This feature is particularly valuable for local businesses seeking visibility in specific geographic areas. The **People Also Ask** (PAA) feature brought in 2015 itself displays a series of related questions that users might find useful. When a question is clicked, it expands to reveal a brief answer and additional related questions. Another wonderful development in SERP Features was seen in 2015 in the form of **Social Media Integration.** Google started integrating **Tweeter** *(now X)* tweets and other social media posts into search results, providing real-time trending topics, events, updates, and social media content related to user queries. Google expanded the types of featured snippets in 2016 by including **paragraph-style snippets, list-style snippets, and table-style snippets** ***(PLT Snippets Family)***. This variation aimed to present information in the most suitable

[28] https://www.wordstream.com/serp

format for different types of queries. Later by 2017, Local packs became prominent, displaying business listings, reviews, and maps for location-based queries on SERPs.

5. **Schema-Semantic & Contextual Snippets Era (Late 2010s):** This time witnessed the urgence of authenticity of a web content through social signals. Social signals such as **likes, shares, follows, views and comments** were put on top priority by all Search Engines for assessing the authenticity of a web document. Rich snippets with **Reviews and Ratings** began appearing in search results in 2016, offering users additional information about products, businesses, and services. This feature is particularly important for e-commerce and local businesses. Google introduced **Rich Cards** in 2016, a visual format for **mobile search results**, providing a more engaging presentation for specific content types like recipes and movies. To facilitate rich experience with these cards, Google introduced **AMP** *(2016)* and **Mobile-First Indexing (MFI)** *(2018)* the projects aiming to speed up mobile page loading times by creating lightweight versions of web pages. There was another genre of SERP features evolving during this era such as for food-related queries, Google displayed **Recipe Cards** in 2016 that include key information like ingredients, preparation steps, and user reviews directly on the search results page. Then Google came up with a specialized **Job Search** feature in 2017 providing detailed job listings directly within the search interface. **Voice Search** capabilities were integrated into SERPs during 2016-18, allowing users to perform searches using voice commands. Google introduced **Video Carousels** in 2018, allowing users to scroll through a selection of videos related to their query without visiting a separate video platform. Video carousels display a horizontal row of video thumbnails related to the search query. Users can scroll through the carousel to find video content directly on the SERP. Google started displaying **FAQ** snippets in 2019, featuring frequently asked questions and answers directly on the SERP, helping users quickly find relevant information. Continuing focus on Social Signals **Review Snippets** were introduced in 2019, which display user-generated reviews and ratings for products, businesses, or services directly on the SERP. This information helps users make informed decisions without clicking through to a specific website.
6. **Conversational & User Experience Snippets Era (2020s):** Google introduced the ability to display **Web Stories** *(a visual storytelling format)* in featured snippets in 2020, providing a more visually engaging way to present information directly on the SERP. In the same year of 2020, Google introduced multifaceted featured snippets, which could provide answers to broader queries by breaking down the information into different facets. This allowed users to explore diverse aspects of a topic without clicking through to a website. In 2022, after initial blast of Chat-GPT by OPEN-AI; it got adopted in SERP View as a new Search Bar Tab in Google and Bing SERP Interfaces.

There are more than **60 SERP Features** active at present in various SERP views as per the context and intent of a Search Query. The Chart-8 enlists the most widely seen features in diverse SERP views which were introduced time to time from pre-1990s to 2020s.

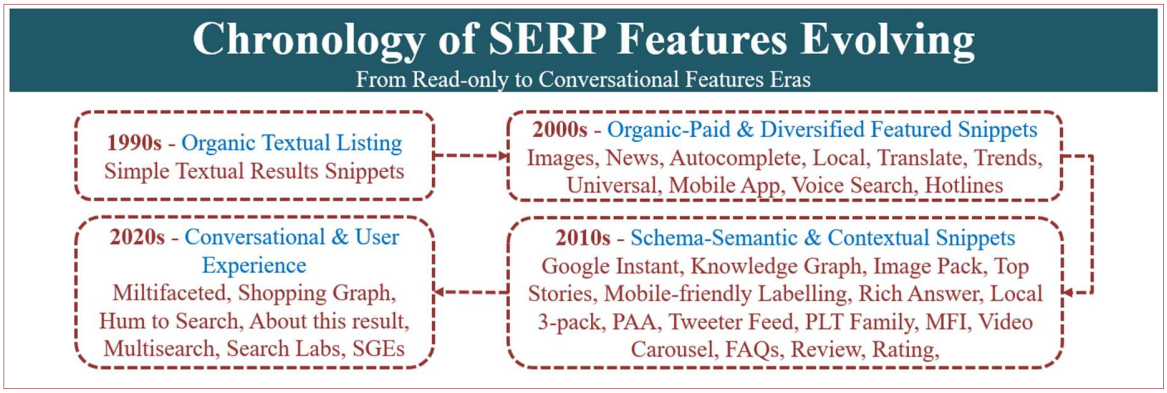


Chart-8: Chronology of SERP Features Evolving

Discussion: Remember these words in title of this chapter **'Information Creation and Presentation'?** Our above review establishes that the basics of philosophies retain their originality in most of cases. What changes with time is the technology and methods of presenting the created information, as here in our case. And so, reflects the dynamism of Search Engines through ongoing continuous refinements, including new SERP features and adjustments to algorithms to enhance further the response of SERPs to match user behaviour and meet their expectations. Tremendous integrations of AI and ML are evolving new capabilities of Information presentation and so is the case for SERP. Another major emphasis of Search Engines is for better user privacy and security, with encrypted search such as Hypertext Transfer Protocol Secure *(HTTPS here onward)* and enhanced privacy settings. Hereby, with this establishment of how vast the theme of this thesis is and now being aware of its so profound legacy and historical background; we have plenty of good reasons and motivations to affirm the choice of the topic of this thesis "**Why the SERP only?".** In the next chapter, we this legacy is linked to philosophical and psychological theories, the integral and complementary gradient around the title of this thesis.

Chapter-2

Psychological Literature Review in context of SERP

2.1. 'Art of phrasing' of Search User and Search Engine:

A two-way neural and semantic understanding: We need to see the way of phrasing through end-to-end process of Information Retrieval, i.e. from a Search Engine by the Search User; it involves the 1) Query Phrasing by Searcher (Search User), 2) Query Interpretation by Search Engine, 3) SERP presented to Searcher in response to their query and 4) grasping of SERP by the Searcher until his/her search requirement is met. Out of these four stages, first and fourth involve the Searchers and second and third involve the Search Engine. This way the art of phrasing is involved on each step of a web search process as shown in Diagram-4 below:

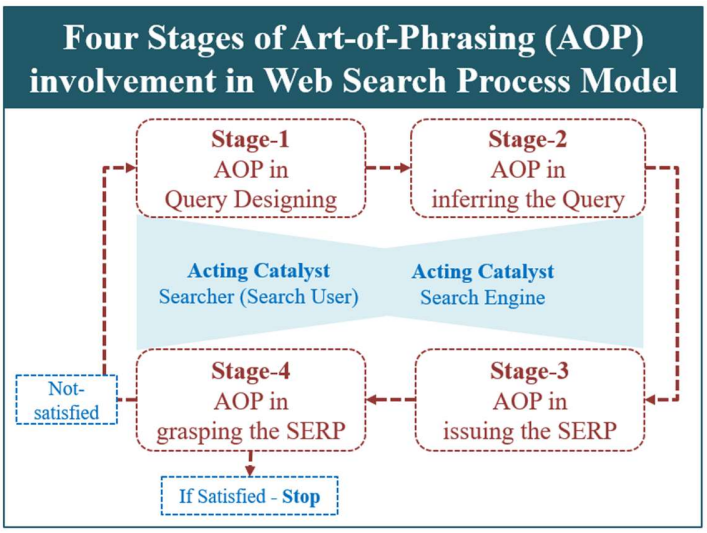


Diagram-4: Web Search Process Model

Morphology, Syntax and Semantics of a Search Query: As any web search journey starts with the very conceiving of a search demand or search idea in the neural system of human brain and that neural system then first tries to connect linguistic NODES in mind and then yields those NODES on web search query box like Google SE. So, let's pick the task from this beginning to first understand how it evolves and then its interpretation and inferring by SEs. In a nutshell, all the SEs theories, technologies, algorithms and interpretation models are designed to decipher the user search terms as formations and expansions of the three linguistic terms i.e. **Words, Phrases** *(a group of words or singular word acting as a grammatical unit)* &

Sentences, [29]*Wiki Phrase,* and their literal natural language meanings are respectively recognized as **Morphology, Syntax and Semantics** as shown in Diagram-5, Appendix-9. Every aspect of this thesis is stitched with either of these three words. Lately, with the start of "voice-search" in information search, the term joins this group is the **phonetics**, shown at the first internal circle of the Diagram-5 itself. However, Natural language utterances are in general highly ambiguous, and a unique interpretation can usually be determined only by considering the constraining influence of the context in which the utterance occurred, *Hirst & St-onge* [1995]. So along with morphology, syntax, and semantics, it is the context of speaking or writing which completes the linguistic sense and purpose of a search term. **This language purpose forms the base to all word game of online searches thus so for this thesis.**

Inferring Linguistic Efficiency of a Search Query by SEs: Linguistic Efficiency here in this context means the gazing the linguistic sense and nuances, semantic matching, cognitive and neural understanding, search intent detection, recall and precision, relevance probability and cultural context behind a search query or say its choice of words and words ordering. For these purposes, Search Users usually use a **single keyword or a phrase or a sentence** for searching information on SEs. In linguistics, **words, and morphemes** *(base and affix)* combine to form larger units such as phrases *(a group of words that have a particular meaning when used together,* [30]*Oxfordlearnerdictionaries* and sentences. This art of combining or arranging words and phrases sensibly is called as **syntax**; the sense which helps in catching the intent of the Searcher behind their search query. **Word phrasing** in search queries is an essential aspect of search engine usage. Word phrasing in search queries is crucial for improving search performance and matching user intent. Search engines employ various techniques such as **query auto-completion, query rewriting**, and inferring user interests to assist users in formulating their queries and retrieving relevant information. Many queries posed to search engines involve phrases, *Williams et al.* [2004]. Phrase queries are a key functionality of modern search engines, *Anand et al.* [2014]. Users typically input a string of words or concepts as their search query, but the results provided by search engines may not necessarily match the user's intent, *Purnamasari et al.* [2016]. To improve search performance, it has been observed that queries based on a dependence model, which include phrases, lead to better results compared to bag-of-words queries, *Kulkarni & Callan* [2010]. Search engines employ various techniques to assist users in formulating their queries. One common technique is **query auto-completion**, which suggests relevant queries based on the user's input, *Tahery & Farzi* [2020]. **Query suggestions** play an important role in modern information retrieval systems, accounting for roughly 30% of all queries, *Bhatia et al.* [2011]. Another approach is to **extract frequent phrases from the best documents** and add them back into the

[29] https://en.wikipedia.org/wiki/Phrase
[30] https://www.oxfordlearnersdictionaries.com/definition/english/phrase_1

search query to augment the existing system, *Ide et al.* [2007]. Additionally, search systems can **infer user interests by observing their interaction** and offer suggestions for query terms or reorganize search results to enhance exploration, *White & Ruthven* [2006]. The use of phrases in queries is a standard capability in most contemporary text retrieval systems, *Bradford R.* [2014]. However, one challenge in today's e-commerce search is the **semantic matching** problem, where relevant items may not contain the exact terms in the user's query, and to address this, **query rewriting techniques** have been proposed to translate the user's query into a more suitable form for retrieval, *Qiu et al.* [2021]. Furthermore, **detecting multilingual and multi-regional intent** in search queries is a critical challenge for serving international users with diverse language and region requirements, *Chang et al.* [2011]. The use of phrases in queries is a standard capability in modern text retrieval systems, and there is ongoing research to address challenges such as **semantic matching and multilingual intent detection**. *Baeza-Yates* [2017] states that Search Engines use **semantic query understanding** to predict intentions behind ambiguous queries, using language prediction, parsing, entity extraction, and semantic ranking techniques. As per *Joseph Sam* [2002] **NeuroGrid** offers an alternative to hierarchical, folder-based file organization and offers distributed search through semantic routing and dynamic adjustment of meta-data. The subcomponents of reading **prosody** *(pattern of stress and intonation in a language)* have been defined to include one or more of the **expressions** *(e.g., pitch and tone),* **syntax** *(e.g., grammatical cues as idea units)*, and **phrasing**/smoothness *(e.g., word grouping), Shhub et al.* [2023], *Hudson et al.* [2008], *Kuhn M. R.* [2005] and *Kuhn et al.* [2018]. *Vinod Sharma* [2023] states that **G-AI** can associate keywords and concepts that are relevant to the researcher's field of study. By exploring these associations, researchers can uncover new connections and potential research angles, aiding in the formulation of more focused questions. Semantic Web uses **related G-AI method** in performing associations and connections. The use of **neural understanding** in word phrasing search queries is an emerging area of research. While there are no specific references available on this topic, there is a related reference that explores the use of brain-inspired techniques in search engine assistants based on **knowledge graphs,** *Zhao et al.* [2023]. This reference discusses the application of knowledge graphs and **brain-inspired algorithms** to enhance search engine assistants, which could potentially have implications for word phrasing in search queries. So, search engines use **neural understanding in interpreting search queries** by rewriting them using neural networks, analysing search behaviour and user profiles, and finding specific interpretations for ambiguous keywords. As per *Wang et al.* [2017] Search engines use **neural understanding** to find specific interpretations for ambiguous keyword queries, reducing the search space and providing more relevant results. *Nogueira & Cho* [2017] states in their research that neural network-based query reformulation system improves document recall by 5-20% compared to strong baselines, with room for further improvements. Diversified top-k keyword query interpretation on knowledge graphs can help users discover meaningful patterns and reduce search space, leading to more relevant results. *Leung et al.*

[2008] states that SEs use neuro understanding by extracting concepts from web-snippets and using them to identify related queries, providing personalized query suggestions with better **precision and recall**. According to *Raghavan et al.* [1989] **recall and precision** are often used to evaluate the effectiveness of information retrieval systems. They are easy to define if there is a single query and if the retrieval result generated for the query, is a linear ordering. **Relevance probability**, mentioned in the literature expected precision, and so forth, are some for this purpose. Furthermore, and the retrieval results when many queries are to be averaged over these queries, some method of interpolation of values at certain preselected recall levels is needed. The currently popular approaches for **handling both a weak ordering and interpolation** are found to be inconsistent, are not easy to interpret. *Khan et al.* [2015] find in their research that Search engines use Fuzzy Neural Networks ***(FNN,** here onward)* for query information filtering, categorizing user requests into categories and sub-categories for more relevant results. *Prakash & Patel* [2015] states that Search engines use query understanding to correct user errors, guide users in query formulation, and infer the precise intent of the user's query. In general, **understanding the cognitive processes** involved in query formulation and information retrieval is crucial for improving search engine performance. Estimating the **difficulty of queries** is an important aspect of information retrieval, *Carmel and Yom-Tov* [2010]. By quantifying the quality of search results retrieved for a query, search engines can better serve customer needs and reduce variance in performance. This understanding can be used to optimize word phrasing in search queries and improve the relevance of search results. Additionally, the **interpretation of search queries** can vary between **strict and vague interpretations**, *Trotman & Lalmas* [2006]. By extending query languages and incorporating different interpretations, search engines can enhance their performance and better **understand user intent.** This understanding can be leveraged to improve word phrasing in search queries and provide more accurate and relevant results. It is also worth noting that most evaluation efforts in the field of search engines have focused on English queries and document collections, *Lazarinis F.* [2007]. However, as search engines become more global, they consider the impact of **different languages and cultural contexts on word phrasing** in search queries. Understanding the **nuances of different languages** and incorporating neuro understanding can help improve the accuracy and effectiveness of search engine results.

Framing Neuro-physiological evidences between Query - SERP - Searcher *(The role of Neural Matching in generating a SERP)*: Art of phrasing is not only used in issuing a search query but in inferring the SERP obtained because of that search query by same query issuer as SERP user/interpreter. So as a next step of the process and taking the cue from query inferring playground elaborated above, a similar game of NEURAL NODES MAPPING begins by each SE from very typing of a Search Query to infer the query first, setting linguistic relevancies as talked above and then extracting the most coherent to query SERP. As evident, for this

purpose Neuro Understanding plays the crucial role in generating SERPs by providing insights into not only the Search Query but alongside searcher behaviour and improving search effectiveness. *We shall be taking in more depth on Searcher Behaviour and Search Effectiveness in later sections of this thesis.* Click models based on Probability Graphical Models ***(PGMs**, here onward)* have been used to **explain and predict user clicks** on search results, *Xu et al.* [2023]. However, the structure of dependencies in PGMs needs to be manually set. An alternative approach to use is **distributed representations**, where user behaviour is modelled as a sequence of vector states associated with a query session, *Borisov et al.* [2016]. This approach allows for a direct understanding of user browsing behaviour without predefined rules. Additionally, **neuro-physiological evidence** generated through modalities such as **Eye-tracking,** Electroencephalogram ***(EEG**, here onward),* and Functional Magnetic Resonance Imaging ***(fMRI**, here onward)* can provide further insights into search behaviour, *Mostafa & Gwizdka* [2016]. Eye tracking studies have revealed that the sequence of fixations, known as the scanpath, provides insights into the user experience with SERPs, aiding in the design of user-friendly layouts, *Lorigo et al.* [2008]. The analysis of eye movements during browsing multiple search results pages is crucial in understanding user activities and behaviour during web searches, *Matsuda et al.* [2009].

Alsaffar et al. [2018] in their study on Visual Search (Eye-tracking) Behaviour amongst Arabic, English and Chinese Search Engine Users represent the aggregated fixation duration from all Arabic, English and Chinese participants in the two PC SERP presentations *(Plain PC SERP Chart-10, Appendix-9 and Rich PC SERP - Chart-9, Appendix-9)*. Their study shows that Arab fixated most often on the beginning of each sentence, English fixated most often on the middle and Chinese scattered their fixations around the page.

By analysing user browsing behaviour and incorporating neuro understanding, search engines can improve the relevance and ranking of search results, *Alnatsheh* [2020] and *Shu et al.* [2020]. *Serrano & Gelenbe* [2018] proposed an Intelligent Search Assistant ***(ISA**, here onward)* that uses a Random Neural Network ***(RnNN**, here onward)* to **record, reorder and present results based on perceived relevance to user needs**, outperforming commercial search tools. Their ISA adapts and learns from **user previous relevance measurements** *(based on Gradient Descent for quality and Reinforcement Learning for higher learning rate)* significantly increasing its quality and improvement within the first iteration. *Ye et al.* [2022] utilized brain signals to estimate the usefulness of non-click results, improving performance and enhancing human-computer interactions. *Iraji et al.* [2015] developed a system for web page retrieval using an adaptive Neuro Fuzzy System ***(NFS**, here onward)* increasing search engine efficiency. *Chen & Décary* [2018] introduced a cognitive-based semantic approach, combining rule-based NLP, a world model, and cognitive frames to improve the relevance, accuracy, and efficiency of information search. Understanding the underlying intent of search queries and user behaviours is

essential in comprehending search engine traffic, *Cordova & Jonsen* [2012]. Furthermore, the impact of search engines on the discovery of new web pages is a significant aspect that influences the composition of SERPs, *Cho & Roy* [2004]. It has been observed that users tend to click on the items listed at the **top of the result list**, indicating a **bias in their use of search engines**, *Keane et al.* [2008]. Moreover, the quality of search engine results, including sponsored and non-sponsored links, significantly affects users' perceptions during web searching, *Jansen & Resnick* [2006]. Neuroscience and immunology research have highlighted the cross **talk between the nervous system and the peripheral immune system**, emphasizing the role of neurotrophic factors and neuroactive cytokines in this interaction, *Sanders & Jones* [2006]. So, neuro understanding is established usually through:

1. **Neural Networks:** NNs are a type of ML model inspired by the structure and functioning of the human brain. In the context of search engines, neural networks, particularly DL Models, have been increasingly used for various tasks, including **NLP and Image Recognition**. They can be employed to enhance language understanding, sentiment analysis, and even the interpretation of complex search queries. **Google's BERT** is an example of a neural network architecture that has been applied to better understand the context and nuances of words in a search query. BERT aims to improve the understanding of the relationship between words and provide more relevant search results.
2. **User Intent Understanding:** Understanding user intent is a critical aspect of generating relevant SERPs. So, search engines strive to comprehend not only the explicit words in a query but also the underlying intent or purpose behind the search. To enhance user intent understanding, search engines utilize various techniques, including **NLP, MLAs, and Pattern Recognition**. These technologies help in deciphering the context of a query, allowing search engines to deliver more precise and contextually relevant results.

Learning of Search Engines from Searcher' Cognitive Feedback *(Searchers Cognitive Feedback contributes to the ongoing evolution of SE Algorithms)***:** The relationship between SERPs and human cognition is complex and multifaceted. Human cognition refers to the mental processes and activities related to acquiring, processing, storing, and using information. The relationship between SERPs and human cognition is intertwined with how searchers perceive, process, and act upon information presented on SERPs. Designing SERPs that align with cognitive processes can enhance the user experience and the effectiveness of information retrieval. These alignments are set throw following methods:

1. **Visually engageable SERP reduces the Cognitive load of searcher:** Cognitive load assessment during web searches helps characterize search tasks and system features based on their demands on the searcher's mental effort, *Xin-Fu* [2010]. SERPs aim to minimize cognitive load by presenting information in a structured manner. Searchers tend to prefer results that are easy to understand and require less cognitive

effort to process. Humans process information *(on SERPs)* through visual perception, quickly scanning titles and snippets to identify relevant content. Thus, SERPs are designed to be visually engaging and easily scannable tending to reduce cognitive load of search users.

2. **Cognitive assessment leads to make relevance judgment of Search Results:** As said in above point search users make rapid judgments about the relevance of search results so they often engage in decision-making processes on SERPs, evaluating which link to click based on perceived relevance, authority, and trustworthiness. Studies have shown that cognitive feedback, such as implicit feedback from browsing behaviours, can accurately predict the relevance of search results, *Yuan et al.* [2020]. Additionally, utilizing cognitive feedback techniques can assist searchers in formulating effective search strategies by capturing behavioural evidence from web searches, *Alnatsheh* [2020].
3. **Searcher Cognitive Ability match with SE's Semantic Understanding:** Search engines use algorithms to understand the semantic meaning of queries and deliver relevant results i.e. through Semantic Understanding. Searchers, in turn, rely on their cognitive abilities to articulate queries that reflect the best of their intent accuracy. Search engines like Seenso.com utilize cognitive frames and a world model to enhance semantic analysis and improve search relevance, *Chen & Décary* [2018]. Semantic search engines, unlike syntactic ones, offer more meaningful and specific results by comprehending queries effectively, *Mana* [2014]. To address the limitations of conventional keyword-based search engines, newer approaches incorporate natural language understanding, semantic analysis, and deductive reasoning to provide accurate answers to user queries, *Zhao* [2008] and *Sembok et al.* [2012]. By employing rule-based NLP, lexical functions, and Meaning Text Theory, search engines can transform and normalize information for better semantic matching, ultimately enhancing the cognitive alignment between searchers and search engines, *Santha & Jayakumar* [2019].
4. **Visual elements or concise information draw searcher attention and focus:** Various **SERP features**, such as featured snippets, knowledge panels, and rich media, are designed to capture users' attention and influence their cognitive focus. *Ahn et al.* [2018] state in their work on "attention variations while surfing SERPs" in regard of attention capturing of search users that the larger the number of alternatives in a SERP, the greater is the attention placed on the top-ranked alternatives by users, they name it as **Attention Adjustment Effect**. In another hypothesis, they assert when consumers perceptually encounter a new category of alternatives due to exposure to different stimuli, they are inclined to renew their attention, naming it as **Attention Renewal Effect**. Attention decay occurs more rapidly after attention renewal than before, known as **Attention Equilibrium Effect**.
5. **Search Engines create Learning and Adaptation Loops from Searcher Feedback:** While Searchers learn from the search experience, adapting their search behaviour based on the relevance of results, same

time the Search engines use their feedback and behaviour to improve algorithms, creating a dynamic learning Feedback Loops. Search engines utilize learning and adaptation loops based on feedback from searchers to enhance search quality, *Dziak et al. GP* [2015]. Various methodologies like supervised learning and relevance feedback are employed to improve search algorithms, *Rahman & Seyal* [2014] and *Lv & Zhai* [2009]. Supervised learning aids in understanding algorithm flows in controlled settings, preventing information overload, *Joachims & Radlinski* [2007]. Relevance feedback optimizes the balance between original queries and feedback information, enhancing retrieval accuracy, *Bendersky et al.* [2018].

6. **Searcher Memory and Recalling previous results choices:** Searchers use their tendency of relying on their memory and recall abilities to remember previous search results or recognize patterns in SERP elements. Visually distinctive **SERP elements** may enhance the recall. *Jonsson & Stahl* [2015] state that previous experience and the trustworthiness of a result, strongly affected if a user would choose to click on that result. Understanding the complexities of cognitive feedback and its impact on search behaviour can lead to improved search performance and learning outcomes, challenging the notion that web searchers have simple information needs, *Gwizdka* [2009].
7. **Cognitive Trust gained through previous evaluations of Search Results:** Cognitive Trust in search results is shaped by a combination of technological trust, cognitive biases, and memory processes, highlighting the complexity of evaluating and trusting search outcomes. *Peterson et al.* [2022] studies show that trust in technology corresponds to greater perceptions of search result quality, with belief in Google's algorithm playing a significant role. Searchers assess the trustworthiness and authority of search results based on cognitive evaluations of source credibility, relevance, and perceived expertise. **But here is a thin line of differentiation between this trust and biasing generated by search habit developed through intakes of unbroken use of a Search Engine should also be interpreted.** Below Chart-11 displays the sub-variable driven from above findings around Variable 1 of this thesis:

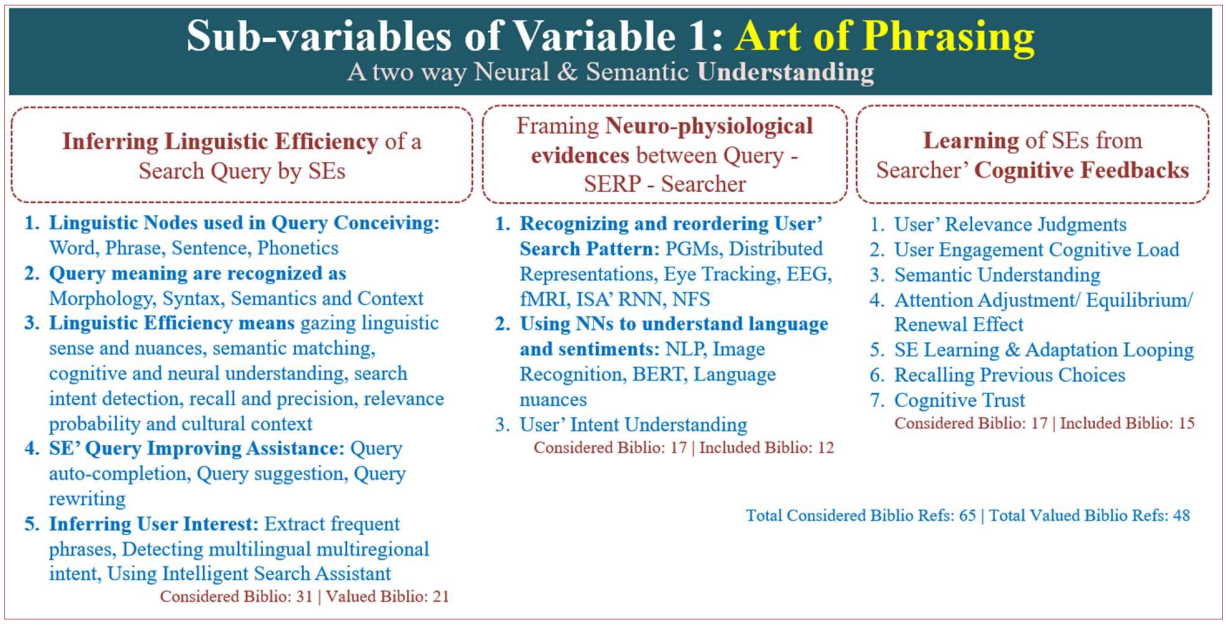


Chart-11: Sub-variables of Variable 1 'Art-of-Phrasing'

Discussion: A potential foundation for hypothesizing emerges from the mutual learning dynamic between search users and search engines, despite varying levels of linguistic proficiency on both sides. Search engines evolve through user interactions, refining SERP presentations based on behavioural feedback, while users enhance their search expertise through repeated exposure. This interplay commences with the user's query formulation and culminates in SERP generation, during which the search engine's self-improvement mechanisms operate concurrently to deliver increasingly personalized results. However, a notable vulnerability arises from the engine's reliance on user behaviour and feedback loops, which introduces the **risk of bias** even in top-ranked results. Consequently, search engines subtly shape users' cognitive frameworks - often without their explicit awareness, except among professionals. As societal norms evolve into ingrained habits, adherence to these norms becomes instinctive for the majority; over the past two decades, search engines have become such a ubiquitous societal fixture, providing abundant opportunities to influence cognitive patterns. An intriguing avenue for exploration would be quantifying the proportional learning exchange: the percentage ratios of knowledge transfer from engine to user versus user to engine.

Discussion Scrutiny: Although the literature lacks direct references to the application of neuroscientific insights in phrasing search queries for search engines, related studies on brain-inspired search assistants and query difficulty estimation underscore the value of cognitive processes in enhancing search performance through optimized phrasing. As examined in the preceding findings, SERP construction is shaped by user behaviour, query intent, and the engine's web page discovery capabilities. Addressing these elements

necessitates interdisciplinary collaboration among neuroscience, computer science, and information science to foster more intuitive and efficient search experiences. Collectively, these investigations illuminate the promise of **neuroscientific understanding** in SERP optimization. A key dimension involves neural network adaptability, which can be modulated by user inputs to refine keyword processing; concepts like **Neural Networks and User Intent Understanding** provide a robust basis for integrating neuroscientific principles into SERP generation. While not identical to "neuro understanding," these approaches leverage neural network analogies and human cognition to advance SERP mechanisms. Moreover, the research informs decisions on phrase selection and sequencing in queries. Users' neuroscientific comprehension and experientially honed cognitive abilities significantly contribute to their "art of phrasing," extending beyond mere lexical proficiency. Thus, **Lingual Command** - the second variable in this thesis—interacts with neuro understanding and cognitive capabilities to underpin query formulation across languages and geographic contexts on any search engine. These three elements serve as pivotal instruments in shaping the Semantics of Web Information Search. Subsequent sections will delve deeper into Lingual Command and Human Cognitive Capabilities as the second and third variables, respectively, of this thesis. Keep these learnings and their importance in purview, a special **Psycho-Cognito Profiling** of Test Participants in our research study is done in Chapter-9.

2.2. Natural Language Ontology *(Search Query and User Intent inferring by Search Engines)*:

Continuing the discussion from previous section, where we concluded the three factors behind the “Art of Phrasing”, here we shall be learning about one of them “Inferring Linguistic Efficiency” in details. This topic is such a vast and impactful one that we have made it the second State of Art (SOA) variable of this thesis. It is divided into three sub-sections as Decoding, Inferring and Creation of Search and SERP Features as follows:

Decoding Search Query using NLP Components *(Role of NLP in improving overall SERP quality)*: The most prominent example of a SERP Feature formed based on decoding of Search Query and Intent is Knowledge Graph *(third sub-section of this section talks about it in detail)*. Similarly, NLP is behind almost every SERP Feature in rendering its relevance quality. NLP enables SEs to understand and interpret user queries, leading to more accurate and relevant search results. Here are some keyways in which SEs use NLP to enhance search results and features:

1. **12 NLP components for interpretation of Search Queries:** NLP through its 12 components *(Tokenization, Word type labelling, Word dependencies, Lemmatization, Parsing labels, Named entity*

analysis and extraction, Salience scoring, Sentiment analysis, Text categorization, Text classification and function, Content type extraction, Identify implicit meaning based on structure) allows SEs to understand the meaning of words, sentences, and texts in user queries, enabling them to provide more relevant results, [31]*Olaf Kopp* [2022]-2.

2. **Classification of Subject and Purpose of Documents:** NLP helps search engines classify the subject and purpose of documents, which aids in matching search queries to relevant content, [31]*Olaf Kopp* [2022]-2.
3. **Entity Analysis in Documents and Search Queries:** NLP is used for entity analysis in documents, search queries, and social media posts, which helps SEs generate featured snippets and answers in voice search, [31]*Olaf Kopp* [2022]-2.
4. **Semantic Search:** NLP enables semantic search, which allows SEs to understand the intent behind complex, multi-part queries and provide more context-based results. Natural Language Search ***(NLS**, here onward) is* the search carried out in everyday language, phrasing questions as you would ask them if you were talking to someone. These queries can be typed into a search engine, spoken aloud with **voice search**, or posed as a question to a **digital assistant** like Speech Interpretation and Recognition Interface ***(SIRI,** here onward)* or Cortana, [32]*Rebecca Sentance* [2016].
5. **Functioning Mechanism of NLP:** As per [33]*Marketbrew-2* NLP is a field of AI that focuses on the interaction between computers and human language. NLP is commonly used in named entity recognition, where SEs use the technology to extract and disambiguate entities that are tied to a knowledge graph. Another technology that fits into NLP is the use of **LLMs**. These models use **deep learning algorithms** to process and understand large amounts of text data, allowing for more accurate and sophisticated language processing. One example of an LLM is GPT-3.

Inferring Search Intent using Natural Languages Understanding *(Adjusting the understanding between User and SEs)***:** Understanding as a psychology has been very challenging journey to define up to exact definition term. Earliest known remarkable work by *A. M. Turing* [1950] tries to define it as a state; while recently *Blaha et al.* [2021] argue it as a process, not an outcome. **Understanding is an ongoing cognitive activity** of acquiring, integrating, and expressing knowledge according to the task or situation at hand. Skimming through different research works and models proposed through several breakdowns in this span of 70 years propels a chronical sophistication delivered to the fields of Cognitive Science, Intelligence Behaviour, Computational Intelligence, Constructive Nature of Conversation, Artificial Intelligence, Machine Learning,

[31] https://www.kopp-online-marketing.com/search-query-processing
[32] https://www.searchenginewatch.com/2016/04/12/everything-you-need-to-know-about-natural-language-search/?amp=1
[33] https://marketbrew.ai/natural-language-processing-and-its-role-in-seo-and-search-engines

Linguistics, Verbal Protocols, Deep Learning, Natural Languages and Explainable Artificial Intelligence *(XAI)* significantly. *Blaha et al.* [2021] further state that understanding is comprised of mechanisms, representations, learning and reasoning. They define it as ability to probe and therefore demonstrate any degree of understanding requires natural languages. Here our research argues that as searching on a Search Engine is a one-to-one interaction between human and machine, so setting up a common and coherent understanding ground between the two with maximum of optimized common factors is what has been a challenge for ITciens and Softiciens. And here to gauze this gap requires the decoding of the query sense and meaning. It is usually achieved by forming scalable algorithms by tiding the nodes between both human expectation and machine response through **natural languages as a tiding thread**. *Michael Hannon* [2021] argues about articulation of **evidence for understanding**. He states that understanding is a cognitive achievement, not gained simply by receiving information; it comes in degrees; understanding manifests itself through abilities or know-how, especially being able to 'grasp' connections. *Hough & Gluck* [2019] recently defined understanding as "The acquisition, organization, and appropriate use of knowledge to produce a response directed toward a goal, when that action is taken with awareness of its perceived purpose". This is perhaps an updated, more general version of Simon's early definitions developed in his efforts to outline the criteria for software programs capable of understanding. Simon emphasized that understanding is "a relation among a system, one or more bodies of knowledge, and a set of tasks the system is expected to perform", *Herbert Simon* [1977] (p. 1070). Simon's incorporation of the task or goal for an intelligent system is an extension of *Moore & Newell* [1974] definition of understanding as a relationship between a system and its appropriate use of knowledge. So, setting the understanding between a search user and a search engine machine involves assessing the **effectiveness of communication and information retrieval** in the search process. Understanding between a searcher and a search engine is an ongoing process of refinement and adaptation, which can be understood and measured using different Click Models such as *Random Click Model, Position-Based Model, Cascade Model, User Browsing Model, Dependent Click Model, Click Chain Model and Dynamic Bayesian Network Model*, *Chuklin et al.-1* [2015]. As technology evolves, search engines strive to better understand and anticipate user needs, creating a more seamless and effective user experience. Thus, the connection between natural language understanding ***(NLU,** here onward)* and SERP algorithms involves several key components. SERP algorithms play a crucial role in determining the ranking and presentation of search results, and the integration of NLP is fundamental to enhancing their effectiveness. Here's how the Search Engines set understanding between search user expectations and the results to presented in a SERP in response to a search query:

1. **Query Expansion, Interpretation and Relevance Understanding:** SERP algorithms use NLP to **expand user queries** by incorporating synonyms, related terms, and contextually relevant variations. This helps in broadening the scope of search results and capturing a more comprehensive range of relevant content. It

involves the evaluation of how well the search engine interprets and understands the user's search query by recognizing user intent, context, and any implicit information within the query. So by analysing the structure and semantics of the query, the search engine can determine the user's underlying goal and provide results that best match that intent. A well-understood query is that assessing its relevance to that of search results directly address the user's information needs, considering both semantic relevance and context. According to *Carpineto & Romano* [2012] the relative ineffectiveness of information retrieval systems is largely caused by the inaccuracy with which a query formed by a few keywords models the actual user information need. One well known method to overcome this limitation is automatic query expansion ***(AQE,** here onward),* whereby the user's original query is augmented by new features with a similar meaning. The advance of AQE techniques has been confirmed by a number of experimental tests on classical benchmarks.

2. **Search Intent catching through Semantic Understanding:** It helps the search engine deliver results that align with the user's actual intent rather than just the literal interpretation of the query by going beyond and around the comprehensive meanings of words and phrases in it. This is measured as an ability of Search Engine semantic understanding. Instead of relying solely on keyword matching, semantic search considers the meaning and context of words, enabling a more nuanced understanding of user intent. SERP algorithms leverage NLU to perform semantic search. Semantic search considers the meaning and context of words, enabling the search engine to understand the relationships between concepts and deliver more contextually relevant results. Various approaches have been proposed to enhance query understanding, such as leveraging detailed intent descriptions for queries, *Mittal & Agarwal* [2019], and automating intent discovery through clustering techniques on unlabelled data sets, *Zhang et al.* [2020]. Additionally, the use of domain-specific embeddings extracted from search logs has shown significant improvements in Named Entity Recognition ***(NER**, here onward)* tasks, surpassing traditional CRF-based parsers *(Conditional Random Field)*, *Zhang et al.* [2020]. Furthermore, learning semantic query annotations tailored to individual query intents has proven to enhance intent-specific approaches, outperforming intent-agnostic methods based on Markov random fields and learning to rank, *Padmasundari & Bangalore* [2018]. These diverse methodologies collectively contribute to advancing semantic understanding for capturing search intent accurately.
3. **Recognising the Entity behind a Web Search Query:** NLU techniques, including **NER, entity disambiguation, and query interpretation** help identify entities *(e.g., people, places, organizations)* mentioned in search queries. This information is used to refine search results and provide more accurate answers, particularly for queries related to specific entities. NER systems like deep neural models extract entity features from word embeddings learned from search logs to improve performance, *Mittal & Agarwal*

[2019]. Entity disambiguation methods utilize deep neural networks to capture textual information in queries and generate weakly supervised training data from web search results, enhancing performance significantly, *Dandekar et al. GP* [2021]. Query interpretation approaches focus on linking query segments to entities in a knowledge base, improving accuracy and efficiency in understanding ambiguous web search queries, *Kasturia et al.* [2021]. These methods collectively contribute to better understanding the semantics of search queries and identifying the entities users are searching for in web searches.

4. **User Behavioural Feedback Analysis in Web Search:** Positive user feedback indicates effective understanding and relevant results. User behavioural feedback analysis in web search involves leveraging various metrics such as reading dwell time, clicks on links, CTR *(Click-through-rate),* mouse movements, bookmarking, sharing, and more to predict the relevance of search results and improve search effectiveness, *Kaushik & Jones* [2021]. Incorporating user behaviour data can significantly enhance the ranking of search results, leading to better user experience, *Alnatsheh* [2020]. Furthermore, understanding user emotions and behaviours on websites can help tailor personalized experiences, enhance user satisfaction, and improve service offerings, *Chimienti et al.* [2022]. Evaluating search systems based on user behaviour modelling is crucial for assessing user satisfaction and system performance accurately, moving beyond traditional batch evaluations, *Eugene et al.* [2019]. By utilizing machine learning, recommender systems, and emotion recognition algorithms, interactive systems can be developed to provide personalized experiences and maximize user satisfaction, *Zhang-1 et al.* [2021].
5. **SERP Personalization through Contextual Understanding:** Having contextual awareness and understanding enhance the personalization of search results to better meet the individual needs of the searcher. This involves considering factors such as user location, device type, and recent search history to tailor results to the specific needs and preferences of the user. NLU allows search engines to grasp the context of a search query. Personalizing SERPs through contextual understanding is crucial for enhancing user experience, *Kota et al.* [2021]. Leveraging user search behaviour and semantic knowledge from domain ontologies, a method was proposed to build contextual user profiles and re-rank search results for personalized web search, *Chebil et al.* [2020]. Similarly, in the job-seeking domain, utilizing job-sessions data, Bidirectional Encoder Representations from Transformers *(**BERT**, here onward)* models were fine-tuned to predict entities in extreme multi-label classification tasks, leading to improved job-search query recommendations and job notifications, *Zhang et al.* [2021]. Moreover, the inductive contextual personalization framework was introduced to infer relations between existing and unseen entities, outperforming baseline methods for tasks like relevant author search and new item recommendation, *Shen et al.* [2021]. These approaches highlight the significance of context-driven personalization in enhancing search experiences across different domains.

6. **Multimodal Search Capabilities:** A comprehensive understanding of multimodal queries contributes to a more inclusive and versatile search experience. It is achieved by exploring whether the search engine can effectively handle different modes of input, including text, voice, images, and videos. Multimodal search engines leverage various data types like text, images, and videos for comprehensive information retrieval, *Solaiman & Bhargava* [2023], *Tomar & Bharti* [2023], *Tautkute et al.* [2019], *Pinho et al.* [2016] & *Shaowei Wang* [2015]. These engines employ deep neural networks, convolutional neural networks, and transformer models to extract features from different modalities efficiently. They also integrate contextual dependencies between modalities using joint neural network architectures, enhancing search accuracy. Additionally, multimodal search engines cater to diverse domains like healthcare, fashion, and software development, offering tailored solutions for medical image retrieval, fashion merchandise search, and code search tasks. By combining multiple input forms such as free-form texts, domain-specific languages, code examples, and execution traces, these engines provide users with convenient and effective search capabilities, significantly improving search accuracy over traditional methods.

NLU and Natural Language Generation (NLG) are interconnected. By integrating NLU and NLG, a shared latent variable can facilitate information exchange between natural language and formal representations, enhancing search result quality. **NLG enables SERP algorithms to understand and process queries in multiple languages**, a crucial aspect for delivering relevant results to users. SEs often update their algorithms to incorporate advancements in NLU, ML and other AI technologies. These updates aim to improve the precision and relevance of search results by better understanding the nuances of natural language.

Creation of SERP Features using NLP Ontologies *(Understanding and matching the queries to relevant content)*: As per [34]*Ontology Wikipedia* definition, "in computer science and artificial intelligence, ontology languages are formal languages used to construct ontologies. They allow the encoding of knowledge about specific domains and often include reasoning rules that support the processing of that knowledge". Research has consistently shown that linguistic expertise significantly enhances search engine results. Since 1996 there have been two methods of searching on search engines; the Keyword Based Search ***(KBS**, here onward)* and Natural Language Search ***(NLS**, here onward)*. Till late 2010s, KBS was outperforming the NLS but along with the advancement of Machine Learning in SERP in early 2010s, NLS is highly preferred by Generation-Z. I have noticed it from my personal experience in my family as well. My family kids use the smartphones for voice searches since 2014. Alongside this *Romanov et al.* [2003] states that the **Linguistic Automaton** *(LINGTON)* system, based on behaviour-based artificial intelligence and machine translation concepts, offers a

[34] https://en.wikipedia.org/wiki/Ontology_language

new approach to NLP and machine translation in the 21st century. *Osipov et al.* [2006]; *Osipov et al.* [2008] and *Krstev et al.* [2008] all emphasize the importance of linguistic knowledge in improving search relevance, precision, and performance. Beside the requirement of specific language mastery, *Marlow et al.* [2008] underscores the impact of language skills for multilingual web searches, highlighting the need to consider the user's proficiency in different languages. These observations and studies collectively demonstrate the critical role of linguistic expertise in enhancing the effectiveness of SERP. Alongside Search Engines, the linguistic expertise of a searcher contributes to better search engine results by enabling search engines to understand and interpret natural language queries more effectively. Search engines like Google use **NLP** for this purpose, which allows them to interpret the meaning of words, sentences, and texts to generate information and knowledge, [35]*Olaf Kopp* [2022]-1. Olaf further adds that NLP enables search engines to interpret search queries, classify the subject and purpose of documents, analyse entities in documents and search queries, and generate featured snippets and answers in voice search. Google's AI models, such as **RankBrain, BERT and MUM** *(Multitask Unified Model)* work together to **understand queries** and **match them to relevant content**, ensuring that important words aren't left out of a query and that the search results are ranked according to relevance, [36]*Matt G. Southern [*2022]. Google' **Hummingbird** was designed to understand the **searcher's intent** and **RankBrain** to **guess the meaning** of the never-before-seen queries leading way towards the semantic search. **Semantic search** denotes search with meaning, as distinguished from lexical search where the search engine looks for literal matches of the query words or variants of them, without understanding the overall meaning of the query, *Hannah et al.* [2016]. Additionally, and as said above **NLS**, which is carried out in everyday language now, allows users to phrase questions as they would ask them in a conversation, and this type of search is key to advancements in technology, including voice search, digital assistants, and smart home hubs, [32]*Rebecca Sentance* [2016]. These queries can be typed into a search engine, spoken aloud with voice search, or posed as a question to a digital assistant like **SIRI or Cortana** *(taken off from market now)*. Below Table-6 explains the difference between **KBS and NLS**:

	Natural Language Search - NLS	**Keyword Based Search - KBS**
Search Platform	SIRI, Cortana, Alexa, Google Assistant	Google, Bing, Yahoo, Baidu, Yandex
Example	Empire State Building Height	How high is the Empire State Building?

Table-6 - Difference between KBS and NLS

[35] https://searchengineland.com/how-google-uses-nlp-to-better-understand-search-queries-content-387340
[36] https://www.searchenginejournal.com/how-google-search-understands-human-language/436565/#close

So here we see that **NLS** is the next generation search model over the **traditional KBS** model, but both are running in user in parallel to each other.

Important Ontological Algorithms *(Query purpose understanding)***:** According to the [37]*Pandu Nayak* [2022] there is another important algorithm used in deciphering a search query called as **Neural Matching**. It is used to match and extract the content of the query. In the Table-7 below are the four AI Search Algorithms put in a comparative study by Pandu:

	RankBrain	**Neural Matching**	**BERT**	**MUM**
Tagline	A smarter **ranking** system *(Google's First AI System)*	A sophisticated **retrieval** engine	A model for understanding meaning and context i.e. **ideation & relevance**	Moving from language to **information understanding**
Launch in	2015	2018	2019	May 2021
AI Model	Deep Learning	Neural Networks	Natural Language Understanding	T5: Text-To-Text Framework
Utility	In understanding and **matching** words to their related concepts. **Rank** or decide the best order for top search results.	Understand fuzzier representations of concepts in queries and pages and match them to one another. It can **decipher** the purpose of the searcher like they look for management tips, leadership tips, personality or more based on popular, colour-based personality guide.	In understanding how combinations of words express different meanings and intents. It comprehends how a combination of words expresses a complex **idea**.	Understanding and generating language. A more intuitive way to search. Understands information across text and images and, in the future, can expand to more modalities like video and audio, [38]*Nayak google.com* [2021].

[37] https://blog.google/products/search/how-ai-powers-great-search-results/
[38] https://blog.google/products/search/introducing-mum/

Functioning	Ranks content by understanding how keywords relate to real-world concepts.	Looks at an entire query or page rather than just keywords. Gives Google a broader understanding of concepts, which expands the amount of content Google can search through.	Allows Google to understand how words can change the meaning of queries when used in a particular sequence.	It can understand information across text, images and more in future. Understands information and world knowledge across dozens of languages and multiple modalities, such as text and images.

Table-7: Comparative Understandings of 4 AI Search Algorithms of Google | Source: [37]*Pandu Nayak* [2022]

Above comparison enhances the importance of linguistic expertise of a searcher as well as the necessity of equally strong ontological ecosystem at SEs end that contribute to better SERP by enabling SEs to understand and process natural language queries more accurately matching to user query intents, leading to relevant and useful search results. We shall be discussing specifically about User Query Intents, its types and purposes in Chapter-4 below.

Formation of Knowledge Graph *(the best example of Intent Inferring)* **using ontological approaches**: KBS starts with breaking down a query into the most important terms, getting rid of unnecessary connecting words like "how", "and", "the", and so on. Infographic-4, Appendix-9 is an example of how Search Engines decipher a query from query text, first by annotating it using NLP then the real information extraction happens to serve the searchers in the form of a Knowledge Graph. Infographic-5, Appendix-9 explains how Google identifies the ontological entities to extracts the information in blocks around different characteristics of a query, after deciphering the query as shown in Infographic-4. Infographic-6, Appendix-9 explains the next step of Knowledge Graph formation after entities recognition as shown in Infographic-5. For this purpose, Google Index is consulted through Data annotation and understanding algorithms to decide piece of contents, content' main entity, ontologies can be assigned to main entity, how entities in content relate to each other, which author, or which entity the content is assigned and properties or attributes to be assigned to entities. These Infographics form a complete sequential flow cycle of Search Algorithms for query deciphering and results picking through the combination of **NLP** and **Semantic Web** technologies. Here's the general overview of the process:

1. **Ontology Selection:** First, the **domains or subjects** that the Knowledge Graph will cover are identified. Then relevant ontologies that define the concepts, relationships, and properties within those domains are selected.
2. **NLP for Text Extraction:** NLP techniques are used to **extract information** from unstructured text sources such as documents, articles, and websites by identifying entities *(e.g., people, organizations, locations)* and relationships mentioned in the text.
3. **Entity Recognising and Linking:** After text extraction NER algorithms are applied to identify and classify entities in the text. Then recognized entities are linked to corresponding concepts in the selected ontologies.
4. **Relationship Extraction:** Further the NLP methods are used to **extract relationships** between entities mentioned in the text. Then these relationships are mapped to the properties defined in the ontologies.
5. **Relationship NODES and Edges Construction:** Relationship **nodes** are formed to connect the extracted entities, their relationships, and properties in a graphical structure. **Nodes** in the graph represent entities, and **Edges** represent relationships between entities; *refer to Infographics 4 and 5 above*.
6. **Ontology Mapping:** Then the **ontology concepts align** with the entities identified through NLP. The concepts and relationships from the selected ontologies are mapped to the Nodes and Edges in the Knowledge Graph.
7. **Alignment and Integration:** If the Knowledge Graph spans **multiple domains**, then information from different ontologies is integrated by aligning the ontologies to a common framework to resolve semantic heterogeneity.
8. **Query Expansion:** The expansion of Searcher queries based on the NLP-based querying to include synonyms and related concepts from the ontologies improve the relevance of search results and facilitate better exploration of the Knowledge Graph.
9. **Continuous Updating:** Regularly update the Knowledge Graph by incorporating new information from updated ontologies or newly available data.
10. **Quality Assurance:** Implement measures to ensure the quality and accuracy of the Knowledge Graph, including validating against ontological constraints and resolving inconsistencies.

This way by combining **NLP with Ontologies**, organizations can leverage the contextual understanding to populate and enrich a Knowledge Graph with structured and semantically meaningful information. This approach facilitates more effective data integration, exploration, and retrieval within the context of specific domains. Below Chart-12 summarizes our findings of Lingual Ontologies in connection to Searcher and Search Engines.

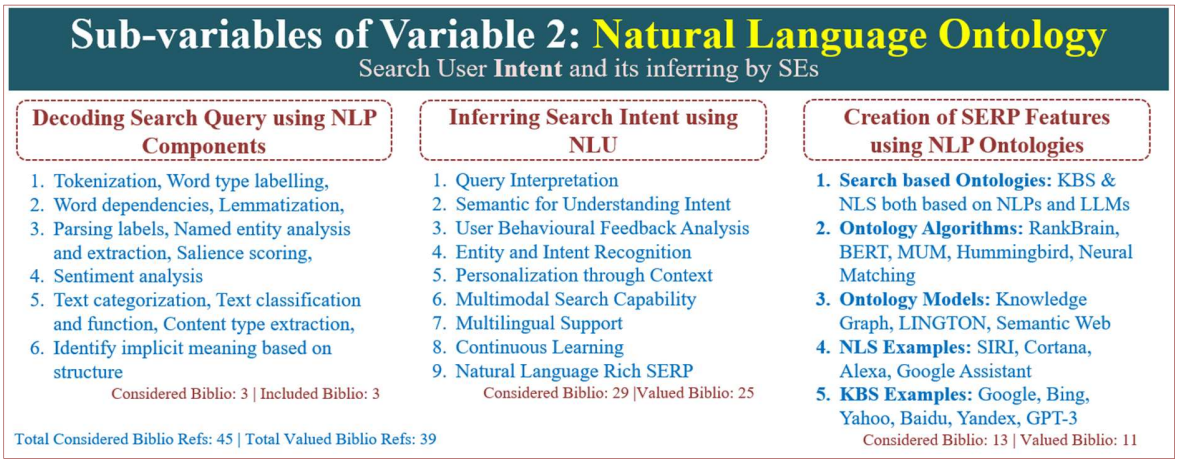


Chart-12: Sub-variables of Variable 2 'Lingual-Ontologies'

Discussion: Possible ground for hypothesis conceiving: As strong as the linguistic knowledge of a search user is, that easier is for SEs to deliver relevant SERP; however, above LR shows that SEs do adequate self-learning to decipher what comes in queries simple or complicated. Interesting can be to find out percentage of search users who rely on search engines itself for their query understanding rather than making efforts in asking in clear words. Or knowing what percentage of search users, do reformulation or iteration around their search queries due to lack of proper linguistic expertise or lesser knowledge of their topic of search *(topicality is talked in detail in next section below)*. It can be interesting to know how the health information seekers in France search, reformulate and iterate their queries to finally get the desired search results.

Discussion Scrutiny: Based on above discussions, models and theories; connection between NLU and SERP algorithms is foundational to the effectiveness of SEs in leveraging user-centric, and contextually fit search experience. NLU churns the whole game here like a spinal bone. As technology progresses, the synergy between SERP algorithms and natural languages continued to meet the ever-changing expectations of users, e.g. Knowledge Graphs. We evaluated above how the search engine leverages knowledge graphs through enhanced understanding. LLMs, NLG and NER play instrumental role in extracting and presenting information in featured snippets and rich results. This presents users with more detailed and relevant information directly on the SERP. Our study sees the interaction between Search Engine and Human as a **language dependent process of understanding** which articulates through mental models and mind methods. Advances in NLU continue to shape the evolution of search engine technologies and enhance the overall search experience. In this section we learnt about Lingual Ontology framing by SEs, in next section we intend to understand how topic of search develop in Searcher' mind.

2.3. Topicality *(Search Context and Cognitive Accuracy)*:

Research has consistently shown that a searcher's topic knowledge significantly influences their search behaviour and performance. *Duggan & Payne* [2008] demonstrated that greater topic knowledge leads to more efficient search processes, including faster decision-making and shorter queries. *Mehala & Bhatia* [2020] and *Panagis et al.* [2006] proposed methods to improve search results by leveraging topic knowledge, with Mehala and Bhatia focusing on **concept-based clustering** and Panagis on a mechanism for **reranking and grouping search results** based on user preferences. These studies collectively highlight the importance of topic knowledge in producing better search results. In the context of online search, the topicality of search results is crucial for providing users with the most current and pertinent information. Here are key aspects related to topicality in the context of a searched query:

1. **Relevance to Query:** Topicality ensures that search results are directly related to the user's query considering factors such as keywords, context, and user intent.
2. **Freshness** *(Timeliness)* **of Information:** The topicality of search results depends on how recent and up to date the information is. Freshness of content is particularly crucial for **topics that evolve over time**, such as news, technology, and current events. So, the topicality of a search query considers the **temporal relevance** of the content, for example by considering the publication date of articles and the recency of information. Freshness of a topic is also measured through the **current popularity and trends associated** with specific queries. **Trending topics** may receive higher visibility in search results, reflecting the real-time interests of users. To acquire about trending, algorithms analyse social media trends, news sources, and other indicators. Adding to flavour the next is the **seasonal relevancy of a topic**, such as holidays, sports events, or weather-related queries. Search engines consider the **seasonality of topics** to deliver results that align with the current time of year. Search engines also recognize the value of **evergreen content** i.e. the information that remains relevant over time. Balancing timely updates with the enduring value of certain topics contributes to a well-rounded set of search results.
3. **User Behaviour and Engagement:** User engagement metrics, such as click-through rates and time spent on a page, can assess, and influence the topicality of search results. Content that is frequently clicked and engages users may be considered more relevant and topical by search algorithms. Higher engagement often indicates that the content effectively meets user expectations.
4. **Social Signals Influence:** Social media plays a role in determining topicality, as trending topics and discussions on platforms like X *(lately Twitter)* and Facebook can influence search results. Search engines may incorporate social signals to gauge the popularity and relevance of content.

5. **User Intent Matching:** Topicality is enhanced when search engines accurately match the user's intent with the content of the search results. Understanding the context and purpose behind a query enables the delivery of more relevant and topical information.
6. **Geographical Relevance:** For location-specific queries search engines consider the user's location to provide information that is contextually appropriate to their geographic area.

Search engines continuously update their algorithms to improve search results. Algorithmic changes may include adjustments to prioritize certain types of content or to better understand emerging trends. Search engines aim to provide diverse perspectives and formats of content related to a topic. This includes articles, images, videos, and other media types, ensuring a comprehensive and well-rounded set of results.

Domain and Topic Knowledge of Search User retrieves better results: The topic knowledge of a search user helps in producing better search results by enabling **more precise search queries** and facilitating the evaluation of search results for relevance and accuracy. Research has shown that individuals with high levels of topic knowledge are **more effective in formulating search queries**, leading to improved information retrieval in online catalogues, *Allen* [1991]. Additionally, search engines, such as **Microsoft's Topics,** utilize AI to automatically search for and identify topics, compiling information and making it available to users, thereby leveraging the knowledge within an organization to enhance search experiences, [39]*Topics Microsoft*. In the context of **Salesforce**, topic filters are used to refine knowledge search results, allowing users to search faster and more effectively by leveraging their topic knowledge, [40]*Topic Filters Salesforce*. The moderating effect of knowledge type on search result modality has been studied, indicating that **different types of knowledge** influence the preference for specific search result modalities, *Pardi et al.* [2023]. Search engines consider **topicality scores** to better understand the relevance of content to a search query, particularly in the context of user-generated content, which can be clouded by other users' content, [41]*Slawski* [2022]. It states that topicality is attributed by three stake holders: **the query producer, the results producer, and the content producer.** Research consistently shows that a user's topic knowledge significantly influences their search behaviour and the quality of their search results, *Liu & Zhang* [2019], *Liu et al.* [2016], *Duggan & Payne* [2008] and *Zhang et al.* [2013]. Users with **high topic knowledge** tend to select more relevant documents, read more documents, and give higher relevance ratings, *Liu & Zhang* [2019]. **Greater topic knowledge leads to faster decisions and shorter queries**, *Duggan & Payne* [2008]. However, the impact of **background domain knowledge** on

[39] https://learn.microsoft.com/en-us/microsoft-365/topics/topic-experiences-overview?view=o365-worldwide
[40] https://help.salesforce.com/s/articleView?id=release-notes.rn_search_esk_topic_filters.htm&release=238&type=5
[41] https://gofishdigital.com/blog/topicality-scores-social-scores-and-user-generated-content/

search performance is also significant, with topic knowledge being more closely linked to individual task performance and background knowledge to overall performance, *Zhang et al.* [2013]. Users with **high levels of domain knowledge** tend to select and view more documents, giving higher relevance ratings to the viewed documents. They also tend to select documents ranking down the search result lists. On the other hand, users with **low levels of domain knowledge** tend to select documents ranking toward the top of the search result lists. **Early-session search behaviours** can be used to predict users' topic knowledge levels, allowing personalization of search for users with different levels of topic knowledge. This helps in designing search systems that can better assist users in finding relevant documents from search results, *Huang et al.* [2017] and *Liu & Zhang* [2019]. The topic knowledge of a search user contributes to better search results by improving query formulation, enabling the utilization of AI for topic identification, refining search results through topic filters, and influencing the evaluation of content based on relevance scores.

Domain knowledge helps in retrieving better search results: A topic expert may not be a domain expert and vice-versa. Topic knowledge can be acquired by anyone without experiencing the same, but domain knowledge comes with direct experience in that domain. But as both are the part of Contextual and Cognitive Accuracy, so it is sensible to discuss this aspect specifically. Domain knowledge empowers users to tailor their search queries, navigate through the vast amount of information available, and critically evaluate the relevance and reliability of search results. Let's find how the domain knowledge contributes to the enhancement of search results:

1. **Contextual Understanding:** Domain knowledge allows users to frame queries in a way that is more aligned with the specific terminology and context of a particular field. This helps search engines better understand the user's intent and return more relevant results.
2. **Refined Query Construction:** Users with domain knowledge are often able to construct more precise and targeted queries. They can use industry-specific terms, acronyms, or jargon that are more likely to appear in relevant documents, leading to more accurate search results.
3. **Identification and Interpretation of Relevant & Authoritative Key Sources:** Understanding the domain helps users identify authoritative sources within the field. This knowledge enables them to pick specific websites, databases, or journals that are known to provide reliable and high-quality information. They can better evaluate the relevance of search results. They can quickly filter out irrelevant information and focus on content that is more likely to be valuable in their contextual and efficient information retrieval. Domain experts are better equipped to evaluate the credibility and accuracy of search results. They can critically assess the information retrieved, identify potential biases, and make informed decisions based on their knowledge.

4. **Awareness of Trends and Updates:** Keeping up with developments in a specific domain allows users to include the latest terms, technologies, or trends in their queries. This ensures that the search results are up-to-date and reflect the current state of knowledge in the field.
5. **Previous Search Experience:** A domain expert has the intensity of being a routine or periodic searcher of that domain online. So, their previous search experiences about their domain influence their search strategies, they may use more specific and refined keywords. On the other hand, those less experienced might start with broader terms and gradually narrow them down.
6. **Social and Cultural Influences:** Social and cultural factors associated with a domain are more accustomed to domain experts. For example, a health domain expert would know the social and cultural beliefs related to that domain in their market and that helps them to sort out appropriate results. Here, different regions or communities may use distinct language variations or slang, affecting the choice of keywords.

Although not necessary that a Topic and Domain expert can be SE Algorithms experts too; but have higher than others probability of being to know to algorithms too. Such knowledge helps them grasp how search engines work, including the algorithms used for ranking results. Similarly, Domain level aware users are often more adept at using advanced search operators of search. These operators allow for more specific and refined searches, leading to better-targeted results.

User' Literary Skills precise Topical Search Terms: Linguistic expertise significantly impacts the use of keywords in online search positively, as it enables individuals to define search topics accurately and select targeted keywords efficiently, *Kemtes et al.* [2003]. This expertise is particularly important in the context of literature digital libraries, where content-driven search keyword suggesters can enhance the quality and user-friendliness of search results, *Bani-Ahmad & Ozsoyoglu* [2008]. Furthermore, skilled readers are better able to evaluate online information, suggesting that literary expertise can also influence the evaluation of search results, *Hahnel et al.* [2018]. However, despite the prevalence of keyword searching, many users lack a basic understanding of how keyword searches are processed, indicating a potential gap in literary expertise, *Hildreth* [1997]. Research has shown that the formation of search keywords in a searcher's mind is influenced by the **semantic gap** between the user's information need and the input of search engines, *Bobed & Mena* [2016]. This gap can be bridged by semantic keyword interpretation, which involves discovering the meaning of the input keywords and combining them in a formal query that captures the user's intended semantics, *Bobed et al.* [2010]. The selection of search keys is also influenced by the quality and availability of thesauri and indexing, with multi-database searches often leading to the use of text words without consulting a thesaurus, *R. Fidel* [1991]. *Tsai & Wu* [2017] paper solves the problem of the semantic gap between human visual features and

understanding perception for an image search, which results in a situation where image search engines cannot understand the users' intentions. They used machine learning to periodically train and update their model for each user search process to adapt to changes in user intent under different spatial and temporal conditions. The formation of search keywords or search phrases in a searcher's mind is a complex process influenced by various factors such as **information needs, language proficiency, and search experience**. Here are key factors that contribute to the creation of search keywords under the influence of literary skills:

1. **Precision in Language:** Topic and literary experts are typically familiar with a wide range of synonyms, understand the nuances of word meanings within different contexts and terminology and jargon associated with a particular topic. So, they may use different terms based on their background, knowledge, and familiarity with the subject. The same concept or item may be described using various terms, here the individual preferences play a role in keyword selection. They are likely to iterate on their queries based on initial search results, continuously improving the precision and effectiveness of their searches. This
2. **Clear Contextual Understanding:** The context of the search and the searcher's intent shape the choice of keywords. Linguistic expertise allows them to use related variations of keywords and adapt their search queries to capture the diverse ways information may be needed. For instance, a person looking for information on "cars" may refine their search based on the context, such as "electric cars," "best family cars," or "car maintenance tips."
3. **Critical Thinking and Analysis:** Literary expertise often involves critical thinking and analytical skills. Searchers with these abilities can carefully evaluate search results, refine their queries based on the content they find, and adjust their approach to better meet their information needs.

However, not as such necessary that a linguistic expert would be equally familiar with use of SEs Search Operators; but still literary experts may be more comfortable in employing advanced search operators, such as quotation marks for exact phrases, asterisks, or minus signs to exclude certain words. This knowledge enables them to fine-tune their queries for more precise and targeted results. We shall study the details of Search Operators in 4th Chapter below. Knowledgeable searchers can quickly filter through search results and identify the most relevant information. Their ability to discern the significance of different sources and content allows them to prioritize and engage with the most valuable resources.

Search Engine's Topical Ontology Methods *(Process of deciding a multifaceted nature of topicality)***:** Search engines decide the topicality of search results around a search query by utilizing various methods such as **topicality scores, user-generated content analysis, and ontology-based semantic search**. As per [41]*Slawski* [2022] Topicality Scores indicate the degree to which the content pertains to the search query, helping search

engines understand the relevance of the content to the user's intent. Slawski further states that User-generated Content is also evaluated for topicality, and its relevance is determined based on topicality scores. Ontology-based semantic search engines utilize ontologies to improve search capabilities and understand the meaning of each word or phrase, contributing to the decision of topical relevance, [42]*SAP.com-1*. Additionally, search engines may use **specialized intent models**, such as the Topicality Relevance-aware Intent Model ***(TRIM**, here onward),* to enhance the understanding of topical relevance in web search, *Wang et al.* [2023]. Such models enable search engines to provide more accurate and relevant search results based on the topicality of the content by considering the similarity between the query and documents, as well as the user's session interaction sequence. Another method involves constructing **query context models** that facilitate topical search. These models identify topical queries and related contextual keywords, which are then used to restrict query results to a particular topic, *Parikshit et al.* [2010]. *Han et al.* [2021] noticed in their research that overall, participants prefer the real search pages over the synthesized ones with a significant higher ratio. It indicates that adding more varieties makes the results less biased but less relevant and consistent to the queries, which hurts users' satisfactions. In addition, when the diversities in the synthesized pages are present at the top 5, participants tend to prefer the real pages. *Gao & Shah* [2020] found in their research on 'fair ranking' that top web search results from search engines are topically biased. Diversity fairness can improve diversity and relevance in retrieval results, but performance of fair ranking strategies depends on fairness constraint and data, the lower the diversity bias, the higher the relevance, diversity and novelty. Additionally, search engines can **automate the selection of topical terms** in a search query, allowing users to quickly reformulate and execute new queries that return more focused search results, [43]*Kumaran* [2012]. Furthermore, topicalization and topicality can be defined in morphosyntactic terms, with topic prominence being determined through **typological correlations across languages**, *Sajid* [2022]. [44]*Christensen & Van Hoff* [2015] add to the flow that a topical search computer system can identify topics from data sources and differentiate them based on their context, allowing for improved search results. The determination of topicality by search engines is a complex process that involves various factors. *Sondhi et al.* [2010] proposes the use of topical search engines, which alter **ambiguous queries** to focus on a specific topic. *Stamou & Ntoulas* [2008] further explores this by discussing the challenges of **identifying user interests** and ranking results, accordingly, using a topical ontology and a ranking function. *Kumar & Lang* [2007] emphasizes the importance of search term formulation and the role of search engine support for topical organization in improving search performance. *Chuklin & Rjike* [2015] highlights the need to consider not only topical relevance but also **perceived and snippet relevance** in

[42] https://help.sap.com/doc/saphelp_nw75/7.5.5/EN-US/48/63917cf29407bee10000000a42189d/content.htm?no_cache=true
[43] https://patents.google.com/patent/US20130204892A1/en
[44] https://patents.google.com/patent/US9875309B2/en

evaluating search engine performance. These studies collectively underscore the **multifaceted nature of topicality** determination in search engines.

Topicality Score Parameters of SEs: As per [45]*Google Cloud Topicality* definition, "Topicality Score is a specific metric used to measure the relevance of a search result to the original query terms. It refers to the relevance of a search result to the original query terms and is calculated based on the importance of each query term, the number of hits *(the number of times a query term appears in the item's content or metadata),* and the type of matches the query term and their variants have with an item indexed in Cloud Search". Topicality Score also coined as Topic Relevancy Score or Currency of Information is defined by *Huang and Soergel* [2004] as Topic Relevance generally referred to as direct topical matching or the direct matching between the overall topic of a relevant document and the overall topic of a search query. Chart-13, Appendix-9 from Huang and Soergel represents the types of Topic Relevance. The computation of topicality specialization and relative topicality specialization allows the identification of specialized objects for topics without any pre-defined categorization of topics or pre-computation of expertise, *Coleman and Thrasher* [2006]. The deciding parameters for topicality score include **lexical generality, specificity of the phrasal expressions, and the share of citations** for an object matching the query term among all citations for all objects matching the query term. So, "**Topicality around the searched query**" refers to the relevance and currency of information related to a particular search query.

Topical modelling: After have a granular insight of the various aspects of topicality and search, it is imperative to understand the ways the topics can be arranged or grouped. As per *Alghamdi & Alfalqi* [2015] Topic Modelling provides a convenient way to analyse big unclassified text. A topic contains a cluster of words that frequently occurs together. A topic modelling can connect words with similar meanings and distinguish between uses of words with multiple meanings. They provide two categories that can be considered under the field of topic modelling. First one discusses the area of **methods of Topic Modelling**, which has four methods and can be considered under this category. These methods are:

1. **Latent Semantic Analysis (LSA):** As per *Dumais* [2005] LSA was designed to overcome the vocabulary mismatch problem between searchers and document creators. The LSA analysis consists of four main steps as "Term-Document Matrix", "Transformed Term-Document Matrix", "Dimension Reduction" and "Retrieval in Reduced Space".
2. Probabilistic Latent Semantic Analysis (PLSA)
3. Latent Dirichlet Allocation (LDA)

[45] https://developers.google.com/cloud-search/docs/guides/improve-search-quality

4. Correlated Topic Model (CTM)

The second category is called **Topic Evolution Model**; it considers an important factor time. In this category, different models are discussed, such as

1. Topic Over Time (TOT)
2. Dynamic Topic Models (DTM)
3. Multiscale Topic Tomography (MTT)
4. Dynamic Topic Correlation Detection (DTCD)
5. Detecting Topic Evolution (DTE)

As this subject by now has sufficient learning required for this thesis and not giving in descriptive details of each of these concepts and models, and we may be using them on appropriate occasions further; conclude this topic by talking one of these methods in details. *Blei et al.* [2003] report results in document modelling, text classification, and collaborative filtering, comparing to a mixture of unigrams model and the probabilistic LSI model using Latent Dirichlet Allocation ***(LDA).*** Their work tries to resolve the problem of modelling text corpora and other collections of discrete data. *Jelodar et al.* [2019] in their synthesis of research development, current trends and intellectual structure of topic modelling in purview of LDA state that any part of the text *(an email, a book chapter, a blog post, a journal article or any kind of unstructured text)* is combined by selecting words from probable baskets of words where each basket corresponds to a topic. The tools and methods like LDA go via this process repeatedly until it stays on the most probable distribution of words into baskets i.e. topics. *Tong & Zhang* [2016] represent two text mining model based on LDA from Wikipedia articles for document topic model and Twitter users' tweets user interest model, both including data collecting, data pre-processing and model training. However, *Potha & Stamatatos* [2018] proposed the intrinsic profile-based author verification method based on LSI for applications in humanities and forensics, demonstrated as more effective and stable than LDA in their context.

SE's Topical Ontologies *(Engineering that decides the ontology of search query topics)***:** After discussing about the capabilities of a Search User with Topicality, here we intend to find out how search engines design the ontology of topics around a search query to improve the accuracy and relevance of search results, facilitate the integration of data from multiple sources, and support the development of personalized and adaptive search experiences, often termed as **Ontological Engineering**. As per [46]*Marketbrew-1,* it is a crucial aspect of SEO and the development of effective information systems. By using ontology, the search engine can analyse and

[46] https://marketbrew.ai/ontology-engineering-for-seo

categorize data, generate relevant keyword ideas for web content, and understand the meaning of each word or phrase, which is essential for SEO. First connected term without which there is no ontology possible is the **Semantics of Topics**. In this regard, semantic search engines aim to bring relevant information together in a comprehensive way, moving beyond traditional keyword-based search, *Keyser* [2012]. Ontology-based semantic search engines utilize ontologies for **classical and semantic search**, contributing to the improvement of search capabilities, *Sayed & Muqrishi* [2017]. The best example of one such ontology is the keyword ontology. As evident from explanation in above section, a **Keyword Ontology** is the base of a knowledge graph that describes relationships between the keywords used in search queries, and it can be used in auto-tagging technology to assist content producers in focusing on writing authoritative content, [47]*Biznology-1* [2018]. Search engines design the ontology of topics around a search query by utilizing **ontological ranking algorithms** such as **term frequency, link analysis** *{PageRank algorithm and Hyperlink-Induced Topic Search (**HITS**) algorithm},* and **keyword frequency** to determine the relevance between the user's query and the data in the computer system or internet, *Hou & Tao* [2019]. Another approach is to utilize ontologies to represent **user profiles and gather their interests and relevance**, allowing for personalized search and reasoning, *Zaid et al.* [2013]. Ontologies enable relationships between keywords and terms to be defined, allowing for the retrieval of desired information by sharing common vocabularies and understanding the meaning of terms in the domain, *T. Shakti et al.* [2012]. Additionally, ontologies can support the specification of search queries and aid in the preparation of the document corpus, facilitating the integration of all accessible data about the safety and performance of a medical device during post-market surveillance, *Uciteli et al.* [2017]. **Formal concept analysis** can also be used to identify conceptual structures among data sets, leading to an ontology that describes concepts and relations in a way suitable for knowledge exchange, *Obitko et al.* [2004]. Search engines design the ontology of topics around a search query by utilizing ontologies to **enhance precision and recall in information retrieval**, *Barros et al.* [1998]. This is achieved using a semantic search engine framework based on ontology, which overcomes traditional search engine shortcomings, *Yong et al.* [2007]. The result is an efficient and accurate search engine that produces results based on the **literal meaning of the query**, *Kavitha et al.* [2015]. By applying semantic capabilities using ontology, search engines can search **conceptually linked keywords and present semantically linked search results**, *Esa et al.* [2010].

Cognitive Evaluation of Topicality *(Relationship between topicality and cognition pertaining to search results)***:** The relationship between topicality and cognition pertaining to search results is a multidimensional multifaceted process. While topicality is a crucial aspect of **relevance**, research has shown that users consider various criteria beyond topicality in assessing the relevance of search results. These criteria can include factors

[47] https://biznology.com/2018/02/what-is-a-keyword-ontology/

such as **credibility, understandability, interest, and sentimentality,** *Taylor* [2011] and *Bruza & Chang* [2014]. *Liu et al.* [2016] found that early search behaviours can predict a user's topic knowledge, suggesting a link between cognition and search behaviour *(we stated it above also)*. *Gerjets & Kammerer* [2010] emphasized the importance of evaluating the topical relevance and information quality of search results, which can be influenced by cognitive biases. *Wu et al.* [2014] further explored this, showing that the number and distribution of relevant results on a SERP can impact search behaviour, with individuals with higher need for cognition engaging more deeply with the results. *Qu et al.* [2010] added that **task type and familiarity** can also influence search behaviours, indicating that cognitive factors play a significant role in how individuals interact with search results. Current cognitive models of web navigation primarily focus on evaluating the topical relevance of search results, neglecting the evaluation of information quality, *Gerjets & Kammerer* [2010]. *Khamsum et al.* [2014] state that users' cognitive styles influenced their information searching strategies, query reformulation behaviour, web navigational styles and information processing approaches. Their **wholist-analytic** and **verbal-imagery** dimensions of Cognitive styles defines the gut relation between User Cognition and Web Search Behaviour. Chart-14, Appendix-9 gives the picturization of their model. **Eye-tracking data and retrospective think-aloud** analysis have shown that specific source and topicality are the most frequently used criteria for relevance judgments. Fixation duration is longer on results judged on topicality, and pupils dilate significantly on the top-ranked result on SERPs, *Gwizdka & Chang* [2020]. **Understanding the meaning of topicality** to users is essential for designing effective approaches to information search and retrieval, *Greisdorf & Connor* [2003]. **Redirection architecture and contextual keywords** can be used to bias search results towards a particular topic or domain, [43]*Kumaran* [2012] and *Sondhi et al.* [2010]. In another interesting and relevant aspect, *Jin et al.* [2019] investigate the relationship between **cognitive abilities and age differences** in information search and the moderating role of task self-relevance by measuring the decision-making processes of participants in both high and low self-relevance decision-making tasks. The authors found that **self-relevance** can moderate the mediation effect of verbal fluency on the relationship between age and information search time; this means that older adults whose **verbal fluency** was limited require relatively more time to search information to make an informed decision. Below Chart-15 lists the findings of Topicality variable of this thesis:

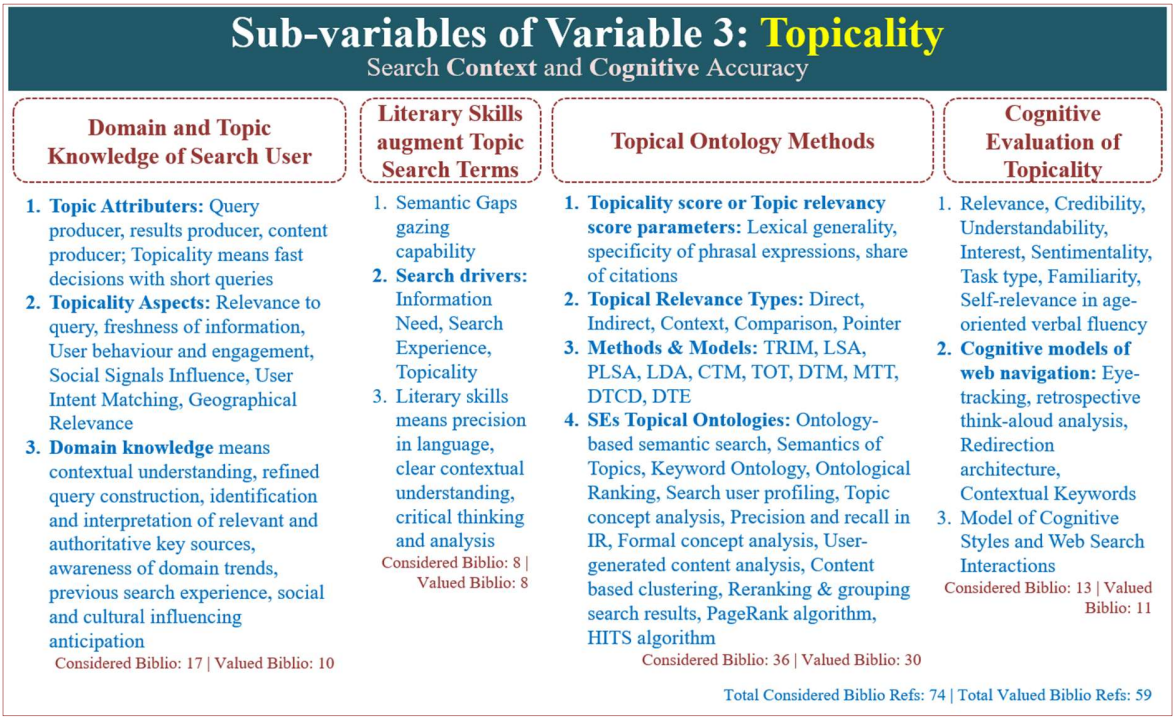


Chart-15: Sub-variables of Variable 3 'Topicality'

Discussion *(Possible ground for hypothesis conceiving)*: Literary expertise brings creativity and precision in search results through query formation or reformation. Domain knowledge brings critical capacity to diversify search results on quality and relevance. Knowledge and expertise combined lead to a better cognitive semantic understanding and sophistication around a topic or subject search.

Discussion Scrutiny: Based on above LR it is easily understandable that the topicality in the context of a searched query is about providing searchers with information that is not only relevant to their query but also contextual and reflective of the latest developments in that topic. Above findings also affirm that the topicality of a search query is a dynamic and multifaceted concept, encompassing not only the recency of information but also the alignment with user intent, diversity of content, and cognitive relevance. And that is why as search engines evolve, they strive to enhance the topical relevance of search results to provide users with the most valuable and current information. Above study also helps in understanding that the formation of search keywords is a dynamic and individualized process influenced by information needs, language choices, context, and the searcher's experience with search engines. As searchers interact with search results, they continuously adapt and refine their queries to obtain the most relevant information. Further to add literary expertise enhances the smart use of keywords in online search by promoting precision, understanding of context, critical thinking, effective use of search operators, and creativity in query construction. These skills empower individuals to

navigate the vast landscape of online information more effectively and extract meaningful insights from search results. Finally, the domain knowledge of a searcher enhances the quality of search results by influencing the refinement of search queries pertaining to cognitive capability, the effective use of alternate terminologies, the ability to discern credible sources, and the optimization of searches for specific needs.

2.4. Semantic Web *(Predicting search user expectations from SERP- Mechanism of Web 3.0 & Web 4.0)*

Semantic Web Approaches: Semantic relatedness refers to the degree to which two concepts are related (or not) and humans are able to easily judge if a pair of concepts are related in some way, *Patwardhan* [2003]. Concepts are commonly represented in dictionaries by word senses, each of which has a definition or gloss that briefly describes its meaning. The types of semantics are talked in the research of *Nasution* [2018] which states that in the logic of language, the phenomenon of relationships between entities is recognized as semantics such as **meronymy, holonomy, hyponymy, synonymy, and polysemy**. *Leacock et al.* [1998] research evidence indicates that local context is superior to topical context as an indicator of word sense when using a statistical classifier. The benefits of adding topical to local context alone depend on syntactic category as well as on the characteristics of the individual word. *Wu & Palmer* [1994] state that the verb semantic representation has great impact on the quality of lexical selection. Human beings have an innate ability to determine if two concepts are related, *Banerjee & Pedrsen* [2003]. Semantic web predicts what a search user wants from their search query by using various techniques. One approach is **semantic query expansion**, which expands the query with additional terms that are semantically like the original query, *Raza et al.* [2021]. This helps to overcome the limitations of ambiguous and limited query terms. Estimating the semantic similarity between text data is one of the challenging and open research problems in the field of NLP, *Chandrasekaran & Mago* [2021]. In the era of 1990', there have been three famous research on Semantic Similarity calculations in NLP. **Resnik's** semantic Process Model Matching ***(PMM)*** is a measure of semantic similarity between ontology terms, *Khurram et al.* [2018] and *Philip Resnik* [1995]. It is based on the information content of a term, which uses the term probability, which further relies on a relationship in the WordNet taxonomy, where each node represents a unique WordNet synset or concept. *Pedrsen et al.* [2004] talk of **WordNet :: Similarity** software that makes it possible to measure the semantic similarity and relatedness between a pair of concepts or synsets. It provides six measures of similarity, and three measures of relatedness, all of which are based on the lexical database WordNet. The second is the **Lin's** measure of semantic similarity based on the information content of ***GO*** (Gene Ontology) terms, *Dekang Lin* [1998]. Third, **Jiang and Conrath** defined a distance measure between

GO terms that is based on the information content, *funsimmat-sss*. In a try of enriching the information-theoretic definition of semantic similarity, *Anna & Francesco* [2021] proposed a measure based on a novel approach that essentially addresses the context of the contrasted concepts, by associating them with their intended senses. Another approach is **semantic query suggestion**, which identifies relevant queries by considering the semantic concepts contained in user queries, *Aytuğ Onan* [2018]. Machine learning algorithms, such as **Random Forests**, can be used to improve the accuracy of semantic query suggestion, *Hong & Sug* [2014]. Additionally, **context awareness technology** can be utilized to predict user interests based on the current situation of the user, *Sharma & Rana* [2020]. By establishing **context and tour ontologies**, the necessary information for the user can be accurately predicted and provided, *Hojgaard et al.* [2016]. *Wechsler et al.* [2004] presents a search system prototype based on **ontologies** and a **cognitive model of human information acquisition**. **Semantic search and query** are a technique that goes beyond matching keywords and phrases. It analyses the **intent, context, and semantics** of a user's query, and returns results that are more aligned with their goals and expectations. It also considers factors such as the user's location, preferences, history, and feedback to personalize and refine the results. Additionally, users can provide feedback, such as **ratings or reviews**, to indicate their satisfaction or dissatisfaction with the results. This approach helps search engines understand the meaning and context of a user's query, thereby enhancing the relevance and trustworthiness of the information retrieved, [48]*Linkedin.com-3,* [49]*Hill* [2016] & [50]*Hill* [2021]. A range of studies have explored the use of web semantics to predict search user responses. *Fathy et al.* [2011] and *Garofalakis et al.* [2008] both propose personalized search approaches that use **ontological user profiles** to model user preferences and re-rank search results. *Lin et al.* [2005] takes a different approach, using **probabilistic latent semantic analysis** to mine user-specific information from web logs and improve search results. *Neelima & Rodda* [2016] paper on web log mining for predicting user behaviour through sessions provides a systemic method to do useful information extraction regarding user behaviour. *Xu et al.* [2016] introduces a method based on user tasks to **perceive user intent** and build a conceptual semantic space model, which can predict user behaviour to some extent. Predicting search user responses with web semantics involves using **semantic query expansion techniques** to improve the retrieval of user-relevant content from the web, *Raza et al.* [2021]. By expanding the query with additional terms that are semantically like the original query, the technique aims to better capture user information needs. However, it is important to consider **individual user interest** in the generation of expansion terms. One approach is to **combine ontology knowledge and user interest** to enhance semantic query expansion, *Dennis & Stone* [2009]. This involves computing the domain of the input query via

[48] https://www.linkedin.com/advice/1/how-can-semantic-search-query-help-users
[49] https://www.hillwebcreations.com/structured-approach-web-content-creation-for-user-conceptual-search/
[50] https://www.hillwebcreations.com/what-is-semantic-search/

ontology, generating expansion terms from the user browsing history, and selecting expansion terms that represent user preferences based on semantic similarity and user feedback. Experimental evaluation shows that incorporating user preferences in semantic query expansion improves the retrieval of personalized content over web search, *Chatterjee et al.* [2021]. A technically sound article from [51]*elastic.co-1* states that semantic search is a search engine technology that **interprets the meaning of words and phrases**. It uses ML and AI to match search intent to semantic meaning, which helps to improve the quality of search results by interpreting natural language more accurately and in context. It further says that when a query is launched, the search engine transforms the query into **embeddings**, which are numerical representations of data and related contexts. They are stored in **vectors**. The **kNN algorithm** *(k-nearest neighbour algorithm)* then matches vectors of existing documents to the query vectors. The semantic search then generates results and ranks them based on conceptual relevance. Diagram-6, Appendix-9 depicts the functioning model of semantic web. [51]*elastic.co-1 and* [52]*Bothra* [2020] define that semantic search also uses **context clues** to determine the meaning of a word across a dataset of millions of examples. It identifies what other words can be used in similar contexts. For example, a search for "football" would mean "soccer" in the USA and "football" in the UK and other parts of the world. Semantic search would distinguish results based on the user's geographical location. Deepika further asserts that by considering the context of queries and the user intent, semantic search engines **filter out irrelevant results** *(contextual filtering)* at the very early stage of the search. They also keep a **tab on users' location, search history, word variation, synonym, acronym, foreign language,** etc. to provide users with what they are seeking for. The technology relies on new AI technologies, domain concepts from preexisting ontologies, independent thesauri, and advanced capabilities to **decipher semantic relationships** between them during runtime, [49]*Hill* [2016]. By understanding the intent and contextual meaning behind a search query, semantic search algorithms deliver more relevant results, leading to a more conversational and personalized search experience, [53]*Roberts* [2023]. *Goyal et al.* [2013] in their research about precision of search engines results in purview of varied search queries highlighting the inability of current traditional query clustering methods unsuitable to query recommendations propose a new query recommendation system by capturing user's preferences. They name it as Query-Document-Concept Clustering, in which documents are used to decouple queries and features in a tripartite graph structure. This method claims better precision, recall and F-measure than the existing approaches. As per Goyal, this approach can be extended to the concept of query-based and/or concept/topic-based document clustering, which will allow the web users to get more precise information as query results.

[51] https://www.elastic.co/what-is/semantic-search
[52] https://www.searchunify.com/blog/an-introduction-to-semantic-search-and-its-relation-to-search-relevance/
[53] https://www.bloomreach.com/en/blog/2019/semantic-search-explained-in-5-minutes

Semantic Web Algorithms: Semantic web algorithms play a crucial role in enhancing SERP by processing and analysing data more readable, meaningful, and understandable for machines to intelligently link, match and retrieve relevant information, *Thakur & Kumari* [2019]. These algorithms focus on metadata, NLP, and semantic frameworks to improve the accuracy and efficiency of SEs, *Alan Clark et al.* [2020]. For instance, and as mentioned in above sections as well, a proposed semantic framework includes phases like crawling, indexing, ranking, and retrieval, utilizing efficient ranking and crawling algorithms to enhance the retrieval process, *Banerjee & Choudhary* [2017]. Additionally, a computer-implemented system uses semantic analysis to rank search results based on discrepancies between documents authored by the same person. Moreover, advanced Page Rank Algorithms incorporate semantics, inlinks, outlinks, and user behaviour data from Google Analytics to provide more relevant and valuable search results. *More et al.* [2021] introduced a novel model for web page ranking that leverages semantic relations, demonstrating its superiority over existing mechanisms. *Pandiarajan et al.* [2020] developed a semantic search engine algorithm based on natural language processing, achieving a high accuracy rate in retrieving relevant documents. *El-gayar et al.* [2015] presented a semantic framework for search engines, incorporating an efficient ranking algorithm and enhanced crawling algorithm to improve retrieval accuracy and speed. Major algorithmic updates by Google like Hummingbird in 2013 and the introduction of the Knowledge Graph in 2012 were significant milestones in the shift towards more semantic, intent-based search, [54]*seozoom.com* [2023] and [55]*Silvers* [2024]. These studies collectively highlight the potential of semantic web algorithms in enhancing the performance of search engines in producing SERPs. Here are some key semantic web algorithms:

1. **RDF Processing Algorithm and Triplestore Indexing and Search Algorithm:** Resource Description Framework ***(RDF**, here onward)* is a foundational framework for representing information on the semantic web. Algorithms for parsing, querying, and manipulating RDF triples *(subject-predicate-object statements)* are essential for extracting meaningful data. RDF Triple itself is known as Triplestore. Triplestores are databases optimized for storing and retrieving RDF triples. Algorithms for indexing and searching triples efficiently are important for quick access to relevant semantic information. RDF processing algorithms aim to enhance querying efficiency in RDF datasets. Various approaches have been proposed to optimize query performance. RDFCSA *(Resource Description Framework Compressed Suffix Array)* introduces a compressed representation of RDF datasets, enabling efficient querying through suffix sorting. Another method involves indexing chain and star-shaped subgraphs to reduce join operations in SPARQL queries, significantly improving query response times. Additionally, a distance-based triple pattern reordering

[54] https://www.seozoom.com/semantic-search-the-key-to-modern-seo/
[55] https://www.algolia.com/blog/ai/the-definitive-guide-to-semantic-search-engines/

algorithm optimizes query execution by identifying common subtrees and leveraging materialized views for multi-query optimization. Trie-based indexing layouts offer space-efficient solutions for triple indexing, outperforming existing methods in terms of space utilization and query execution speed. Furthermore, exploring multidimensional indexing for RDF triples presents an alternative approach, demonstrating improved query processing for star group patterns.

2. **SPARQL Query Processing:** SPARQL Protocol and RDF Query Language ***(SPARQL)*** is the query language for querying RDF datasets. Algorithms for processing SPARQL queries efficiently are crucial for retrieving relevant information from semantic web databases.
3. **OWL Reasoning Algorithms:** Web Ontology Language ***(OWL)*** is used for representing ontologies that define relationships and properties of entities. Reasoning algorithms for OWL enable inference, classification, and consistency checking, enhancing the understanding of semantic relationships.
4. **NER Algorithms:** NER algorithms identify and classify entities *(such as people, places, and organizations)* within text. These algorithms are essential for extracting semantic information from unstructured data on the web.
5. **Graph Algorithms:** Algorithms for graph processing are crucial for analysing the relationships and connections between entities in a semantic graph. This includes algorithms for traversing graphs, finding paths, and identifying patterns.
6. **Semantic Similarity Algorithms:** Semantic similarity measures quantify the similarity between concepts based on their meanings. These algorithms are used in tasks such as **matching and ranking resources** or documents based on their semantic relevance.
7. **Ontology Alignment Algorithms:** Ontology alignment involves matching equivalent or related concepts across different ontologies. Algorithms for ontology alignment help **integrate information from diverse sources** by establishing correspondences between terms.
8. **Schema Matching Algorithms:** Schema matching is the process of identifying correspondences between different data schemas. Algorithms in this category help **integrate and link data from heterogeneous sources by aligning their schemas**.
9. **Semantic Link Prediction Algorithms:** These algorithms predict the likelihood of a semantic link between entities based on existing relationships. They are useful for completing missing information in a semantic graph.
10. **Temporal Reasoning Algorithms:** Temporal reasoning algorithms consider the time dimension in semantic data, allowing the modelling and analysis of changes and evolution over time in the semantic web.

Semantic Web Evolution *(SERP is a parallel journey to W3C, presently assimilates with Semantic Web which is basic to Web 3.0 with further merger in Web 4.0)***:** The magical word Semantic used so often in this chapter and will be used throughout this thesis, has a wonderful series of its evolution. This word is so semantically stick to this thesis that without it mentioning this thesis can't even be imagined. Be it the journey of Search Engines, their models, their SERP updates everything grown with a reflection in a mirror named as World Wide Web ***(W3)***. Semantic Web is usually termed to Web 3.0 series of W3 journey *(here Web 3.0 is not to be confused with W3, both represent different entities)*. So, what has been the Web 1.0, Web 2.0 then further beyond Web 3.0; let's know about it. The web has evolved through different generations, namely Web 1.0, Web 2.0, Web 3.0 and its journey is even predicted by experts up to Web 6.0. Web 1.0 refers to period from 1989 to 2004 better knowns as the **era of static web pages** mostly, when the users of web were consumers and not web content producers, [56]*Graham & Balachander* and [57]*Carter* [2015]. Web 1.0, also known as the **Classic Web**, focused on connecting and retrieving information from the internet *Nath* [2022]. Web 2.0 introduced the ability for users to **not only consume but also create** and deliver content, leading to significant growth in social media and e-commerce, *Kumar* [2022], *Karan* [2013] & *Nath et al.* [2014]. Web 3.0, also known as the Semantic Web, aims to **enable computers to generate and process new information**, utilizing AI and ML technologies. It combines the best elements of Web 1.0 and Web 2.0, offering a more intuitive and engaging user experience, as well as more efficient and cost-effective solutions for retailers. The journey from Web 1.0 to Web 3.0 represents a transition towards a more **decentralized system**, with Web 3.0 promising far-reaching impacts on the future of e-commerce and overall web-use. As per [58]*Wiki Semantic Web*, World Wide Web Consortium ***(W3C)*** proposes Semantic Web i.e., Web 3.0 technology to **make internet data machine-readable.** *Agarwal P.* [2012]'s Infographic-7, Appendix-9 depicts the differences between Web 1.0, Web 2.0, Web 3.0. [59]Faiz*ar* [2018] presents a periodic timeline of Semantics of Information Connections in Chart-16, Appendix-9. Going to future, *Patel* [2013] depicts about the Web 5.0 in his summary stating that Web 1.0 as **web of cognition** i.e. **read only web**, Web 2.0 as **people-centric and participative web** i.e. **read-write web**, Web 3.0 as **web of knowledge connection** i.e. **read-write-execution web**, Web 4.0 as **Ultra-Intelligent Electronic Agent** i.e. **read-write-execution web with concurrency** and Web 5.0 as a **quasi-emotive web** described as fifth generation web in his paper. *Patel* [2013] and *Singh & Gulati* [2011] both provide overviews and comparisons of these web generations, highlighting the shift in technology, networking, and user engagement. [60]*Kollmann & Lomberg* [2010] emphasize the role of e-business in this evolution, with Web 3.0

[56] https://firstmonday.org/ojs/index.php/fm/article/view/2125/1972
[57] https://www.techradar.com/news/internet/web/is-web-1-0-making-a-big-comeback-1291121
[58] https://en.wikipedia.org/wiki/Semantic_Web
[59] https://ahmadfaizar.blogspot.com/2018/08/evolution-of-web-web-10-web-20-web-30.html
[60] https://www.igi-global.com/chapter/web-web-web/41283

characterized by demand-oriented systems and individualized products and services. *Onday* [2019] extends this discussion, introducing the concept of **Web 6.0** and the potential for further evolution in the future.

Out of this all understanding, the SERP can be called as the byproduct of Web 3.0. So how this journey from Web 1.0 to Web 3.0 has been, we can better understand from comparative Table-8, Appendix-9. **Web 4.0 and its impact on SERP:** Web 4.0 *(synonymously known as Symbiotic Web, Web of Things, Web Social Computing)* is the latest evolution of the web, is characterized by **pervasive computing, ubiquitous computing, and symbiotic web technologies**, *Almeida* [2017]. This new paradigm, which integrates big data, augmented reality, machine-to-machine communication, cloud computing, and AI, is expected to have a significant impact on SERPs, *Asaad Khaleel* [2021]. The increased interactivity and collaboration in Web 4.0 will likely lead to more personalized and intuitive search results, *França et al.* [2021]. However, it may also present challenges such as the potential for increased web-page spam and the need for effective measures to combat it, *Dom et al.* [2005]. Web 4.0 *(the fourth generation of the W3)* is intended for a more collaborative, interactive, and user-friendly approach to web development. It aims to make the web more immersive and intelligent, with features such as ubiquitous computing, user control over data, and semantic search, [61]*Agile District* [2022] and [62]*Strivemindz.com* [2023]. This evolution impacts the SERPs by enabling improved search results, better personalization, and enhanced understanding of user intent, leading to more accurate and relevant search outcomes, [61]*Agile District* [2022] and [63]*BairesDev.com*. The increased interconnectivity and integration of online platforms and services in Web 4.0 also contribute to a more seamless and tailored user experience on SERPs, [63]*BairesDev.com*. Web 4.0 technology, which includes **IoT, AI, and blockchain**, has the potential to revolutionize e-commerce platforms by enhancing service quality and customer experience, *Martasari* [2023]. Semantic search engines, using technologies like RDF and OWL, can overcome limitations of traditional search engines by providing accurate and relevant results based on logical reasoning and understanding of web content, *Rajkumar et al.* [2012]. The search engine algorithm called PCF-Engine uses probability-based similarity function to assess the accuracy of facts provided by web pages, improving the trustworthiness of search results, *Srikantaiah et al.* [2012]. Chart-17, Appendix-9 represents the five characteristics and their traits for Web 4.0. Chart-18, Appendix-9 is a visualization of how the journey of W3C has been from Web 1.0 to Web 4.0.

[61] https://www.linkedin.com/pulse/web-40-explained-brief-agiledistrict
[62] https://www.strivemindz.com/blog/what-is-web-4-0/
[63] https://www.bairesdev.com/blog/web-4-0-its-time-to-talk-about-it/

Semantic relation between Web 3.0 and Web 4.0: The Semantic Web, also known as Web 3.0, is an extension of the current web that focuses on giving data well-defined meaning, enabling better cooperation between computers and people, *Agarwal P.* [2012] and *Szeredi et al.* [2014]. Web 4.0, on the other hand, is a more complex evolution of the web paradigm, incorporating multiple models, technologies, and social relationships *Almeida* [2017]. The relation between the two lies in the fact that **Web 4.0 builds upon the principles of the Semantic Web,** using its core principles of associating meaning with data and advanced semantic techniques for search, integration, and analysis, [64]*Sheth & Thirunarayan* [2012]. Web 3.0 focuses on providing machine-readable content on the web, aiming to decrease human tasks and decisions and leave them to machines. It brings about semantic integration and data interoperability, allowing for content personalization and a more intelligent web, [65]*Raghavan* [2023], [66]*cronuts-digital* and *Aghaei et.al* [2012]. On the other hand, Web 4.0 is characterized by a more collaborative and interactive approach to web development. It aims to make the internet more user-friendly, efficient, personalized, interactive, intelligent, and collaborative, with features such as social networking, blogs, wikis, and video-sharing sites. Web 4.0 is also known as **the symbiotic web**, aiming to blur the line between human and machine, **creating a virtual world** where users can fully immerse themselves and live their daily lives, [65]*Raghavan* [2023], [66]*cronuts-digital* and [61]*Agile District* [2022]. Therefore, Web 3.0 and Web 4.0 represent distinct stages in the evolution of the web, with Web 4.0 building upon the advancements of Web 3.0 to create a more interactive and collaborative online environment. Semantic web (Web 3.0) and Web 4.0 are also related in terms of **their use of semantic technologies**. Semantic technologies, such as knowledge graphs, play a crucial role in both Web 3.0 and Web 4.0. Web 3.0 utilizes semantic extraction and communication to achieve decentralized semantic sharing and precise information transfer, *Lin et al.* [2022]. It focuses on enabling participants to read, write, and own content in a decentralized manner, driven by **blockchain, semantic communication, edge computing, and artificial intelligence**. On the other hand, Web 4.0, also known as **Knowledge 4.0**, leverages semantic technologies, particularly knowledge graphs, to manage knowledge in the digital economy, *Rettinger et al.* [2018]. Knowledge graphs enable the representation of unstructured content, like text documents, in a machine-processable format, facilitating value creation and interoperability in cyber-physical systems. Therefore, **both Web 3.0 and Web 4.0 rely on semantic technologies** to enhance information sharing, knowledge management, and value creation. The Chart-19, Appendix-9 draws a depiction of timeline for surge of Web 1.0 Web 2.0 and Web 3.0 while displaying their corresponding nomenclature. SAS-29: Products of Web 1.0; SAS-30: Products of Web 2.0 and SAS-31:

[64] https://corescholar.libraries.wright.edu/books/70/
[65] https://acodez.in/evolution-of-the-world-wide-web/
[66] https://cronuts.digital/en/diferencias-entre-web-10-web-20-web-30-web40/

Products of Web 3.0 | Source: *Gan et al.* [2023] in the Appendix-1 gives a visual idea about the different products and platforms of Web 1.0, Web 2.0 and Web 3.0.

Differences between Web 3.0 and Web 4.0: Web 3.0, the next phase of the internet, is characterized by decentralized architecture, intelligent applications, and a more user-friendly experience, *Daniel Harris* [2008], *Gan et al.* [2023], *R-S-Kumar* [2022]. It combines the best elements of Web 2.0 and Web 1.0, with new features such as data customization and blockchain technology, *R-S-Kumar* [2022]. However, it is still in the development stage and faces challenges such as technical implementation and regulatory issues, *Fan et al.* [2023]. The transition to Web 3.0 is expected to have a significant impact on e-commerce and other sectors, *R-S-Kumar* [2022]. The differences between Web 3.0 and Web 4.0 are not explicitly discussed in these papers, but it can be inferred that Web 4.0 will likely build upon the decentralized and intelligent features of Web 3.0, possibly incorporating advancements in AI, 5G, and other emerging technologies, *Fan et al.* [2023]. Web 3.0 is characterized by decentralized data ownership, AI, and blockchain technology, enabling more extensive data searches and direct communication between applications, [67]*William* [2022]. On the other hand, Web 4.0, also referred to as the Symbiotic Web, is still an evolving concept with no clear definition. It is envisioned to enable symbiotic interaction between people and machines, potentially incorporating advanced interfaces such as mind-controlled interfaces and making web material easier for computers to interpret and load with better quality and speed, [67]*William* [2022]. While Web 3.0 is already defined by its characteristics, Web 4.0 is still a subject of ongoing development and speculation. *Gupta et al.* [2023] also has more or less similar kind of observations. As per them, Web 3.0 and Web 4.0 are both next-generation web architectures that aim to bring about significant changes in various fields. Web 3.0 is characterized by its decentralization, security, and intelligence, and it incorporates technologies such as blockchain, artificial intelligence, semantic web, and decentralization. It has implications for e-commerce, social networking, finance, healthcare, and education, and has attracted substantial research interest globally. On the other hand, Web 4.0 is associated with the digital marketing domain and refers to the use of the internet, computer communication technologies, and digital interactive media to achieve marketing goals. It utilizes advanced computer networking technology to find new markets and customers in a cost-effective manner. **Digital marketing 4.0** is considered an evolution of marketing, with demand management at its core, and it focuses on timely, relevant, individual, and cost-effective communication with consumers, *Vasiliu-Feltes* [2023].

[67] https://blog.temok.com/web-30-vs-web-40/

In our observation of above literature review, **Web 4.0** is more pertaining toward creation of a **virtual world** while **metaverse** can be related to **Web 5.0**. We shall discuss about metaverse in the last section of Chapter-3. Based on above reviews, below Chart-20 sums-up the sub-variables of Semantic Web:

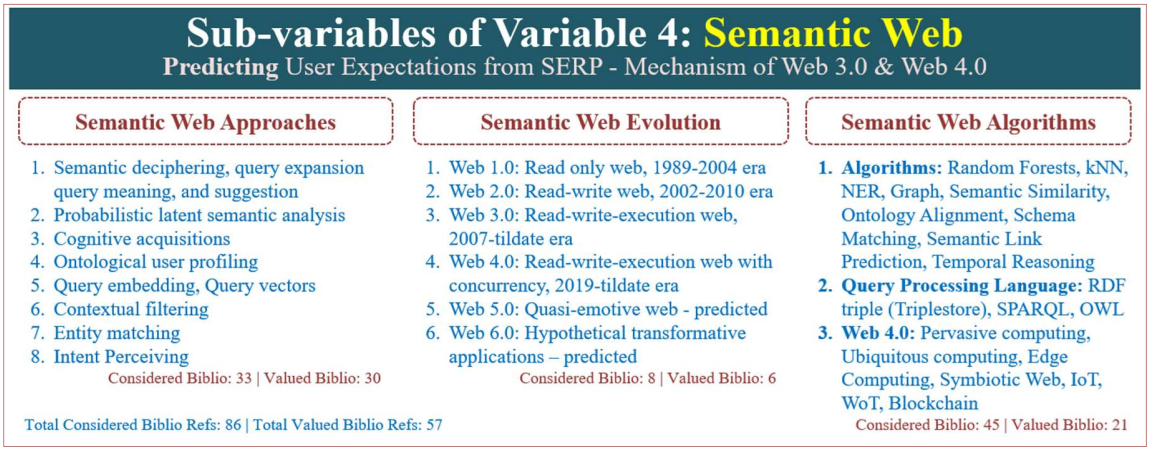


Chart-20: Sub-variables of Variable 6 'Semantic Web'

Discussion *(Possible ground for hypothesis conceiving)*: Most enviable from Table-11 the search engines focus has developed deep-down up to serving individuals by peeping into their SERP behaviour so predicting everything possible around Search User is a custom for them now. Search Engines breathe with their search users' yawning. Setting this thesis between Web 3.0 and Web 4.0 can be a tricky one but that is doing what gives more sentimental richness of this thesis. So, checking people's knowledge and understanding about Web 4.0 is very crucial in finding of this thesis, a ground for another hypothetical question.

Discussion: Current emphasis of Search Engines is increasingly on semantic search and understanding the relationships between entities. Optimizing content to align with entity-based search and semantic relevance is crucial at present. Google's focus on personalization and predictive results suggests a shift toward understanding user intent up to nuanced understanding. This involves adapting content to align with potential user intent and preferences.

2.5. SERP Personalisation *(Tangible Behaviour that drives the Personalization)*

In this section we intend to address search user' tangible behaviour i.e. on-screen behaviour, after receiving the SERP; which can be stored and analysed directly in the form of clicks, scrolling, bounce-backs, or factors directly impacting their on-screen activities, such as reformulation of the query, using features of SERP

interface and how their tangible behaviour drives search engine to produce personalized SERP. We shall study the treatment of these feedbacks by search engine in producing improvised SERP as a custom. Both can be categorized by relating the output as "**output due to search user' intelligence**" versus "**output due to the Search Engine** *(inhouse SERP features and algorithms)* **and Content Creators** *(outsourced)* **for Search Engines**".

SERP Personalization Factors: Research on the personalization factors of SERP has identified several key elements. *Seig et al.* [2007] emphasizes the importance of user context, proposing the use of **ontological user profiles** to tailor search results to individual interests and preferences. *Khopkar et al.* [2003] found that personalization features, such as email, business and financial information, and entertainment listings, were available on a limited number of search engine websites. *Jayanthi & Jayakumar* [2011] proposed an integrated **page ranking algorithm** that considers user preferences in content and link to improve result relevance. *Hannak et al.* [2013] developed a methodology to measure the **extent of personalization in SERP**, finding that it is primarily influenced by a **user's logged-in account and IP address**. Personalization factors of search engine results pages include **tailoring search results for different users who have issued the same query,** *Rashed et al.* [2022]. This personalization can lead to the **Filter Bubble algorithmic bias**, where certain users are unable to access relevant content based on the search engine's algorithm, *Singhal et al.* [2020] & Ge *et al.* [2019]. Research in this area often involves creating different **Google accounts** with varying features and **comparing the search results** obtained from each account. Other factors that can influence personalization include geographic location, logging in and out of a Google account, different networks *(home, phone, university),* different search engines, and different web browsers. Additionally, a hierarchical recurrent neural network **(HRNN)** can be used to exploit sequential information and generate user profiles for personalization. Search users prefer to see search results in their preferred language and relevant to their country i.e. **Language and Market Country**, [68]*Membrillo* [2020]. Search engines consider an individual's **search history and browsing behaviour** to personalize the search results. This means that the results may vary for different users, even if they use the same search language, [68]*Membrillo* [2020] & [69]*Marketbrew-3*. Information such as a user's location and the device they are using is also used to personalize search results, [68]*Membrillo* [2020]. The impact of website personalization on search engine rankings can be both positive and negative, depending on how it is implemented. Tailoring the **user experience on a website** can also influence the personalized search results, [70]*Mehta* [2023] & [71]*Torres* [2021-2023]. SERP personalization influences the specificity and customization of

[68] https://www.forbes.com/councils/forbesagencycouncil/2020/01/23/how-to-incorporate-personalization-into-your-seo-strategy/
[69] https://marketbrew.ai/how-search-engine-personalization-affects-rankings
[70] https://abmatic.ai/blog/impact-of-website-personalization-on-search-engine-optimization
[71] https://thegray.company/blog/personalization-seo

search results based on individual user preferences, behaviours, and historical search data. Search engines employ various algorithms and techniques to tailor SERPs to better meet the unique needs of each user. In the purview of above arguments and findindgs, here are key factors how the SERP personalization is met:

1. **SERP takes relevance to User' Search History and Behaviour:** Search engines track the search history of users to understand their interests and preferences. **Previous search queries, clicked links**, **information-seeking behaviour and engagement patterns** help personalize future search results, presenting content that aligns with the user's historical behaviour. If a user frequently searches for specific topics or **clicks on types of content**, the SERP becomes more specific by prioritizing similar results. The more closely a search result aligns with a user's historical search patterns, the more specific and tailored the result will be to the user's interests. This can include **suggested articles, products, or websites** based on the user's preferences and behaviour, enhancing the relevance of the search results.
2. **Location-Based Personalization:** Search engines often use geolocation data to personalize results based on the user's location. **Localized results**, such as nearby businesses, services, or events, are prioritized to provide more relevant information based on the user's geographical context.
3. **Device-Specific Personalization:** SERPs may vary based on the device used for the search. Search engines consider the **type of device** *(desktop, tablet, mobile)* to optimize the presentation of results for the specific **screen size** and user experience. For example, mobile users may see results optimized for smaller screens and **touch interactions**, while desktop users may see results formatted for larger screens and **mouse-driven interactions**. With the advancement of technology, intention of ease the user experiences and machine learning programs have resulted in numerous Mobile Digital gadgets such as Smartphones, Tablets/iPads from initially where the journey started with Desktop in 80-90s to laptops in 00-10s. According to *Rimm-Kaufman Group* Mobile searches snatched the 31% of search market from desktop and laptops by 2014, *Moran & Hunt* [2014] - Page 41. Mobile searches here mean the searches on move or in market while shopping. A study conducted by *MediaPost* in 2014s says that more than 50% of people *(43% as per InsightExpress)* use the mobiles in product research before asking a salesperson in shops or shopping malls. This shift is both need based, and style based, *Moran & Hunt* [2014] - Page 42.
4. **Temporal Relevance:** Personalization considers the temporal aspect of a user's search behaviour. Recent searches and trending topics are given more weight, resulting in more specific and timely results that align with the user's current interests.
5. **Social Media Integration:** Integration with social media data provides additional contextual information which impacts the result specificity. Some search engines integrate social media data into their algorithms for personalized results. Information from a user's social profiles, connections, and activities may lead to more personalized and specific recommendations that align with the user's social circles and interests.

Social signals, such as content shared by the user's social connections, may influence SERP personalization.

6. **Content Engagement Signals (Feedback):** Search engines analyse user engagement signals, such as **click-through pattern, click-through rates**, **dwell time,** and **bounce rates**, to refine SERP personalization. High engagement with specific types of content may influence the ranking and presentation of similar content in future searches. Content that aligns with a user's preferences and engages them is more likely to be promoted in personalized results. The amount of time a user spends on a webpage, as well as their interactions with the content *(such as scrolling and engagement with multimedia elements),* influences SERP personalization. Content that has historically captured the user's attention is given preference.
7. **Personalized Advertisements:** Advertisements displayed on the SERP may be personalized based on the user's **browsing history, interests, and online behaviour**. This targeted advertising approach aims to present users with more relevant and engaging ads.
8. **Customization Based on User Settings:** Users can often customize their search preferences and settings. Search engines consider these user-defined preferences, such as **content in a specific language or from a particular region and search history settings**, including preferences for **safe search, content types, and display preferences** when personalizing the SERP. Personalization takes these user-defined settings into account, refining the specificity of the presented results.
9. **Predictive Analytics and Machine Learning:** Advanced algorithms, including ML and predictive analytics, play a role in anticipating the user's interests and accordingly the SERP personalization. These algorithms learn from user behaviour over time, predicting what specific content a user is likely to find relevant and presenting it with greater specificity.
10. **User Demographics:** While search engines prioritize user privacy, demographic information may play a role in SERP personalization. Factors like **age, gender, and interests** inferred from user behaviour contribute to a more nuanced personalization.
11. **Search Intent Recognition:** Analysing the user's search intent helps personalize results. Understanding whether the user is looking for **information, products, or local services** allows the SERP to present more relevant content aligned with the user's goals.

SERP User Satisfaction Parameters: These parameters can be seen in two aspects, one the overall SERP Satisfaction, two the impact of various SERP Features and Snippets influencing the user's satisfaction.

Search User Satisfaction related to overall SERP: A range of studies have explored the measurement of user satisfaction with search engine results. *Yu et al.* [2012] proposed a metric named 'Repetition rate involved in Discounted Cumulative Gain ***(RDCG)'*** that incorporates user behaviour, which was found to be more effective than traditional metrics such as Mean Reciprocal Rank ***(MRR)***, Mean Average Precision ***(MAP)*** and Normalized Discounted Cumulative Gain ***(NDCG)***. They further discussed the challenges of measuring user satisfaction and the potential for different evaluation approaches. *Dan & Davison* [2016] highlighted the importance of predicting and characterizing user satisfaction and discussed methods for improving search engines based on user behaviour. They state that understanding and predicting search user satisfaction is crucial for improving search engine ranking models and user retention rates. *Awadallah & White* [2013] emphasized the need for personalized models of satisfaction, as user behaviour can vary significantly. These studies collectively underscore the complexity of measuring user satisfaction and the need for a nuanced, multifaceted approach. One approach is to analyse user interaction data, such as **mouse cursor movement**, to predict satisfaction. This can be done by constructing detailed universal interaction sequences and extracting informative subsequences from them, *Mehrotra et al.* [2017]. *Liu et al.* [2014] propose a two-stage SERP result skimming model for web search catching the expected user mannerism around selecting and clicking a particular search result. Chart-21, Appendix-9 describes the naturalistic procedure of treating a search result by a search user. Another method involves utilizing a **long-short time memory network** to learn the feature representation of the search behaviour sequence and establish a user satisfaction classifier, *Ling & Alin* [2019]. Additionally, user satisfaction can be evaluated based on the types of search results and the experiences of users in using those results, *JungHwa* [2020]. In the context of image search, both offline metrics based on **topical relevance and image quality judgments**, as well as online metrics based on **image click information**, can be used to assess user satisfaction, *Zhang et al.* [2018]. Search Satisfaction Surveys collect feedback from users after they perform a search. They can include questions about the **relevance, speed, and ease of use** of the search results, as well as overall satisfaction with the search experience, [72]*Kersting* [2024]. *Palanisamy & Liu* [2019] measured the satisfaction by various factors such as **response time, URL correctness, results display, overall impression, and website stability**. They state that the quality of a search engine is measured by the number of websites collected by the search engine, accuracy of search results, display of search results in a sorted order, relevance of search results, easy to remember URL of search engine, less ads in the search page, and the short loading time of search results. Their research model for users' search satisfaction evolved from the literature. The model is empirically tested and validated by obtaining survey-inputs from 101 students from a Canadian University. **The factors affecting users' search satisfaction are user characteristics, search characteristics, search engine characteristics, and clicking behaviour of advertisements.** The data analysis

[72] https://www.linkedin.com/advice/0/how-do-you-measure-user-satisfaction-your-1c

validates the impact of website features, search engine contents and search problems on user's search satisfaction. User's search behaviour includes the frequency of different search data input, navigational search and the usage of different devices positively influences user's search satisfaction. Besides, the advertisements clicking negatively influences user's satisfaction. *Pundkar & Baheti [*2015] propose a time ranking algorithm-based mechanism in their ontological research with the purpose of increasing the Search Satisfaction level of users; their proposal says that not only rank the pages based on its importance but also based on time user spends on each page. SERP results contribute to customer *(searcher's)* satisfaction by providing users with relevant, reliable, and user-friendly information in response to their search queries. Here are several ways in which SERP results ensure customer satisfaction:

1. **Relevance to Search Query:** SERPs aim to deliver results that directly address the user's search query. When the displayed results are closely aligned with the **user's Intent and Location** it increases the **likelihood of satisfaction** as users find the information they were seeking.
2. **Local Relevance:** For users seeking local information, such as businesses, services, or events, SERPs that provide accurate and relevant local results contribute to satisfaction. Localized information ensures that users find what they need in their immediate vicinity.
3. **Trustworthiness *(also know as Accuracy Rate)*:** Search engines prioritize content from reputable sources and authoritative websites. When users see results from **trustworthy sites,** they are more likely to feel confident in the accuracy and reliability of the information, contributing to overall satisfaction. Trust represents the degree to which a particular information source can be trusted to provide accurate and correct information, and is measured as information source relative trustworthiness, *Houmb et al.* [2006].
4. **User-Friendliness:** The layout and design of the SERP contribute to user satisfaction. **Clear presentation, easy navigation, and well-organized snippets** make it more efficient for users to find the information they need, enhancing the overall user experience.
5. **Mobile-Friendliness:** As mobile device usage continues to rise; search engines prioritize mobile-friendly results. SERPs that are optimized for various screen sizes and devices contribute to a positive user experience, leading to higher satisfaction among mobile users.
6. **Loading Time:** Speed is crucial for customer satisfaction. SERPs that load quickly and efficiently contribute to a positive user experience. Slow-loading pages or delays in displaying search results can negatively impact satisfaction. However, as per *Yu et al.* [2012] most of search engine users **think highly of the accuracy and comprehensiveness of search results** and not care about search speed and information refresh rate by contrast perhaps in the shadow of accuracy and comprehensiveness.
7. **Diversity *(Featured Results)*:** SERPs often include rich snippets, featured snippets, and multimedia blocks/carousels *(such as images, videos, news),* Zero Click Answer Results that provide users with a more

comprehensive and visually appealing preview of the content, [73]*Sullivan* [2007]. This helps users quickly assess the relevance of the results, leading to higher satisfaction.

8. **Personalization:** SERP personalization tailors results to individual user preferences, increasing the likelihood of presenting content that is personally relevant. Personalized results contribute to customer satisfaction by delivering a more customized and user-centric experience.
9. **Clear Call-to-Action (CTA):** When users are looking for specific actions, such as making a purchase, downloading a file, or visiting a website, a clear and visible call-to-action in the SERP contributes to satisfaction. Users appreciate results that make it easy to take the next step.
10. **Accessibility Features:** SERPs that incorporate accessibility features, such as **alt text for images** or options for screen readers, contribute to a positive experience for users with disabilities. Accessibility considerations enhance satisfaction by ensuring inclusivity.
11. **Consistency Across Devices:** Users often switch between devices, and a consistent experience across different platforms contributes to satisfaction. SERPs that maintain a consistent look and functionality across desktop and mobile platforms enhance user satisfaction.
12. **Feedback and Ratings Integration:** Integration of user feedback, ratings, and reviews into SERPs can help users make informed decisions. Search engines that display user-generated content contribute to customer satisfaction by providing social validation and insights from other users.

User's Utility, Overall scope of Results and Freshness of results are other crucial factors which a Search User seeks from every SERP. These are talked in detail in Chapter-4.

Search User Satisfaction related to Individual SERP Snippets: After having a structured insight about overall satisfaction of SERPs in above section, in this section we intend to focus on specific SERP Snippets and their satisfaction probability. Research on user satisfaction with different SERP snippets has yielded mixed results. *Koyena* [2021] found that **summary snippets** can enhance search success and user trust, while *Jiang & Liamruk* [2020] reported **higher satisfaction with snippet information** over other categories. Jiang and Liamruk further found that participants were more satisfied with SERPs displaying snippet information compared to those displaying published date and citation count information. However, *Maxwell et al.* [2017] discovered a **discrepancy between user perception and performance**, with **longer summaries** not necessarily leading to better outcomes. *Elbedweihy et al.* [2012] further highlighted the importance of considering user type, with casual users preferring a form-based query approach. These findings underscore the need for a nuanced approach to measuring user satisfaction with SERP snippets, considering both user

[73] https://searchengineland.com/google-20-google-universal-search-11232

preferences and actual performance. *Moffat* [2022] & *Bink et al.* [2023] observed that the **richness of snippets** can capture user attention, although ranking remains the most critical factor in relevance perception. As per [74]*Mmeje* [2024], **Rich Snippets, which add a visual layer to an existing result**, have been shown to increase click-through rates. And **Answer-contained snippets**, which provide a concise answer to the user's query, have also been found to improve the effectiveness of search engines. *Bink et al.* [2022] also researched to understand the influence of **featured snippets** on users' credibility judgments and their impact on user satisfaction. Featured snippets have been found to draw more attention, reduce the time spent searching, and increase user satisfaction. *Mao et al.* [2016] showed that **answer-contained snippets** significantly outperformed traditional methods in terms of relevance judgments and information satisfaction evaluations. These findings suggest that the design of SERP interfaces should prioritize providing valuable information without overwhelming users with excessive information categories, and that the content displayed in snippets can have a significant impact on user satisfaction and search experience. One another common method is to conduct a user satisfaction study and collect feedback from participants on their satisfaction with the snippets. Another approach is to use metrics such as **click-through rates and dwell time** to evaluate the effectiveness of different snippets, [75]*Neil Patel-3*. Patel further signifies to note that user satisfaction is subjective and can vary depending on the user's information needs and search context. Therefore, it is essential to consider multiple metrics and conduct user studies to gain a comprehensive understanding of user satisfaction with different SERP snippets. SERP features play a significant role in enhancing customer satisfaction by providing users with diverse, informative, and visually engaging elements beyond traditional organic search results. SERP features contribute to SERP personalization by tailoring the presentation of diverse elements to match the individual preferences, behaviours, and context of each user. Here's how various SERP features contribute to customer satisfaction:

1. **Featured Snippets: Users appreciate quick and relevant information**, leading to higher satisfaction as they can obtain answers without having to click through to a website. The content displayed in a featured snippet is selected based on its relevance to the user's query and historical search behaviour, contributing to a personalized presentation of information.
2. **Rich Snippets:** Rich snippets display structured data, such as ratings, reviews, and pricing information, providing users with a **more personalized and detailed preview** of search results. The inclusion of these snippets is influenced by the user's preferences and the context of their search. Users can quickly assess the relevance and credibility of a result, contributing to a more informed and satisfying search experience.
3. **Knowledge Graphs:** Knowledge graphs present organized information related to entities, events, or topics. Users benefit from a **visual representation of interconnected information**, enhancing their understanding

[74] https://moz.com/learn/seo/serp-features
[75] https://neilpatel.com/blog/serps-guide/

and satisfaction with the depth of content available. SERPs may personalize knowledge graphs based on the user's search history and preferences, ensuring that the displayed information aligns with their interests and previous interactions. *Ensan & Al-Obeidat* [2019] states that the growing availability of knowledge graphs has motivated researchers within the information retrieval community to consider exploiting knowledge graph entities within the ad hoc document retrieval process.

4. **Local Packs:** Local packs display maps and local business listings, providing users with easily accessible information about nearby services. This feature contributes to SERP personalization by presenting results that are relevant to the **user's immediate geographic context**.
5. **Top Stories (News):** Top Stories feature displays the latest news articles, and the selection of articles can be influenced by the user's interests, search history, and the relevance of the **news topic to their information appetite profile**.
6. **Images and Videos:** Users often appreciate a **visual preview of content**, leading to increased engagement and satisfaction. Visual content, such as images and videos, may be personalized based on the user's preferences and engagement history.
7. **Site or Deep Links:** This feature contributes to satisfaction by saving users time and helping them find the most relevant information quickly. Site links provide additional navigation options beneath a search result. The selection of site links can be influenced by the user's past interactions with the website, aiming to present links that are most likely to be relevant to their interests.
8. **Reviews and Ratings:** Reviews and ratings displayed directly on the SERP **contribute to user confidence in making decisions**. Users can quickly assess the reputation and quality of a product, service, or business, leading to increased satisfaction with their decision-making process. If a user frequently engages with content containing reviews or ratings, the SERP may prioritize results with similar characteristics.
9. **People Also Ask (PAA):** The People Also Ask feature presents related questions and answers on the SERP. This feature contributes to customer **satisfaction by anticipating additional queries** and providing users with more information on the topic of interest and the selection of these queries may be influenced by the user's search history and the context of their current query.
10. **X (Twitter) Feeds:** Integrating real-time X feeds in the SERP allows users to access the latest social media discussions related to their query. This feature contributes to **satisfaction by providing a dynamic and current perspective on the topic**. The SERP may display tweets from accounts or topics that align with the user's preferences.
11. **AMP (Accelerated Mobile Pages):** AMP results, which prioritize fast-loading mobile-friendly pages, contribute to personalized SERPs by considering the user's device preferences and historical interactions with mobile-friendly content.

12. **Breadcrumb Navigation:** Breadcrumb navigation can be personalized by considering the user's past interactions with specific website hierarchies. The SERP may present breadcrumb trails that align with the user's previous navigation patterns.

SERP Personalization Algorithms *(the calculative algorithmic approaches)*: There are two SERP algorithms designed for personalization of SERP. The first algorithm is the Personalized PageRank ***(PPR)*** algorithm, which is used for ranking the SERPs, *Yao et al.* [2021] & *Shinde et al.* [2020]. The PPR algorithm focuses on improving the quality of web search by implementing profile-based personalization in both active and passive manners, *Vijayalakshmi et al.* [2021]. Vijayalakshmi further adds that **Machine learning algorithms analyse** the actual content of posts to provide more personalized feeds based on the user's recent history. The second algorithm is the Reinforcement Learning-based framework for Personalized Search ***(RLPS framework)***, which utilizes reinforcement learning to track sequential interactions between the user and search engine and continuously update the personalized ranking model based on the user's real-time feedback, *Shobby et al.* [2015]. Within the RLPS framework, two models are implemented: **RLPS-L** *(L stands for **Listwise** here)*, which trains the ranking model with document lists, and **RLPS-H** *(H stands for **Hierarchical** here)*, which improves model training by introducing document pairs and a layered structure, *Jayanthi & Jayakumar* [2012] and *Yao et al.* [2021]. Some algorithms prioritize previously visited websites and local news sites, but not necessarily websites featuring the user's ideology, *Ge et al.* [2019]. Sequential information, such as the **order of issued queries** and more recent sessions, is also considered to infer the user's real interests, *Shinde et al.* [2020]. *Hu Nai* [2002] proposed a **high-efficiency recommendation** algorithm that includes dimensionality reduction and item-based recommendation methods. *Venugopal et al.* [2009] introduced an adaptive weight update algorithm that captures user interests and dynamically updates a user profile to refine search queries. *Makeev et al.* [2014] study the usefulness of personalized features for improving the quality of aggregated search ranking on Yandex by training a number of machine-learned blending algorithms through blending of user search history. They proposed a framework with three classes of personalized features as aggregated search need, certain vertical preference (most important of three) and vertical navigationality, which allowed them to build the ranking function outperforming their competitive baseline. *Chirita et al.* [2004] developed a platform for personalized ranking, automating the hub selection process and integrating user surfing behaviour analysis. *Huang et al.* [2004] presented an online adaptive method for personalization, which collects user interests from their profile and measures similarity at the semantic level. These algorithms aim to enhance the accuracy and efficiency of personalized SERPs. [76]*searchenginejournal-2025* sums-up a complete list of Google Algorithms which turned a complete picture of SERP in regard of SERP personalization alongside SERP

[76] https://www.searchenginejournal.com/google-algorithm-history/

quality and relevance maintenance; these algorithms have been *Florida, Big Daddy, Jagger, Vince, Caffeine, Panda, Freshness Algorithm, Page Layout Algorithm, Venice Update, Penguin, Exact Match Domain (**EMD**), Payday, Hummingbird, Pigeon, Mobilegeddon, Quality Updates, RankBrain and Fred*. Above findings envisage that Search engines employ various algorithms to personalize SERPs based on individual user preferences, behaviours, and context. While the specific details of these algorithms are proprietary and closely guarded by search engine companies, there are several general types of **calculative algorithmic approaches** commonly used for SERP personalization as below:

1. **Relevance Algorithms:** These algorithms aim to tailor search results to individual users' preferences, *Geetha Rani* [2013], *Sontag et al.* [2012], *Rathod et al.* [2015]. These algorithms utilize various approaches such as word-correlativity networks, conditional relevance judgments, user profiles, and generative models of relevance to enhance the accuracy and effectiveness of search results for each user, *Xiao-xing* [2008]. By incorporating user interests, preferences, and search history, these algorithms can significantly improve retrieval performance, particularly for ambiguous queries, *Golbus et al.* [2014]. However, the challenge lies in balancing search quality with privacy concerns, as personalized search often involves collecting and aggregating user information, which can raise privacy issues. By understanding what types of content, a user finds relevant, the algorithm aims to present more personalized and tailored results.
2. **Collaborative Filtering:** Collaborative filtering-based recommendation algorithms are widely utilized in recommendation systems to address data sparsity issues. These algorithms aim to capture user interests accurately by establishing scoring relationships between user models and evaluating resources at different levels. By calculating similarities and generating recommendation lists, these algorithms provide accurate and fast personalized recommendations, proving to be effective with high accuracy and application value. Additionally, studies have explored enhancing e-commerce personalized recommendation systems through improved collaborative filtering algorithms, enabling more precise user need mining and accurate commodity information recommendations. Collaborative filtering algorithms consider the preferences and behaviours of users with similar profiles. If two users share similar interests or search histories, the algorithm may recommend content that one user has found relevant to the other.
3. **Content-Based Filtering:** Content-based filtering algorithms focus on the characteristics of the content itself and how well it matches a user's preferences. This can include analysing keywords, topics, and other content attributes to personalize search results.
4. **Machine Learning Models:** MLAs, such as neural networks and decision trees, are increasingly used for SERP personalization. These models can learn and adapt over time, considering a wide range of factors to predict the user's preferences and deliver more personalized results.

5. **Contextual Algorithms:** By understanding the context in which a search is performed, such as the user's location, device type, and time of day, the algorithm can adjust the presentation of results to better suit the user's immediate needs.
6. **Temporal Algorithms:** By leveraging user query data's temporal aspects, these algorithms can better understand current user needs and intent, leading to more accurate and efficient personalized recommendations. Temporal algorithms prioritize recent and timely information. They consider the recency of a user's searches and interactions to present content that aligns with the user's current interests, ensuring that the SERP reflects the latest developments.
7. **Session-Based Algorithms:** Session-based algorithms consider a user's search session, analysing the sequence of queries and interactions. By understanding the progression of a user's information-seeking behaviour during a session, these algorithms personalize results to match the evolving context.
8. **Implicit Feedback Algorithms:** Implicit feedback algorithms analyse user behaviour, such as **clicks, scroll depth, and dwell time**, to infer preferences. By interpreting implicit signals, the algorithm refines its understanding of what content is most engaging and relevant to the user.
9. **Explicit Feedback Algorithms:** Explicit feedback algorithms consider direct feedback from users, such as **likes, shares,** and explicit preferences set by the user. This explicit input helps refine the algorithm's understanding of the user's preferences and interests.
10. **Personalization based on Social Signals:** Some algorithms incorporate social signals from a user's social media accounts to personalize results. Information from social networks may influence the ranking and presentation of content on the SERP. Below Chart-22 depicts the sub-variables of SERP Personalization studied above:

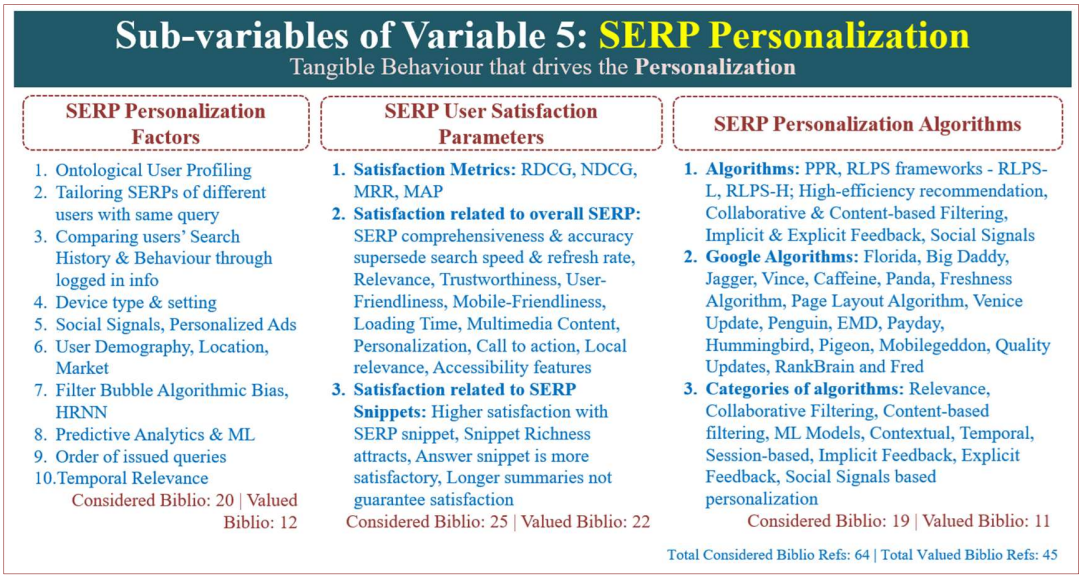


Chart-22: Sub-variables of Variable 5 'SERP Personalization'

Discussion *(Possible ground for hypothesis conceiving)*: Most of search users think high of the accuracy and comprehensiveness of search results, and not care about search speed and information refresh rate by contrast. This finding suggests that when quality and relevancy is of top priority, people can adjust delay in SERP display. Users like crisp content, longer content doesn't guarantee a desired satisfaction.

Discussion Scrutiny: The goal of personalization is to continuously improve the relevance and satisfaction of search results for individual users based on their history, behaviour, location, and other contextual factors. Search engines prioritize user privacy and implement measures to ensure that personalization is done responsibly and transparently. While personalized results can enhance relevance and specificity of results, users should be mindful of potential limitations, such as the formation of filter bubbles, and occasionally explore diverse search strategies to access a broader range of information. SERP features contribute to SERP personalization by dynamically presenting information that aligns with the user's interests, search history, and contextual factors. Through the customization of featured elements, search engines aim to create a more personalized and relevant search experience for individual users.

2.6. Search Skills *(Persistent Intangible Search Behaviour reduces Intent Ambiguity)*:

This variable revolves around three axes of **Expertise, Experience and Exposure**. Expertise in terms of technical capabilities, experience in terms of knowledge up to shrewdness level and exposure in terms of follow-ups of information sources and algorithms. The scale of these **3Es** determines the intensity of search skills of a search user, which ultimately leads to their intangible behaviour in accessing the best of information through their search query formulation and reformulation. These 3Es make a setup of cognitive layering on pattern of **'mental models'** in user' mind. So, we term it as the another "mental model" set of Information Search Skills. These mental models vary person to person and so vary the user search behaviour. This search behaviour is generated due to their cognitive abilities rather than what they act due to what is produced to them. And estimating about this behaviour is near to most difficult task because no one can see or assume it in real measurable terms thus it can be counted as an **intangible search behaviour.** Its intensity and accuracy depend on the tendency of pursuing the 3E aspects of information acquiring and search engine updates. Let's try to understand these 3Es one by one.

Experience - Information Search Mental Models: Mental models of information search on search engines have been studied by researchers for over two decades. As per [77]*Thurow* [2009] mental model is an explanation of a person's **thought process** about how something works in the real world, faithfully representing root motivations and matching behaviours. According to a study conducted in 2011, users' mental models of Web Search Engines remain largely inaccurate and incomplete, despite improvements in their alignment with designer's conceptualizations, *Mlilo & Thatcher* [2011]. Another study conducted in 2009 found that everyone has a mental model about how a website or a search engine works, and no one person has the same mental model as another person, [77]*Thurow* [2009]. Studies have also shown that users' mental models of search engines are relatively unsophisticated, which can hinder their ability to use search engines efficiently and effectively, *Thomas et al.* [2019]-2 *& Thomas et al.* [2019]. Several cognitive processes, such as **memory, attention, comprehension, problem solving, executive control, and decision making**, are involved in the process of information search on the internet, *Karanam et Oostendorp [*2020]. Here in the context of this thesis a mental model can be attributed to a capacity of compiling skills and updates, understanding search nuances and their modalities. **Computational cognitive models** can be used to simulate the effects of cognitive factors and improve our understanding of information search behaviour, *Wai-Tat Fu* [2020]. These models can generate explicit predictions on user behaviour and complement existing computational techniques for information retrieval, *Karanam et al.* [2017]. Computational cognitive models like **CoLiDeS+** have been shown to predict user behaviour significantly better than traditional models, especially on difficult tasks, *Nakhoda et al.* [2016]. Adjusting and developing mental model completeness scales can help improve users' interactions with search engines and facilitate the systematic development of IR systems. Another article on Search Engine Land explains that a mental model is an explanation of a person's thought process about how something works in the real world, faithfully representing root motivations and matching behaviours, [77]*Thurow* [2009]. The mental models of information search on search engines also refer to the internal representations and understanding that users have about how search engines work. Research indicates that users expect search to have three components: a box to type words, a search button, and a list of top results on a new page, known as the SERP, [78]*Nielsen* [2005]. While there have been efforts to understand and elicit generic mental models of search engines from users, it is important to consider these mental models critically and not take them at face value, *Khoo & Hall* [2012]. Research on the mental models of information search on search engines reveals a range of factors that influence searchers' understanding and use of these tools. *Murtuza et al.* [2010] observed in their research that users consume content proportional to popularity but otherwise do not differentiate between head and tail items; appreciating the tail less than the head. *Thomas et al.* [2019]-2 found that while searchers have a

[77] https://searchengineland.com/seo-searcher-mental-models-27949
[78] https://www.nngroup.com/articles/mental-models-for-search/

good grasp of concepts like **popularity and personalization**, they struggle with understanding **recency and diversity** and often have misconceptions about payment for ranking. This suggests a need for more explanatory annotations in search results. Similarly, *Chanceaux et al.* [2012] developed a computational cognitive model that simulates the **oculomotor scanpath of web users** during information search, highlighting the **importance of visual, semantic, and memory processes** in this activity. *Slone* [2002] further emphasizes the role of **experience, motivation, and mental models** in determining search approaches and sources used. These studies collectively underscore the need for search engines to better understand and cater to the mental models of their users to improve the effectiveness and efficiency of information search. Now when we know the concept of mental models in search, let's move on to understand how different mental models i.e. search users access the information.

The intangible behaviour that determines the ease of information accessibility on search engines: 3E factors are increasingly recognized as fundamental to the way search engines evaluate and rank contents up to the exactitude of a search query. Google's Search Quality Rater Guidelines recent update includes E-E-A-T *(initially E-A-T)* as essential components for assessing the quality of content. E-E-A-T stands for Experience, Expertise, Authoritativeness, and Trustworthiness. On the cognitive parameters, I try to assume third E i.e. Exposure a combination of Authoritativeness, and Trustworthiness, because both are attained majorly through Exposure to Internet world. So, to me E-E-A-T can be justified with 3E simply. Research has shown that experts outperform novices in cognitive and perceptual tasks, indicating the **impact of expertise on search behaviour**. Additionally, the development of expertise is characterized by the acquisition of specific skills and knowledge in a domain, which leads to changes in thinking and learning processes. The ease of information accessibility on search engines is influenced by a range of **intangible factors**. *Arrue et al. [*2007] highlights the importance of web accessibility in search engine results, particularly for users with disabilities. *Lopatovska et al.* [2012] identifies the value of search engine features such as convenience, independence, privacy, and the negative impact of being unable to use search engines. *Daniel Rose* [2006] underscores the need for search interfaces to align with user information-seeking behaviour, suggesting that a better understanding of user goals and context can enhance the effectiveness of search interfaces. The ease of information accessibility on search engines is determined by the **design of their user interfaces**. It is essential for search engines to offer user interfaces that are easy to use and accessible to all, *Gray & Fu* [2000]. The accessibility and usability of search tools are crucial factors in ensuring that users can find specific and appropriate information on the internet, *Lawrence & Giles* [1999]. The goal is to improve interface design based on individual feedback from users, *Buzzi et al.* [2004]. The ease of information accessibility on search engines is determined by a variety of factors, including the **quality of the content, relevance to the search query, user engagement, website**

authority, website speed, mobile-friendliness, and security, [79]*Unesco.com-1* [2023]. These factors are used by search engines to rank websites and determine which ones to display at the top of the SERP. In addition to these factors, **information accessibility** is also influenced by the perception of easy access to information. Studies have shown that users are more likely to remember how to access information than the details of the information itself, *Esther Kang* [2023]. This is because ubiquitous internet access has provided easy access to information and has influenced users' attention and knowledge management, *Esther Kang* [2023]. The intangible behaviour that determines the ease of information accessibility on search engines is related to the **concept of web accessibility**. This includes factors such as the ease of navigation, the availability of alternative text for images, and the compatibility with assistive technologies. Research has shown that web accessibility not only benefits people with disabilities but also provides options that are useful to all users in various situations, such as those using mobile devices, older individuals, or people in challenging environmental conditions, [80]*w3.org-wai-1*. Therefore, the intangible behaviour that influences the ease of information accessibility on search engines is closely tied to the principles of web accessibility and its impact on user experience. Furthermore, as per the research made by Microsoft Team of *White et al.* [2009] **Domain experts** search differently than people with little or no domain knowledge *(we talked about this aspect in previous topic above too)*. In the above section we found that domain experts employ different search strategies and are more successful in finding what they are looking for than non-experts. *Ryen While et al.* [2009] also state that Search Expertise is knowledge of search process thus **not necessary that a domain expert can equally be a search expert too**. Searcher skills, expertise and exposure play a crucial role in making it more accessible to the most relevant and precise information on search engines.

So, depending on the technical, software and lexical aspects of search, a search user's search skills should be weighed in three streams. First, the expertise in utilising the **Information Search Mental Model**; second the **Critical and Smart usability of domain knowledge** and third the **Consistent follow-up on SE' Algorithmic updates**. Here's how these skills contribute to an enhanced search experience:

Expertise - Linguistic Nuances and Contextual Refining *(Technical know-how of Search User in Information Retrieval from Search Engines and interpretation)*:

1. **Addressing Search Intent Ambiguity:** Now when we have talked of "Mental Models" and "Intangible Search Behaviour"; which are interchangeably connected to **ambiguity of Search Intent** as well. As higher the 3Es level as lesser will be the **Intent ambiguity** and vice-versa; means 3E level is inversely

[79] https://www.unesco.org/en/ifap/information-accessibility
[80] https://www.w3.org/WAI/business-case/

proportional to Intent ambiguity. This section investigates the researched solutions to intent ambiguity proposed by various researchers. A range of studies have proposed methods to address the issue of ambiguous search intent on search engines. *Hu et al.* [2015] and *Zamani et al.* [2020] both focus on **diversifying search results** to cater to multiple potential intents, with Hu et al. introducing a **hierarchical structure** for user intents and Zamani et al. suggesting the use of **clarifying questions**. *Radlinski et al.* [2010] and *Awad et al.* [2012] take a different approach, inferring intent from user behaviour and **reformulating queries** to better match user intentions. Users often enter ambiguous search terms without considering the potential for multiple interpretations. For example, Chart-23, Appendix-9 shows the process of how Search intents are identified by Search Engines through 3 Steps **Intent Identification process** involving **Expand, Filter and Cluster steps**. To address this issue, several methods have been proposed. One approach is to provide clarification to the user by presenting clarification data through a user interface, allowing them to specify their intent, *Bernhardt et al.* [2013]. Another method involves resolving ambiguity and rearranging search results based on user preferences using algorithms like **Dynamic Page Rank,** *Jain et al.* [2013]. These studies collectively highlight the importance of understanding and addressing search intent ambiguity to improve the effectiveness of search engines. *Mehala & Goyal* [2011] proposed related concepts with different meaning selection and recommendation as query suggestions to tackle the ambiguity between query and concept i.e. result. Using weight functions and comparison through TFIDF method in a middleware mechanism between user and search engine, they estimated the relevance between the query and concepts with gain in average precision in this relevance setting. Their approach can be extended to cluster the documents returned by the search engine and to cluster search engine queries. Category labelling can be done automatically and can be used in clustering of queries and automatic hierarchy building. Additionally, search **result diversification** aims to identify the different aspects of an ambiguous query and retrieve relevant documents for each aspect, increasing the chances of satisfying the user's information need, *Maarten* [2014]. Furthermore, incorporating the context of the information need and user intention can improve the **autocomplete feature** of search engines, leading to more effective results, *Gajendragadkar & Joshi* [2017]. *Yano et al.* [2016] state in their model the **quantification of query** ambiguity using **topic distributions** to understand user intent and improve search engine performance, as shown in their Chart-24, Appendix-9. Search engines handle ambiguous queries by analysing the query and understanding its intent using natural language processing and machine learning, [81]*Ali-LN*. They consider factors such as **query modifiers, entities, context, location, device, language, and previous searches to infer the user's intent**, [81]*Ali-LN*. Google distinguishes between three types of interpretation for ambiguous searches: **dominant, common, and subordinate interpretations**,

[81] https://www.linkedin.com/advice/0/how-does-search-algorithm-handle-ambiguous-dfnif

[82]*Julia-Ryte.* To optimize for search intent, it's important to examine ambiguous intent and create content that aligns with the various user intents for a particular query, [83]*Claire* [2023]. Understanding and addressing search intent ambiguity is essential for improving the effectiveness of search engine optimization and providing users with the most relevant and accurate results, [81]*Ali-LN* & [82]*Julia-Ryte*. Addressing search intent ambiguity is crucial for obtaining relevant results on search engines. *Mark Sanderson* [2008] states that while searching on relationship between ambiguity and information retrieval effectiveness finds that ambiguous proper nouns are ambiguous across different forms, e.g. person names, places, titles, etc, with no single type accounting for most occurrences; he shows that the class is more ambiguous than other word types and is commonly used in queries. *Xu & Ye* [2018] present a work on understanding the query and mining intention ambiguity around it to produce hybrid relevant novel query recommendation algorithm based on the query-URL pairs, which achieves superior performance compared to other algorithms. The sequences of queries issued by a user within a short time have same intention, therefore, query co-occurrence in the same session has query relevance and can be used to produce a recommendation. Clicked URLs in a query log show query intention, to some extent. Their model combines "query relevance based on query information" and "query relevance based on URL information" and results a "hybrid query relevance" which is highly useful in reducing the ambiguity. Below are some strategies to refine searches and cater to different potential intents:

a. **Use of Descriptive Keywords:** Using specific and descriptive keywords clearly convey your intent. For example, if you're looking for information about apple as a fruit, include terms like "fruit," "nutrition," or "recipes" to distinguish it from results related to the tech company. Example: "apple fruit nutrition" or "apple fruit recipes".
b. **Specify Context:** Including additional context or qualifiers helps in narrow down the search. This can include location, industry, or specific attributes. Example: "apple computer company" or "apple fruit health benefits".
c. **Utilizing Question Phrases:** Framing your query as a question to provide additional context helps the search engine understand your intent. Example: "What are the health benefits of apple fruit?".
d. **Use of Verbs for Action-Oriented Queries:** If you're looking for instructions, guides, or tutorials, use action-oriented verbs to refine your search. Example: "How to grow apple trees" or "Guide to baking apple pie".

[82] https://en.ryte.com/magazine/search-intent-seo
[83] https://searchengineland.com/optimize-search-intent-tips-430857

e. **Explore Synonyms and Related Terms:** Experiment with different synonyms and related terms to cover various aspects of your query. This helps capture different potential intents. Example: "smartphone reviews" or "mobile device comparisons".

f. **Consider Search Suggestions:** As you start typing your query, pay attention to the autosuggestions provided by the search engine. These suggestions can guide you toward commonly searched phrases and intents.

g. **Filter Results:** Use search engine filters to refine results based on categories such as images, news, videos, or specific timeframes. This can help cater to different types of intents.

h. **Check Related Searches:** Review the "Related Searches" section at the bottom of the search results page. This can provide ideas for refining your search based on alternative intents.

i. **Review Snippets and Previews:** Scan through the snippets or previews of search results to quickly gauge if the content aligns with your intended purpose. This can save time by avoiding irrelevant pages.

j. **Be Prepared to Iterate:** If your initial search doesn't yield the desired results, be open to refining and iterating your query. Adjust keywords, use different qualifiers, or try alternative phrasing until you find the information you're looking for.

2. **Identifying the Contextual Discrepancies and Search Nuances:** Just like Bookmarks feature in the browser, *Krishna Bharat* [2000] brought a complementary extension named "**SearchPad**" to serve the function like Bookmarks but with intent of recording retaining the results, search context of SERP in hand. This suggests that explicit availability of search context might partially compensate for non-relevant pages in the ranking. Contextual search is a form of optimizing search results based on context provided by the user and the computer being used to enter the query. Contextual search attempts to increase the precision of results based on how valuable they are to individual users. Research has explored various methods to refine contextual discrepancies and nuances in search engine queries. *Chesnevar et al.* [2006] and *Barathi & Valli* [2011] both emphasize the importance of user context in enhancing search results, with Chesnevar proposing the use of **context-specific terms** and a **defeasible argumentation framework**, and Barathi introducing a method that captures user context using **WordNet** and joint keyword occurrences. *Gao et al.* [2010] suggests using user-input contextual terms for query disambiguation, while *Challam et al.* [2007] focuses on personalizing search results using **ontology-based contextual profiles**. Refining of contextual discrepancies can also be achieved through various other methods proposed in the literature. When it is about measuring the semantic similarity between words and concepts, *Jiang & Conrath* [1997] in their approach combined the lexical taxonomy structure with corpus statistical information so that the semantic distance between nodes in the semantic space constructed by the taxonomy can be better qualified with the computational evidence derived from distributional analysis of corpus data. This approach is highly

applicable in word sense disambiguation. One approach is to capture context information to improve **image retrieval accuracy**, *Almasri* [2014]. Another method involves using **user feedback logs** to determine query keywords and refine search queries based on **historical query results,** *Xie & Ou* [2010]. Additionally, a context-based knowledge refinding system called **Improvised ReFinder** allows users to refind files and websites based on previous access context, *Sree & Santhanalakshmi* [2014]. **Contextual retrieval systems** can also capture implicit and explicit user data to develop personal contextual profiles, which can be shared to create a shared contextual knowledge base and improve search results and user experience, *Limbu et al.* [2009]. Furthermore, mining context-sensitive query reformulations and injecting new relevant documents into search results based on user context can also help refine search queries and improve retrieval quality, *Shokouhi et al.* [2015]. When users search on search engines, addressing contextual discrepancies and query refinement is essential for providing accurate results. Search engines utilize various techniques to understand and respond to query refinement and contextual discrepancies. These techniques include **recognizing session context, analysing user behaviour, and employing query refinement methods** such as autocomplete, suggestions, related searches, filters, facets, and personalization, [84]*Tunkelang* [2017] & [81]*Ali-LN*. There are patented methods and apparatus for determining user contexts and refining context-based queries in search engines, *Taylor* [2012] & *Xie & Ou* [2010]. By leveraging these approaches, search engines can better understand and respond to users' search intent, ultimately enhancing the search experience.

3. **Understanding and Identifying the Linguistic Nuances:** The next connecting factor as a search skill is of **linguistic expertise**, which we have defined as one of the six variables of this thesis as "**Natural Language Ontology**". We have talked about it in depth in its section above; but here we shall concentrate to look into its implications in terms of nuances; which users usually commit and how these can be refined. Here to add-up further arguments on this aspect in context of Search Skills, search engine users often refine their queries to gain more useful information. **Query reformulation** is a crucial aspect of search engine usability. Users' **reformulation behaviours** have been investigated, revealing that **intent ambiguity** can lead to longer search sessions. *Jiang et al.* [2014] relate the longer sessions to user focus shifting from top results to lower-ranked ones, and results become less attractive over time due to factors such as decreased performance, decreased novelty, and user persistence. Longer session can also be an indication of user having higher cognitive capacity of user or SERPs lack required quality. This lacking can be due to the linguistic unclarity or knowledge level of the domain a search user may have. The terms added or deleted during reformulation are influenced by clicked results more than skipped ones. Users' specification actions are inspired by result snippets or landing pages, while generalization behaviours are impacted by result

[84] https://queryunderstanding.com/session-context-4af0a355c94a

titles, *Chen et al.* [2019]. A search enhancement system can create different narrow searches using **synonyms, hyponyms, hypernyms, and antonyms**, allowing users to explore different nuances of meaning in their search phrase, *Flitcroft* [2011]. **Machine learning** methods can be used to classify relevant sources to a query, improving the user experience by automating the selection of different data sources, *Okland* [2011]. A method for refining a search involves **comparing search result** *(you can read about various such methods run by Google and Bing in 4th chapter of this thesis)* quantities and determining information gain from the comparison, *Osipov et al.* [2006]. Semantic search, utilizing linguistic knowledge, can improve the semantic relevance of search engines compared to traditional search engines, [85]*Lavallee-Google-Patent* [2013]. A range of studies have explored the use of linguistic knowledge to refine search queries and improve search precision. *Osipov et al.* [2008, 2006] both emphasize the importance of linguistic approaches in this context, with the latter specifically highlighting the role of refined linguistic processing tools in semantic search. *Johnson et al.* [2004] presents an algorithm for refining search queries using **Boolean expressions and latent semantic analysis**, which has been shown to improve both **recall and precision**. *García-Serrano et al.* [2001] proposes a natural language interface for search engines, which enriches user queries and utilizes semantic domain information to enhance the search process. These studies collectively highlight the potential for linguistic nuances to be refined by search users, leading to more effective and relevant search results. Based on above findings and discussions, we can sum-up the nuances refining methods at user end as below:

a. **Effective and Refined Search Query Formulation:** Skilled searchers are adept at formulating precise and well-crafted search queries. Their ability to choose the right keywords and phrases helps search engines understand their intent accurately, leading to more relevant results. Experienced searchers are likely to formulate more refined and specific search queries. Their expertise allows them to use industry-specific terms and advanced search operators, contributing to the retrieval of more precise and relevant results.
b. **Context Recognition backed by Language Proficiency:** Search Expertise often correlates with language proficiency. Searchers with a strong command of the language can more accurately express their information needs, leading to better search results. They can also navigate through diverse sources, including those in different languages, to access authentic information. Skilled searchers consider the context of their queries, recognizing that the meaning of words can vary based on context. This contextual awareness helps them construct queries that capture the nuanced aspects of their information needs, resulting in more accurate results.

[85] https://patents.google.com/patent/US20140257811A1/en

c. **Refinement Through Iterative Searches:** Skilled searchers often engage in iterative searches, adjusting their queries based on the information they find. This iterative process allows them to refine their searches progressively, leading to a more focused and precise exploration of the topic.

d. **Topic Familiarity and Terminology - Understanding of Domain-Specific Context:** Searcher skills include a deep understanding of the topic and its terminology. As talked in detail in "Topicality" section of this thesis, topic knowledge enables searchers to use industry-specific or domain-specific terms, improving the accuracy of their search queries and enhancing the relevance of the results. Expert searchers possess a deep understanding of the domain-specific context surrounding their queries. This contextual awareness enables them to refine their searches based on the unique characteristics and terminology of the field, contributing to the retrieval of more authentic results.

e. **Critical Thinking, Evaluation and Problem-Solving for better search results:** The ability to think critically and solve problems is a valuable skill in the search process. Skilled searchers can adapt and adjust their strategies, overcoming challenges and refining their approach to uncover the most relevant and precise information. Searchers with expertise are skilled at critically evaluating search results. They can analyse the relevance, currency, and depth of information presented in various sources, helping them prioritize and select the most authentic and reliable content. *Thomas et al.* [2014] picturise beautifully the usual mechanism of end-to-end search process in Chart-25, Appendix-9.

f. **Information Evaluation Skill:** Searchers with skills in information evaluation can quickly assess the relevance and credibility of search results. Experienced searchers can quickly assess the credibility of sources and content. Their exposure to various domains allows them to distinguish between reliable information and potentially biased or untrustworthy sources, resulting in the retrieval of more authentic results.

4. **SE Technics for refining Linguistic Nuances:** Now let's see how this refining is done by Search Engines:

a. **Use Quotation Marks:** To search for an exact phrase, enclose it in quotation marks. This ensures that the search engine looks for the specific words in the given order. Example: "climate change impact on biodiversity"

b. **Include Synonyms:** Include synonyms of your keywords to broaden or narrow your search. This is especially useful when a term might have multiple meanings. Example: renewable energy alternatives, sustainable energy options

c. **Utilize Boolean Operators:** Use operators like AND, OR, and NOT to connect or exclude specific words. This helps in crafting more precise searches. Example: renewable energy AND government policies. Similarly, the minus operator can be used to exclude specific terms from search results.

d. **Specify File Type:** If you are looking for a specific type of document, add "filetype:" followed by the file extension. For example, "filetype:pdf" for PDF documents. Example: climate change impacts filetype:ppt
e. **Site-Specific Searches:** To search within a specific website, use "site:" followed by the domain. This is helpful when you want information from a particular source. Example: site:wikipedia.org global warming
f. **Use Wildcards:** If you're uncertain about a word or phrase, use an asterisk (*) as a wildcard. This can help when searching for variations of a term. Example: sustainable * practices
g. **Limit by Time:** If you want the most recent information, use tools that allow you to filter results by date or use specific date ranges in your search query. Example: climate change news 2023
h. **Explore Related Terms:** Scroll down to the bottom of the search results page to find related searches. This can lead you to alternative terms or concepts related to your query.
i. **Check Spelling and Grammar:** Ensure correct spelling and grammar in your search queries. Incorrectly spelled words can lead to irrelevant results.
j. **Effective Use of Filters and Advanced Search Features:** Expert searchers leverage filters and advanced search features offered by search engines. These tools help them narrow down results based on factors like publication date, source type, or geographical location, content type enhancing the authenticity of the information retrieved. This helps them quickly access the most recent and relevant information that aligns with their specific criteria.

Exposure - Web Sources and Algorithms Tracking *(Keeping track of new Search Engine Algorithms and Search Trends enables search users)*: Expertise and experience should be adapted in package with Exposure. In context of this thesis Exposure means the consistent follow-up of news and updates regarding Search Engines. Tracking new Search Engine Algorithms and Search Trends is crucial for search users, *Duclos* [2019], *Woloszko* [2020], *Gillespie* [2016]. Algorithms play a significant role in organizing social interactions, producing knowledge, and shaping public discourse, *Fang & Chen* [2016]. Understanding trending algorithms helps individuals navigate social media by identifying popular topics broadly relevant to the public, *Yadav et al.* [2010]. Additionally, monitoring search patterns through algorithms can aid in trend surveillance, such as tracking epidemic outbreaks, by analysing search query logs. By staying informed about the latest algorithms and search trends, search users can enhance their search experience, access relevant information efficiently, and adapt to fast-changing events and topics. This knowledge empowers users to make informed decisions, find desired information effectively, and improve the overall functionality of search engines based on their

behaviour and requirements. Below are key instruments which enables users to garner best of web search through a consistent track of Search Algorithms development and search trends:

1. **Familiarity with Authoritative Web Sources becomes culture:** When Web Search Users keeps a consistent follow up with web search trends and sources, they become accustomed to various web sources such as websites, blogs, video channels etc. And when it comes to the specific domain of a search user, they become a voluntary expert of it.
2. **Consistency in search updates and trends follow-up augments the knowledge of Trustworthy Web Sources:** With accustomed to web search culture gains a search user the arbitrary knowledge of specific web content contributors with their frequency, knowledge, and expertise, which help them to get bookmarked by their followers. Search users memorise their influencers on fingertips.
3. **Understanding of Search Algorithms:** Search engine algorithms are designed to deliver the most relevant content based on user queries. Searchers with expertise understand how these algorithms work, enabling them to optimize their queries for better alignment with the ranking criteria, leading to more authentic results.
4. **Utilization of Academic and Research Databases:** Searchers with expertise in academic or research domains are familiar with specialized databases. They may utilize these resources to access scholarly articles and publications, contributing to the retrieval of authoritative and authentic information.
5. **Utilization of Multiple Search Engines:** Skilled searchers may explore multiple search engines or specialized databases to access diverse sources of information. This multi-platform approach enhances the chances of finding the most relevant and comprehensive content. It helps in comparing the results quality of multiple search engines and puts you in a better situation of annotating their results quality and relevance to your needs.
6. **Awareness of Specialized Search Engines:** Expert searchers are often aware of and may use specialized search engines or databases that cater to specific industries or disciplines. This knowledge expands the range of potential sources and contributes to retrieving more authentic and in-depth information.
7. **Continued use of Search Engines:** It helps to traverse with the development journey of search engines. Whenever new SERP Features are introduced or the oldies are taken off practice, any refurbishing done in the overall aesthetic of Search Interface, you come to know earliest. And you user behaviour flow feels a continual cognitive update.

A consistent follow up with web search practices, methods and algorithms as explained above help people to transform their search habits and needs like a culture practice. Here to be noted from the first two points of above 7 points that why I said at beginning of this section that why I sum-up Authoritativeness and

Trustworthiness in Exposure *(one of 3Es, first 2 Es i.e. Expertise and Experience in common agreement with that of Google's E-E-A-T)*. Below Chart-26 depicts the sub-variables of Search Skills:



Chart-26: Sub-variables of Variable 4 'Search Skills'

Discussion *(Possible ground for hypothesis conceiving)*: The first thing first to establish from above LR is the direct relation between the Web Search Acumen and search intent ambiguity. Above LR has cleared that search skills helps impactfully in reducing the intent ambiguity, consequently making the search user to express their search intent with clarity in their search queries. The second establishment here can be of enhanced understanding of query nuances and better linguistic command, which ultimately helps to manage your search with ease at "peace of mind". Third benefit appears as a better contextual sense of designing a search query au grace de search skills. In the above LR, we come to know about the importance of follow-up on web search algorithms and trends, but a deep researched study on quantitative impact of same on searchers can lead to new believes and information. Also, an argument can be opened about defining EEAT with 3E only.

Discussion Scrutiny: Search skills help in framing better to best of the contexts for search. As we discussed above a person with high level of 3Es is the most suitable prospect for searching with highest possible intent and mental clarity and least ambiguity. However, contexts have their specific characteristics, which one needs to know by defining the search context as a special skill. Searcher' search skills significantly contribute to the accessibility of the most relevant and precise information on search engines. Through effective query formulation, the use of advanced search features, and a nuanced understanding of topics, skilled searchers optimize their search processes to find the information quickly and efficiently. The expertise and exposure of a searcher significantly impact the retrieval of authentic results from a search engine. Through refined queries, an

understanding of search algorithms, recognition of credible sources, and effective use of search features, expert searchers optimize the search process to access the most trustworthy and authentic information available.

2.7. Thesis Research Variables: Table-9 is the initial conceiving of a framework model in mind, which is presented in below table based on above FLR. Above FLR has helped in putting our conceived variables in order and viable relation to their corresponding sub-variables and other complementary aspects.

SERP Thesis Research Variables

State of Art Variables | Psychological Variables | Sub-Variables

Research Thesis Variables

Sr.	State-of-Art Variables (A Set)	Psychological Variables (B Set)	Sub-variables	Variable in Single Term	Facilitator vs. Player	SERP Cycle
1	Art of Phrasing	Neural & Semantic Understanding	Inferring Linguistic Efficiency Neuro Physiological Evidences Cognitive Feedback Learning	Understanding	**Player:** Search User **Facilitator:** SE	Query Interpretation & Deciphering
2	Natural Language Ontology	Intent(s) Precision	Decoding Search Query Inferring Search Intent Creation of SERP Features	Intents	**Player:** SE **Facilitator:** Search User	Query & Content Mapping
3	Topicality	Search Context and Cognitive Accuracy	User' Domain and Topic Knowledge User' Literary Skills SE's Topical Ontology Cognitive Evaluation of Topicality	Cognition	**Player:** Search User **Facilitator:** SE	
4	Semantic Web	Intuitive Matching & Predictability	Semantic Web Approaches Semantic Web Evolution Semantic Web Algorithms	Intuition	**Player:** SE **Facilitator:** Search User	Sorting & Picking Contextual Results
5	SERP Personalization	Tangible Behavioural Traits	SERP Personalization Factors SERP Search Users Satisfaction SERP Personalization Algorithms	Personalisation	**Player:** SE **Facilitator:** Search User	
6	Search Skills	Intangible Behaviour Traits	Experience - Information Search Mental Models Expertise - Linguistic Nuances & Contexts Exposure - Web Sources & Algorithms Tracking	Persistence	**Player:** Search User **Facilitator:** SE	Content Accessibility & Clicking

Table-9: Six variables of this thesis

Psychological Behavioural Model of SERP Behaviour: After having the above 6 variables for this thesis in Table-9 above, we below in Diagram-8 draw the first central model of this thesis. The Diagram-8 below presents what we have studied so far in this chapter completely, the variables, sub-variables of the psychological model of this thesis in one picture. To know about each term used in this model, is elaborated in its corresponding variable and their sub-variable sections above.



Mutual give and gain of three stakeholders of Web 3.0 + Web 4.0 SERP

Stakeholder	Give	Gain
Search User	Search Feedback	Information Needs
Content Creator	Content, Information	Revenue, Recognition
Search Engine	SERP, Information	Search Feedback, Revenue

Diagram-8: Psychological model of SERP Behaviour

Chapter-3

Technical Literature Review in the context of SERP

3.1. Data Science Technologies used in generation of SERPs:

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data, [86]*DS Wikipedia 2*, and apply knowledge and actionable insights from data across a broad range of application domains. Data science is related to **data mining, machine learning and big data**, [87]*CE ibm.com 1.* Data science is a "concept to unify **statistics, data analysis, informatics, and their related methods**" to "understand and analyse actual phenomena" with data, [88]*Vasant Dhar* [2013]. It uses techniques and theories drawn from many fields within the context of **mathematics, statistics, computer science, information science, and domain knowledge.** However, data science is different from computer science and information science. Turing Award winner Jim Gray imagined data science as a "**fourth paradigm**" of science *(empirical, theoretical, computational, and now data-driven)* and asserted that "everything about science is changing because of the impact of information technology" and the data deluge, *Hayashi et al.* [1998] and *Tony Hey* [2009]. As simulations and experiments yield ever more data, a **fourth paradigm** is emerging, consisting of the techniques and technologies needed to perform data-intensive science, *Gorden et al.* [2009].

SERPs production involves the use of various types of data sciences. One approach is **Web Content Mining**, which helps users extract useful information from SERPs. This technique involves mechanisms such as Web Data Extraction using Similarity Function ***(WDES)*** and Web Data Integration using Cosine Similarity ***(WDICS)***, *Venugopal & Srikantaiah* [2020]. Another approach is Clustering Search Results ***(CSR),*** which groups similar Search Result Records ***(SRRs)*** into clusters based on similarity, allowing users to focus on their specific demands, *Abinaya & Jaishanthi* [2014] and *Azzopardi et al.* [2020]. Additionally, there is a need to consider **item heterogeneity in SERPs**, as modern SERPs consist of different item types *(e.g., news, web, entity)*. Data-driven metrics based on item type can provide more accurate estimates of user behaviour and performance on SERPs, *Mudassar et al.* [2015]. Search engines use data such as a **website's security, credibility, traffic, backlinks, and the searcher's location and search history** to generate SERPs. **Google**

[86] https://en.wikipedia.org/wiki/Data_science
[87] https://www.ibm.com/cloud/learn/machine-learning
[88] https://cacm.acm.org/research/data-science-and-prediction/

Dataset Search, for example, relies on an open ecosystem and aggregates metadata to enable dataset discovery, *Natasha et al.* [2019]. The influence of different item types on user behaviour in SERPs is explored through a data-driven evaluation metric, *Azzopardi et al.* [2020]. The impact of various page features on search engine result rankings is also analysed, with the scoring function found to be difficult to approximate using observed features, *Bifet et al.* [2005]. The generation of SERPs involves a range of technical aspects, including **web page classification, hyperlink analysis, and data mining**. *Fiol-Roig et al.* [2011] highlights the use of **automatic web page classifiers**, which can assist search engines in providing relevant results. The types of **data sciences used in SERP** production include the application of various principles such as **data splitting, content generation, and prediction algorithms**, [89]*Imarticus* [2018] & [90]*Sharma* [2021]. *Henzinger* [2004] discusses the use of **PageRank technology**, an **eigenvector-based analysis** of hyperlink structure, to produce relevant results. *J. Pei* [2009] emphasizes the role of data mining in **context-aware search and query suggestion**, particularly in large-scale search engines. *Banerjee & Dua* [2022] introduces a framework for ranking **big data** in search engines, which includes processes such as Extracting, Transforming, and Integrating ***(ETI),*** **ontology graphs**, and indexing **RDF** *(Resource Description Framework)* using the wide-column **NoSQL method**. SERP generation involves various technical aspects of data science, such as **recording user-predefined destination** labels and sorting methods corresponding to these labels. The selected destination label is used to sort the search results and display them to the user, *Xiaofang* [2018]. Another method involves saving search results in a **search domain library**, analysing and structuring the results, and linking them with related search domains. Statistical analysis of previous searches is used to anticipate user actions and generate search results, *Valentinovich et al.* [2018]. Additionally, data regarding user activities on a website indexed in a search engine is stored and used to calculate a weighting factor that influences the website's position in search results, *McLellan* [2011]. Data models are also generated to give **semantic meaning to unstructured or structured data** received by search engines, allowing for more effective querying and result generation, *Neels et al.* [2013]. These technical aspects contribute to the efficient generation of SERPs. These studies collectively underscore the importance of data science techniques in enhancing the relevance and efficiency of SERPs. Data science in the context of SERP production and SEO also involves the use of various tools, algorithms, and machine learning principles to eliminate guesswork from SEO and make the SEO data more actionable through visualizations, [91]*Bhattacharya* [2019]. The different types of SERP features that are generated because of data science applications include **knowledge graph features, rich snippets, and paid results**, which are influenced by the type of search query entered, such as navigational, informational, or transactional queries,

[89] https://imarticus.org/blog/how-data-sciences-principles-play-an-important-role-in-search-engines/
[90] https://www.datasciencecentral.com/how-to-use-data-science-for-search-engine-optimization/
[91] https://www.searchenginewatch.com/2019/09/04/improve-seo-using-data-science/

[92]*mailchimp.com-1*. Based on above findings and associated readings, here below is the list of key algorithms, models and data engineering methods used in SERP generation:

1. **Relevance Models, Information Retrieval and Predictions:** As found in above review, data science is applied to build **relevance models** that assess the relevance of a webpage to a particular search query. These models analyse a wide range of features, including content quality, user engagement, and contextual information, to estimate the likelihood of a page being relevant to a user's needs. **Information Retrieval** focuses on retrieving relevant information from a large collection of data, such as the vast index of webpages maintained by search engines. IR techniques are used to rank webpages in SERPs, ensuring that the most relevant and authoritative results are displayed first. These techniques consider various factors, including **keyword relevance, backlinks, domain authority, page authority and the overall quality of the content** on the webpage. Search engines **build massive indices** of webpages, organizing them in a way that allows for efficient retrieval of relevant results. IR techniques, such as 'term frequency-inverse document frequency' ***(TF-IDF)*** and **Cosine Similarity**, are used to rank webpages based on their relevance to search queries. Data scientists engage in **feature engineering** to identify and select relevant features that contribute to the prediction of search result relevance. *Bruce Croft* [2019] by continuing the discussion of interaction in the past two Salton Award lectures from Jarvelin and Belkin states that the key message of their talk is that effective information access requires interaction between the user and the system, where both play a role. Making it base he emphasizes on two crucial areas of interaction as **question-answering and conversational search** for effective IR.
2. **User Behaviour Data Analysis and Personalization:** Data scientists analyse user behaviour data, including **click-through rates *(CTR),* bounce rates, search queries, webpage content and dwell time**, to understand how users interact with search results. This information is crucial for understanding user preferences, identifying trends in search patterns, and developing targeted advertising campaigns. It is used to refine ranking algorithms, developing personalized ranking models and improve the overall relevance of search results based on user preferences. By analysing user preferences, search history, and behaviour, data scientists can create personalized ranking algorithms that tailor search results to individual users, enhancing the overall user experience. Data mining also helps search engines detect and prevent spam, malware, and other malicious content from appearing in SERPs.
3. **Query Intent Modelling and NLPs:** Data science is used to model and understand user query intent. By analysing patterns in search queries and user interactions, data scientists can identify the underlying intent behind a user's search, enabling search engines to deliver more contextually relevant results. NLP is essential for understanding the intent and meaning of user queries. It allows search engines to analyse text,

[92] https://mailchimp.com/fr/marketing-glossary/serp/

identify keywords, and extract relevant information, enabling them to match queries with the most appropriate webpages. NLP techniques are also used to generate summaries of webpages, provide concise answers to questions, and personalize search results based on a user's past interactions and preferences. NLP algorithms **extract relevant keywords** from search queries and webpage content. This helps determine the semantic similarity between queries and webpages, enabling search engines to match queries with the most appropriate results. NLP and IR techniques are combined to refine user queries, suggesting relevant keywords and identifying synonyms to improve the accuracy of search results. NLP techniques analyse the **tone and sentiment expressed** in search queries and webpage content to understand user intent and the emotional context of the information. This helps search engines provide more relevant and personalized results.

4. **A/B Testing Ranking Experiments:** A/B testing and other experimentation techniques are employed by data scientists to evaluate the impact of changes to the ranking algorithms. By conducting controlled experiments, data scientists can assess the effectiveness of proposed modifications and iteratively refine the ranking models. Search engines constantly process massive amounts of data to optimize their algorithms and improve the quality of SERPs. Big data analytics tools enable **real-time analysis** of user behaviour, search trends, and webpage content to make informed decisions promptly.
5. **Machine Learning for Ranking and Ranking Models:** Machine learning techniques, such as **supervised learning and reinforcement learning**, are applied to train ranking models. ML algorithms are trained on historical massive datasets of search queries, webpages, labelled examples and user interactions to learn patterns and relationships between these data points. This allows search engines to make predictions about user behaviour, identify trends in search queries, and optimize their algorithms to improve the quality of SERPs. ML is also used to generate personalized recommendations, suggest related search terms, and implement new features based on user feedback. ML algorithms are trained on massive datasets of search queries, webpages, and user interactions to learn complex patterns and relationships between these data points. These **Results Ranking Models** are used to rank webpages in SERPs, ensuring that the most relevant and authoritative results are displayed first. ML models are used to personalize SERPs for individual users based on their past search history, interests, and location. This tailoring helps deliver more relevant and engaging results for each user. ML algorithms are used to identify and provide direct answers to certain types of search queries, such as factual inquiries or simple questions. This "**zero-click search**" feature saves users time and effort by providing immediate answers without requiring them to visit individual websites.
6. **Data Quality and Cleaning, Data Visualisation, Continuous Learning, Adapting to Evolving Trends and Big Data Analytics:** The collected data is often noisy and inconsistent, requiring data cleaning

techniques to remove errors, duplicates, and irrelevant information. Ensuring the **quality and cleanliness of the data** used to train and evaluate ranking models is a crucial aspect of data science in SERP ranking. Data scientists work to identify and address issues related to data accuracy, completeness, and relevance. ML models are used to **detect anomalies** in search patterns or webpage content that may indicate spam, malware, or other malicious activities. This helps search engines maintain the integrity of SERPs and protect users from harmful content. Data science helps search engines adapt to **evolving trends** in user behaviour, content creation, and technology. By continuously analysing new data, data scientists can identify emerging patterns and adjust ranking algorithms to provide up-to-date and relevant search results. Search engines use **data visualization techniques** to create charts, graphs, and infographics that **illustrate search trends**, user behaviour, and website traffic. This information is valuable for understanding user preferences, identifying emerging topics, and optimizing search results. Data visualization is integrated into SERPs to provide users with interactive experiences, such as expandable summaries, visual representations of data, or interactive maps. This enhances the user experience and makes it easier to understand complex topics. Continuous learning from new data allows algorithms to stay relevant and responsive to emerging patterns. With the ever-increasing volume of data generated on the internet, search engines must employ **Big Data Analytics** techniques to handle and process this vast amount of information efficiently. Big Data Analytics tools allow search engines to analyse **petabytes of data** in real time, enabling them to make informed decisions about how to rank webpages, personalize SERPs, and improve the overall user experience.

7. **Ethical Considerations:** Data scientists are actively involved in addressing biases in search results and ensuring fairness. Techniques such as **debiasing and fairness-aware algorithms** are employed to mitigate potential biases in SERP rankings.

Discussion: These data science disciplines work together to power the complex algorithms that drive search engine rankings and result generation. By leveraging data-driven insights, ML techniques, and user behaviour analysis, data scientists contribute to the ongoing improvement of SE algorithms, ensuring that users receive the most relevant and valuable results for their queries. By employing these technical aspects of data science, search engines can provide users with relevant, accurate, and personalized search results that meet their needs and expectations. As data science techniques continue evolving, SERPs will become even more sophisticated and capable of delivering customized experiences for each individual user. Chart-28 below summarizes the facts and findings in above research.

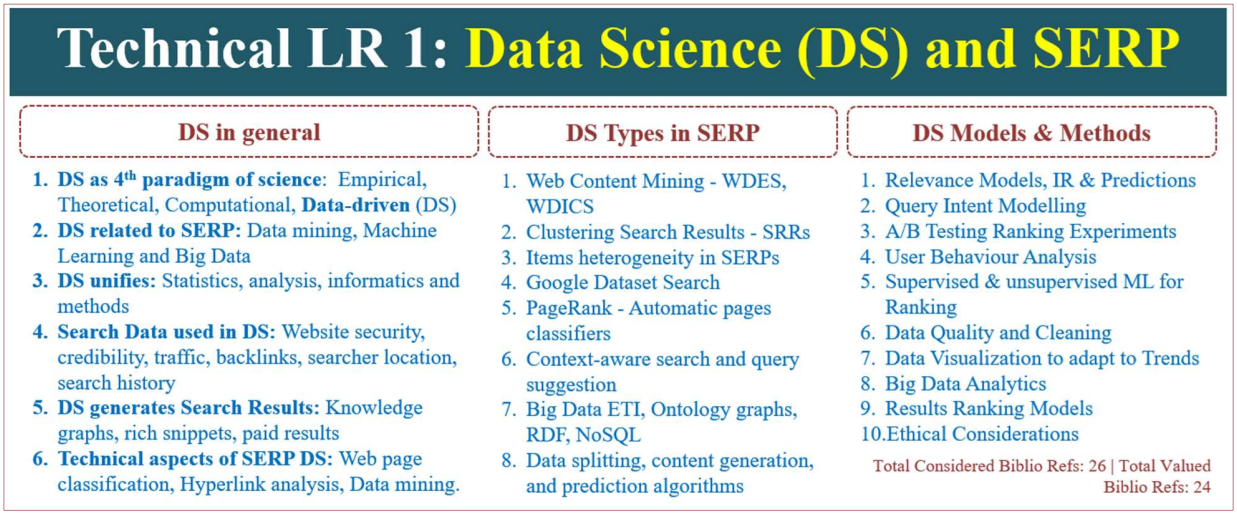


Chart-28: Data Science and SERP

3.2. Grand perspective of Artificial Intelligence and its specificity for SERP:

[93]*Copeland* [2021] states that the earliest substantial work in the field of artificial intelligence was done in the mid-20th century by the British logician and computer pioneer Alan Mathison Turing. **In 1935, Turing** described an abstract computing machine consisting of a limitless memory and a scanner that moves back and forth through the memory, symbol by symbol, reading what it finds and writing further symbols. Since the development of the digital computer in the 1940s, it has been demonstrated that computers can be programmed to carry out very complex tasks, as for example, discovering proofs for mathematical theorems or playing chess with great proficiency. As mentioned in Chapter-1 Artificial intelligence was found as an academic discipline in 1956, and in the years since has experienced several waves of optimism, followed by disappointment and the loss of funding *(known as an "AI Winter")*, further followed by new approaches, success, and renewed funding, *Crevier Daniel* [1993]; *McCorduck Pamela* [2004]; *Russell & Norvig* [2003]; [94]*USNRC 1* [1999]; *Newquist 1* [1994]; [95]*Clark Jack* [2015]. AI research has tried and discarded many different approaches since its founding, including **simulating the brain, modelling human problem solving, formal logic, large databases of knowledge and imitating animal behaviour**, [95]*Clark Jack* [2015]. AI applications include advanced **Web Search Engines** *(e.g., Google, Bing, Yandex, Baidu, Yahoo),* **Recommendation Systems** *(used by YouTube, Amazon, Netflix),* **Understanding Human Speech** *(such as SIRI, Alexa),* **Self-driving cars** *(e.g., Tesla),* **Automated decision-making and competing** at the highest level in strategic game systems *(such as chess,*

[93] https://www.britannica.com/technology/artificial-intelligence
[94] https://nap.nationalacademies.org/read/6323/chapter/11#210
[95] https://www.bloomberg.com/news/articles/2015-07-01/musk-backed-group-probes-risks-behind-artificial-intelligence

Go), [96]*Mike Anderson* [2021]. A signified study about AI Algorithms by *De Silva & Alahakoon* [2022] took my attention shown in Chart-29, Appendix-9. Not only it gives a spectacular picture of how the Algorithms are developed and deployed but also constitute many examples of **algorithms which are used in SERP**, such as ANN, RNN, LSTM in prediction, Naïve Bayes and CNN in Content Classification for sentimental analysis; k-means and Autoencoders in associating topics and segments of texts, and GAN in content optimization. In the field of Healthcare Automation as per [97]*automationedge.com-1* [2023] the five top trending AI applications of 2024 are Artificial Intelligence and Robotics, Internet of Things **(IoT),** Telemedicine, Healthcare Equity and Self-Service Kiosks. As the nature of this thesis is B2C oriented i.e. the SERP used by the end users so out of these five applications the G-AI in context of SERP shall be of our prime interest, with possibility of extending this list to similar kind of other applications. As per [98]*AI sas.com 1* most AI examples from chess-playing computers to self-driving cars, healthcare devices to retail, manufacturing to banking rely heavily on DL and NLP. **So Artificial Intelligence (AI) is:**

1. The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings, [93]*Copeland* [2021].
2. Intelligence demonstrated by machines, as opposed to natural intelligence displayed by animals including humans, [99]*AI Wikipedia 1.*
3. The simulation of human intelligence processes by machines, especially computer systems, [100]*Ed Burns et al.*
4. Makes possible for machines to learn from experience, adjust to new inputs and perform human-like tasks, [98]*AI sas.com 1.*
5. The science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable, *John McCarthy* [2007].
6. Leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind, [101]*Stryker & Kavlakoglu*.

Relation of AI and SERP: AI is playing an increasingly important role in SERPs. Here are some of the ways AI is being used to improve SERPs:

[96] https://deepmind.com/research/case-studies/alphago-the-story-so-far
[97] https://automationedge.com/home-health-care-automation/blogs/top-7-healthcare-trends-to-watch-for-in-2025/
[98] https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html
[99] https://en.wikipedia.org/wiki/Artificial_intelligence
[100] https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence
[101] https://www.ibm.com/think/topics/artificial-intelligence

1. **Understanding Search Intent:** One of the most important tasks for a search engine is to understand what a user is looking for when they type in a query. AI can help with this by using NLP to analyse the query and identify the user's intent. For example, if a user searches for "best restaurants in Paris," AI can understand that the user is looking for a list of restaurants, not just a definition of the word "restaurant."
2. **Ranking Results:** AI is also being used to rank websites in SERPs. Traditional SEs use algorithms that consider factors such as keyword relevance, backlinks, and domain authority to determine a website's ranking. AI can be used to supplement these algorithms by considering additional factors such as user behaviour, sentiment analysis, and the overall quality of the content on a website.
3. **Personalizing Results:** AI can personalize SERPs for individual users. This can be done by using data about a user's past searches, location, and interests to tailor the results to their needs. For example, if a user has previously searched for information about travel to France, AI can rank websites about French travel at the top of their SERPs.
4. **Generating summaries:** AI can generate summaries of webpages, helpful for users who are short on time or who want to get a quick overview of a topic before clicking on the link.
5. **Providing answers directly:** In some cases, AI can provide answers to questions directly in the SERPs. This is known as "**zero-click search**" and it can save users a lot of time. For example, if a user searches for "what is the capital of France," AI can provide the answer directly in the SERP, without the user having to click on any links.

Discussion: AI technology continues to develop, we can expect to see more innovative ways in which it is used to improve SERPs, while assisting the goal of creating SEs that are more helpful, faster, and more personalized for users. Chart-30 below summarizes the facts and findings in above review.

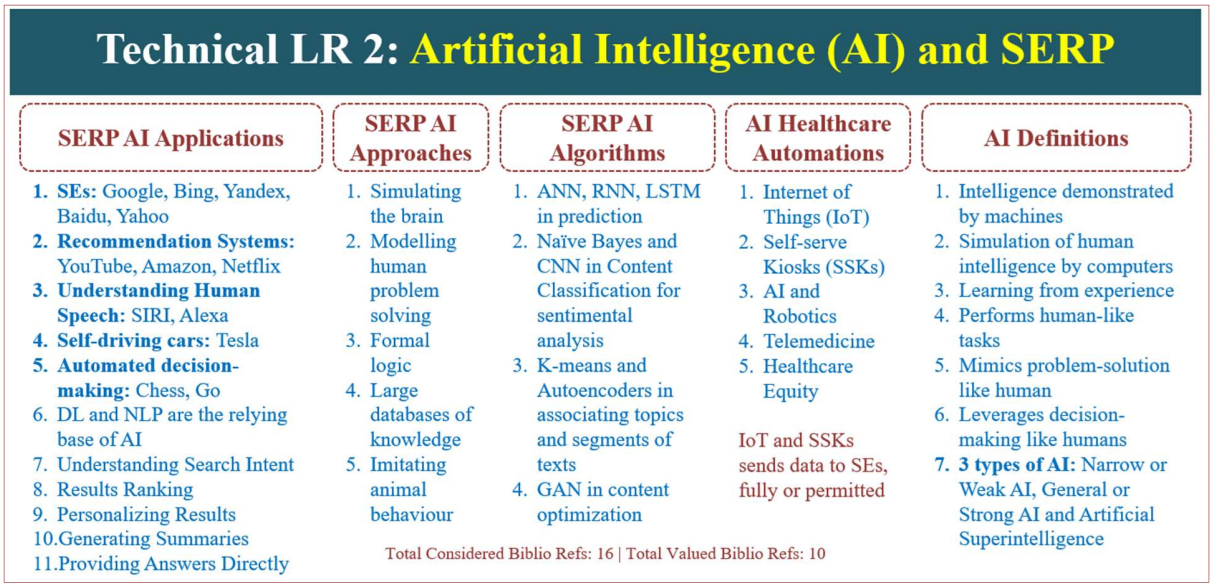


Chart-30: Artificial Intelligence and SERP

3.3. Search Engine Machine Learning (SEML) and Neural Networks:

Out of the three types of AI i.e. **Narrow or Weak AI, General or Strong AI and Artificial Superintelligence**, as per [102]*Kevin Rowe* [2021] Machine Learning contributes practically the most in **Strong AI**. Continuing the journey of AI, in the first decades of the 21st century, rigorous mathematical statistical machine learning has dominated the field, and this technique has been highly successful, helping to solve many challenging problems throughout industry and academia as per *Russell & Norvig* [2003]; [94]*USNRC 1* [1999]; *Newquist 1* [1994]; [95]*Clark Jack* [2015] and *Kurzweil Ray* [2005]. Machine learning is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, thus perform tasks without explicit instructions, [103]*ML-Wiki*. Machine Learning in Search Engine is used for Search ranking, Query and URL/document Understanding, Search Features, Crawling and User Classification, [104]*Dandekar* [2016]. As per [105]*Wroe* [2019] machine learning allows a program to gain new information without requiring programmers to alter the code themselves or supervise the program at every step. As per [106]*Grigorov* [2019] there are eight principles of machine learning and models used in Search Engines as:

1. Analysis of search and classification of phrases.

[102] https://www.searchenginejournal.com/ml-things-we-know/408882/
[103] https://en.wikipedia.org/wiki/Machine_learning
[104] https://medium.com/@nikhilbd/what-are-some-uses-of-machine-learning-in-search-engines-5770f534d46b
[105] https://www.embeddedcomputing.com/technology/ai-machine-learning/how-google-search-probably-uses-machine-learning
[106] https://serpact.com/8-machine-learning-principles-and-models-used-by-search-engines/

2. Find patterns for URLs and page content.
3. Better identification of word connections.
4. Identification of synonyms.
5. Identifying the similarities between the words in a search query.
6. Search images to understand photos.
7. Custom alerts based on specific requests.
8. Identifying new alerts.

Generally, it is seen as the part of Artificial Intelligence; but there are two different expert claims on it. On one axis *Sindhu et al.* [2020] define **ML as subset of AI**, while the other axis of Deep Learning Research by [107]*Sebastian Rashcka* [2020] states **ML and AI as subset to each other**. But after analysing views of numerous experts and research; **majority considers ML as subset of AI**. So, we find it convenient going with the majority. Adding further [107]*Sebastian Rashcka* [2020] states that "*The study of computer algorithms that can improve automatically through experience and using data. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that* ***systems can learn from data, identify patterns and make decisions with minimal human intervention***". As per [108]*ML sas.com 2,* Machine learning is a branch of AI and computer science which focuses on the **use of data and algorithms to imitate the way that humans learn**, gradually improving its accuracy.

ML SERP gets base from Neural Networks: As per *Hopfield* [1982] computational properties of use of biological organisms or to the construction of computers can emerge as collective properties of systems having many simple equivalent components *(or neurons).* A neural network is a network or circuit of neurons, or in a modern sense, an **artificial neural network composed of artificial neurons or nodes**, [109]*NN Wikipedia 2*. It has three layers as **input layer, the hidden layer and the output layer.** This neural network is basically the structural composition of the **cognitive power capability of a human brain**. In today's modern time, this mapping is instrumental in understanding how the Machine Learning system put to work. We shall use this analogy in our thesis while we shall be talking on "**Search Engine Machine Learning Algorithms**" in particular. When to talk about how the neural network models function, [110]*Mitchell Waldrop* [2019] depicts it through Lucy Reading-Ikkanda' Diagram shown in Diagram-9, Appendix-9 below. He says that "Neural network" models of AI process signals by sending them through a network of nodes analogous to neurons.

[107] https://github.com/rasbt/stat453-deep-learning-ss20/blob/master/L01-intro/L01-intro_slides.pdf
[108] https://www.sas.com/en_us/insights/analytics/machine-learning.html
[109] https://en.wikipedia.org/wiki/Neural_network
[110] https://www.pnas.org/content/116/4/1074/tab-figures-data

Signals pass from node to node along links, analogues of the synaptic junctions between neurons. "Learning" improves the outcome by adjusting the weights that amplify or damp the signals each link carries. Nodes are typically arranged in a series of layers that are roughly analogous to different processing centers in the cortex. Today's computers can handle "deep learning" networks with dozens of layers. Image credit: Lucy Reading-Ikkanda (artist).

ML State-of-Art System for SERP: Building good state-of-the-art ML solutions and systems is an iterative process that requires a combination of domain knowledge, data expertise, and a solid understanding of machine learning principles. It uses **Automatically Tune Results** *(Collaborative Filtering, Learning to Rank, Ground Truth),* **Personalize Search** *(Capture User Signals, Recommendations),* **Cluster and Classify Content** *(Find Similar Users, Find Outlier),* **Understanding User Query Intent** *(Head/Tail Analysis, Phrase and Spelling Correction, Synonym Detection),* **See and Understand User Behaviour** *(What Users are searching for, Experiment and A/B Testing)* methods to serve the users satisfaction, [111]*ML Lucidworks.com 1*; *Yao-2 et al.* [2021] exploit the relatedness and dependency between search model and recommendation model of a search engine to ultimately give a USER ***(Unified Search and Recommendation)*** Model. Their USER model mines user interest from the integrated sequence and accomplish the two tasks in a unified way by outperforming the state-of-art separate baselines. In a 2019 Webmaster Central Office Hours discussion, *Emmanuel et al.* [2019] states **ML is a small part of the big picture,** Google's [112]*John Mueller* [2019] references how machine learning helps Google's engineers better understand various issues, but he's careful to note that, "machine learning isn't just this one black box that does everything for you where you feed the internet on one side and the other side comes out search results." More recently, in a May 2021 Office Hours discussion, [113]*Google SEO YT* [2021] explained that machine learning may adjust the weight of various ranking signals. As per 107*Sebastian Rashcka* [2020] a good ML system should pass through to:

1. Data preparation capabilities
2. Algorithms - basic and advanced
3. Automation and iterative processes
4. Scalability
5. Ensemble modelling

[111] https://lucidworks.com/blog/17382
[112] https://www.youtube.com/watch?v=5QxYWMEZT3A&t=2267s
[113] https://www.youtube.com/watch?v=0h8YslnIuuY&t=250s

However, it is not limited to just Sebastian list, rather creating good ML systems involves a combination of skills, processes, and considerations. Here are key requirements for building effective ML systems:

1. **Clear Problem Definition:** A well-defined problem statement guides the entire ML development process to solve the aimed problem.
2. **Data Cleansing and Processing:**
   i) **Quality Data:** Gathering high-quality, representative data for training, validation, and testing helps ML models to leverage performance.
   ii) **Data Preprocessing:** Feature scaling, normalization, and other preprocessing steps to handle missing values, outliers, and inconsistencies are often necessary to prepare the data for model training.
3. **ML Modelling:**
   i) **Feature Engineering:** Extracting relevant features from the data or create new features to enhance the model's ability to capture patterns and relationships.
   ii) **Model Selection:** Choosing an appropriate ML algorithm or model architecture based on the nature of the problem *(classification, regression, clustering, etc.)* and the characteristics of the data.
   iii) **Hyperparameter Tuning:** Fine-tuning the hyperparameters of the selected model to optimize its performance. This may involve using techniques like grid search or random search.
   iv) **Model Training:** Training the model using the training dataset. Monitor and analyse performance metrics during training to identify potential issues, such as overfitting or underfitting.
4. **Quality Robustness:**
   i) **Validation and Testing:** Validating the model on a separate dataset to assess its generalization performance. Testing the model on an independent test dataset to evaluate its ability to make accurate predictions on new, unseen data.
   ii) **Evaluation Metrics:** Choosing appropriate evaluation metrics based on the nature of the problem. Common metrics include **accuracy, precision, recall, F1 score, and Mean Squared Error *(MSE),*** depending on whether the task is classification or regression.
   iii) **Interpretability:** Ensuring that the ML model is interpretable, especially in applications where understanding the decision-making process is crucial. This is important for gaining user trust and meeting regulatory requirements.
   iv) **Scalability:** Design ML systems to be scalable, capable of handling larger datasets and increasing computational demands as needed. Consider scalability from both a training and inference perspective.
   v) **Regularization and Optimization:** Implement regularization techniques to prevent overfitting. Optimize the model for efficiency and resource utilization, especially for deployment in production environments.

vi) **Robustness and Security:** Enhance the robustness of the model by addressing issues such as adversarial attacks or noisy data. Implement security measures to protect against potential vulnerabilities in the ML system.

vii) **Ethical Considerations:** Consider ethical implications related to data privacy, bias, and fairness. Implement measures to address and mitigate biases in the data and model predictions.

5. **Documentation:** Document the entire ML development process, including data sources, preprocessing steps, model architecture, hyperparameters, and evaluation results. This documentation is crucial for reproducibility and collaboration.
6. **Continuous Monitoring and Maintenance:** Establish a system for continuous monitoring of model performance in production. Implement mechanisms to retrain the model as new data becomes available and to address any degradation in performance over time.

Purposes of SEML for SERP:

a) **Content (URL / Document / Page) Pattern Detection and Understanding:** It helps identify spam or duplicate content, which often is **low-quality content** such as **outbound links** to unrelated pages, in **excess use of stop-words** or synonyms, repetition of identified **"spam" keywords** etc. As per [114]*Dave Davies* [2021], Google uses "**human quality raters**" to identify new patterns of these kinds and keep on feeding these patterns in form of algorithms under their machine learning program, which helps in consistent weeding out of such low-quality content before it appears to their searchers. This often falls under the "**supervised machine learning**". Understanding what **types of a page** it is. E.g. blog vs. news site vs. forum etc. ML uses **Page Classification and Spam Detection**, [102]*Kevin Rowe* [2021]. **Junk / Low-quality URL Detection** by looking into over optimization of anchors text, multiple outbound links to unrelated pages, [115]*Tie-Yan Liu* [2009]. Detecting entities, such as people or places in the page content. Figuring out **relationships between entities** detect entity /relationship and sentiments in a document.

b) **Deciphering the Query Language Pattern:** Search Engine like Google uses algorithm knowns as "**BERT**" to decipher the language patterns from search queries based on NLP. According to [116]*Patent google.com 3*, BERT is designed to pre-train deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture

[114] https://www.searchenginejournal.com/google-eat/quality-raters-guidelines/
[115] https://www.nowpublishers.com/article/Details/INR-016
[116] https://patents.google.com/patent/US20050102282A1/en

modifications. According to *Devlin et al.* [2018] BERT helps Google understand natural language better, particularly in conversational search. As estimation, it impacts around 10% of queries and lot of organic rankings and featured snippets. It enables Search Engines to be able to recognize how similar one piece of text is to another and their deeper meanings, [110]*Mitchell Waldrop* [2019]. **Identification of connections and similarities between the words** in a search query technic detects the true sense and intent of a query, [111]*ML Lucidworks.com 1*.

c) **Identification of New Signals:** According to Google's Gary Illyes in a 2019 Reddit AMA, "**RankBrain of Google** is a PR-sexy machine learning ranking component that uses historical search data to predict what would a user most likely click on for a previously unseen query.", [117]*Matt Southern* [2019]. RankBrain has made the job of Human Engineer very easy while Engineers work it runs in the background. Along with other wonderful tasks it primarily handles the first-time appearing queries from understanding them to categorizing them. Learning to Rank for Information Retrieval ***(LRIR)*** is a task to automatically construct a ranking model using training data, such that the model can **sort new objects** according to their degrees of relevance, preference, or importance, [117]*Matt Southern* [2019]. Search Engines have multiple phases of ranking that happen in series, such as initial retrieval, primary ranking, contextual ranking, personalized ranking etc. And it is all done using the "Learning to Rank" systems, [102]*Kevin Rowe* [2021].

d) **SE Interface Featuring:** Search Engine Interface plays the most impactful role in convincing the search users through its various SERP Features designed to serve the specific needs and taste of the user behaviour. For example, **Sitelinks, Related Searches, Knowledge Graph Data** *(for some search engines)* supplements the information provided in more user-friendly elaboration.

e) **Ad Quality & Targeting Improvements:** Google puts their commitment of serving with better relevant and quality Ads; like they do for their Organic results. According to [118]*Patent google.com 4,* Google uses ML to improve "otherwise weak statistical model" in its ads. Subsequently as per [119]*Patent google.com 5,* its Ad Rank is influenced by machine learning system. As per [120]*Ad rank google.com 6,* the Ad Rank Threshold *(the context of person's search)* is fed into the system on a keyword-by-keyword basis to determine what thresholds are considered on each Keyword. The Ad Rank Threshold is determined by **Ad Quality** *(expected CTR, Ad Relevance and Landing Page Relevance*), Ad Position, User Signals and attributes such as location and device type, the topic and the nature of the search and related auctions.

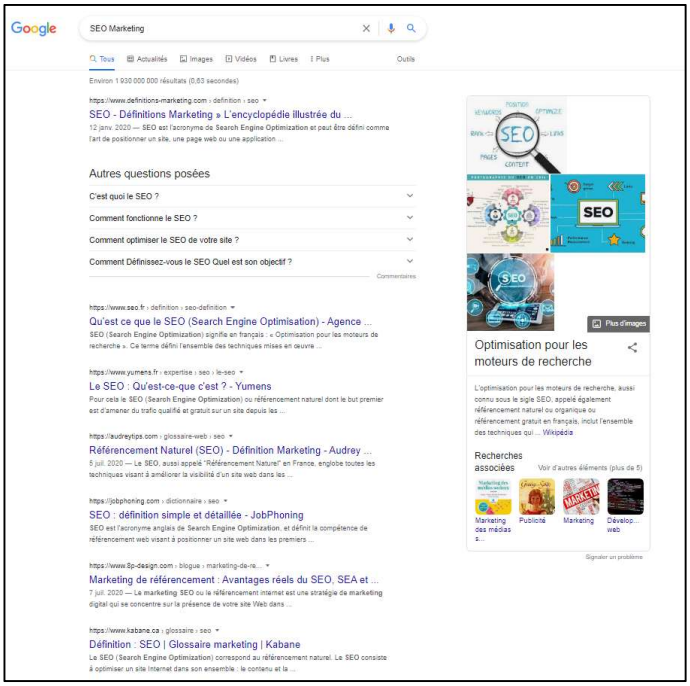
f) **Crawling**: Crawlers use ML to figure the **optimal rate to crawl** a particular URL based on its importance, how often it is updated etc, [102]*Kevin Rowe* [2021].

---

[117] https://www.searchenginejournal.com/googles-gary-illyes-explains-how-rankbrain-works/292381/
[118] https://patents.google.com/patent/US20070156887
[119] https://patents.google.com/patent/US9773256
[120] https://support.google.com/google-ads/answer/1752122?hl=en

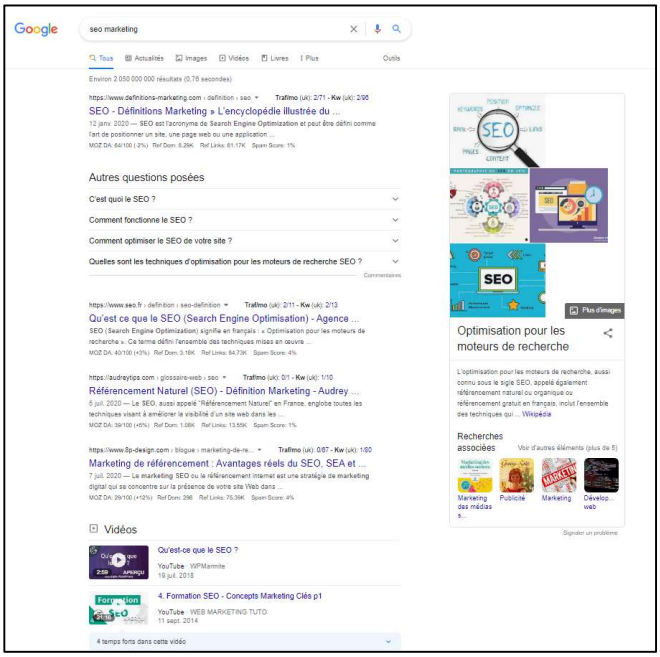
g) **User Classification:** ML is used for personalized search based on the kind of a user can be, [102]*Kevin Rowe* [2021].

h) **Personalized Search Query vs. Incognito Window Query:** Google's personalized search patent, [121]*AI-ML google.com 2*, states that, "personalized search generates different search results to different users of the search engine based on their interests and past behaviour." A wonderful way of testing this is doing search using incognito window of a browser and general window. For example, if we search "SEO Marketing" in general window of Google Chrome *(it produces personalised results as per your browser cookies tracking)* versus incognito window of Google Chrome *(see the two screenshots below, Screenshot-1 and Screenshot-2)*, we see noticeable variations in both SERPs.



Screenshot-1: SERP obtained in Incognito Window of Google Chrome

[121] https://cloud.google.com/products/ai



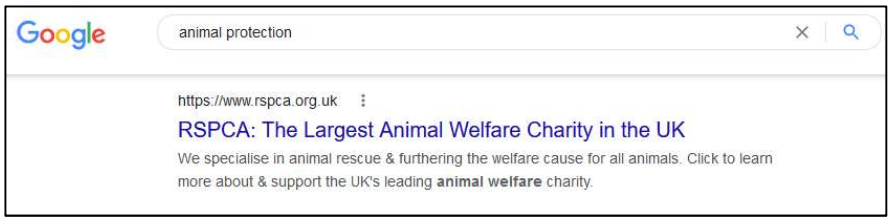
Screenshot-2: SERP obtained in General Window of Google Chrome

This scenario happens in case of "Auto-suggest" feature too. Incognito window won't show you auto-suggest results from web, while general window will show.

i) **Search Query Intent Understanding - Analysis of Search and Classification of Phrases**: As per [102]*Kevin Rowe* [2021] Search engines run different classifiers on the search query known as Query Clarification. E.g. **Detecting navigational vs. informational vs. transactional queries.** Or **news queries vs. local intent queries vs. shopping queries** etc. using:

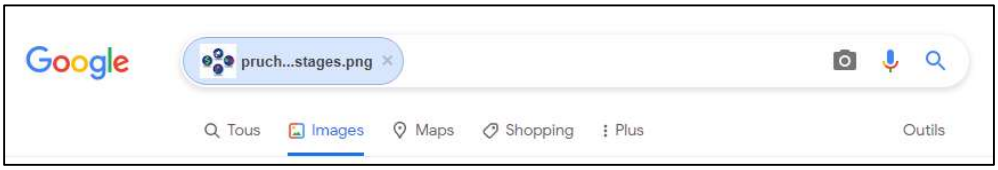
- Spelling suggestion / correction,
- Query Synonyms Identification or Expansion and
- Intent disambiguation.

For example, when you search for [eagles], the intent is to find Eagles the band or Philadelphia Eagles or the bird *(or all 3)*. Search engines use **synonym detection technic** to detect when two queries are alike and used in similar contexts or contain synonyms. It evaluates which query is more effective in terms of click-through rates and replace one with the other, [111]*ML Lucidworks.com 1*. Many a times we don't see the keywords in search snippets, but their synonyms and it all happens due to machine learning. For example, if we search [Animal Protection] we can see results with [Animal Welfare] too; just like Screenshot-3 below.



Screenshot-3: We queried [Animal Protection] but got result for [Animal Welfare]

j) **Image Search to Understand Photos:** As per [122]*Brent Csutoras* [2019] average 4050 photos are uploaded on Facebook in a second and as per [123]*Cooper Smith* [2013] average 1132 photos are uploaded on Instagram in a second. It makes around 350 million photos uploaded on FB alone in 24 hours. Analysing and cataloguing these many uploads is almost arduous job. Machine Learning analyses colour and shape patterns and pairing of existing schema data of photos to help the search engine understand them accurately. It not only enables Google Search Engine to catalogue the images but empowers the reverse image search too. Reverse image search feature of the Search Engines enables users to search by image instead of textual query. This feature is accessible from Google Image section as shown in below Screenshot-4.



Screenshot-4: SERP obtained in General Window of Google Chrome

ML Technology and Algorithmic Models for SERP: Machine learning technologies used in search engine results generation include **machine learned models** trained to predict user actions and responses to search results, *Koushik et al.* [2023], **machine learning components** that optimize search engine results **based on user information and search requests**, *Qi et al.* [2019], and a **machine learning framework** that improves search quality through automated personalization based on user search history, *Madden* [2018]. These technologies utilize features extracted from queries, user profiles, and information about potential search results to rank and personalize search results for users. **CNNs and DNNs** are also used as classifiers to improve image classification and object recognition in search engine results, *Abinaya & Jaishanthi* [2014]. Additionally, **XML technologies** are employed to **cluster search results** from web databases based on similarity and assign

[122] https://www.searchenginejournal.com/bert-explained-what-you-need-to-know-about-googles-new-algorithm/337247/
[123] https://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9?r=US&IR=T

meaningful labels, *Yoganarasimhan* [2020]. *Karwa & Honmane* [2019] and *Koragoankar et al.* [2023] both propose the use of machine learning in search engines, with Karwa focusing on the use of ML techniques to provide more relevant web pages, and Koragoankar emphasizing the use of NLP techniques to enhance search results. *Buchholz et al.* [2004] presents a system that allows users to re-rank search results using machine learning, while *Gungor* [2013] proposes a machine learning approach for displaying query results in a more effective way. Some of the key ML technologies used in search engine result generation include **pattern detection, generative AI, AI for search rankings, and NLP**. Search engines use ML for **pattern detection** to **identify spam or low-quality content**, which helps improve the quality of search results, [102]*Kevin Rowe* [2021]. **Generative AI** is used to **generate more relevant and informative results**. It can understand user queries more deeply and generate results based on the context of the query, *Deshmukh* [2023]. **Complex AI algorithms are** used to rank webpages, videos, and other content in search results. These algorithms prioritize different factors, such as keywords, content quality, and user experience, to determine the most relevant results for a given query, [124]*Berkel* [2023]. **Advanced NLP** is used to analyse the **context and nuance of search** queries, enabling search engines to provide more accurate and relevant responses, especially in conversational search experiences, [125]*Kevin Watts* [2023]. These ML technologies are transforming the way search engines generate results, making the search experience more personalized, informative, and efficient for users. Based on above FLR the major ML technologies used in SERP generation can be identified as follows:

1. **Ranking Models:**
   - **Learning to Rank:** This ML algorithm ranks webpages in SERPs based on their relevance to search queries, considering **keyword relevance, backlinks, domain authority, and user behavior** factors.
   - **RankBrain:** As talked above it is Google's **neural network-based ML model** that helps understand the intent behind search queries and rank webpages accordingly. It can NLP to identify patterns in search queries that traditional algorithms may miss.
   - **Factorization Machines:** Factorization Machines are ML models that can handle high-dimensional data, such as the vast amount of information collected by search engines. They are used to improve the accuracy of ranking algorithms by capturing complex relationships between different factors.
2. **NLP Models:**
   - **Text Classification:** NLP algorithms classify webpage content into different categories, such as news, articles, or product reviews. This helps SEs provide more relevant results for specific topics.
   - **Sentiment Analysis:** NLP techniques are used to analyse the sentiment expressed in search queries and webpage content. This helps SEs understand user's emotional state and provide more relevant results.

[124] https://www.schemaapp.com/schema-markup/is-google-generative-ai-search-an-opportunity-or-threat-for-seos/
[125] https://www.linkedin.com/pulse/future-search-generative-ai-kevin-watts

- **Question Answering:** ML models identify and answer questions directly within SERPs. This **"zero-click" SERP feature** saves users time and effort by providing immediate answers without requiring them to visit individual websites.

3. **Information Retrieval Models:**
   - **Latent Semantic Indexing (LSI):** LSI is an IR technique that uses statistical methods to identify relationships between words and concepts. This helps search engines understand the semantic meaning of webpage content and match queries with more relevant results.
   - **Vector Space Models:** Vector space models represent **text documents as vectors of numbers**, enabling search engines to calculate the similarity between queries and webpages. This is used for tasks like query refinement and relevance ranking.
4. **Recommendation System Models:**
   - **Collaborative Filtering:** This ML technique uses past user behaviour and preferences to **recommend similar websites or products**. It is commonly used in **e-commerce platforms and personalized search features.**
   - **Content-Based Filtering:** This technique analyses the content of webpages and suggests other **webpages with similar characteristics**. It is used in **news recommendations and personalized search results.**
5. **Anomaly Detection Models:**
   - **Fraud Detection:** ML models are used to identify and flag suspicious or fraudulent websites that may be trying to manipulate search results. This helps maintain the integrity of SERPs and protect users from harmful content.
   - **Spam Detection:** ML techniques are used to identify and filter out spammy or irrelevant websites that may be trying to game search algorithms. This ensures that only high-quality and relevant content appears in SERPs.

These are just a few examples of the ML technologies used in search engine results generation. As ML techniques continue to evolve, we can expect even more robust and personalized search experiences in the future.

Differences between AI and ML: as per [126]*Satavisa Pati* [2021]; [127]*Geeksforgeeks.org* and [128]*Bernard Marr* are as shown in *Table-10* below:

[126] https://www.analyticsinsight.net/the-difference-between-artificial-intelligence-and-machine-learning/

Sr.	Artificial Intelligence (AI)	Machine learning (ML)
1	AI is a technology which enables a machine to simulate human behavior.	ML is a subset of AI which allows a machine to automatically learn from past data without programming explicitly.
2	The goal of AI is to make a smart computer system like humans to solve complex problems.	The goal of ML is to allow machines to learn from data so that they can give accurate output.
3	In AI, we make intelligent systems to perform any task like a human.	In ML, we teach machines with data to perform a particular task and give an accurate result.
4	ML and DL are the two main subsets of AI.	DL is a main subset of ML.
5	AI has a very wide range of scope.	ML has a limited scope in comparison of AI.
6	AI is working to create an intelligent system which can perform various complex tasks.	ML is working to create machines that can perform only those specific tasks for which they are trained.
7	AI system is concerned about maximizing the chances of success.	ML is mainly concerned about accuracy and patterns.
8	The main applications of AI are Siri, customer support using chatbots, Expert System, Online game playing, intelligent humanoid robot, etc.	The main applications of ML are Online recommender system, Google search algorithms, Facebook auto friend tagging suggestions, etc.
9	Based on capabilities, AI can be divided into three types, which are, Weak AI, General AI, and Strong AI.	ML can be divided into mainly three types that are Supervised learning, Unsupervised learning, and Reinforcement learning.
10	It includes learning, reasoning, and self-correction.	It includes learning and self-correction when introduced with new data.

[127] https://www.geeksforgeeks.org/artificial-intelligence/artificial-intelligence-vs-machine-learning-vs-deep-learning/
[128] https://bernardmarr.com/what-is-the-difference-between-artificial-intelligence-and-machine-learning/

11	AI completely deals with structured, semi-structured, and unstructured data.	ML deals with structured and semi-structured data.

Table-10: Difference between AI and ML

Discussion: AI is a broader concept that encompasses the creation of intelligent systems, while ML is a specific approach within AI that focuses on enabling machines to learn from data. ML is a tool and technique both used to achieve the deeper and dedicated goals of AI. Neural Network forms the base of most of technics and technologies used in ML particularly concerning the **SERP**. ML constitutes something for every stage of SERP cycle, starting from Search Query input up to SERP production; at the backstage of this process various NNs synthesize the results. Chart-31 below summarizes the facts and findings in above research.

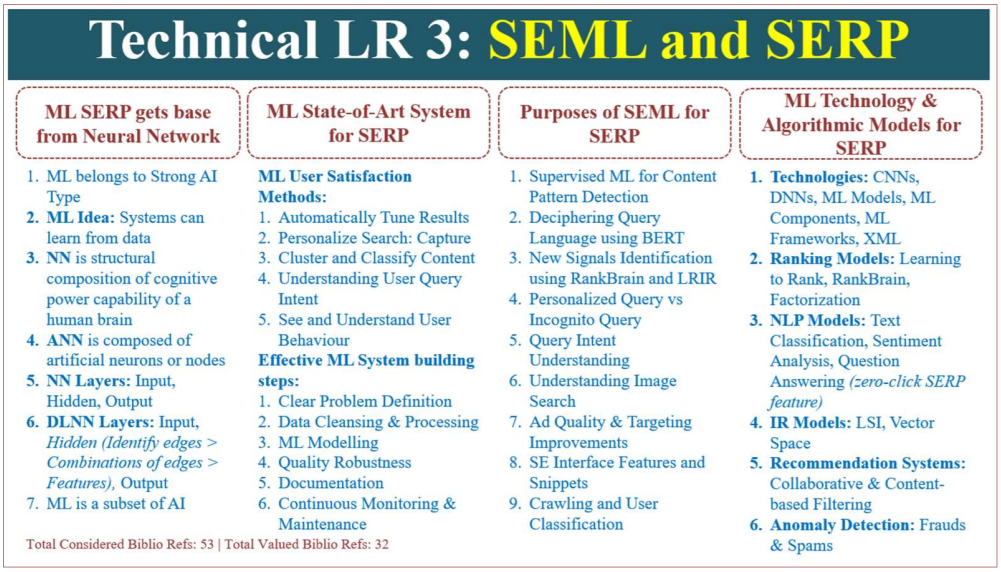


Chart-31: SEML and SERP

3.4. Deep Learning in SERP Production:

SERP features and mechanisms backed by DL: To get a visual idea about Deep Learning composition and modelling, refer to right side 'Deep Learning Neural Network' section of Diagram-9, Appendix-9 and also talked in previous section. Deep learning plays a crucial role in search engine results production, particularly in enhancing retrieval and ranking performance, *Ye et al.* [2015]. It enables the development of more sophisticated and efficient data mining and machine learning solutions to improve the quality of search and recommendation systems. Deep learning models can process complicated, large-scale, and real-time data, provide personalized

content and enhancing the customer experience. It is particularly helpful in NLP tasks, such as **query understanding, retrieval and ranking, and language generation**, which are essential components of search and recommender systems, *Li et al.* [2022]. Deep learning mechanisms can be used to **extract features from search queries, generate word vectors, and map** them to words from a corpus of data, resulting in more accurate and relevant search results, *Venkata* [2020]. Additionally, deep learning approaches have been applied to **query auto completion systems, improving their accuracy and scalability**, and making progress towards **neural query auto-completion** in production search engines, *Long et al.* [2020]. In the context of search systems, deep natural language processing has been identified as a powerful tool for improving search performance, *Guo et al.* [2019]. Furthermore, deep reinforcement learning has been proposed as a promising approach for continuously updating information seeking strategies and maximizing long-term user rewards in search, recommendation, and online advertising, *Zhao et al.* [2018]. Deep Learning enables search engines to **analyse vast amounts of data, learn various tasks, and provide more accurate and relevant results**. Google's **RankBrain algorithm, for instance, was one of the first to exhibit deep learning** capabilities, and it has since been integrated into the search process. Deep learning technology, based on **artificial neural networks (ANN)**, allows machines to learn without human supervision, ultimately aiming to create successful searches that satisfy users, [129]*ZhaiCreativoMedia*. So, we can see from above findings that Deep learning is playing an increasingly important role in the production of SERPs. Here are some of the ways deep learning is being used to produce SERPs:

1. **Understanding search intent and context:** Deep learning can be used to analyse natural language queries and identify the underlying intent behind the search. This helps search engines provide more relevant and useful results. For example, if a user searches for "best restaurants in Paris," deep learning can understand that the user is looking for a list of restaurants, not just a definition of the word "restaurant." This way Deep learning can analyse the context of a user's search query, considering factors such as the time of day, location, and current events to provide more relevant and personalized results. Deep learning models can analyse the nuances of search queries, going beyond just matching keywords. They can decipher the user's intent, whether it's finding specific information, comparing products, or seeking directions. This deeper understanding of user intent enables search engines to deliver more relevant and helpful results.
2. **Ranking results:** Deep learning can be used to learn complex patterns and relationships between queries and webpages. This information can then be used to rank webpages in SERPs, ensuring that the most relevant and authoritative results are displayed first.
3. **Personalization as per User History and Behaviour:** Deep learning can be used to personalize SERPs for individual users. This is done by analysing a user's past search history, interests, and location to tailor

[129] https://www.datasciencecentral.com/deep-learning-and-google-search-the-future-of-seo/

the results to their needs. For example, if a user has previously searched for information about travel to France, deep learning can rank websites about French travel at the top of their SERPs. This means that users receive results tailored to their individual interests, past search history, and browsing patterns. This personalized approach enhances the user experience and ensures that users find the information they need most effectively.

4. **Generating summaries:** Deep learning can be used to summarize webpages, providing users with a quick overview of the information without requiring them to read the entire article.
5. **Answering questions directly:** Deep learning can be used to answer questions directly in SERPs. This "zero-click search" feature saves users time and effort by providing immediate answers without requiring them to visit individual websites.
6. **Identifying and ranking reliable sources:** Deep learning can be used to identify and rank reliable sources of information. This helps ensure that users are directed to accurate and trustworthy information.
7. **Creating interactive experiences:** Deep learning can be used to create interactive experiences within SERPs, such as expandable summaries, visual representations of data, or interactive maps. This enhances the user experience and makes it easier to understand complex topics.
8. **Generating code snippets:** Deep learning can generate code snippets based on user queries, making it easier for programmers and developers to find and use relevant code segments. This can streamline the development process and improve productivity.
9. **Transcribing audio and video:** Deep learning can transcribe audio and video content into text, providing a searchable transcript that users can easily access. This is particularly useful for educational videos or lectures where users want to quickly find specific information.
10. **Understanding Content Semantics:** Traditional search algorithms relied on keywords and links to assess the relevance of websites. However, deep learning models can grasp the meaning and context of text, enabling them to evaluate the quality and relevance of content at a much deeper level. This allows search engines to surface more informative and engaging content that truly addresses user queries.
11. **Entity Recognition and Contextual Understanding:** Deep learning algorithms can identify and extract entities from text, such as people, places, and products. This enables search engines to understand the context of search queries and provide more precise results. For instance, if a user searches for "Michael Jordan," the algorithm can identify him as a basketball player and suggest relevant results related to his career achievements.
12. **Real-Time Search and Dynamic Results:** Deep learning algorithms can process vast amounts of data in real-time, enabling search engines to provide dynamic and up-to-date results. For example, if a user

searches for a current event, the algorithm can incorporate real-time news feeds and social media updates to provide the most relevant and accurate information.

13. **Spam Detection and Content Quality Assessment:** Deep learning models can identify and filter out low-quality, spammy, or misleading content from search results. This ensures that users are presented with reliable and trustworthy information. Additionally, deep learning can help assess the quality of content, such as its originality, accuracy, and overall value, further enhancing the search experience.

These are just a few examples of how deep learning is being used to improve SERPs. As deep learning technology continues to develop, we can expect even more innovative and transformative applications in this field.

DL simulates the learning process of human brain: Deep learning simulates the learning process of the human brain by using neural networks with multiple layers to process data *(simulated neurons)* and make predictions. These networks attempt to mimic the behaviour of the human brain by recognizing, classifying, and describing objects within the data, learn by processing large amounts of data and adjusting the connections between neurons to improve performance. While deep learning excels at processing large volumes of data and making accurate predictions, it is still far from matching the human brain's ability to process and learn from information, [130]*dl-connectedworld.com-1 &* [131]*dl-ibm.com-3*. The **human brain outperforms deep learning models in intuitive feature extraction, pattern recognition, and language comprehension**. However, deep learning models thrive on large-scale data processing and computational power, making them suitable for handling complexity, [132]*Sohan Kanse* [2023]. The understanding of artificial deep learning systems may contribute to our understanding of how humans and animals learn and remember, as human brains are deep neural networks that process information in a parallel, distributed, and context-sensitive fashion, *McClelland & Botvinick* [2020]. The success of deep learning in tasks such as image recognition and speech processing is attributed to its ability to mimic the distributed patterns of brain activity, *Manish Gupta* [2022]. However, there are still **differences between deep learning and the brain**. While **deep learning networks have uniform neurons, the brain's neurons have complex structures** that may play a role in supporting deep learning, *Zhang & Duan* [2022]. To make deep learning more biologically realistic, researchers have designed artificial neurons with compartments like the structure of real neurons, which have shown improved learning capabilities, *Arana et al.* [2023]. Overall, deep learning models aim to replicate the learning mechanisms

[130] https://connectedworld.com/deep-learning-vs-the-human-brain/
[131] https://www.ibm.com/topics/deep-learning
[132] https://www.linkedin.com/pulse/deep-learning-vs-human-brain-classification-complexity-sohan-kanse

observed in the human brain, but there are ongoing efforts to further align the two, *Kamruzzaman & Charles* [2019] & *Guerguiev et al.* [2017]. Deep learning, a field of AI, aims to replicate the human brain's ability to process complex data, learn quickly, and solve various tasks, *Dandan Mo* [2012]. This is achieved through the use of deep architectures, such as **CNNs** and Deep Belief Networks (**DBNs**), which have shown promising performance in AI tasks like computer vision and speech recognition. The use of **convolutional deep belief networks in deep robotics** has further advanced the simulation of the human brain, particularly in the identification of brain diseases, *Kumar et al.* [2018]. However, the inner workings of deep learning are not fully understood, which can impact user trust, especially in production settings, *Le et al.* [2020]. Despite this, deep learning continues to be a key research focus, with the potential to significantly impact various fields, *Liu & Chen* [2019]. DL is a subset of machine learning that involves neural networks with multiple layers **DNN,** also known as **ANNs.** The depth in deep learning comes from the multiple layers through which the data is transformed. So, it is clearly sayable that DL's value lies in its ability to automatically learn complex representations from data, solve intricate problems, and contribute to advancements across industries, making it a powerful and transformative technology in the field of artificial intelligence.

G-AI models development using DL: Generative Artificial Intelligence **(G-AI)** models can be developed using DL techniques. Gen-AI is particularly suited for **Bayesian Computation** and can be applied to easily simulated Bayesian models. The process involves generating a large training dataset, finding a summary statistic for dimensionality reduction, and using a DNN to uncover the inverse Bayes map between parameters and data. This allows for **high-dimensional regression** with feature selection and nonlinearity, equivalent to DL. The advantage of Gen-AI is its **density-free nature**, which avoids the need for Markov Chain Monte Carlo **(MCMC)** simulation of the posterior. Deep quantile neural networks are proposed as a general framework for inference and decision making. Examples of Gen-AI applications include **traffic flow prediction and image generation**, *Polson et Soklov* [2023], *Xia Dong* [2023] & *Xingcheng Xu* [2023]. As per [133] *World Bank* [2023] G-AI models, such as SERP, can be developed using deep learning techniques. Deep generative models **(DGMs)** are neural networks with multiple hidden layers that are trained to approximate complex, high-dimensional probability distributions. DGMs have been successfully used to estimate the likelihood of observations and generate new samples from the underlying distribution. The literature on DGMs is vast and growing rapidly, with approaches such as normalizing flows **(NF),** variational autoencoders **(VAE),** and generative adversarial networks **(GAN)** being popular. These models have been applied to various domains, including **image generation, voice synthesis, and movie generation**. However, there are still challenges in designing and training DGMs effectively and understanding their effectiveness. Researchers are actively

[133] https://openknowledge.worldbank.org/entities/publication/4f623641-ba34-4f0d-9a7d-105f02a5ee00

working on advancing the theoretical understanding of DGMs and exploring their connections with optimal transport. *Salakhutdinov* [2009] and *Sohn et al.* [2015] both explore the use of deep generative models in AI. Salakhutdinov focuses on the efficient learning of high-level feature representations, while Sohn develops a deep conditional generative model for structured output prediction. *Lee et al.* [2018] presents a Graphical user interface (**GUI)-based** software tool for developing deep neural network models, addressing the challenge of model design. *Xu et al.* [2015] provides an overview of deep generative models, including their applications in **image processing, speech recognition, and information retrieval**. G-AI models, such as **GPT-3** and generative adversarial networks **(GANs),** are developed using deep learning and play prominent role in SERP generation. These models are powerful tools for producing various types of content, including text, imagery, audio, and synthetic data. They can be fine-tuned for various tasks and adapted to different domains and industries with proper training and fine-tuning. While generative AI has made significant progress, there are still challenges related to accuracy, bias, and the need to carefully select and manage the data used to train these models. Generative AI has practical applications in different areas, including product development, content creation, and customizing applications, [134]*Rebecca* [2025], [135]*George Lawton-techtarget,* [136]*Damian Brady* [2023] & [137]*altexsoft.com* [2022]. So, Deep learning plays a crucial role in the development of generative AI models.

DL and ML relation with each other: *Alaskar & Saba* [2021] and *Bhattacharjee* [2019] states that Machine learning and deep learning are closely related fields within the broader domain of data science and artificial intelligence. Machine learning, which involves training computers to learn from data, is a foundational concept that underpins deep learning. While both fields have their own unique applications, they are interconnected and often used in conjunction with each other, *John Cherrie* [2016] and *Shinde & Shah* [2018]. Deep learning is a subset of machine learning that uses deep neural networks with multiple layers to learn and extract features from complex data, *Shanthamallu & Spanias* [2023]. **Neural networks** are the foundation of deep learning, and they are used to model the relationships between input data and output predictions, *Frederick Douglass* [2023]. Deep learning algorithms allow these networks to learn and improve over time by adjusting the **weights and biases of the neurons**, *Jin Hu* [2023]. Machine learning, on the other hand, is a broader category of algorithms that can learn patterns from data and make predictions or decisions without being explicitly programmed, *Fernández & Peters* [2023]. Deep learning is a particular branch of machine learning that makes use of

[134] https://www.coveo.com/blog/generative-models/
[135] https://www.techtarget.com/searchenterpriseai/definition/generative-AI
[136] https://github.blog/2023-04-07-what-developers-need-to-know-about-generative-ai/
[137] https://www.altexsoft.com/blog/generative-ai/

machine learning functionality and moves beyond its capabilities, *Y Yamana* [2023]. As per [138]*Arne Wolfewicz* [2020], [139]*Coursera Staff* [2023] and [140]*Patrick Grieve* [2023] deep learning is a specialized form of machine learning that relies on neural networks to process data and make predictions, while machine learning is a broader field that encompasses various techniques for enabling machines to learn from data.

So, Machine Learning is a broader concept that encompasses various techniques and algorithms designed to enable computers to learn from data and make decisions or predictions without being explicitly programmed. Machine learning techniques can be categorized into three main types:

1. **Supervised Learning:** In supervised learning, the algorithm is trained on a labelled dataset, where the input data is paired with corresponding output labels. The goal is to learn a mapping from inputs to outputs, making predictions on new, unseen data.
2. **Unsupervised Learning:** Unsupervised learning involves working with unlabelled data to discover patterns or relationships within the data. Clustering and dimensionality reduction are common tasks in unsupervised learning.
3. **Reinforcement Learning:** Reinforcement learning involves training an agent to make decisions in an environment by receiving feedback in the form of rewards or penalties. The agent learns to take actions that maximize cumulative rewards over time.

Deep Learning: Deep learning is a specialized subfield of machine learning that focuses on the use of artificial neural networks (ANNs), particularly deep neural networks (DNNs), to model and solve complex problems. DNNs consist of multiple layers (deep architectures), allowing them to automatically learn hierarchical representations of data. Key characteristics of deep learning include:

1. **Neural Networks:** Deep learning models are based on artificial neural networks inspired by the structure and function of the human brain. These networks consist of layers of interconnected nodes (neurons) that process information.
2. **Deep Neural Networks:** Deep neural networks have multiple hidden layers between the input and output layers. The depth of these networks enables them to capture intricate features and representations of data.
3. **Feature Learning:** Deep learning excels at feature learning, automatically discovering and extracting relevant features from raw data. This eliminates the need for manual feature engineering in many cases.

[138] https://medium.com/levity/deep-learning-vs-machine-learning-whats-the-difference-e367803bb96d
[139] https://www.coursera.org/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide
[140] https://www.zendesk.co.uk/blog/machine-learning-and-deep-learning/

4. **Representation Learning:** Deep learning models automatically learn hierarchical representations of data, allowing them to understand and represent complex patterns and abstractions.
5. **Applications:** Deep learning has demonstrated remarkable success in various applications, **including image and speech recognition, natural language processing, computer vision, and playing strategic games**.

Discussion: Here by now, one very important point to understand is that by going through this section along with the previous section of SEML, it becomes evident that why Neural Network is superset to Deep Learning. Deep Learning as clear from its name is a further deeper and more complex concept within Neural Network. This section should further help the readers to understand how DL applications and models help in processing query requests. But to come up with something concrete as a further research possibility pertaining to SERP, it becomes imperative to traverse across further sub-sets of DL as G-AI and LLMs. Chart-32 below summarizes the facts and findings in above research.

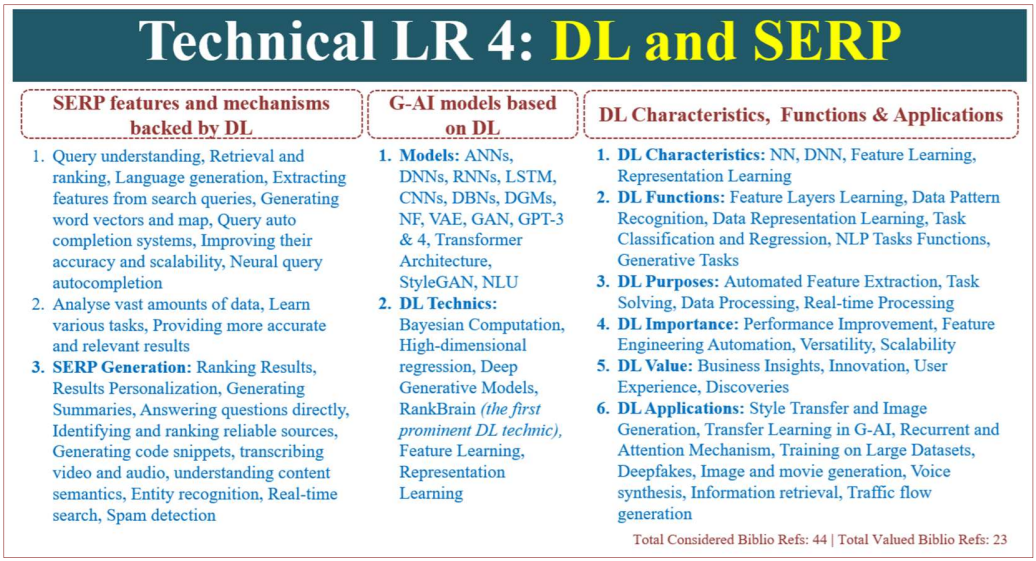


Chart-32: Deep Learning and SERP

3.5. Generative-AI and LLMs are the core of SERP:

G-AI is the branch of AI that can generate text, images, or other data using generative models, often in response to prompts. G-AI models learn patterns and structure of input training data from large datasets and can then

create novel outputs that resemble the training data, [141]*G-AI-SAP*. These characters of G-AI make it the ideal fit to needs of search engines to enhance the search experience. The roots of G-AI can be traced back to early work in ML and AI. Early models focused on rule-based systems and basic pattern recognition. As read in Chapter-1 above, the breakthrough in G-AI is seen as the introduction of GANs by **Ian Goodfellow** and his colleagues in 2014. The use of G-AI in SEs involves the application of GANs to distinguish between relevant and irrelevant search results, *Zhihan Lv* [2023]. Search engines like Google and Bing are incorporating G-AI technologies, such as Google's Search Generative Experience **(SGE)** and **Bing Chat**, to generate information for users and provide a more conversational and personalized search experience. G-AI in search involves using AI to help users uncover more insights about topics, process information more efficiently, and **generate images or text**. However, it also presents challenges such as misinformation and potential decreases in website traffic due to users finding information directly on the search results page. *Galanin et al.* [2020] propose a relevance evaluation approach based on GANs, which can **distinguish between relevant and irrelevant search results**. This approach improves the quality of search by utilizing fully automated reference results selection and multi-layered generator and discriminator networks with dense layers. The G-AI models generate reference results that are potentially relevant to the search space, enhancing the search experience, *Capra & Arguello* [2023]. As said above also, companies like Microsoft and Google are incorporating G-AI technology into their search systems, such as the "**new Bing**" search system, to transform how people search for information, *Ted Selker* [2023]. These G-AI chat systems, like **ChatGPT**, are integrated with existing search tools to provide innovative ways for users to engage with online information, *Cohen & Queen* [2023]. *Colton et al.* [2021] explores the use of a GANs to produce images that reflect a given text prompt, with potential applications in search result page design. *Zhao et al.* [2005] presents a technique for automatically producing wrappers to extract search result records from dynamically generated result pages, using visual content features and HTML tag structures. *Sontag et al.* [2012] introduces a generative model of relevance for **personalizing web search results** to specific users, based on their search history. These studies collectively demonstrate the diverse applications of G-AI in search engine result generation. G-AI refers to a class of AI systems that can **generate new content**, whether it be **text, images, audio, or other forms of data**. These systems use various techniques, including **machine learning and deep learning**, to **understand patterns in existing data and then create new**, similar content.

How G-AI Works: G-AI models work by first being trained on a large dataset relevant to the desired output type (e.g., text, images), [135]*George Lawton-techtarget*. During training, the model learns the underlying patterns and distributions in the data, [141]*G-AI-SAP and* [135]*George Lawton-techtarget*. When given a prompt or input, the

[141] https://www.sap.com/products/artificial-intelligence/what-is-generative-ai.html

model can then generate new content by sampling from the learned distributions while maintaining key characteristics of the training data. The core principles behind generative AI involve deep learning techniques like neural networks that can effectively model and learn complex data distributions, [141]*G-AI-SAP*. The training process tunes the model's parameters to capture the relevant patterns for a particular use case or domain.

Applications, Impact and Ethical Concerns: Generative AI has numerous applications across various industries, [141]*G-AI-SAP and* [135]*George Lawton-techtarget*, including:

1. Content creation *(writing, art, music, marketing materials)*
2. Software development *(code generation)*
3. Product design and prototyping
4. Scientific research *(drug discovery, material design)*
5. Entertainment *(video games, movies)*
6. Data augmentation for training other AI models

While still an emerging field, G-AI is already transforming how content is created and automating certain creative tasks, there exists the concerns around potential misuse, bias, and the broader societal impacts of this technology, [142]*G-AI-Wiki*. G-AI holds great promise in various applications, from creative endeavours to problem-solving tasks. As the field progresses, researchers and developers are working to strike a balance between pushing the boundaries of what AI can create and ensuring responsible and ethical use.

Transformative potential of G-AI in enhancing the SERP: G-AI models, such as ChatGPT, Bard, and CoPilot, are rapidly advancing technologies that autonomously create new content like text, images, audio, and video, *Zhihan Lv* [2023], [133]*World Bank* [2023] and *Weisz et al.* [2023]. These AI systems are revolutionizing various industries by enhancing search experiences, reshaping information generation methods, and boosting software engineering productivity, *Louridas & Ebert* [2023]. The G-AI market is projected to grow significantly, from 1.5 billion dollars in 2021 to 6.5 billion dollars by 2026, with a compound annual growth rate of 34.9%, [133]*World Bank* [2023]. To ensure safe and productive use of G-AI, principles for designing applications have been proposed, focusing on factors like multiple outcomes, exploration, control, and guarding against potential harms. These advancements highlight the transformative potential of G-AI in driving innovation and creativity across various sectors. G-AI is used in various ways to enhance SERPs.

General G-AI Models and Technologies:

[142] https://en.wikipedia.org/wiki/Generative_artificial_intelligence

1. **Markov Models:** Markov models, which use **probability distributions** to model sequences of events, played a role in early generative models. These models were used for tasks like text generation, but they had limitations in capturing complex dependencies. The thesis of *Sebastian* [2019] on Hidden Markov Model presents a detailed work on how the user behaviour probabilities can be distributed and calculated.
2. **Neural Networks:** With the advent of NNs and the development of deep learning techniques, G-AI saw significant advancements. **RNNs** and **LSTM** networks became popular for **sequence generation tasks.**
3. **The Generative Adversarial Networks (GANs):** GANs consist of a **generator and a discriminator network** that are trained simultaneously. The generator creates data, and the discriminator evaluates its authenticity. This adversarial training process results in the generator producing high-quality, realistic content. GANs consist of two neural networks - a generator that creates new data instances, and a discriminator that evaluates whether the generated data is real or fake, [141]*G-AI-SAP*. Through this adversarial training process, GANs can generate highly realistic images, videos, and other data types.
4. **The Variational Autoencoders (VAEs):** VAEs are another class of generative models that **use variational inference** to learn a probabilistic mapping between input data and a latent space. VAEs have been applied to tasks like image generation and data synthesis. VAEs encode input data into a compressed latent space representation, and then decode this representation to generate new data instances with some randomness introduced, [141]*G-AI-SAP*. They are commonly used for generating images, music, and other continuous data.
5. **Transformer-based models:** These models, like GPT-3 *(Generative Pre-trained Transformer)* and GPT-4, use a transformer architecture to generate coherent text by considering the entire context of the input, [141]*G-AI-SAP*. They are widely used for language tasks like translation, summarization, and open-ended text generation. In the field of NLP, G-AI has been applied to tasks such as text completion, dialogue generation, and language translation. Models like OpenAI's GPT have gained attention for their ability to generate coherent and contextually relevant text.
6. **StyleGAN:** G-AI has made notable contributions to **image generation and style transfer**. Models like **StyleGAN** can generate highly realistic images, and neural style transfer techniques allow the transformation of images in the style of famous artworks.
7. **Creative Applications:** G-AI has found **creative applications** in creative fields, including art, music, and design. Artists and musicians are exploring ways to collaborate with AI systems to generate novel and innovative creations.
8. **Deepfakes (worries about Ethical Considerations):** The rise of G-AI has brought about discussions and debates regarding ethical considerations. Concerns include the potential misuse of AI-generated content for

malicious purposes, deepfakes, and the ethical implications of AI systems creating content that may be indistinguishable from human-generated content.

9. Other notable generative models include autoregressive models, normalizing flow models, and diffusion models, [141]*G-AI-SAP*.
10. **Continual Advancements:** The field of G-AI continues to evolve rapidly. Ongoing research focuses on improving the quality, diversity, and controllability of generated content. Researchers are exploring ways to address challenges such as biases in generated content and improving the interpretability of generative models.

G-AI and Large Language Models (LLMs): LLMs and G-AI have garnered significant attention due to their transformative potential, *Caines et al.* [2023], *Ma et al.* [2023], *Douglas* [2023] and *Bariah et al.* [2023]. LLMs like GPT series and ChatGPT exhibit remarkable capabilities in text generation, automated grading, error correction, and even enhancing speech recognition systems. The ethical use and disclosure of LLMs and G-AI have become crucial topics, leading to initiatives like CANGARU *(Generative Artificial Intelligence, and Natural Large Language Models for Accountable Reporting and Use Guidelines)* to establish guidelines for responsible reporting and application in academia, *Cacciamani et al.* [2023]. The integration of LLMs in various domains, such as wireless networks, is envisioned to revolutionize technology by enabling autonomous systems and paving the way for artificial general intelligence (AGI) empowered networks. LLMs and G-AI represent a paradigm shift in technology with vast potential and ethical considerations. LLMs are a type of artificial intelligence model that is trained on vast amounts of text data to understand and generate human-like language, [143]*LLM-Databricks* [2024] and [144]*Dee* [2023]. They use deep learning techniques, particularly the transformer architecture, to capture the statistical patterns and relationships in text. They excel at NLP tasks like text generation, translation, summarization, and question answering, [143]*LLM-Databricks* [2024] and [144]*Dee* [2023]. Popular examples of LLMs include but not limited to GPT-3, GPT-4, BERT, and LaMDA, [144]*Dee* [2023]. LLMs can generate coherent and contextually relevant text outputs when given a prompt or input, [143]*LLM-Databricks* [2024]. They are "foundation models" that can be fine-tuned for specific language tasks and domains, [143]*LLM-Databricks* [2024] and [145]*Bell* [2023]. When we connect GAI and LLMs, GAI is a broader category that includes LLMs as well as other generative models like variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models, [143]*LLM-Databricks* [2024] and [144]*Dee* [2023]. In addition to text generation, generative AI models can create images, music, 3D models, and other data types,

[143] https://docs.databricks.com/en/generative-ai/generative-ai.html
[144] https://www.algolia.com/blog/ai/large-language-models-llms-vs-generative-ai-whats-the-difference/
[145] https://appian.com/blog/acp/process-automation/generative-ai-vs-large-language-models.html

[144]*Dee* [2023] and [145]*Bell* [2023]. Generative AI has applications in content creation, data augmentation, drug discovery, and various creative domains, [143]*LLM-Databricks* [2024] and [144]*Dee* [2023].

Relationship between LLMs and Generative AI: LLMs are a powerful form of G-AI specifically designed for language tasks, while G-AI is a broader field that includes LLMs as well as other models for generating various types of data, [143]*LLM-Databricks* [2024], [144]*Dee* [2023] and [145]*Bell* [2023]. LLMs are a specific type of G-AI model focused on text generation and language understanding, [143]*LLM-Databricks* [2024] and [144]*Dee* [2023]. They are a key component of G-AI systems for tasks involving natural language. As said above, G-AI encompasses a wider range of models and techniques beyond just LLMs, [143]*LLM-Databricks* [2024] and [144]*Dee* [2023]. For example, GANs and VAEs are used for generating images, while other models like WaveNet and MusicVAE are used for audio generation, [144]*Dee* [2023]. In many cases, LLMs and other G-AI models are combined or used together to create multimodal generative systems that can handle multiple data types, [144]*Dee* [2023] and [145]*Bell* [2023]. For instance, an LLM could be used alongside an image generation model to create captions or descriptions for generated images.

G-AI' Large Language Models (LLMs) use in SERPs generation: However, many of the items in this section we have discussed at various occasions in above sections of this thesis, but some are the first timer entry in this thesis. LLMs are increasingly utilized in search engine tasks, including document retrieval and information compression, *Ziems et al.* [2023]. LLMs like GPT-3.5 and GPT-4 exhibit the ability to generate URLs for document retrieval based on human instructions, achieving high accuracy in open-domain question answering tasks, *Gilbert et al.* [2023]. Additionally, research explores the potential of LLMs in compressing and reconstructing text effectively, enabling the processing of larger sets of information than current limits allow, *Suri & Ziaee* [2023]. Suri & Ziaee further add-up that despite lacking cognitive and affective processes, LLMs have shown biases and decision effects like humans when responding to prompts, suggesting a role of language in generating such effects. Overall, LLMs play a significant role in enhancing search engine capabilities and information processing tasks. And here intention is to provide a concrete explanation in context of LLMs. G-AI models that are generally relevant in the context of SERPs and information retrieval include BERT of Google, GPT of OpenAI, T5 *(Text-To-Text Transfer Transformer)* of Google, XLNet (*eXtreme Language Understanding NETwork*) by Google Brain and Carnegie Mellon University, is another transformer-based model that **outperforms BERT; i**t considers bidirectional context while maintaining an autoregressive approach. XLNet, like BERT, contributes to the understanding of contextual relationships within language,

improving the relevance of search results, [146]*saturncloud.io-1*; and RoBERTa *(Robustly optimized BERT approach)*, a variant of BERT that was developed by **Facebook AI.** Here to note that **RoBERTa and XLNet** are widely used in various NLP tasks, including but not limited to **Sentiment Analysis, Text Classification, Named Entity Recognition and Question Answering.** It's important to note that search engines utilize a combination of models and algorithms to determine the ranking and presentation of search results. These may include not only generative models but also a variety of other machine learning techniques, rule-based systems, and heuristics.

Key functions of LLMs in SERP Generation:

1. **Information Gathering and Result Retrieval:** LLMs can act as "generative retrievers" to gather relevant information from various sources like websites, databases, and knowledge bases in response to a user's query, *Zhang et al.* [2024] and *Ziems et al.* [2023]. They can generate URLs or identifiers to retrieve the most pertinent documents, webpages, or data points to include in the SERP.
2. **Layout and Content Organization:** Approaches like GenSERP use LLMs to dynamically organize and layout the retrieved search results into a coherent SERP design, *Zhang et al.* [2024]. Zhang et al. add further the LLM can propose different content layouts, populate them with relevant items (text, images, etc.), and optimize the ranking and presentation of these items for the best user experience.
3. **Answer Generation and Summarization:** LLMs can generate direct answers, summaries, or abstracts from the retrieved search results, providing more focused and condensed information upfront in the SERP, [147]*Marketbrew-4, Zhang et al.* [2024] and [148]*Shorten & Cardenas* [2023]. This can enhance the search experience by quickly surfacing key information without users having to click through multiple links.
4. **Multimodal Integration:** LLMs combined with computer vision models can integrate and present different data modalities like text, images, and videos in a unified and contextually relevant SERP layout, *Zhang et al.* [2024] and [148]*Shorten & Cardenas* [2023].
5. **Personalization and Reranking:** By understanding the user's intent and context from the query, LLMs can personalize and rerank search results to prioritize the most relevant information for that specific user, [147]*Marketbrew-4 and* [148]*Shorten & Cardenas* [2023]. While still an emerging area, leveraging LLMs allows search engines to move beyond static ranked lists to more dynamic, AI-powered SERP generation tailored to each user's information needs, [147]*Marketbrew-4, Zhang et al.* [2024] and [148]*Shorten & Cardenas* [2023].

[146] https://saturncloud.io/glossary/nlp-transformers-beyond-bert-roberta-xlnet/
[147] https://marketbrew.ai/unleashing-the-power-of-large-language-models-for-seo
[148] https://weaviate.io/blog/llms-and-search

However, challenges remain around quality, safety, and scalability of these generative approaches, *Ziems et al.* [2023].

LLM' Chat GPT and SERP Connect: After having an intriguing insight of G-AI and its relationship with LLMs, talking about Chat GPT in connection to both; answers many questions emerged with commencement of Open AI in Nov. 2022. One thing the most common in both of SERP and Chat GPT is that both use G-AI and LLMs up to intensive levels. And evolution of Chat-GPT as the most prominent application in LLMs makes it another reason to give a special consideration to this subject in this thesis. So here below is the appropriate assessment necessary in context of this thesis:

Possibility of integrating Chat GPT in SERPs: Recent studies have explored the potential of integrating ChatGPT, a LLM powered chatbot, into SERP. *Rayna et al.* [2023] found that ChatGPT can improve user experience in terms of usefulness, enjoyment, and satisfaction and search performance, particularly for straightforward questions and general solutions. *Sun et al.* [2023] further highlighted the potential of ChatGPT for relevance ranking in information retrieval, with **competitive results compared to supervised methods**. ChatGPT has been found to excel in answering straightforward questions and providing general solutions, although it may fall short in fact-checking tasks, *Cabra & Arguello* [2023]. Users perceive ChatGPT's responses as having higher information quality compared to traditional search engines like Google, [149]*Magruder et al.* [2023]. However, it is important to note that ChatGPT may also lead to overreliance and generate or replicate misinformation, which can yield inconsistent results, *Bhatia* [2023]. The integration of chat systems with existing search tools has been investigated, and users have reported integrating AI chat into their search process, *Zhao et al.* [2023]. While there are concerns about the factual accuracy of large conversational AI models like ChatGPT, researchers and developers are called upon to improve transparency and factual correctness. In the environment of these speculations, the integration of ChatGPT in SERPs is already underway. **Microsoft's Bing** search engine has integrated with **OpenAI's ChatGPT**, allowing users to receive more timely and up-to-date answers within their chat conversations, [150]*Matt G.* Southern [2023]. OpenAI has also officially launched Bing-powered search for ChatGPT users, making it a fully integrated, real-time, multimedia generative search engine, [151]*Sawers* [2023]. This integration enables ChatGPT to ground its responses in search and web data, providing a more interactive and personalized search experience, [152]*Sangita-packtpub*. Additionally, Microsoft has developed a new AI model, the **Prometheus model**, which is more

[149] https://www.arthroplastyjournal.org/article/S0883-5403(23)00666-6/fulltext
[150] https://www.searchenginejournal.com/bing-integrates-with-chatgpt/487638/
[151] https://techcrunch.com/2023/10/18/openai-formally-brings-bing-powered-search-to-chatgpt-as-dall-e-3-arrives-in-beta/
[152] https://www.packtpub.com/article-hub/chatgpt-for-search-engines

powerful than ChatGPT and customized for improved relevance, accuracy, and safety in search results, [153]*Microsoft-AIVideo*. But there is no threat to any kind of replacement of Search Engine by Chat GPT; Search Engine are going to stay in the scene for longer or in parallel with more strong purpose than now.

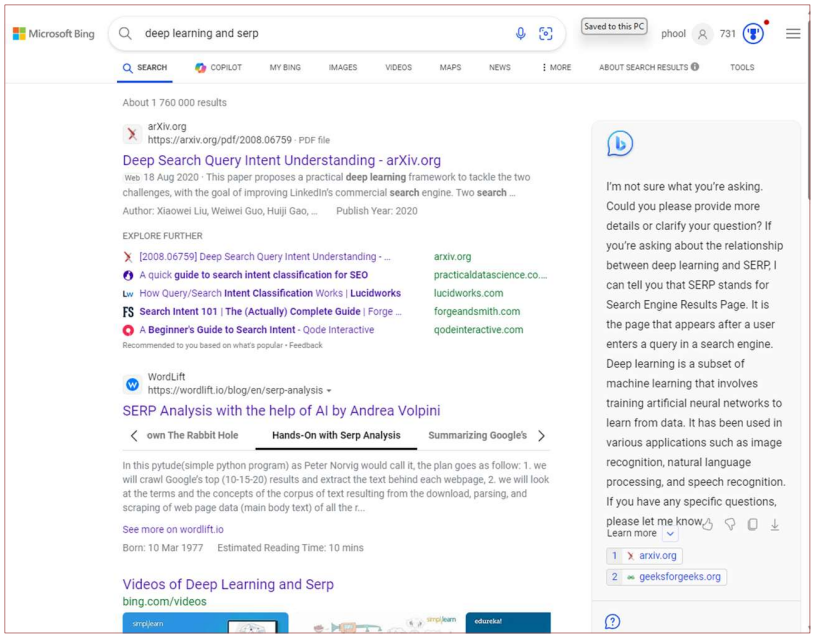
Possibility of replacing of Search Engine by Chat GPT: Taking further from last line of previous paragraph, *AlAfnan et al.* [2023] found that **ChatGPT can provide accurate and reliable information, particularly for theory-based questions**, but also highlighted the **risk of unethical use**. *Ruiyun et al.* [2023] compared ChatGPT and Google Search, noting that ChatGPT excels in answering straightforward questions and providing general solutions, but may lead to overreliance and misinformation. *Arif et al.* [2023] also expressed same kind of uncertainties while discussing that the use of ChatGPT in medical education, cautioning against its potential to generate papers **lacking critical thinking** and clinical reasoning. *Sorecau & Sorecau* [2023] proposed an alternative application to ChatGPT that uses reliable sources to enhance the learning process. ChatGPT has the potential to replace traditional search engines like Google due to its ability to generate human-like responses to natural language queries and provide a wide range of information based on its training dataset from various sources, *Salmon* [2023]. ChatGPT's integration with search engines like Bing combines keyword search with interactive conversations, offering a more convenient and efficient user experience, *Thomas Hügle* [2023]. However, there are concerns about the impact of relying solely on ChatGPT for information retrieval. It may lead to a **decline in students' abilities to conduct rigorous literature searches, synthesize evidence, and attribute the work of others**, *Asraful Haque* [2023]. Additionally, the use of ChatGPT for generating content raises **issues of plagiarism and the erosion of students' accountability to learn**, *Abinaya & Gowthami* [2023]. [154]*Joshua Hardwick* [2023]-2 states in his recent study that Google can deal with an influx of bad content if it wants, so it doesn't care about AI content generated through tools like Chat GPT. So, taking to this reference, it is easily statable that **ChatGPT can be another Content Creator like billions of users, but can't be an alternative to Search Engine.** While some educators see ChatGPT as a tool comparable to search engines, it is important to critically evaluate the accuracy of the information generated and reflect on its use. And considering criticality, the possibility of **replacing a search engine with ChatGPT is currently unlikely.** While it can answer questions and provide information, it is not designed to replace traditional search engines like Google. ChatGPT's strength lies in generating knowledge in a conversational manner, but it cannot directly answer all types of questions and lacks the ability to connect to the internet for real-time information retrieval. Therefore, it is improbable that ChatGPT will replace search engines soon, [155]*Wojciech Urban* [2023],

[153] https://news.microsoft.com/the-new-Bing/?msockid=1ad76e1cf04364620f277bf7f1f7654d
[154] https://ahrefs.com/blog/google-ai-content-guidelines/
[155] https://www.boldare.com/blog/will-chatgpt-replace-google-search-engine/

[156]*Rosenberg* [2023], [157]*Tristan Greene* [2023], [158]*Varn Bristol* [2023], [159]*Rebecca Sentance* [2022] and [160]*Dustin Coates* [2023]. However as per *Kaushik & Jones* [2021] conversational search enable more effective or efficient engagement with web search engines, reduces the cognitive load on the searcher necessary to achieve their objective, while improving search effectiveness.

Seeing Chat GPT as another Search Engine Feature: So, what now if Chat GPT is not going to replace or it is improbable as per just above discussion that SEs can't be replaced by Chat GPT? **Well can ChatGPT be a fit as another SERP Feature?** Let's try to explore. ChatGPT is not just another search engine feature, but a powerful conversational tool that has the potential to automate various tasks and has been trained on a large dataset to generate coherent and human-like responses, *Sahin et al.* [2023]. Generative AI tools like ChatGPT are expected to transform the way people engage with online information, and companies like Microsoft and Google are incorporating chat and generative AI technology into their search systems, *Asraful Haque* [2023]. It can generate natural language responses and can be used for customer service interactions and creating engaging content for virtual agents, *Abinaya & Gowthami* [2023]. Here to note that while some speculate about its potential to impact search marketing, it is important to note that ChatGPT does not have internet access and cannot retrieve information from the web. Therefore, **it is not a substitute for traditional search engines like Google, that is near-most confirmed.** Instead, either it is more of an autonomous tool or a probable SERP Feature that can be used to assist in creating conversational content adjacent right to traditional SERP view, just like in Screenshot-5 below. This screenshot is of Bing Search Engine where upon querying, it is showing Chat-GPT with relevant search query conversation chat text in right side, in parallel of traditional SERP results to its left. Surprisingly, it has been allotted slot of famous "Knowledge Graphs" which was seen in same result at second place beneath ChatGPT result. Can it be taken from here that in time to come, it is going to stay here permanently? However, one may need to be logged in, before seeing this, as with general search it is optional as of the day. Not only this, rather it has found place in vertical search filters thumbnails strip, just see the second item named "COPILOT" beneath the search box in same screenshot. From here, this much has become feasible that yes, Chat-GPT can be adjusted as another SERP Feature and should be started calling so if no such understanding is yet set among Search Users; practically it is there. It is well suited here and there shouldn't be any harm in continuing like this; rather it enhances the purpose and beauty of Search Engine in this display, ultimately leading to more user satisfaction.

[156] https://www.linkedin.com/pulse/chatgpt-replace-search-engines-zack-rosenberg
[157] https://blogs.lse.ac.uk/impactofsocialsciences/2023/01/27/chatgpt-will-not-replace-google-search/
[158] https://bristolcreativeindustries.com/how-chat-gpt-will-impact-search-marketing-varns-view/
[159] https://econsultancy.com/chatgpt-search-engine-threat/
[160] https://www.algolia.com/blog/ai/why-chatgpt-wont-replace-search-engines-any-time-soon/



Screenshot-5: ChatGPT as another SERP Feature | Source: www.bing.com

Commonality in SERP and Chat GPT: ChatGPT and SERPs are interconnected because they both aim to provide users with relevant and useful information. As said just in above assessment that integrating conversational AI into search experiences can enhance user engagement, personalization, and the overall search process, bridging the gap between traditional search results and interactive, natural language conversations. Let's see how SERPs and ChatGPT are interconnected with many common or adjacent purposes:

1. **User Intent Understanding:** Both SERPs and conversational AI models aim to understand and serve the user's intent effectively. In the past, SERPs primarily focused on providing relevant web pages based on search queries. ChatGPT, on the other hand, seeks to engage in a natural language conversation and provide direct responses to user queries. As conversational AI advances, it can assist users in more complex and interactive ways, potentially reducing the need for users to navigate SERPs.
2. **Voice Search:** As voice search becomes more prevalent, both SERPs and conversational AI need to adapt. SERPs are increasingly designed to handle voice-activated queries by providing concise, spoken responses. Conversational AI models like ChatGPT are inherently designed for voice interactions, making them an important tool for voice search applications.
3. **SERP Features:** SERPs have evolved to include various features like knowledge panels, featured snippets, and People Also Ask (PAA) boxes to provide users with quick, direct answers to their queries. ChatGPT can be integrated into SERPs to enhance these features by providing more detailed, conversational responses and explanations.

4. **User Assistance:** While SERPs are excellent at providing information from websites, conversational AI models like ChatGPT can offer a more interactive and personalized form of assistance. For example, a user searching for a recipe can not only find a list of recipes on SERPs but also engage in a conversation with ChatGPT to get additional tips, ingredient substitutions, and cooking instructions.
5. **Search Engine Integration:** Search engines may integrate conversational AI models like ChatGPT into their platforms to offer a more holistic search experience. For instance, when a user searches for a complex topic, the search engine could provide both traditional search results and a ChatGPT-powered chat window for more in-depth discussions and explanations, just like Screenshot-5 above.
6. **Content Creation and Optimization:** Content creators and SEO professionals may use conversational AI models to generate content that performs well on SERPs. ChatGPT can assist in creating engaging and informative content that ranks high on search engines.
7. **Feedback Loop:** Conversational AI models like ChatGPT can continuously learn from user interactions, improving their ability to answer questions and help effectively. This feedback loop can inform search engine algorithms, making SERPs more accurate and responsive to user needs.

Collaboration of SERPs and Chat GPT: Til recent, there is no direct collaboration between traditional SERPs and ChatGPT is observed. Both are yet treating each other as distinct technologies serving different purposes. However, there are ways in which elements of these technologies could potentially complement each other or be used together. Here are some considerations:

1. **Integration in Applications:** Developers can integrate ChatGPT into applications, websites, or platforms that also display search results. For example, a website could utilize ChatGPT for a **conversational help desk** while still displaying traditional search results on the same page.
2. **Contextual Assistance:** ChatGPT can provide contextual assistance alongside search results. Users may **ask clarifying questions or seek additional information** related to the search results they receive.
3. **Enhanced User Interactions:** ChatGPT's natural language understanding capabilities can enhance user interactions with search engines. Users may express queries in a conversational manner, and ChatGPT can help refine and interpret those queries before passing them to a search engine.
4. **Customized User Experiences:** Websites or applications could offer a customized user experience by combining ChatGPT's conversational capabilities with search engine results. For instance, a user might ask ChatGPT for personalized recommendations based on search results.
5. **Task Automation:** ChatGPT can be employed to automate certain tasks or actions based on search queries. For example, users could ask ChatGPT to bookmark or save specific search results for later reference.

6. **User Guidance:** ChatGPT could guide users through the search process, helping them formulate better queries, providing explanations about search results, or suggesting relevant follow-up queries.
7. **Experimental Projects:** Developers and researchers may explore experimental projects that combine elements of both technologies to create novel user experiences. For instance, a platform could leverage ChatGPT to assist users in navigating and understanding search results.

Discussion: Above Technical Literature Review presents the trajectory of how GA-I embraces LLMs and Chat-GPT into a single platform of SERP, how they are interconnected, dependent and independent of each other. If to go on technical line only, distinction between relevant and irrelevant results, personalization of web results using LLMs, understanding patterns in existing data and creating new data for SERPs can be highly relevant topics. Also, as evident from above findings and interpretations, integrating ChatGPT with SEs could **enhance the user experience** by allowing users to interact conversationally with search results, seek clarification, or receive personalized guidance. ChatGPT can offer **contextual assistance** alongside to search results. ChatGPT can **automate certain tasks** or actions based on user queries, providing a more interactive and dynamic experience beyond what traditional SERPs offer. So, seeing above said dynamism of ChatGPT, it is likely that users would love to accommodate its display in the SERP as another SERP feature or menu tab and so is reflected in above LR as well. Above LR also established that while ChatGPT is not a direct replacement for traditional SERPs, still it can be seen as an innovative feature that brings a conversational and interactive SERP Feature to user interactions. Chart-33 below summarizes the facts and findings in above research.

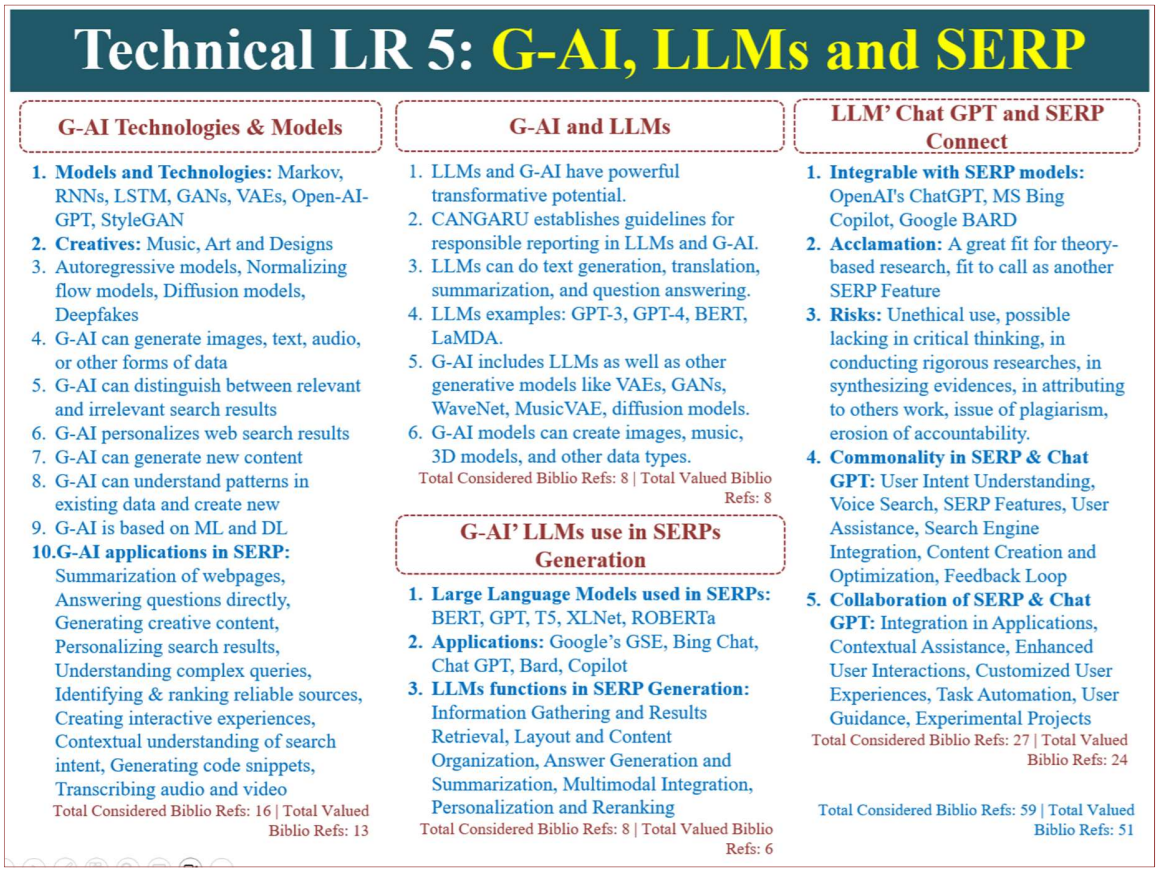


Chart-33: Chat-GPT and SERP

3.6. MetaSearch Engine and SERP:

Though, as such discussion of this topic can be debatable to include in this thesis, but to give a widespread coverage to the master entity of this thesis i.e. Search Engine and giving a hand on awareness about upcoming picture of Search; its conclusion gives the sense of completion of the subject. So, let's have a glimpse of how this thesis would be relevant to people who would be working in future on MetaSearch Engines and SERP and would require a connect between it and its previous generations.

Metaverse Virtual Indexing: The term "metaverse" was first used by sci-fi writer Neal Stephenson in his 1992 novel "Snow Crash." in which he describes a 3D virtual world that people could, in a sense, occupy, [161] *Wilson* [2023]. As metaverse is rapidly evolving into a new frontier for user interaction and content consumption, the search engines are poised to play a crucial role in navigating this immersive digital realm. The

[161] https://www.searchenginejournal.com/metaverse/484803/

connection between SERPs and Metaverse Technology is a complex and evolving one. In this regard, [162]*Kaj Leroy* [2023] at Forbes writes that the metaverse is a term that has been around for decades but has recently gained significant attention as tech companies like Facebook, Microsoft, and Roblox have invested in **shared virtual spaces** where people can interact with each other and digital objects in real time. He further adds that the metaverse is not just a single platform or game, but a **network of interconnected virtual environments where users can move freely, regardless of the device or platform they use**. Leroy arguments that though there is no direct connection between SERPs and the metaverse technology, as the metaverse continues to grow it is possible that **search engines will need to adapt to the new ways in which people interact with digital content in the metaverse**. For example, search engines may need to **index virtual objects** and environments in the metaverse or provide new ways for users to search for and discover content in virtual spaces. *Eno et al.* [2009] highlights the potential for a **cross-world search engine**, which could potentially integrate **virtual world content into traditional search results**. This concept is further explored by *Cheng et al.* [2022], who emphasizes the role of metaverse technologies in driving application scenarios, potentially influencing the content and presentation of SERPs. Work by *Yang et al.* [2023] on Metaverse Search system is a sophisticated presentation of how the Metaverse Searches are shaping up and how the Search Engines still going to be its integral part as shown in Chart-34, Appendix-9. See "Search Engine" mentioned in "Search Systems row" on both sides. Having, clue from, it appears imperative that Search Engine and SERP are going through the Metaverse as well.

Metaverse Virtual Environment: Adding further to flow, the use of **metasearch engines**, as discussed by *Vijaya & Chander* [2018] and *Liu Li* [2003], could also be relevant in this context, as they could potentially aggregate and present both **traditional and metaverse-related search results** in a unified manner. In the metaverse, which is a virtual reality space where users can interact with a computer-generated environment, the role of search engines becomes crucial. Metaverse search engines differ significantly from traditional search engines, as they incorporate spatial understanding and prioritize interactivity, [163]*thatware.co-1.* Thatware further states that unlike traditional search engines that rely on page rankings and keywords, **metaverse search engines provide search results based on a user's location within the virtual space**. As the metaverse promises to provide users with extensive, interconnected virtual ecosystems, the significance of metaverse search engines and their impact on search results pages is expected to grow, [164]*Kieron Allen* [2022]. The evolving nature of the metaverse and its influence on search and digital marketing present new challenges and

[162] https://www.forbes.com/councils/forbesbusinesscouncil/2023/04/25/the-metaverse-a-new-frontier-in-technology-and-interconnectivity/
[163] https://thatware.co/metaverse-seo/
[164] https://accelerationeconomy.com/metaverse/do-search-engines-have-a-place-in-the-metaverse/

opportunities for businesses and marketers, [165]*Lisa Buyer* [2021]. Therefore, the connection between SERPs and metaverse technology is an area of increasing importance and is likely to shape the future of search and information discovery. However, the specific impact of metaverse technology on the content and presentation of SERPs remains an area for further research. While traditional SERPs display primarily text-based results, the **metaverse SERP display** requires a more sophisticated approach that caters to the unique characteristics of virtual environments, these characteristics can be listed as below:

1. **Spatial Understanding and Location-Based Search:** Unlike the **flat and static nature of the internet, the metaverse is a three-dimensional space** where users can move around and interact with virtual objects. This presents an opportunity for search engines to **incorporate spatial understanding**, enabling users to search for content based on their location within the metaverse. For instance, a user could search for "the nearest virtual museum" or "a coffee shop with a breathtaking view."
2. **Interactivity and Voice Search:** The metaverse demands a more interactive search experience than traditional text-based queries. Users should be able to physically explore virtual spaces while interacting with objects and engaging in real-time conversations. Search engines can facilitate this by incorporating interactive elements into search results, such as **virtual buttons or links,** allowing users to directly access relevant content without leaving the virtual environment. **Voice search** will also become increasingly important, enabling users to navigate and search with natural language while immersed in the metaverse.
3. **AI and ML for Personalized Recommendations:** The metaverse is a vast and ever-growing space, making it essential for search engines to provide personalized recommendations based on user preferences and behaviour. AI and machine learning algorithms can analyse user activity, interests, and interactions within the metaverse to dynamically tailor search results. This ensures that users receive relevant and engaging content tailored to their specific needs and interests.
4. **Interoperability and Cross-World Search:** As the metaverse continues to develop, it's likely that users will seamlessly transition between different virtual worlds. To accommodate this, search engines will need to be interoperable, allowing users to search for content across multiple metaverse platforms. This will create a unified virtual search experience, making it easier for users to discover and access information regardless of the virtual space they're currently in.

Integration of Virtual Assets of Metaverse into SERP: The connection between SERP and Metaverse Technology lays in the evolving nature of online IR, user interactions, and the integration of virtual experiences. It should be understood as below:

[165] https://www.searchenginejournal.com/marketing-in-the-metaverse/416813/

1. **Information Retrieval in the Metaverse:** Traditional search engines may evolve to index and retrieve information from these virtual spaces. Search engines might need to adapt their algorithms to understand and rank content within the Metaverse, considering factors like **virtual location, user interactions, and the dynamic nature of virtual worlds.**
2. **Augmented Reality (AR) Search:** As the Metaverse incorporates AR, users might use visual search features to find information within their augmented reality environment. This could include searching for virtual objects, places, or information overlaid on the real world.
3. **Integration of Virtual Assets:** Search results may include virtual assets or items from the Metaverse. For example, users searching for a specific item may see not only web links but also **virtual marketplace listings** or 3D models that can be interacted with in the Metaverse.
4. **User Profiles and Personalization:** Metaverse technology could provide more data points about user preferences and behaviour. Search engines may utilize this information to tailor SERPs more effectively, delivering personalized virtual experiences based on users' Metaverse activities.
5. **Semantic Search in Virtual Environments:** With the integration of NLP and semantic search capabilities, users might interact with the Metaverse using conversational queries. Search engines could provide more context-aware results, understanding the nuances of virtual conversations.
6. **Digital Identity and Reputation Management:** As users create digital personas and identities within the Metaverse, search engines may play a role in indexing and displaying relevant information about these digital entities. Reputation management within virtual spaces could become a crucial aspect.
7. **Cross-Platform Search:** The Metaverse spans multiple platforms and applications. Search engines might evolve to provide cross-platform search capabilities, allowing users to find information seamlessly across different virtual environments and the traditional web.
8. **Dynamic and Real-Time Updates:** The dynamic nature of the Metaverse requires search engines to provide real-time updates and reflections of changes within virtual spaces. This could include updates on virtual events, changes in virtual landscapes, or real-time interactions among users.

Discussion: While talking about the survival of search and search marketing, [166]*Warner* [2022] argues that search is fundamentally about facilitating data retrieval. It's about helping people find the right information at the right time *(okay and selling them stuff during the process).* That existed before the internet, and it will continue to exist as the internet evolves. That is why the integration of SERPs with Metaverse technology involves adapting search algorithms to the unique aspects of virtual environments, providing users with relevant and personalized information in the context of their Metaverse experiences. Above review shows that SEs are

[166] https://www.click.co.uk/insights/search-marketing-and-the-metaverse/

undergoing a transformation paradigm to adapt to the evolving metaverse landscape. By incorporating spatial understanding, interactivity, AI-powered personalization, and interoperability, SEs can provide a seamless and intuitive experience for users navigating this immersive virtual world. As the metaverse matures, SEs will play a pivotal role in connecting users with the information, experiences, and communities that matter to them. Chart-35 below summarizes the facts and findings in above research.

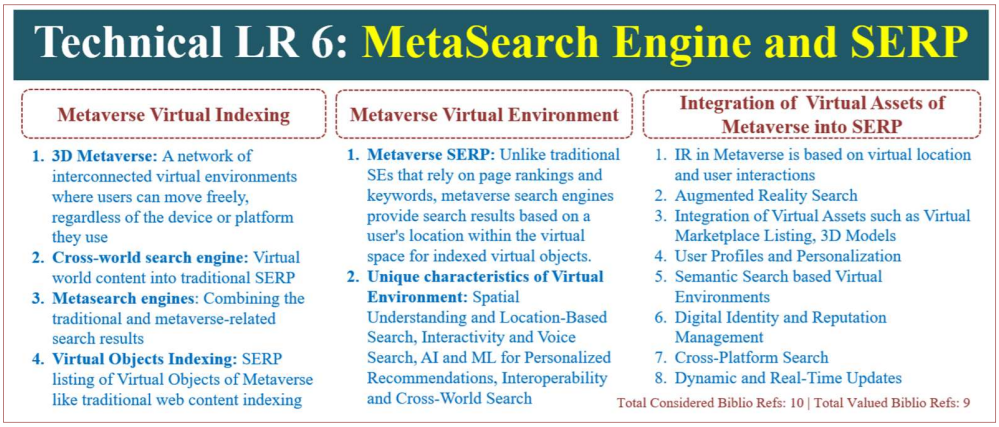


Chart-35: MetaSearch Engine and SERP

Overall Chapter Discussion: Based on the literature, research and facts accumulated in this chapter; a grand picture of the technical concepts sitting in the core of this thesis can be visualized in Diagram-10 below:

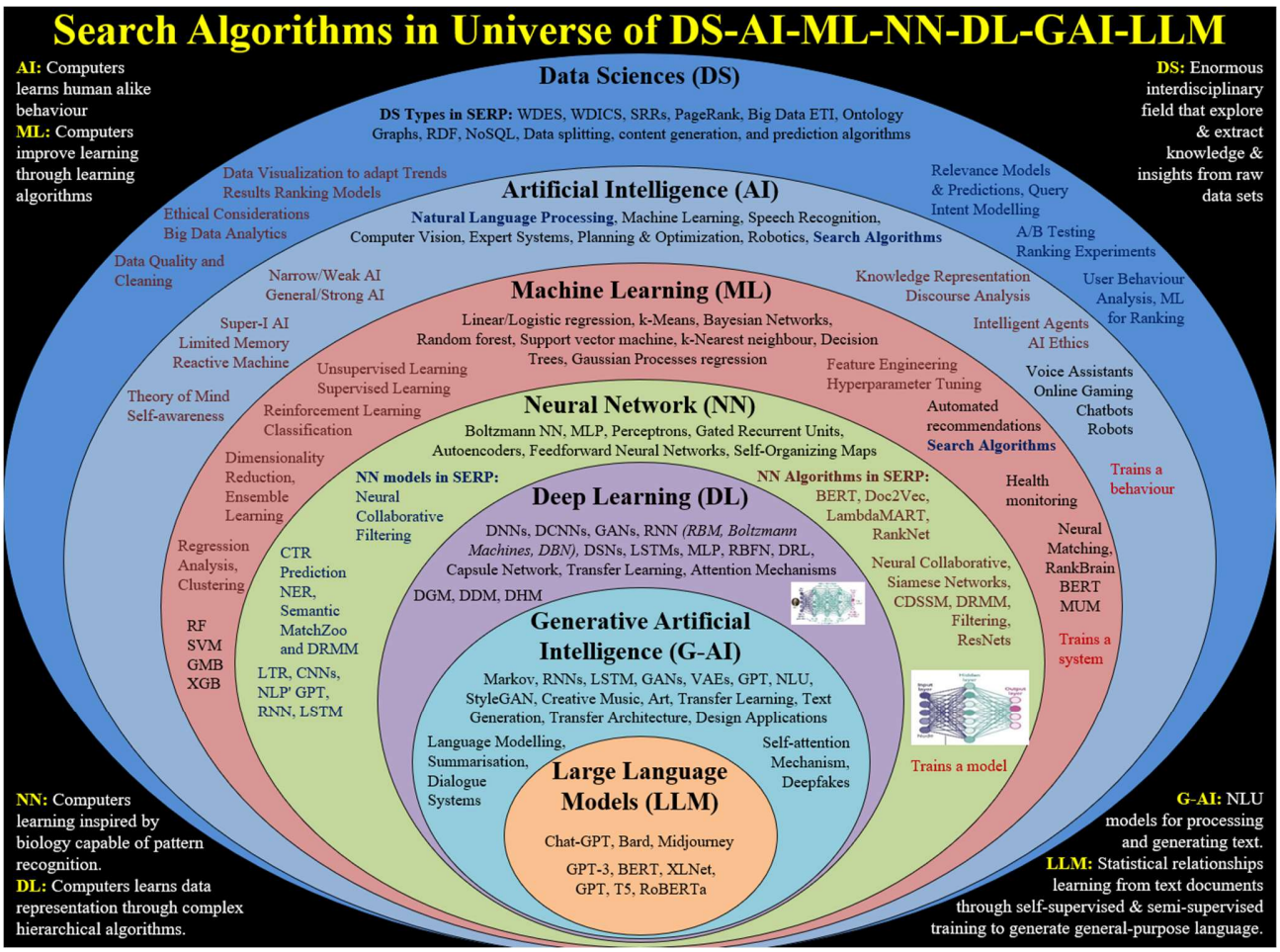


Diagram-10: Search Algorithms in Universe of DS-AI-ML-NN-DL-GAI-LLM

However, one may talk a ton on DS, AI, ML, NN, DL, G-AI, LLMs and Metaverse in context of SERPs in varied dimensions and aspects up to imaginations. Here, the purpose of knowing about these terms and their technical prodigy in the field of SERP and all together provide a required profound foreground to the thematic of this thesis. Going through this should put its reader in a thinking and decisive position to take its next route. This acquiring of the psychological and technical knowledge and awareness enthuses the confidence and motivation for taking this review to the real playground now, which is played by the three stakeholders of the SERP, which are the Search, the Searcher and the Search Producers i.e. Search Engine itself and the Content Producers. Let's proceed to next chapter with these three stakeholders.

Chapter-4

Web Search Dynamics

4.1. Web Search User at Mind SERP Model of Google:

Chart-36 below depicts a process model of how the indexed content is processed between the moments of asking a search query from search bar of a search engine and getting the search results in the form of a unique SERP every time.

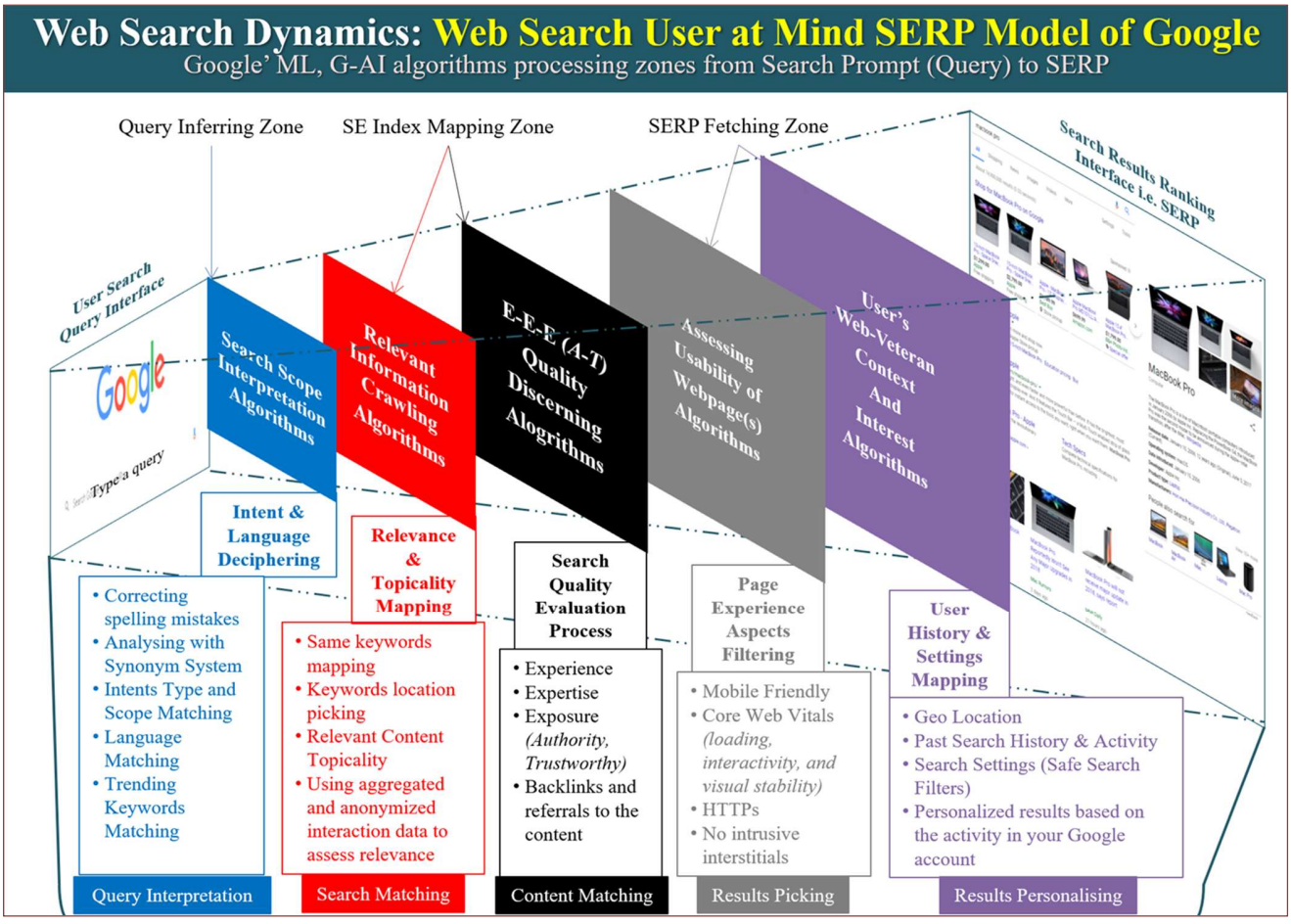


Chart-36: Web Search User at Mind SERP Model of Google; **Source:** [167]*Google.com-1;* [168]*Google.com-2,* [169]*Google.com-3;* [170]*Google.com-4;* [171]*Google.com-5;* [172]*Google.com-6,* Chart-37, Chart-38

[167] https://www.google.com/search/howsearchworks/
[168] https://www.google.com/search/howsearchworks/our-approach/
[169] https://www.google.com/search/howsearchworks/how-search-works/organizing-information/
[170] https://www.google.com/search/howsearchworks/how-search-works/ranking-results/

Having skimmed through to the history, psychology, and technical literature related to the theme of this thesis; and having the variables and sub-variables defined; it is time to move towards the shaping up the things towards the research question of this thesis. It is the moment to justify the hypothesis grounds conceived in the thought of this thesis. The mechanism of Online Search revolves around a **Search User**, **Search** i.e. the SERP Interface and the **Search Producer** i.e. the Search Engines and the Content Creators. Beside the sources mentioned under Chart-36 *(just above),* the other major sources used in defining Chart-36 are as shown in Chart-37 and Chart-38 in Appendix-9.

However, every Search Engine differs from the other on scale of selective parameters but Google being the biggest worldwide, we can call it as a representation of majority of them if not all. And Chart-36 makes a better fit on this aspect, which shows a standard search request process flow of "**Five Algorithmic Zones**". This flow starts from intake of a 'Search Term' *(or 'Keyword' or 'Search Query' or 'Search String' or 'Keywords String' or 'Search Phrase' or 'Search Question')* and ends with generation of the 'SERP' i.e. **Search Engine Results Page**. Our aim of this chapter shall be to articulate our literature review for these five algorithmic zones *(each phase has a definite set of different algorithms)* adjustable into following three processing zones:

- **Query Inferring Zone:** Search Scope Interpretation Algorithms - Intent & Language Deciphering - Query Interpretation
- **SE Index Mapping Zone:**
  - Relevant Information Crawling Algorithms - Relevance & Topicality Mapping - Search Matching
  - E.E.E. *(A.T.)* Quality Discerning Algorithms - Search Quality Evaluation Process - Content Matching
- **SERP Fetching Zone:**
  - Assessing Usability of Webpage(s) Algorithms - Page Experience Aspects Filtering - Results Picking
  - User's Web-veteran Context and Interest Algorithms - User History and Settings Mapping - Results Personalising

Based on above picturization of technical aspects and steps behind generating a SERP, its functional steps and stages can be fragmented as per the Table-11 below:

Three Processing Stages of a SERP Making

[171] https://developers.google.com/search/docs/advanced/experience/page-experience#signals
[172] https://www.google.com/preferences

Web-products classification	Search Query Processing	SERP Fetching
Listing > Crawling > Parsing > Indexing > Rendering > Topical Grouping > 3E Qualitative Classification	Query Inferring > Refining > Intent interpretation > Matching Query and Index > Relevancy Scoring > 3E Quality Assessing	Results Ranking > SERP Fetching > Search-As-Learning (SAL) Loop Run

Table-11: Three Processing Stages of a SERP Making

In the second part of this chapter, we shall try to understand what proportion of this journey is made up of or say backed-up by G-AI, and how ML is descriptive as a sub-set of G-AI in context of this thesis, in reference and coherency with Diagram-10 of *Chapter-3*. While progressing in this direction, a critical angle shall also be put in the development of this chapter, keeping Foreground Literature Review ***(FLR)*** done in Chapter-2 and Chapter-3 in purview; so that it could then make our Synthesized Literature Review ***(SLR)*** easier in Chapter-5. So this chapter can be seen as a bridge between its previous two and following one.

Classification of indexed web-products: All the five zones shown in Chart-36 show a rigorous focus of Google on their search users. As per *Varma & Jain* [2010] a search engine delivers search results by considering the relevance of the results to the user's search query and their historical clicking data. Their Google Patent Model Chart-39, Appendix-9 depicts the search response model.

The search engine generates a plurality of results based on the search query and ranks them according to their **original relevancy score**, *Zhang Jiayin* [2013]. It then calculates a **click relevancy score** for each result based on the historical clicking data and re-ranks the results accordingly, *Seshadri et al.* [2006]. Additionally, the search engine can **associate categories with documents** based on their relevance to a base category and deliver search results that indicate the associated categories, *El-Yaniv & Friedman* [2008]. The search engine also enables the user to receive information directly from a content provider without infringing on the provider's rights, handling any limitations or payments between the user and the content provider, *Nadav Gur* [2012]. Furthermore, the search engine tracks parameters associated with the **user's implicit response**, such as delay, speed, volume, tone, and movement, to improve the perception of the user's preferences and modify search results accordingly. Search engines are constantly evolving to meet the needs of their users, who typically expect instant, relevant results and this is achieved through a variety of methods, including the **integration of results from different collections** and the use of **user intent to guide the delivery** of results, *Lewandowski* [2008]. Interestingly, the computational problem faced by human memory and search engines is similar, with

PageRank, a component of Google's search engine, outperforming other predictors of human performance on a fluency task, *Griffiths et al.* [2007]. However, users' visual attention to query terms on SERPs varies, suggesting a need for further understanding of their behaviours, *Dickerhoof et al.* [2014]. To address this, SEs are employing strategies such as **query expansion, user search history, and re-ranking** to provide relevant results, *Inthiran et al.* [2010]. The **analogy of the mind as a search engine** is often used metaphorically to explain how the mind filters information based on what an individual is searching for, [173]*Axios-Steemit* [2018]. However, Axios declares very next that this concept is not a literal representation of how search engines deliver results in response to a search query. Here's a breakdown of how the search user is considered by the search engine to meet the search query expectations, alongside Search Engine' role of a guide as well:

1. **Query Understanding using NLP:** SEs utilize NLP to understand the context and semantics of user queries. This involves analysing the structure and meaning of the words in the query.
2. **User Intent Recognition using Intent Modelling:** SEs employ complex algorithms to model user intent. This involves predicting what the user is looking for based on the query, considering factors like context, language nuances, and search history.
3. **Contextual Relevance for Personalization:** SEs personalize results based on the user's search history, preferences, and behaviour by considering the individual user's context.
4. **Semantic Search by Understanding Concepts:** SEs use semantic search to understand the relationships between words and concepts. This enables them to deliver results that are conceptually related to the user's query, even if specific terms are not used.
5. **User Behaviour Analysis:** SEs analyse user behaviour, including CTR, to assess the relevance of search results. High CTR on a result suggests user satisfaction, while low CTR may indicate a mismatch.
6. **Diverse Content Types using Multimedia Integration:** SEs include various content types such as text, images, videos, and featured snippets in response to diverse user preferences and expectations.
7. **Featured Snippets and Direct Answers for Quick Information Delivery:** Featured snippets aim to provide direct answers to user queries at the top of the SERP, recognizing the user's desire for quick and concise information.
8. **Location-Based Results for Local Relevance:** For location-based queries, SEs consider the user's geographical location to deliver results that are relevant to their local context.
9. **Mobile-Friendly Presentation with Responsive Design:** Recognizing the prevalence of mobile devices, SEs ensure that results are presented in a mobile-friendly format, considering factors like page load speed and adaptability to different screen sizes.

[173] https://steemit.com/psychology/@axios/how-to-use-the-google-search-engine-of-your-mind-to-find-what-you-want-most-in-life

10. **Continuous Improvement with Algorithm Updates:** SEs regularly update their algorithms to improve the relevance and quality of search results, considering user feedback and emerging search patterns.
11. **User Feedback Mechanisms through Feedback Loops:** SEs incorporate user feedback mechanisms to understand user satisfaction and refine their algorithms. Feedback on the quality of results helps in continuous improvement.
12. **User Privacy Considerations for Privacy Controls:** Recognizing the importance of user privacy, SEs offer settings and controls that allow users to manage their privacy preferences and data sharing.

Anatomy of a Search Engine: The Chart-36 understanding is inherited from the understanding of the Anatomy of an SE. Here below are three of such selective depiction reference models which supports the Web Search User at Mind SERP Model of Google of this thesis. However, my understanding goes beyond these references as an extension with add-up of other missing components in these models. Chart-40, Appendix-9 is a work of [174]*Kaviani Siavosh* [2019] which presents a five steps standard pattern of search engine working as follows:

1. **Crawling:** It allocates the new web content from Search Console Listings.
2. **Indexing:** Getting the crawled webpages parsed and classified.
3. **Processing:** Engine compares the query string with indexed pages.
4. **Calculation of relevance:** Calculation of each page relevance in index to search string.
5. **Ranking:** Displaying the best results in SERP

All of these steps are talked in detail in Chart-44. Another Search Engine working model is presented by [175]*Scriptol.com* [2010] as shown in Chart-41, Appendix-9. Although the components shown in Chart-41 have a wonderful explanation of the Anatomy of a Search Engine, but it misses on "User Feedback" Part. Another scientific research-based model is presented by *Gupta & Sharma* [2014] as in Chart-42, Appendix-9. However, this chart also needs to be updated with latest updates due to new algorithms and G-AI evolution. As per [176]*Dave Davies-2* [2020] of Search Engine Journal about the functioning of Searches through algorithms, "A search algorithm is a massive collection of other algorithms, each with its own purpose and task.". Search Engines use "entity" to recognize the various content. Each page of a website or blog or any other digital product is considered as an entity and that website is taken as "a family of dependent entities". Context and

[174] https://avengering.com/en/how-the-search-engines-works/
[175] https://www.scriptol.com/web/google-anatomy.php
[176] https://www.searchenginejournal.com/search-engines/algorithms/

Entities in combine forms the major chunk of the different types of algorithms behind a SERP Algorithm. A detailed update on "Anatomy" is in the next chapter on Literature Review Synthesis.

Search Engines Comparative and Competitive Analysis: Jansen & Spink [2006] conducted research on four parameters of session length, query length, query complexity, and content viewed among the 9 web search engines of USA and Europe. Their results shows (1) users are viewing fewer result pages, (2) searchers on US-based web search engines use more query operators than searchers on European-based search engines, (3) there are statistically significant differences in the use of Boolean operators and result pages viewed, and (4) one cannot necessarily apply results from studies of one particular web search engine to another web search engine. The widespread use of web search engines, employment of simple queries, and decreased viewing of result pages may have resulted from algorithmic enhancements by web search engine companies. As *Acharya* et al. [2008] and *Garoufallou* [2012] users preferred mainly international search engines rather than Greek search engines. In a comparison study for better response time done on Google, Yahoo, AlltheWeb, Gigablast, Zworks and AltaVista and Bing/MSN by *Edosomwan* [2010] Google appears to be the best search engine in terms of both average precision (70%) and average response time (2s). Gigablast and AlltheWeb performed the worst overall in this study. More of comparative, and competitive analysis of search engines involve evaluating their features, performance, and market positions in relation to each other. Let's explore these perspectives:

1. **Comparative Analysis:**

Aspect	Google	Bing
Strengths	**Market Dominance:** Google is the most widely used search engine globally, with a vast market share. **Advanced Algorithms:** Utilizes sophisticated algorithms, including **BERT**, for improved understanding of user intent. **Rich SERP Features:** Offers a variety of rich features on the SERP, such as featured snippets, knowledge graphs, and image carousels.	**Visual Appeal:** Bing is known for its visually appealing homepage images and daily changing backgrounds. **Integration with Microsoft Services:** Seamlessly integrates with Microsoft products and services. **Video Search:** Offers a unique video homepage and provides video previews directly on the SERP.

Weakness	**Privacy Concerns:** Google has faced criticism for its data collection practices, raising privacy concerns. **Algorithm Updates:** Frequent algorithm updates can impact website rankings, leading to volatility.	**Smaller Market Share:** Bing has a smaller market share compared to Google. **Algorithmic Differences:** Differences in ranking algorithms may lead to variations in search results.
Searcher Interface	Clean and minimalist interface with a focus on simplicity.	Emphasizes visually rich backgrounds and multimedia content on the homepage.
Personalization	Utilizes extensive personalization based on user search history and preferences.	Offers some level of personalization but is perceived as less personalized than Google.
SERP Presentation	Presents search results in a straightforward list format.	Integrates visual elements, such as image and video previews, directly on the SERP.

Table-12: Comparative Analysis of Search Engines

2. Competitive Analysis:

Aspects	Google	Bing	Yahoo
Market Share	Dominates the search engine market with the highest share.	Holds a significant but smaller market share.	Uses Bing's search technology; its share is substantially smaller.
Revenue Model	Primarily relies on advertising revenue from Google Ads	Also generates revenue through advertising but with a smaller share of the market.	Like Bing, relies on advertising for revenue.
User Demographics	Widely used across demographics, from casual users to professionals.	Has been reported to have a slightly older user base and is popular in specific industries.	Attracts users through its portal services, including news and email.
Innovations and Features	Known for frequent innovations, integrating AI and ML into search	Innovates with visual elements and is integrated with Microsoft's suite of	Focuses on portal services, news, and email, with fewer innovations in search

	algorithms.	products.	technology.

Table-13: Competitive Analysis of Search Engines

Alongside above, *Reyes-Menendez et al.* [2018] presents an elaborative classification of type of SEs as shown in Table-14, Appendix-9.

Discussion: Google's Patent Model shown in Chart-39 mentions three calculations viz-a-viz **Original Relevancy Score, Predicted Number of Clicks, Click Relevance Score** and these come from the universal scaling of the content. If a web product like a web page passes the scores, then it gets place in SERP ranking otherwise not. Which means, along with quality and user satisfaction, the changing trends are the major factor with which Google decides about SERP. And Google's this stand is justified by other researchers quoted above. This LR shows that the user experience is central to this SERP generation process and search engines employ various techniques to interpret the user query, user perspective, user intent, current trends before delivering every SERP. Through advance technologies like NLP, ML, and personalization, search engines aim to create a user-centric experience that provides relevant, timely, and diverse information in response to search queries. The comparative, and competitive analysis provide insights into the strengths, weaknesses, and unique features of different search engines. While Google dominates the market, Bing and Yahoo cater to specific user preferences and offer alternate approaches to search and information retrieval. Each search engine has its strengths and weaknesses, and user choice often depends on personal preferences, demographics, and specific use cases. Chart-43 below summarizes the facts and findings in above research.

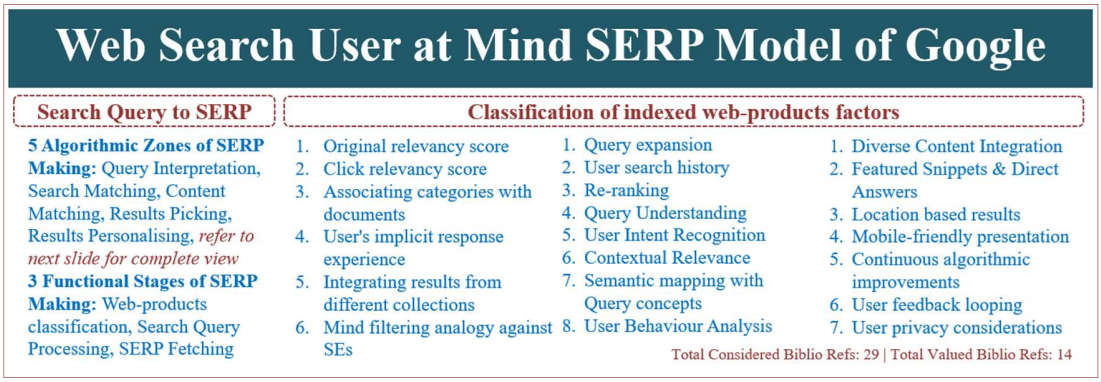


Chart-43: Search User at Mind Model of Google: A Synthesis

4.2. Three Stakeholders of SERP

For a descriptive understanding of SERP production, this whole process is attributed in three stakeholders. Search Engine and Content Creators, two of the stakeholders work as players of the process to facilitate the SERP as end-product to the third stakeholder i.e. the Search User. Chart-44 below articulates the whole SERP production process:

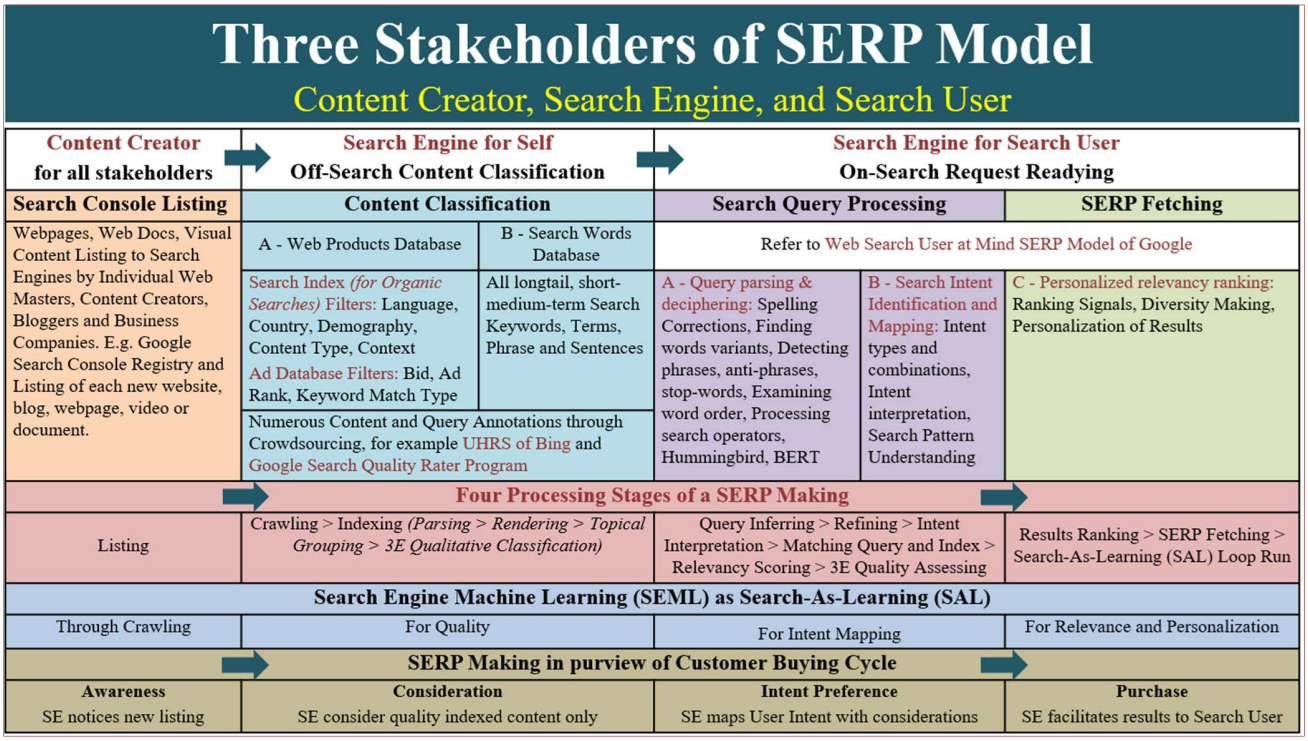


Chart-44: Three Stakeholders of SERP Model

Elaboration of Chart-44: The model emerges through Chart-44 can be taken as an all-in-one standardized sequential visualization of concepts covered in previous chapters, and sections developed in this chapter based on this model. It has four concepts displayed in four horizontal layers, which are associative to adjacent alongside individual existence. It is prepared in-tune with Facilitator vs. Player column of *'Table-9: Six variables of this thesis'*; its broad aperture is conceived from the *'Diagram-8: Psychological model of SERP Behaviour'* and its 'Refer to' cell besieges the complete 'Chart-36: *Web Search User at Mind SERP Model of Google*' in it. This way this model presents a profound picture of **Web Search Dynamics.** All the three stakeholders are put in sequential order of process steps flow of a SERP Making i.e. first the content creator, then the search engine and lastly the search user. '**Four Processing Stages of a SERP Making**' row depicts all the technical and algorithmic processing stages indulged sequentially in SERP Production. **SEML as SAL** happens in four stages, its wonderful depiction fits well to this model, see the second last item at bottom. And then the very interestingly this model in seen in purview of **Customer Buying Cycle**, in last bottom row.

Just like an honest sales representative comes prepare with 'Homework' to give satisfactory service delivery to the client; so does a Search Engine' in advance of receiving a search request. The 'Content Creator' and 'Search Engine for Self" section-cells of first horizontal layer in Chart-44 articulate the 'Homework' of a Search Engine. This first layer is the real '**Three Stakeholders of SERP Model**'; which we are going to talk in detail in left-to-right process-flow in this chapter here below. However and as the core catalyst of the question of this thesis is the search user, and Chart-36, Diagram-8 and Table-9 are very well sufficient in complementing all required dimensions of this core catalyst to call it a saturated coverage of the concept; still these three stakeholder depict the direct play-field of a search user; which is incomplete without understanding the chemistry of all players who put to work together for the search users in SERP production flow. So, it become imperative to grasp the full picture of the game to finally term it as robust research. Here we proceed with detailed description of Chart-44:

Web Content Creation by Content Creators for all stakeholders: Referring to Chat-45, in its first vertical top from left is the Content Creator, who lists their Content with Search Engines like Google via their Google Search Console Account. Search Console queues up all such new or updated entries / registries for Search Crawlers to pick and classify them for its second vertical in sequence. As the title of this topic suggests 'for all stakeholders', this listed content is usually created by the first stakeholder the content creator, indexed on the Search Engine the second stakeholder, for the use of the web search user, the third stakeholder; so for all stakeholders.

How the web content creators are categorized by search engines: The categorization of web content is essential for delivering accurate and meaningful search results. For this purpose, the web content creators are categorized by SEs based on various factors. One approach is to evaluate both the content and the content creator. This involves using content evaluation by evaluators who watch the created content and assessing the creator's skills and abilities. The **evaluated content creators** are then **matched with requested content creation** when a search is made, *Hurwitz et al.* [2015]. Another approach is to consider the **topic-authority of the original author**. This means that SEs not only consider the link-structure of the web graph but also consider the authority of the author in a particular topic. Algorithms are used to **transfer topic-authority between webpages**, which can improve the visibility of certain types of content, *Kyu & Jee* [2017]. Additionally, search engines may use a **hierarchical talent assessment** process to assess original content creators. *Lakshminarayana* [2009] proposed a categorization system based on page weights, which resulted in four different types of web pages i.e. (a) **Simple type** (b) **Axis shifted** (c) **Fluctuated** and (d) **Oscillating** types.

Lakshminarayana estimated two different page weights (a) Page Retaining Weight (**PRW**) and (b) Page Forwarding Weight (**PFW**) for a web page and grouped for categorization. The assessments of the creators and their content items are monitored to ensure objectivity in the search results for talented original content creators, *Ali & LaPaugh* [2013] and *Robinson & Myricks* [2011]. Search engines categorize web content creators based on the **relevance and quality of their content**. The goal of content creators and publishers is to produce content that will rank well in search results and attract traffic to their website, [177]*AIContentfy team* [2024]. SEs play a significant role in the discovery of new content, and content creators are now a days more focused on producing content that will rank well in search results, rather than content that's solely focused on providing value to the user. The categorization is based on the SE's assessment of the relevance and quality of the content to the user's query, which influences the content's ranking in search results, [178]*Britney Muller Moz.com*. Therefore, SEs categorize web content creators based on the **perceived relevance and quality of their content** to ensure that the most relevant and valuable information is presented to users. *Patel & Pandya* [2017] focused on categorizing web pages based on the **context of the content**, particularly in the context of child development and parenting. *Pu et al.* [2002] developed a query categorization approach to classify web query terms into broad subject categories, which was found to be efficient and adaptable to the dynamic web environment. *Lindemann & Littig* [2011] discussed the challenges of identifying relevant web sites and proposed a classification system at the super-genre level. These studies collectively highlight the importance of categorization in improving SE performance and user experience. SEs use various algorithms and methods to categorize web content and determine its relevance to user queries. Beside the universal approach of **Google' E-E-A-T *(as per this thesis 3E)*** below are some categories in which web content and content creators are categorized by SEs:

1. **Content Type:** SEs categorize web content based on its type, such as text, images, videos, or interactive elements. This classification helps the SE understand the diverse formats of content available on the web.
2. **Topicality** also known as Topic Relevance: Content creators are categorized based on the topics they cover. SEs analyse the textual content of web pages to determine the primary themes and topics addressed. This categorization aids in matching relevant content to user queries.
3. **Keywords and Phrases:** SEs analyse the keywords and phrases used in web content to identify the main themes. Content creators are categorized based on the relevance of their content to specific keywords, helping in search result ranking.

[177] https://aicontentfy.com/en/blog/importance-of-search-engines-in-content-distribution#
[178] https://moz.com/beginners-guide-to-seo/how-search-engines-operate

4. **Page Head Structure with Metadata Tags:** Metadata, including title tags, meta descriptions, and image alt attributes, provides additional context about the content. SEs use this information to categorize and understand the nature of the content.
5. **Page Body Structure with HTML Textual Markups:** HTML markup, such as header tags (H1, H2, etc.) and semantic elements, helps SEs understand the structure and hierarchy of content. Proper markup aids in categorizing content accurately.
6. **Multimedia Content HTML Attributes:** For images and videos, SEs analyse attributes such as image alt text, captions, and video transcripts to understand the content. This helps in categorizing multimedia content creators appropriately.
7. **Structured Data and Schema Markup for Featured Snippets Listing:** Web content that includes structured data and schema markup provides additional context about entities, events, relationships and the most important to seize a place in any of SERP features that can be generated using schema.org markup such as Rich Snippets, Knowledge Graphs, Breadcrumb Mark-up, Events, Reviews and Ratings, Rich Results like FAQs, How-To, Local Business, Video Carousels, Recipe Carousels, Job Posting etc.
8. **Mobile Friendliness:** With the increasing use of mobile devices, SEs categorize web content creators based on the mobile friendliness of their websites. Mobile-responsive design is considered a positive factor in search result ranking.
9. **Authority or Trustworthiness:** SEs assess the authority and credibility of web content creators based on the quantity and quality of backlinks pointing to their content. This influences the categorization of content creators as authoritative sources. Other parameters of measuring the trustworthiness are Page Authority *(Overall strength);* Social Signals such as Shares, Likes, Comments *(Content creators with higher engagement metrics may be categorized as more relevant or valuable);* User Engagement metrics such as CTR, Bounce Rate, Time spent on Page; Content Quality in terms of information and structure of the content; Relevance to search queries, Freshness, Multimedia Content, Mobile-friendliness, Site Speed, Domain Rank, Site Structure and Accessibility.
10. **Geographic Relevance:** Some search queries are location-specific, and SEs consider the geographic relevance of content. So, the web content creators and content may be categorized based on geographic focus or relevance to specific region.
11. **Freshness and Update Frequency:** Regularly updated content is often considered more relevant. Search engines may categorize content creators based on the frequency with which they update their content. The key elements through which Search Engines read the recency of web contents are Publication Date, Freshness (Content and Query oriented both), News and Trending Topics, Temporal Trends, Social Media Mentions, Crawl Frequency and Event-based queries.

Above categories can be seen in a tabulated form under three main categories as in Table-15 below:

Categorical classification of Content Creators and their Content		
Lexical	**Technical**	**Contextual**
1. Content Type 2. Topicality 3. Keyword and Phrases	1. Page Head Structure with Metadata Tags 2. Page Body Structure with HTML Textual Markups 3. Multimedia Content HTML Attributes 4. Structured Data and Schema Markup for Featured Snippets Listing 5. Mobile Friendliness	1. Authority or Trustworthiness 2. Geographic Relevance 3. Freshness and Update Frequency

Table-15: Categorical classification of Content Creators and their Content

Off-Search Content Classification by Search Engine for Self: This topic is in refence to the second vertical of the Chart-44. To deliver the best of SERP quality, SEs do rigorous homework back the scene quite often termed as **'Search-As-Learning'** *(we shall be reading this topic in depth in 4th section entitled 'Web Search Intelligence' of this chapter).* This topic has two sub-topics as evident from the Chart-44, the 'Web Products Database' and 'Search Words Database'. Let's discuss both as below:

A - Web Products Database: This database classifies the different types of listed web products such as websites, webpages, blogs, video, and documents. Each SE maintains its own such database with strict privacy and security measures. A SE keeps **initial results ready** to treat a search query at the first instance of receiving the query, *Romero et al.* [2019]. These initial results are generated based on the search query and provide a starting point for further processing. The search engine **identifies a topic associated** with the search query using the initial results and determines additional results that match the topic, *Landon Prieur* [2018]. These additional results are provided along with the initial results as search results for the query, *Lena Wiese* [2009]. This approach allows the search engine to provide relevant and comprehensive results to the user right from the beginning of their search, *Hood et al.* [2005]. Search engines maintain a variety of resources to handle search queries. *Cambazoglu et al.* [2012] proposed **cache-based query processing** as a solution for scenarios where a search engine cannot generate answers. This approach uses previously computed query results to create synthetic results for new queries. *Kucukyilmaz* [2021] evaluates **temporal variations in query** submissions and

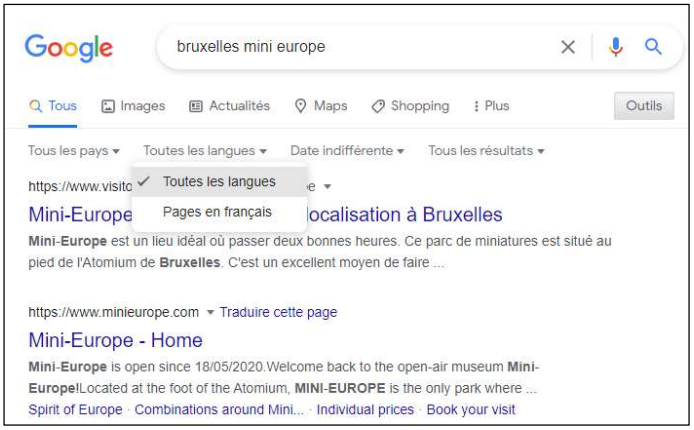
exploit them in order to improve the result caching performance. Conducted experiments demonstrate that the proposed caching framework improves the hit rate of a search engine result cache up to 3.31% and query response time up to 7.27% with respect to the state-of-the-art techniques. *Moore et al.* [2011] identified **three sequential positions of query repair** *(query composition moment, 'did you mean' query suggestions on SERP display and after SERP inspection user changes the query)* in search engine interactions, with users and search engines both able to initiate query repair. *Jain & Mishne* [2010] suggested organizing query completions based on **high-level user intent**, which can improve user satisfaction by reducing the effort needed to identify relevant suggestions. An SE keeps a massive index of web pages and processes search queries through crawling, indexing, and ranking to provide users with the most relevant results, [179]*Du Vall & Preston* [2023]. The SE under this vertical of Chart-44 typically executes following classifications of the content:

1. **Crawling (Spidering) and Indexing:** The search engine discovers and stores web pages through a process called crawling and indexing, [178]*Britney Muller Moz.com*. This allows the search engine to have a vast collection of web pages ready to be searched. The automated crawlers or spiders' programs systematically browse the web and discover new and updated content. Crawlers start by visiting known web pages and following links to other pages, ending by creating an index of the content they find., this process is termed as "crawling" or "spidering"; and these data-gathering programs used by search engines are called **spiders or crawlers.** As the crawlers move through web pages, they discover and collect information about the content on those pages. This information includes text, images, metadata, and links. Once a web product *(webpage or document)* is crawled then starts its **Parsing** also know as **Rendering,** which involves taking the HTML, JS and CSS information to generate how the page will appear to desktop and/or mobile users. The content discovered **(Content Extraction)** by crawlers after parsing is indexed, meaning it is organized and stored in a massive database. As per *Moran & Hunt* [2014], it turns out a lot of preparation lead up to that "magical moment" when the words are searched - organic search experts call that preparation **INDEXING.** Indexing involves analysing the content of web pages, identifying keywords, and creating a structured representation that facilitates quick retrieval. Indexing creates an organic search index, a special database that holds a list of all the words on all the pages on the web. **Bots** store collected information through Crawling and Rendering into the Index, tagging them with relevant search terms. **Indexing by any Search Engine is done in two folds.** In one-fold all the "**Search Words**" on which all Search Keywords and Search Terms are listed and in another fold all the "**Webpages**" are listed. As per [167]*Google.com-1*, the Google Search index contains hundreds of billions of webpages and is well over 100,000,000 gigabytes in size. It's like the index in the back of a book with an entry for every word seen on every webpage we index. When Content Creators list a webpage, they add it to the entries for all the words it contains. For the

[179] https://www.wwt.com/article/implementing-search-an-overview-to-building-a-search-engine

organic searches it is called as "**Search Index**" and for the paid ads it is called as "**Ad Database**". Basically, the back-up on the support of which the "**Language Decipher Model**" works is the "Indexing" itself but indexing happen at the stage of "**Machine Learning**". As said above Indexing happens in two folds, so in more sophisticated sense these two folds can be related to "**Relevant Information Crawling Algorithm**" and "**E.E.A.T. (3E) Quality Discerning Algorithm**" segments of Chart-36 above, which in simple words we can term as "**Search Matching**" and "**Content Matching**" respectively. Alongside the algorithms, the SEs get the web content classification assessed by Crowdsourcing as part of SEML *(read about it in detail in next section)* on Relevance, Quality, Freshness, Authority, Context, Scope, Uniqueness, Language, Market. In parallel to the Web Content Classification, relevant keywords and key phrases **(Keywords Extraction)** are also extracted from indexed content in separate Search Words Database. This helps SEs understand the topics covered by each page. All these exercises are done using "**Filters**". Filters are basically the search operations that rapidly remove the content that shouldn't be shown. These filters filter out the irrelevant "keyword matches" and "webpages" both. Here below is how the **common search filters** work in choosing matches of keywords as per their type:

a. **Language Filters:** These filters are put as per the language of the query and the types of languages preferred in the user location i.e. their IP Address or Language set in computer and other devices.



Screenshot-6: Google SERP with language filter

For countries like Japan and China results are filtered for their national language, while for a country like Belgium content is filtered for a minimum of three languages simultaneously i.e. Flemish, French and English. For example, if we search from Lille, France on Google Belgium, we get results on two languages as shown in Screenshot-6 above. First result is in French and the second is in English. If we do same search by sitting in Belgium, we can get third result in Flemish also. These languages are detected by reading the **Ads language set** *(for the case of paid results)* and by reading the **IP Address**

*(of the user or their computer **operating system language** set for the organic results);* otherwise, users tell the Search Engine on screen itself like the Screenshot-6.

b. **Country Filters:** This filter is mostly used by the Searchers in the Buying stage i.e. the Searchers with Transactional Intent. They don't want to fall in currency exchange rate complications, thus prefer local vendors over the others if available. So, the Search Engines filter the webpages on the basis of their domain extensions such as .fr for France, .uk for England and so on. For the case of Google, their Search Console Tool is the primary base of indexing sites and pages as per regional or global domain *(.com)* and that is why Search Console asks for your complete website name before creating you a console account for your website.

c. **Adult Content Filter:** Search Engines use this filter to filter out the extraneous egregious scatological sexual content such as pornography, child abuse; while keeping the explicit scientific and information-based content such as breast cancer, adult education material on body parts.

d. **Organic Search Filters:** We come here from the "Query Analysis" stage with a decided list of the keywords to fetch the results for; this list is matched with all relevant web documents as shown below in Chart-45.

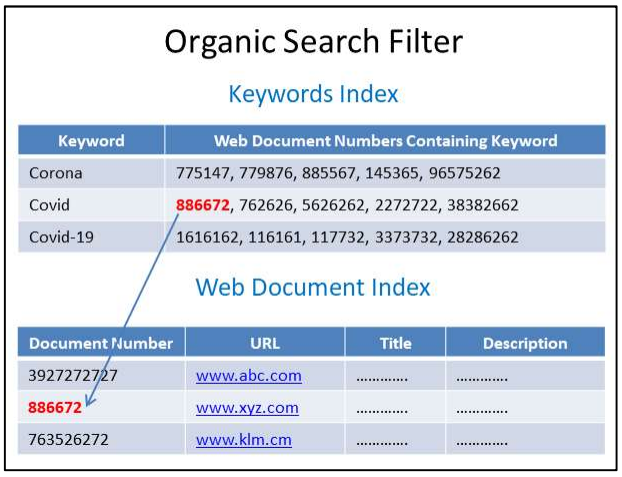


Chart-45: Organic Search Filter

All the keyword variants are listed in an numerical order as shown in above chart, having list of documents mentioned. They are matched to the index of web documents to fetch the relevant ones.

e. **Ads Search Filters:** These are operated usually by using tools like Google-Ads for Google Search Engine. Keywords Match Types like exact match, broad match, phrase match; and Bid methods are used in such filtering.

Now when we know that what the Search Engines keep ready in terms of Keywords and Webpages Indices at backend invisible to search user and visible to search user filters on interface of a search engine; let's see the next about what other preparations Search Engines keep ready to treat a search demand.

B - Search Words Database: SEs use various techniques to organize and store information about the words and meanings within search queries. *Zamir & Etzioni* [1998] have been in very first of those researchers who emphasized the need of **documents clustering** as an alternative method of organizing retrieval results. They proposed the clustering based on short snippets returned by web search engines and named this method as Suffix Tree Clustering ***(STC).*** *Kruse et al.* [2005] and *Osinski & Weiss* [2005] both highlight the use of algorithms to enhance search engines, with Kruse focusing on **WordNet**-based wrappers and Osinski & Weiss on the **Lingo algorithm** for **clustering search results**. *Gaurkhede & Dhotre* [2014] emphasizes the importance of organizing **user search histories** to improve search engine performance, while *Pradhan et al.* [2014] introduces a web search engine-based approach to **measure semantic similarity** between words. These studies collectively underscore the role of algorithms and user data in the organization and storage of information within search queries. Another method involves creating **semantic links between terms in an inverted index**, which forms a semantic network preserving the internal structure of stored documents and enabling structural queries, *Gerasimov et al.* [2014]. As per [180]*Wiggers* [2020] Google says it can now index individual passages from webpages, as opposed to whole pages. When this rolls out fully, Google claims it will improve roughly 7% of search queries across all languages. A complementary AI component will help Search capture the nuances of webpage content, ostensibly leading to a wider range of results for search queries. Another approach involves using a segmented structure in the search engine database, with segment managers updating term matrices to generate a published search matrix, *Sokolan et al.* [2010]. Furthermore, a storage structure and method involve storing guide word information in a database, allowing for easy maintenance and repair by the existing system manager, *Woo & Gu* [2007]. Search engines use semantic topic clustering to understand the context and meaning of search queries. This process involves organizing search queries into meaningful topics to provide more relevant search results. Google has developed techniques such as **word co-occurrence clustering and weighted bigraph clustering** to discover semantically meaningful topics in search queries, *Kong et al.* [2016]. According to *Carpineto et al.* [2009] Web clustering engines organize search results by topic. By showing the most likely meanings for any given request, search results clustering narrows the semantic mismatching between the user need behind the request and the list of results returned by a plain search engine. The technology of Web clustering engines has reached a level of maturity in which a rich body of research has been developed, and several commercial systems are being deployed. *Hearst & Pedersen* [1996]

[180] https://venturebeat.com/2020/10/15/google-details-how-its-using-ai-and-machine-learning-to-improve-search/

presented the Scatter/Gather approach to document clustering which could produce significant, improvements over similarity search ranking alone. They stated that along with the use of a very large text collection, the most important difference from other in their approach is that it produced clusters that are tailored to characteristics of the query, rather than assuming that clusters play a one-size-fits-all, classificational role. Topic clusters, also known as content clusters, are groups of webpages used to establish authority around a particular subject, and each cluster includes a pillar page that provides a broad overview of the topic and links to cluster pages covering associated subtopics, *Rachel Handley* [2023]. Semantic keyword clustering automatically organizes related keywords into clusters based on their meaning and context, which helps streamline keyword research and optimize content for search engines, [181]*Colton Miller* [quattr.com]. Thus, **semantic topic clustering** based on search terms helps search engines understand the relationships between different topics and provide more relevant and comprehensive search results. [31]*Olaf Kopp* [2022]-2 mentions following methods of Query Processing:

1. **Classification of the search term in a thematic ontology:** Improving semantic topic clustering for search queries with word co-occurrence and bipartite graph co-clustering.
2. **RankBrain is an entity-based search query processor:** Identification of acronym expansions, Query rewriting with entity detection, Methods, systems, and media for interpreting queries, Query generation using structural similarity between documents.
3. **Identification of entities in search queries** (Name Entity Recognition): Selecting content using entity properties, associating an entity with a search query, providing search results based on a compositional query.
4. Semantic enrichment of search queries
5. Clustering query refinements by inferred user intent.

When referring to the indexing phenomenon in the context of search words content, it involves following **iterations**:

1. **Indexing Search Queries Content:** This involves understanding the words used in search queries and associating them with relevant meanings and contexts.
2. **Word Analysis:** SEs use algorithms to analyse the words in search queries. This includes **recognizing synonyms, understanding variations, and considering the context** in which the words are used.
3. **Meaningful Semantic Indexing:** SEs create a meaningful index of search queries by categorizing and organizing the words based on their **semantic meanings**.

[181] https://www.quattr.com/optimize-content/why-focus-on-keyword-clusters

4. **Relevance to Indexed Content Algorithms:** SEs employ specific algorithms to determine the **relevance of search queries to the indexed web content**. This helps in providing users with accurate and contextually appropriate results.
5. **User Intent Deciphering:** Understanding and recognising user intent is a crucial aspect of indexing search words. SEs aim to decipher what users are looking for based on the language used in their queries and then match it with relevant indexed content.

Search Words Indexing Phenomenon: The indexing phenomenon related to search words content focuses on the intricate process of analysing and organizing the words within queries to enhance the accuracy and relevance of SERP. **It involves a technical processing of organizing and storing information about the words and meanings within search queries** around key components as below:

1. **Tokenization:** When a search query is received, the first step is to break it down into individual units called as **tokens.** Tokenization involves separating the query into meaningful **chunks,** which are usually individual words. This process helps in the subsequent analysis of the query.
2. **Stop-word Removal:** Common words like "the", "and", or "is" *(known as stop-words)* may not contribute much to the meaning of a search query. So, SEs often remove them to focus on the more relevant keywords. We talk about stop-words in next section below.
3. **Stemming and Lemmatization:** To ensure that variations of words are considered, stemming and lemmatization are applied. **Stemming** reduces words to their root form, while **lemmatization** involves transforming words to their base or dictionary form. This helps in capturing the core meaning of the words.
4. **Indexing:** The processed tokens are then indexed, creating a structured database of terms and their associated documents or web pages. Each term in the index is linked to the documents where it appears, allowing for efficient retrieval during search queries.
5. **Inverted Index:** SEs often use an inverted index data structure. In this structure, each term in the index points to a list of documents where the term occurs. This allows for fast retrieval of relevant documents when a user submits a search query.
6. **Term Frequency-Inverse Document Frequency (TF-IDF):** TF-IDF is a statistical measure used to evaluate the importance of a term within a document relative to its occurrence in a collection of documents. It helps in ranking the relevance of documents to a particular search query. *Fu & Yu* [2020] proposed a method of improving TF-IDF algorithm to propose a word location in text-based representation method to distinguish the words' texts. Their method found more effective for subsequent text categorization operations.

7. **Semantic Analysis NLP:** Advanced SEs employ NLP techniques for semantic analysis. This involves understanding the **context and meaning of words** in a query to provide more accurate and contextually relevant search results.
8. **Machine Learning and Ranking Algorithms:** SEs often use MLAs to continuously improve the relevance of search results. These algorithms **learn from user behaviour and feedback** to refine the ranking of documents based on their perceived relevance to specific queries.

Algorithms associated with Search Words Indexing Phenomenon: By employing above said technical processes, SEs efficiently organize and store information about the words and meanings within search queries, leading to improved accuracy and relevance. There is **relevant algorithm names associated with each of these processes** for organizing and storing information within search queries. Let's see which of the technical processes run independently and which are associated with algorithm(s) as below:

1. **Tokenization:** It is a foundational step in NLP; there is no algorithm as such is used at this stage.
2. **Stop-word Removal:** It is like a basic simple filtering step often implemented using predefined stop-words list tables with no dependency on algorithms.
3. **Stemming and Lemmatization:** Algorithms like **Porter Stemmer or Snowball Stemmer** are used for stemming. And Algorithms like **WordNet Lemmatizer or spaCy** are used for lemmatization.
4. **Indexing: Inverted Index Construction** technic is used at this stage with independent functioning requiring no algorithm.
5. **Inverted Index:** Inverted Index is a data structure rather than a specific algorithm, but algorithms like **TF-IDF** are commonly associated with it.
6. **TF-IDF:** TF-IDF is a specific algorithm used for evaluating the importance of a term within a document relative to its occurrence in a collection of documents.
7. **Semantic Analysis and NLP:** Algorithms for semantic analysis and NLP include **Word Embeddings** *(e.g., **Word2Vec, GloVe, ELMo**)*, Named Entity Recognition (**NER**), and syntactic parsing algorithms (e.g., **constituency parsing, dependency parsing**). For understanding context, algorithms like **BERT** are increasingly used.
8. **Machine Learning and Ranking Algorithms:** MLAs for ranking include various models such as **logistic regression, Support Vector Machines (SVM),** and more recently, deep learning models like **neural networks**. Ranking algorithms also include relevance models like Best Match 25 ***(BM25)*** and learning-to-rank algorithms such as **RankNet or LambdaMART**.

These algorithmic techniques collectively contribute to the effective deciphering, processing, and indexing information around search words within search queries to further store in typically designed databases for efficient management and information retrieval. The Inverted Indexes are commonly stored in **Inverted Index Databases** optimized for quick lookups and retrieval. These databases are designed to efficiently handle the mapping of terms to the documents or web pages where they appear, to further enable rapid retrieval of documents or content relevant to specific search queries. These databases organize information in a way that facilitates efficient matching of terms to documents, enhancing the speed and accuracy of search engine results.

On-Search Request Readying by Search Engine for Search User: A SERP is the result of a search query put by a user with some intent in mind. *Lung et al.* [2011] elaborate the method of SERP results sorting by a search engine like Google in their research. Chart-46, Appendix-9 describes it in a chart flow pattern. The sorting of results process includes:

a) removal of punctuations from the query
b) finding identical alias output in search engine database
c) if found then send the results as a string to search engine interface else fragment the query keywords further into low and high priority pattern in cyclic method until reaches a best possible SERP.

After having a fundamental of the SERP making overview in Chart-46, let's talk about the query processing first. Moving through the top horizontal layer of Chart-44, it is now about the third vertical i.e. Search Query Processing. This group of processing steps can be divided in two segments, the Query Parsing and Deciphering, and the Search Intent Identification and Mapping. The 'Query Interpretation', 'Search Matching' and 'Content Matching' segments of 'Chart-36: *Web Search User at Mind SERP Model of Google'* shall be talked in detail here. Let's investigate both segments one by one as below:

A - Search Query Parsing and Deciphering: This segment delves on the thought of 'Are our web searchers expert enough to retrieve what they seek single handedly?' Obvious answer shall be 'NO'. [182]*Google.com-SERP* gives a very crisp definition to this thought as, "*With the vast amount of information available, finding what you need would be nearly impossible without some help sorting through it.* ***Google's ranking systems*** *are designed to do just that help through hundreds of billions of webpages and other content in our Search index to present the most relevant, useful results in a fraction of a second*". Adding further to this complex challenge, Search Users can be named as the pesky searchers, the curious searchers, demanding searchers, dull searcher, least learned searcher and as many varieties as can be imagined. The search query variations around a single

[182] https://www.google.com/search/howsearchworks/how-search-works/ranking-results/

word make a satisfactory search task stiffer. Can anyone expect a single user to be this much enormously autonomous to deal with these challenges and stiffness? In 99% cases answer shall be 'No'. For example: A laptop, notebook, personal computer, pc; all means a single product at the end; but not necessarily that all shall give same results fitting to user's needs and expectations. Another example can be of Orange, which can be a Fruit, a Colour, a French Telecom Company. And certainly, to reach to the right product, these words need additional explanatory words, quite often called as **affiliated words,** for example for a computer affiliate word can be desktop, laptop, or pc. Then comes the **in-market words or accessory words** very helpful in recognizing the potentially linked products, for example of a computer again keyboard, CPU, motherboard, mouse etc. are in-market words. Then comes the **second-degree affiliation** *(semi-out market)* words like tablets, smartphones to laptop to a query of Computer. Then as next complexity comes the point of having understood on out-of-market search words known as **negative keywords**. Adding to the flow is the importance of understanding about complexities as per the **Web Search Stages**. [183]*AIDA-b2bframeworks* describes the "**Searching in Stages**" using C.P. Russel AIDA *(Attention, Interest, Desire, Action)* model, the purchase path starts with awareness about a brand, product, or service. Then evaluating them all available or accessible in searches. Then comes the third stage of making the preferences and finally do the purchasing. AIDA model states that a searcher usually starts with a broader perspective, and it refines with stages of the purchasing path. Similarly, third pattern of Search cycle can appropriately be set in the [184]*Lucidchart-1'* search decision making cycle as shown in Chart-47, Appendix-9. Searching information on internet also passes through these stages and at each stage the type and length of query is rectified to be more specific to reach to the final need goal. Seeing the Lucidchart and as said above, can a web searcher know all patterns of variations around a search? Rare Chances! So this much varied entities around a product name, domain ranges and search stages can either be handled by a search expert which are hardly 1% or with some help for the 99% web search users. And, to sort these complexities of a web searcher along with facilitating to them as a guide, SEs facilitates the Query Inferring and Deciphering. To handle these challenges, SE processes the search query to **infer it and understand the user's intent behind it**. This involves **parsing the query** and sending it to the index server for processing, [185]*Catherine Dee* [2022]. Google has developed query processing algorithms such as **Hummingbird and BERT** to better understand the context of the search queries, [186]*Anand Srinivasan* [2022]. With the help of query processing algorithms, the task of analysing and deciphering a search query begins. The keywords are interpreted on their meanings and synonyms first then analysed against Search Engine' keywords database. This process is principally based on **"Intent and Language Deciphering"** in "**Query**

[183] https://www.b2bframeworks.com/frameworks/aida
[184] https://www.lucidchart.com/blog/decision-making-process-steps
[185] https://www.algolia.com/blog/product/what-is-a-search-query-and-how-is-it-processed-by-a-search-engine/
[186] https://www.minttwist.com/blog/googles-query-processing-algorithms-and-what-they-mean-for-your-seo-strategy/

Interpretation" section of Chart-36. SE treats each word in a keyword string as separate "search term" and analyse it individually, combined as the whole search term, and thinks of its possible variants. The prime focus in this whole process is given to following variants:

a) **Correcting Spellings**: Search Engines do it by "**Correcting while typing or typed**", "**Auto-correction Prompts**", "**Suggestion while typing**" or "**Did you mean …...?**" prompts beneath the Query Typing Bar or after obtaining results in SERP result, wherever SE finds unclarity in understanding search user's intent and query meanings both. But SE cannot be all time right in such suggestions especially in case of popular brands. In such cases users must pay extra attention if their keyword was taken as per their intended type or not.

b) **Finding Words Variants:** SEs take the singular, plural, synonyms, or conjugation basically as the same or similar meaning. Lower- and Upper-Case typing is generally ignored by SEs.

c) **Detecting Phrases, Anti-phrases and Stop-Words:** As talked in the first variable of this thesis "Art of Phrasing" section, **Phrase** is the group of words that together makes sense about a fact or idea; so they are slightly different from single word searches. SEs work extra in first understanding Phrase searches then get the results accordingly. Here we shall be talking with reference to the work by *Moran & Hunt* [2014] - Page 65. In these phrases itself, the task for SEs becomes double fold difficult when **Anti-phrases** are also present. For example: 'What is the value of pie?' *(previously phrase needed to be enclosed in double quotation mark)*; in this phrase term 'what', 'is', 'the' and up-to some extent 'of' are of no help in getting to right list of results. SEs ignore all of these **Stop-words** and focus on just two 'value' and 'pie'. But if a phrase is like 'meaning of what' or a **brand name** like 'Who - World Health Organization' or 'The Who - a rock band group'; in these situations, 'what' and 'who' become the prime intent; about which the user either wants to know dictionary meaning and its grammatical importance or searching for a brand name. It becomes most typical for a Search Engine if your brand is not well known, and it carries words like 'who' or 'Where's Waldo'. To settle with these complications Search Engine like Google puts a huge focus on making their keywords analysis more conversational. And that is where and how the **Hummingbird** and later the **BERT** fix these issues up-to big extent. The extremely common grammatical articles, prepositions, conjunctions, helping-verbs etc. like 'a' or 'of' are called as **Stop-Words**. Whereas in old days search engines used to ignore them completely but with upgradation of search algorithms *(based on continued research)* they understand the importance of these stop-words, especially and as explained above if they were used directly in Brand Names.

d) **Examining Words Order:** *Moran & Hunt* [2014] - Page 65 assert in their book that the word orders often result in different meaning completely, for example "Joe Little" and "Little Joe". If you will search them by changing their orders, a complete different set of results will be displayed. This situation especially

addressable in cases of very long search phrases and exact name cases like Brands. For example: « Little Joe versus Joe Little; or enterprise management storage system can be searched as storage management enterprise system.

e) **Processing Search Operators:** In the search term "Big Brother -tv" means asking for an organization in same name instead of a TV show. "+the white house" means person is asking for the results with "the" stop-word including in it. We talk about this topic in further details under "*Search Expressions and Operators*" section of this chapter.

f) **Identifying Search Intent:** For your content to appear, you need to have the type of content that is being returned for your targeted keywords. Seeing this as a vast topic, we talk about this topic in further details under "Search Intent Identification and Mapping" just next below.

B - Search Intent Identification and Mapping: Understanding a search query is a critical aspect of search engine functionality, involving the interpretation of the user's input to provide relevant and accurate search results. As per [187]*Jayson DeMers* [2013] at the heart of every search on Google, there's some kind of intention. A user wants answers, resources, information, reviews of a product, and much more. [188]*Herradura-cornerstone-digital* states that Google's algorithm standards have become much more focused on what those searching for information really want. While defining the taxonomy of classic information retrieval *Broder* [2002] defines it as inherently predicated on users searching for information, the so called "information need". But the need behind a web search is often not informational, it might be navigational *(give me the url of the site I want to reach)* or transactional *(show me sites where I can perform a certain transaction, e.g. shop, download a file, or find a map)*. Further to this, Google' mission as per [167]*Google.com-1* is to "Organise the world's information and make it universally accessible and useful." And to achieve it Google always provides its searchers with the most relevant result for their query. Doing this is not so easy job, as it seems in saying. To achieve this perfection, it takes skimming of lot of search-pattern-loops and their analysis on search intentions. Before entering the in-depth explanations and understanding of intents, here are the prima-facie **methods of interpreting a query intent**(s):

1. **Query Structure Analysis:** It involves the understanding whether the query is short-tail or long-tail, indicating specificity. Short-tail queries may be more general, while long-tail queries often reflect a more detailed or specific information need.
2. **User Intent Inferring through Prediction and Classification:** This is about predicting the purpose of a query the user wants to achieve through a query. By analysing patterns and historical data, algorithms

[187] https://blog.hubspot.com/marketing/importance-of-user-intent-in-seo
[188] https://cornerstone-digital.com.au/the-psychology-of-search-understanding-user-intent-for-better-seo/

predict the probable intent of the user. After predicting the intent, the adjacent next task is of categorizing queries into different types based on user objectives. Algorithms classify queries into predefined categories to better match them with suitable result types.

3. **Search Ambiguity Resolution:** Every second query is dealt with this issue. To resolve this addressing ambiguous terms or phrases in the query is used. Algorithms consider surrounding context, user history, and common interpretations to resolve ambiguities and refine search results.
4. **Query' Temporal Context:** A sizeable number of queries are time sensitive such as queries related to news and events. These time-related elements in a query are identified and processed to understand if the query is related to current events, historical information, or time-sensitive topics.
5. **User Feedback Incorporation:** Here the past user behaviour in terms of clicks, historical searches and search trends of a user profile are incorporated in assessing the probable search intent(s). Continuous learning from user clicks, dwell time, and explicit feedback not only helps to answer the demanded query but also refine algorithms and improve the accuracy for future search results.
6. **Search query intent classification:** Query intent classification involves mapping queries into categories based on their navigational, informational, or transactional intent, as well as other factors like geographic locality or topical categories, *Sia et al.* [2012]. There are techniques for providing geographically related search results in map interfaces that consider the intent behind the user's query and the abstract entities it maps to, *Ghanekar et al.* [2010].
7. **Query intent mining:** Query intent mining aims to identify subtopics covered by the queries, such as explicit subqueries or implicit representations like document clusters or entities, *Dou & Guo* [2020].
8. **Search results diversification:** Search result diversification is another approach that aims to offer diverse documents covering multiple intents, using techniques like measuring the similarity of documents based on their intent coverage and constructing an intent graph, *Su et al.* [2021].

Above methods give a glance of the mechanism of identifying search intent(s) behind search queries. And to understand this mechanism, we should know of the types of Intents at first instance. So, let's understand the web search intent types next as below.

Types of Web Search Intents: Chart-48, Appendix-9 from [189]*Jovicic* [2023] below at SEMrush and [190]*Neil Patel-1* presents four types of search intents with example terms and corresponding SERP Features from where

[189] https://www.semrush.com/blog/keyword-intent/
[190] http://neilpatel.com/blog/i-lost-31-of-my-search-traffic-but-my-sales-went-up-by-73/

these intents can be identified on SERP screen. [191]*Edwin Toonen* [2021] also writes about these four search intents. Although, I am not in agreement with this classification, because I see **Entertainment / Watching** and **Mapigation / Destination** as fifth and sixth major category. Let's first review more of the existing literature to crosscheck if any other researcher thinks like me on this subject. Then, I shall justify why these two categories deserve independent recognition as web search intent type. Search intent specificity is more crucial for the case of Commercial Intent and Transactional Intent. Chart-49, Appendix-9 presents a pairing of search intents with stages of Product Purchase Cycle. Research on search intent types and the purchase cycle has identified several key factors. *Nysveen & Pedersen* [2005] found that the degree of goal-oriented search mode is influenced by product risk, product involvement, and internet experience, which in turn affects purchase intention. Adding to the flow, *Shim et al.* [2001] presents a theoretical model for predicting online purchase intentions, see Chart-50, Appendix-9. *Kim & Lee* [2004] identified utilitarian and hedonic values of internet search, perceived benefits and risks of online shopping, and internet purchase experience as key predictors of online search intention, which in turn influences purchase intention, refer to Chart-51, Appendix-9. *Guo & Agichtein* [2010] introduced a model that uses user interactions with search results to distinguish between **"research" and "purchase" variants of commercial intent,** which can be useful for improving search effectiveness. Guo and Agichtein further say that fine-grained user interactions with search results, such as mouse movements and scrolling, can provide insights into search intent and improve the effectiveness of web search.

Comparative study of four Intents: The Chart-52, Appendix-9 shows percentage share of four types of search intents against Number of Keywords; here the share of Informational Intent leads with highest share, followed by Transactional, Commercial and Navigational. Above findings are asserted in, *Truran et. al* [2011] research as well. His findings suggest that search engine results for informational queries are significantly more stable than the results obtained using transactional, navigational, or commercial queries. Research by *Rose & Levinson* [2004] presents a comparative classification of queries searched by intent type and sub-type. Chart-53, Appendix-9 findings figured out that Navigational searches are the least with highest 15.3% while informational searches are the highest with 63%. This data clarifies the searchers purposes frequency and search interests.

Identifying Intent using Intention Map: While talking about the classification of webpages as per the intent type, *A. K. Dani* [2008] states that the informational class includes pages which contain lots of textual material to be read up. The informative pages are generally not the home pages of sites and have a high URL depth. Further, it can be observed that the query keyword occurs more frequently in the latter part of the url not

[191] https://yoast.com/search-intent/

including the domain name. This is also true for the transactional URLs but for navigational pages, the query term frequently is the domain name of the web site, or it occurs frequently in the domain name of the URL. *Park et al.* [2010] proposed a method, Chart-54, Appendix-9, to mine user intentions and build an intention map, leading to improved user satisfaction. *Tsukuda et al.* [2013] focused on the estimation of intent types, achieving a 64.4% accuracy in distinguishing between informational and navigational intents. *Ahmadvand et al.* [2020] introduced JointMap, a model that simultaneously learns commercial vs. non-commercial intent and relevant product categories, significantly improving intent prediction and product category mapping. *Garigliotti & Balog* [2018] addressed the construction of a knowledge base of entity-oriented search intents, demonstrating the ability to generate high-quality data. These studies collectively highlight the potential of intent types and maps in enhancing search engine performance and user experience. [192]*Olaf Kopp-Intents* [2022] has a highly descriptive study on search intent types. He sees 4 major intent types and 12 micro-intent types as shown in Infographic-8, Appendix-9. Believing his intuitive guts, he further distributes these micro intents according to the customer journey stage as shown in below Infographic-9, Appendix-9. As per *Kori et al.* [2015] search intentions are broadly divided into **Verify and Discover** *(I would call it the purpose based intent types)* and puts the navigational and informational intent types under it, as shown in below Chart-56, Appendix-9 for the contextual search engines. Their experiment was to test and compare the contextual search engines output to that of existing search engines and it was found to be same on both. As per the above FLR, we can say that only [192]*Olaf Kopp-Intents* [2022] categorizes Entertainment as sub-type of Informational Intent and Location as sub-type of Navigational Intent; but none as major Intent Category. So here we present the **six major intent types** as below:

a) **Navigational Intent:** People with this intent want to **find** a specific website or blog or a URL but don't remember the exact reach about. They may have heard of or had visited or heard of its company so searching for its website or say web-presence. Till recent times it was just about navigating the official website but now a days specially after the evolution of social media, new designed machine learning programs such as Google and Bing are considering the official social pages or profiles of companies also as a navigational intent, a secondary intent; of course, the primary being the direct official website. For example: 'Linkedin', 'SEMrush Technical Keywords Course', 'similarweb.com', 'Facebook login', 'lemonde facebook' {chase online banking}, {fnac.fr}, or {best buy} etc. Direct URLs are also in the search trends reason being typing it in the search box is psychologically more convenient especially for less technical sound searchers. Auto-correct feature provided by the search engines' search box makes this approach more trustable. Search Engine indexing programs are solely designed to meet this requirement of reaching directly to the official website or profile of a searcher.

[192] https://searchengineland.com/micro-intents-seo-content-journey-mapping-386544

b) **Informational Intent:** People with this intent usually carry a **question or topic** in mind or want to learn or gain knowledge about on a topic in maximum possible depth and they don't have the idea where exactly it could be. The research could be for broad or specific information, but the goal is always to discover additional facts, answer a question, or answer a series of questions. A searcher may be doing research for a college paper. It can be for general queries, which result provides the broadest range of authoritative sources and for specific queries, which result brings the user exactly what they are looking for. For exemple {Macron}, {prix du meilleur réalisateur du festival du film de caen}, or {qu'est-ce que le paludisme}, or "who is Nicolas Sarkozy ?", "Paris airport directions", "Barak Obama", "Olympic 2021 stats", "Python" etc. Informational Keywords don't have a single right answer contrary to Navigational Keywords.

c) **Transactional Intent:** People looking to **do** something like buying products online; online services like book something like a ticket of train, flight, cinema or hotel; download a document file, audio file, video file, photo etc.; filling up a subscription form of a newsletter, an account opening or a filling an online survey. The query intent is to obtain a product or service, or to download an application or multimedia file, calculate an equation, etc. The searcher is looking for a resource or tool (download, buy, obtain, chat) rather than specific information. A searcher may need to calculate how much currency to bring with them on a trip to France. Which result brings for example "buy apple macbook", "auchen coupon", "samsung galaxy a6 furbished", "Air France premium price" etc. This intent category is the biggest money making for Search Engines as the most of paid ads happen with sale-purchase i.e. transactional intent. This category pushes Search Engines to have special Shopping Blocks and optimized aggregated easy-to-use selection experience straight in the SERP results such as flights price listing blocks. *Chapelle et al.* [2011] proposes five types of sub-intents under Transactional Intent category as follows:

1. **Buying Guide Intent:** It is when someone try to retrieve a document that explains how to buy a particular category of products. It might contain main features and considerations, important terminology, **where and how to purchase**.
2. **Reviews Intent:** Where people retrieve documents for evaluations of products that could include **ratings, recommendations, and explanations of the author's** point of view of the product.
3. **Support Intent:** Here people retrieve documents containing technical detail about a product that helps a user in using the item. For example, **manuals, troubleshooting pages, tutorials, and warranties**.
4. **Official product homepage Intent:** It is about retrieving a document that describes the specific product at the **manufacturer's domain**.

5. **Shopping site/purchase Intent:** It is when people retrieve documents from sites that give the user an opportunity to **purchase the product online**. There can also be a ‘‘general’’ intent defined as a catch all for all user needs.

Ever since the beginning of the work on understanding the Intents of Searchers by Search Engines till late years of 2020’ above three intents were floating every here and there. In mid-2020 the fourth intent pop-up in the market as below:

d) **Commercial Intent:** People doing pre-purchase **research/investigation/comparison/review** around a product or service prior to finally make up their mind for purchasing it. They are in process of weighing up their options by reading reviews and comparisons. For example, “best multi-vitamins”, “Windows or iOS”, “Moz review”, “top hotels in Dubai” etc.

e) **Entertainment Intent:** Such as **watching** a video, listening to an audio, playing game online. Research lacks in justifying their concrete category out of the four major intents. Are they navigational or informational; certainly not much to do with transactional and commercial either; because both are driven with buying or reviewing, which even not that lengthy one as watching a movie for hours. So, this is a huge ground of thinking about why not it is yet categorized a separate Intent type i.e. the Entertainment Intent.

f) **Mapigation / Destination Intent:** The person looks for a specific map help like address and its concrete location on Google Map to say; to know about a landscape or want to get traveling road/rail/flight itinerary along with instructions between two destinations. Most of research yet seen sums-up their explanation around “navigational intent” as finding a website or official brand. Should it be categorized under informational intent? Then first thing why it is not seen mentioning of Mapigation in it yet? Also, it is more about visual search instead of texture search which starts and ends up with reading and that’s all. People keep on searching for ways on drive for hours and keep their GPSs work with changing location search in every such second search. So certainly, it should be either defined justifiably in either of the Navigational or Informational intent otherwise Mapigation can be another major category around it. Major because, if a fragmentation like Chart-54 is done among 6 Intent types as mentioned here; certainly, these two i.e. Entertainment and Mapigation will have a minimum of double-digit search share.

Along with above six major intents, there are some more intents of mix type or coexisting in same search, which can be defined as sub-intents or even as major intents, which I am not certain to do for the moment but here is a list:

1. **Local Intent:** It is about finding information about local businesses or services. For example, “Restaurants near me" or "Hardware store in [city]."

2. **Temporal Intent:** It is about seeking information related to a specific time or date. For example, "Weather forecast for tomorrow" or "Historical events in 2020."
3. **Health Intent:** It is about looking for health-related appointments, follow-ups, and tracking. For example, "Doctor appointment" or "Doctor nearby me" or "Hospital in the city". *As talked in the 'Introduction' of this thesis, alongwith four standard search categories; the Health Search Intent category is also at the prime focus of this thesis. We shall be elaborating and testing it in next chapters with the six health search intents chosen.*
4. **Social Intent:** It is about interacting with others or seeking social content. For example, "Connect with friends online" or "Trending topics on social media."

These are general categories, and users often have mixed intents within a single search session. Search engines use sophisticated algorithms to decipher and fulfil these diverse intents, presenting users with a range of relevant results based on their search objectives. When it comes to the mixed intents, it refers to finding multiple intents probability with a search query. Let's understand it as below.

Queries with multiple intents: Many queries have multiple possible intents, especially the short terms and general search queries. The most popular intent should be fulfilled at the top of the page. When multiple intents are a possibility, search engine yield results on same pattern and it takes a lot of research, analytical and algorithmic work for them to present results in such a manner that user gets them in order of the most likely intent, second most likely intent, and so forth when deciding which result is better for them. Catering to a variety of intents does not automatically make one SERP satisfactory rather it requires the guts to sense the intuition and intent of the user. Search intent combinations can be explored and combined in interactive information-seeking tasks. *Strahl et al.* [2021] introduced an interactive information retrieval system that allows users to perform multiple simultaneous search queries and combine them using intersection and difference operations. *Raff et al.* [2014] proposed an architecture that derives user intent from a search query and chooses the appropriate search method based on the derived intent, such as local search, non-local search, or both. *Ruotsalo et al.* [2018] introduced interactive intent modelling, a technique that models a user's evolving search intents and visualizes them as keywords for interaction, allowing users to provide feedback and improve the intent estimate. *Taraknath et al.* [2017] developed a technology for ranking search results based on customer intent, using different combinations of coefficients and scores associated with product entries. Their model is shown in Chart-57, Appendix-9. *Xiao et al.* [2011] evaluated the effectiveness of methods for query intent mining and results diversification in web search, showing improvements in subtopic mining and document

ranking. A range of studies have explored the complex nature of search intent and the potential for combining different factors to improve its understanding. *Xia et al.* [2013] emphasizes the importance of diversity in ranking search intents, proposing a model that combines relevance and diversity scores. *Mitsui et al.* [2017] focus on predicting information seeking intentions from search behaviours, demonstrating the potential of a simple classification model. *Ruotsalo et al.* [2018] introduce interactive intent modelling for exploratory search, which allows for the visualization and improvement of user search intents. *Wu et al.* [2010] takes a different approach, using a classification framework based on linguistic analysis and search engine feedback to identify user intents in web queries. These studies collectively highlight the potential of combining various factors, such as **relevance, diversity, search behaviours, and linguistic analysis**, to improve the understanding and prediction of search intent. *Jiang Zong* [2015] and *Wu et al.* [2010] both emphasize the importance of using multiple features, such as query text and web knowledge, to improve intent identification accuracy. These studies collectively highlight the potential of combining various features and techniques to effectively identify search queries with multiple intents. To identify search queries with multiple intents, different approaches have been proposed in the literature. *Maji et al.* [2020] propose a **latent variable generative model** along with a **novel data dependent regularization technique** for identifying multiple intents in tail queries of e-commerce search engines. *Gupta et al.* [2023] propose a **deep neural network-based ambiguity prediction model** to analyse input-output strings and map them to different properties responsible for multiple intents. *Burchell et al.* [2020] focus on multi-sentence questions ***(MSQs)*** and identify different question discourse relations between subparts of MSQs to determine speaker intents. *Hashemi & Zamani* [2022] argue that as long as an accurate and complete set of intent representations is generated, the model should not be penalized for the sequence of intent representations. They propose a **stochastic permutation invariant approach** for optimizing such networks. *Mohammadi et al.* [2020] introduces a **clustering model that leverages Google results and machine learning** to detect query intent, while *Mendoza & Zamora* [2009] focuses on high-precision **query classification using support vector machines**. These studies collectively highlight the potential of combining various features and techniques to effectively identify search queries with multiple intents. Identifying search queries with multiple intents, also known as "**mixed intents**," can be challenging but is crucial for search engines to provide relevant results. Here are some strategies and considerations for **identifying search queries with multiple intents**:

1. **Context Analysis:** Examine the context of the query to understand if it encompasses different topics or purposes. For example, in the query "apple," the user might be looking for information about the fruit, the tech company, or other related topics.

2. **Semantic Analysis:** Look for phrases within the query that might indicate multiple intents. For example, in the query "best laptop for gaming and productivity," the user may have dual intents related to gaming and productivity.
3. **Intent Keywords:** Identify keywords that suggest different intents within the query. For example, in the query "exercise routine for weight loss and muscle gain," the intent keywords are "weight loss" and "muscle gain."
4. **Modifiers and Connectors:** Pay attention to modifiers like "and," "or", or "vs." that may indicate a combination of intents. For example, in the query "Paris travel guide or budget hotels," the user is interested in both travel guides and budget hotels.
5. **Query Length and Complexity:** Longer and more complex queries may contain multiple intents. For example, the query "healthy recipes for breakfast, easy to make at home" suggests a combined intent of healthiness and ease of preparation.
6. **Topic Variance:** Notice if the query spans across diverse topics. For example, in the query "climate change effects and solutions," the user is interested in both the impacts and potential solutions related to climate change.
7. **User Behaviour Analysis:** Analyse user behaviour, such as clicks on diverse search results, to understand mixed intents. For example, if a user clicks on both news articles and product reviews for the query "iPhone 13," it may indicate a mixed intent of staying updated and making a purchase decision.
8. **Query Reformulation:** If users perform multiple iterative searches in a session, consider the evolution of their queries to identify changing intents. For example, a user might start with a general query like "best smartphones" and later refine it to "best budget smartphones under $500."
9. **Utilize NLP Techniques:** Leverage NLP algorithms to analyse the structure and semantics of the query. For example, NLP can help identify nuanced intents in queries like "healthy vegan recipes without soy."
10. **User Feedback:** Consider explicit feedback or user interactions to refine understanding. For example, if a user refines their query or clicks on specific results, it provides clues about their evolving intents.

Identifying search queries with multiple intents requires a combination of linguistic analysis, user behaviour observation, and algorithmic processing. Search engines continually refine their capabilities to better understand and respond to the complex and evolving information needs reflected in user queries. Below table presents comprehensive **Search Intent Combinations:**

Intent Combinations	Inclusion	Exclusion	Searcher Looks For

Single Intent Query	No ambiguity in what the user looks for	Broad navigational queries such as (amazon.com) are not considered single intent queries	A particular piece of information A very specific webpage or document Intent for purchasing a particular product
Intent Combinations	**Example Case**		
Multi Intents Navigational Query	Navigating a single website such as {amazon.com}		
Multi Intents Informational Query	{Amazon} can be a website, river, USS Amazon Sailing Vessel		
Navigational + Transactional Intents	Buying a jewellery from amazon.com		
Navigational + Informational Intents	1. Navigating & getting information on Family Guy from a website 2. The user is seeking information on installing WordPress *(informational intent)* and may also have a navigational intent to find relevant guides or tutorials. For example, "How to install WordPress"		
Informational + Transactional Intents	1. Information + Downloading for Family Guy 2. The user is looking for information on the best smartphones in 2023 (informational intent) and has a transactional intent to purchase one online. For example, "Best smartphones 2023 buy online"		
Commercial + Navigational	The user is conducting a commercial investigation by seeking reviews of the Samsung Galaxy S21 (commercial investigation intent) and has a navigational intent to find stores nearby for potential purchase. For example, "Reviews of Samsung Galaxy S21 stores near me"		
Local + Transactional	The user has a local intent to find Italian restaurants nearby (local intent) and a transactional intent to place an online order. For example, "Italian restaurants near me order online"		
Informational + Temporal	The user is interested in information on recent advancements in space exploration (informational intent) and has a temporal intent, seeking the latest updates for the year 2023. For example, "Space exploration advancements 2023"		
Transactional +	The user is looking for transactional information on holiday travel deals		

Temporal	(transactional intent) with a temporal focus on the year 2023. For example, "Holiday travel deals 2023"
Informational + Transactional + Navigational	The user is interested in obtaining information on healthy recipes for weight loss (informational intent), has a transactional intent to order ingredients online, and may also have a navigational intent to find specific online stores or platforms. For example, "Healthy recipes for weight loss order ingredients online"
Commercial + Transactional	The user is conducting a commercial investigation by comparing prices of 4K TVs (commercial investigation intent) and has a transactional intent to make a purchase. For example, "Compare prices of 4K TVs buy now"
Local + Navigational + Temporal	The user is interested in local events happening in San Francisco (local intent), has a navigational intent to view them on a map, and has a temporal focus on this weekend. For example, "Events in San Francisco this weekend map"
Transactional + Navigational + Temporal	The user has a transactional intent to buy concert tickets for Taylor Swift (transactional intent), a navigational intent to find tour dates, and a temporal intent for the year 2023. For example, "Buy concert tickets for Taylor Swift tour dates 2023"

Table-16: Search Intent Combinations

Search intent combinations refer to scenarios where a user's search query encompasses multiple intentions or purposes. Users often express complex information needs that involve different aspects or dimensions. Understanding these intent combinations helps search engines deliver more precise and contextually relevant results to meet users' multifaceted information needs.

Understanding the Intent of misspelled Search Queries: Misspellings in search queries are common and can be caused by various factors such as fast typing, different languages, or ambiguous intentions. Search engines handle misspellings effectively through techniques such as spelling correction, fuzzy matching, and query suggestion, [193]*LN-Search-Engine* [2023]. Spelling correction involves detecting and correcting misspelled words using various techniques such as dictionaries, rules, statistics, machine learning, or user feedback. Fuzzy matching is used to find documents like a query, even if they do not match exactly, while query suggestion offers alternative or related queries to the user to guide them in refining their search. Google has also

[193] https://www.linkedin.com/advice/0/what-most-effective-ways-handle-misspellings-search-mlknc

implemented deep learning and NLP to address misspellings, including the use of keyboard adjacency to correct mistyping and the application of its BERT model to better interpret search intent, [194]*MacLeod* [2021]. In e-commerce, typo tolerance is used to handle misspelled searches, ultimately improving the precision and recall of search results. Overall, search engines employ a combination of linguistic, statistical, and machine learning techniques to effectively handle misspellings in search queries. Query understanding is a fundamental problem in information retrieval. Understanding the intent of misspelled search queries is challenging due to the errors or vocabulary gap between the user's intended query and the query issued to the search engine. Query spelling correction plays a crucial role in improving search engine performance. Early works focused on edit distance-based correction, while modern approaches address more complex and realistic scenarios involving multiple types of spelling errors. Additionally, intent detection in chatbot applications also faces challenges with spelling errors. To compensate for the performance degradation caused by spelling errors, training data can be augmented with automatically generated erroneous samples. This approach has shown to significantly improve intent detection accuracy in the presence of spelling errors, *Dou & Guo* [2020] and *Zhang et al.* [2020]. *Roy et al.* [2015] and *Donato* [2011] both emphasize the importance of identifying user intent in search queries, with Roy proposing a classification of words as content or intent, and Donato conducting a study to gain insights into user intent formulation. *Duan & Hsu* [2011] and *Cao et al.* [2009] focus on the technical aspects, with Duan developing a model for online spelling correction and Cao incorporating context information into query classification. These studies collectively contribute to a deeper understanding of user intent in search queries and the development of effective methods for handling misspelled queries. If Search Engine encounters a misspelled query, it calls for what the likely correct spelling may have been. Sometimes the search engine will catch the misspelling and offer a corrected spelling and/or a recourse link. For example:

- {all stete} query is likely misspelled and {all state} may have been intended
- {goggle} query is likely to be a misspelling of {google}, but {goggles} is also possible

Interpreting the intent of the misspelled queries: Interpreting the intent of misspelled queries can be a challenging but important aspect of search engine optimization and user experience. Users may make typos or spelling mistakes while entering search queries, and search engines aim to understand the intended meaning to provide relevant results. Here are some strategies to interpret the intent of misspelled queries:

1. **Autocorrection:** Many search engines use **autocorrection algorithms** to identify and correct common spelling mistakes. This involves comparing the entered query to a dictionary or dataset of correctly spelled

[194] https://www.wordtracker.com/blog/keyword-research/google-outlines-new-approach-to-misspellings-in-search-using-deep-learning

words. Autocorrection algorithms attempt to interpret the intent while maintaining the user's original intent by suggesting or substituting corrected versions of misspelled words.

2. **Semantic Understanding:** Search engines may employ semantic analysis to understand the **context of the query**, considering the surrounding words and the overall structure of the sentence. By analysing the semantic context, search engines can infer the user's intent even if there are misspelled words.
3. **User Intent Models:** Search engines often build models of user intent based on historical search data *(behavioural analysis).* By analysing similar queries and user behaviour, the system can predict the likely intent behind a misspelled query. If the user is logged in, the search engine may **leverage personalization features** to understand the user's preferences and historical intent.
4. **Query Expansion:** Search engines may employ **query expansion techniques** to broaden the search based on potential corrections or related terms. By expanding the query, the search engine aims to capture the user's intent even if there are spelling errors.
5. **Fuzzy Matching:** Fuzzy matching algorithms can be used to find approximate matches for misspelled words. This involves considering variations, phonetic similarity, or keyboard proximity. Fuzzy matching helps align the misspelled query with the closest correct term, preserving the user's intended meaning.
6. **User Feedback:** Search engines may **incorporate user feedback**, such as clicks on corrected results or reformulated queries, to learn from user interactions and improve future interpretations of similar misspelled queries. If users **click on autocorrected results**, it provides valuable insights into whether the correction accurately captured their intent.
7. **Multimodal Search:** In some cases, search engines may incorporate multimodal search, combining modalities of text with other inputs modalities like images or voice. This can help disambiguate misspelled queries.
8. **Sensitivity to Variations:** Search engines are designed to be somewhat **tolerant of variations** in spelling. This includes handling common misspellings and variations in language usage.

Interpreting the intent of misspelled queries is an ongoing challenge, and search engines continually refine their algorithms to improve accuracy. The goal is to provide users with the most relevant results, even when there are minor errors in the input query.

Interpreting the likeliness of Search Intent: The term "intent likeliness" refers to the probability or likelihood that a user's search query is indicative of a specific intent or purpose. In the context of search engines, understanding user intent is crucial for delivering relevant and accurate search results. Intent likeliness involves assessing how probable it is that a given query aligns with a particular user intent category. Search engines

employ various algorithms and techniques to understand and infer user intent based on the entered queries. Understanding the likelihood of search intents by interpreting a search query is a fundamental problem in information retrieval. Different tasks have been proposed to understand users' search queries, such as query classification or query clustering. However, understanding a search query at the intent class/cluster level may result in the loss of detailed information. To address this, researchers have proposed novel tasks like Query-to-Intent-Description (Q2ID) task, which aims to generate a natural language intent description based on both relevant and irrelevant documents of a given query, *Zhang et al.* [2020]. They further investigate that existing approaches for understanding query intents can be classified into **query intent classification** and **query intent mining**, which involve mapping queries into categories or finding subtopics covered by the queries. Accurate query intent prediction is crucial for search engine success, and various deep learning models have been experimented to achieve this, *Dou & Guo* [2020]. *Roy et al.* [2015] and *Das et al.* [2013] all propose methods for classifying user intent in web search queries, with Roy focusing on the detection of intent words and Das using past user interactions with search results. *Donato* [2011] provides insights into how users formulate their intents into queries through an experimental study. More recently, *Liu et al.* [2020] presents a deep learning framework for predicting user intents in search queries, demonstrating its effectiveness in both offline and online evaluations. However, there is not as such concrete idea till date research present around measuring intent likeliness as there are various parameters seem yet to explore and interpret in justifiable extent. **Query intent likeliness** should be studied on following parameters:

1. **Query Analysis for intent(s) likeliness:** Examine the keywords in the query to identify specific terms or phrases that indicate the user's probable intent. Consider the context of the query, including surrounding words and the overall structure of the sentence.
2. **Intent Probabilities Classification Models:** Search engines use machine learning models to classify queries into different intent categories. These models are trained on vast datasets to recognize patterns and associations between queries and user intents. These models assign probabilities to different intent categories, indicating the likelihood that a given query aligns with a particular user intent.
3. **User Behaviour and Preferences Analysis:** Analyse user behaviour, including past search queries, clicks, and interactions with search results. Consider patterns in user behaviour to understand common preferences and intents associated with certain types of queries.
4. **Assessing the Semantic and Contextual Understanding:** Search engines leverage semantic analysis to understand the meaning and context of words in a query. Infer the likely intent by considering the broader context of the user's search history, location, and device.

5. **Multifaceted Intent Variations and Hierarchy:** Recognize that a single query may have multiple possible intents. Some queries may be ambiguous and open to interpretation. Understand that intents can exist in a hierarchy, with more specific intents falling under broader categories.
6. **Long-Tail Queries and Intent Refinement:** Longer and more specific queries (long-tail queries) often indicate a higher likelihood of a specific intent. Users expressing detailed information needs are likely looking for precise and relevant results.
7. **Action-Oriented Terms with Transactional or Commercial Intent:** The presence of transactional keywords (e.g., "buy," "order," "subscribe") suggests a higher likelihood of a transactional intent. Transactional keywords may also indicate an intent to research or compare products or services before deciding.
8. **Time Sensitivity and Search Trend Analysis:** Queries with temporal elements (e.g., "today," "2023," "latest") may suggest a temporal intent, indicating a desire for up-to-date information. Consider historical search patterns to understand temporal intent variations.
9. **Geographical Indicators:** Queries containing location-specific terms (e.g., "near me," city names) may signal a local intent. Users often have specific intents related to their geographical context.
10. **Intent Refinement and Feedback:** Observe how users interact with search results, including whether they refine their queries or click on specific results. Learn from user feedback and interactions to continuously refine and improve intent interpretation.

Interpreting the likelihood of search intent is a dynamic process that involves considering multiple factors and leveraging advanced algorithms to deliver more relevant and personalized search results. Search engines continually evolve their models to enhance their ability to understand and respond to user intents accurately.

Understanding the Scope of a Search Intent: Intent behind a query is complete if its scope is also understood by the search engine. Research on web search queries has identified several factors that can help determine the specificity of a query. *Hafernik & Jansen* [2013] found that query length and parts-of-speech usage can distinguish between narrow and general queries. *Athukorala et al.* [2014] proposed a model for predicting the subjective specificity of search results based on user behaviour. *Wang et al.* [2009] focused on extracting broad latent query aspects from search session logs to aid users in reformulating their queries. *Freund & Toms* [2005] highlighted the brevity of web queries, which can lead to frustration and low search success rates. A narrow scope query is underspecified or lacks senses that the user did not think of. On the other hand, a broad scope query represents one sense or information need that can aid users in reformulating their queries. Broad latent aspects can be extracted from query reformulations found in historical search session logs. These aspects can be

used to optimize the search engine's response to a given query. Extracting broad latent aspects leads to significant gains in terms of accuracy, click-through data, and consistency among aspects predicted for similar queries, *Yoon & Pantel* [2018]. Additionally, the subjective specificity of search results can be predicted based on simple observables such as result-clicks. This allows differentiation between levels of subjective result specificity based on the user's current information need, *Wang et al.* [2009]. The narrow scope and broad scope of a search query refer to the specificity of the search terms used and the resulting inclusiveness of the search results. For example, in the context of PubMed Clinical Queries, the "Broad" and "Narrow" versions are tuned to maximize either retrieval of a wider range of results (broad) or retrieval of more specific results (narrow), *Sperr Jr.* [2018]. Understanding the difference between narrow and broad search scopes is important, as it impacts the specificity and inclusiveness of the search results. This distinction is particularly relevant in fields such as clinical research, where the level of certainty or specificity in the search results is crucial, *Bemmelen & Plough* [2008]. In the context of search intent, the choice between a narrow or broad scope can also impact the relevance of the search results to the user's goal. It is essential to consider the user's intent when determining the appropriate scope for a search query, [195]*Brian Dean* [2022]. So basically, there can be two types of queries in the context of scope (usability):

a) **Broad Scope Queries:** Search Engines fetches general results from a variety of authoritative sources. For example, if the query is {business suit}, then a variety of online retailers would be ideal. It would be less helpful to provide multiple results from the same domain. Other possible relevant returns include tips for buying a suit, tips about how a suit should fit, and/or articles or reviews about suit-buying experiences.
b) **Narrow Scope Queries** call for specific web results and features. It goes two ways:
   i) It can mean **consistency of subject**, as in the example {How to cook pizza}. Relevant web results may include a diverse range of domains *(recipe blog articles, recipe sites, YouTube videos, etc.)*, but the subject matter for all results should be methods of cooking a pizza.
   ii) It can mean **consistency of domain**, as in the example {BBC coverage of world cup}. To best satisfy the intent, relevant web results should come from the domain "BBC," but the subject of the web results and features can address a diverse range of World Cup topics.

Discussion: The major argument of above study can be the proposal of two new major search intent types i.e. Entertainment and Mapigation / Destination. The literature review indicates that not much of attention with psychological test and taste ever given to think this angle. As per the theory of purpose and emotions, it is reasonable to annotate the user behaviour from this perspective. Certainly, there are huge possibilities of having

[195] https://backlinko.com/hub/seo/search-intent

a totally different mindset while thinking about these both intents. This thesis intends to give answer to this proposal.

C - SERP Fetching *(The mechanism of surfacing the web content on SERP)***:** As per first three verticals of first layer of Chart-44 by now we have understood the three phenomena of SERP Making, the Content Listing, Content Classification, Search Query Inferring and Intent Identification. This topic deals with the last and fourth vertical of this horizontal, the SERP Fetching. In this vertical SE involves multiple steps to deliver relevant and accurate results to users. The web content database prepared through Listing, Crawling, Rendering (Parsing) and indexing of discovered content works as the referral point where each search query is taken through to fetch the relevant SERPs. It starts with Query request and follows the steps below:

1. **Ranking & Positioning:** Ranking Algorithms determine **(Algorithmic assessment)** the placement of the content on a SERP, based on how well it satisfies the user's query. These algorithms consider various factors, including keyword relevance, content quality, user engagement metrics, page authority, and more. Each search engine has its proprietary ranking algorithms. The ranked pages are ranked for SERP **(SERP order determining)** based on their perceived relevance and quality. The top results are considered the most relevant to the user's query. Although, **Positioning** is a sub-part of Ranking, but it is a specific task that requires special dedicated efforts. Positioning is an adjustment for a position in a Featured SERP Result Block, for example a top position in just Organic Results or Local Maps block and on SERP Interface top, middle or bottom. This is where the serious and sincere hard work of the web content master or its SEO specialist is determined if it shall get a good ranking in top 3 Organic SERP Results, in Paid Result Blocks or in Featured Organic Results Blocks. The search engine **ranks the results based on their relevance** to the query. This ensures that the most relevant results are presented to the user, [178]*Britney Muller Moz.com*. For example, for Organic Search Engine *(News topics sort on freshness, most of on relevance)* and Product Search Engine *(List based on pricing, availability in stock, time - 1 week or 1 month)*. *Bendersky et al.* [2011] talk of the quality-biased ranking method based on state of art sequential dependence retrieval model that promotes documents containing high-quality content, and penalizes low-quality documents, beside link-based measures such as PageRank of Google. Other ranking factors count for:
    a) **Content Factors:** Such as Strength of links to a webpage, social media activity
    b) **Ranking Organic Search Matches based on Content Quality:** Link Popularity *(Social Media Activity, Human Rating, Freshness, Page Style, and Site Organization)*, PageRank
    c) **Ranking Organic Search Matches based on Keywords:** Keywords in Title, Headings and emphasized text, Body text, Description; Universal Keyword Factors *(Content Theme, Keyword Intent,*

Contextual Relevancy, Term Rarity, and Term Proximity) and number of occurrences of the words from keyword.

d) **Ranking Paid Search Matches:** Higher bidder auction, Ad Quality Score *(Different types of CTRs, Different types of content quality, Different contexts)*

2. **Search Results Retrieval (Fetching) to Display:** By this stage results are ready with **algorithm's assessment** and top-ranking pages **(SERP Order)** decided. Based on the ranking algorithm, the search engine generates a SERP that includes a list of web pages deemed most relevant to the user's query. The SERP typically consists of organic results, paid advertisements, and additional features like featured snippets, knowledge panels, and rich results.

a) **SERP Display:** The search engine generates the SERP **(Searcher Interface)** by displaying the top-ranked results along with other features and information. The SERP may include organic search results, paid advertisements, featured snippets, knowledge graphs, and other relevant features. Search engines often enhance the SERP with **rich features**, such as featured snippets that provide direct answers to queries, local packs for location-based queries, and image carousels.

b) **SERP Features:** In addition to organic search results, SERPs may include various features such as featured snippets, knowledge graphs, local packs, image carousels, and more. These features provide users with additional information and options directly on the search results page.

c) **User Interaction and Feedback:** User interactions with the search results, such as clicks, dwell time, and bounce rate, provide valuable feedback to search engines. This data is used to refine ranking algorithms and improve the overall relevance of search results over time. Based on user interactions, the search engine continuously **refines** its understanding of relevance and may adjust ranking algorithms to improve future search results.

d) **Mobile and Local Considerations:** SEs consider the growing prevalence of mobile devices and the importance of local relevance. Mobile-friendly pages and local search factors may influence the ranking of results, especially for users searching on mobile devices.

e) **Content Freshness:** Fresh and regularly updated content is often given preference in search rankings. Search engines aim to provide users with the most current and relevant information, particularly for queries related to recent events, news, or trends.

f) **Webmaster Guidelines and Regular Algorithm updates:** Webmasters and content creators can optimize their pages to improve visibility on SERPs. Following search engine guidelines, implementing proper on-page SEO (search engine optimization), creating high-quality content, and building reputable backlinks are common strategies. Search engines regularly update their algorithms

to adapt to changes in technology, user behaviour, and emerging trends. Algorithm updates aim to improve the relevance and quality of search results.

g) **Privacy and Security:** Search engines prioritize secure websites (using HTTPS) and may indicate the security status of a website on the SERP to enhance user privacy. Search engines are increasingly incorporating privacy-focused measures, such as anonymizing user data and providing options for users to control their privacy settings.

3. **Search Engine Self-Learning Loop Run:** As the SERP is generated SE runs its loop of self-learning, researching, and updating on each new, updated or trending search query by following methods:
   a) **User Interaction:** Users interact with the SERP by clicking on organic results, advertisements, or featured snippets based on their preferences and needs.
   b) **Click-Through Rates and User Behaviour:** The search engine monitors user behaviour, including click-through rates, dwell time, and bounce rates. This data helps refine the ranking algorithm over time and improve the relevance of search results.
   c) **Continuous Crawling and Updating:** Search engines continuously crawl the web to discover new content and update their index. This ensures that the SERPs reflect the latest information available on the internet.
   d) **Personalization:** Some search engines personalize SERPs based on a user's search history, preferences, and location. Personalization aims to deliver more tailored and relevant results to individual users.
4. **SERP Ranking Signals a Search Engine uses around an indexed web document:** Continuing the base developed in above discussion, it makes it very easy pathway to grasp how Search engines use complex algorithms to determine the relevancy of search results based on various factors. While the exact algorithms are proprietary and constantly evolving, here are some key elements that typically influence the relevancy of search results:
   a) **Keyword Relevance:** Search engines analyse the presence and placement of keywords in a webpage's content, title, headings, and meta tags. Pages with relevant keywords are considered more likely to match a Searcher's query.
   b) **Content Quality:** High-quality, informative, and engaging content is given preference. Search engines assess the depth and uniqueness of content to determine its relevance to a searcher's search.
   c) **Page Authority:** The authority of a webpage is often measured by the number and quality of incoming links. Pages with more authoritative backlinks are considered more valuable and relevant.

d) **Searcher Experience:** Search engines consider factors like page loading speed, mobile-friendliness, and overall searcher experience. A positive searcher experience is considered indicative of a relevant and reliable webpage.

e) **Relevance of External Links (REL):** Outbound links from a webpage to other reputable sites can enhance its credibility. Similarly, inbound links from authoritative sources contribute to a page's relevance.

f) **User Engagement Metrics:** Metrics like click-through rate (CTR), bounce rate, and time spent on page are used to assess how users interact with search results. Pages with higher engagement are often considered more relevant.

g) **Freshness of Content:** Search engines may prioritize recent and updated content, especially for queries that require current information. Timely and relevant updates can improve a page's ranking.

h) **Social Signals:** Social media activity, such as shares and likes, can be considered a signal of a page's relevance. While the impact of social signals may vary, they are sometimes factored into search engine algorithms.

i) **Structured Data:** Websites that use structured data markup (e.g., schema.org) can provide additional context to search engines, helping them understand the content and present rich snippets in search results.

j) **Geographical Relevance:** For location-based queries, search engines consider the geographical relevance of content. Local businesses and services may be given priority for queries with local intent.

5. **Diversity of results in a SERP:** Which algorithms are used for this task? The diversity of results across the entire SERP is influenced by algorithms designed to provide a well-rounded and relevant set of results to users. Search engines aim to offer diversity in terms of content type, sources, and formats to cater to different user preferences. While the specific algorithms are proprietary, here are some general concepts and factors that contribute to result diversity:

   a) **Relevance Algorithms:** Search engines employ algorithms that assess the relevance of a webpage to a given query. These algorithms consider factors such as keyword matching, user intent, and content quality to determine the overall relevance of a page.

   b) **Content Type Diversity:** Algorithms are designed to understand and diversify the types of content that appear in search results. This includes a mix of webpages, images, videos, news articles, and other relevant content formats based on the nature of the query.

   c) **Source Diversity:** To avoid over-reliance on a single source, search engines aim for diversity in the origins of search results. This means results may come from a variety of websites, ensuring a broad representation of information sources.

d) **Domain Diversity:** Search engines consider the diversity of domains in the SERP to prevent a single domain from dominating the results. This helps in providing users with a more balanced and varied set of perspectives on a given topic.

e) **Freshness and Recency:** Algorithms consider the recency and freshness of content. For certain queries, recent information may be prioritized to ensure users receive the most up-to-date results. This contributes to the dynamic nature of the SERP.

f) **Personalization:** Personalized search algorithms consider individual user preferences, search history, and behaviour. While personalization tailors result to a user's interests, search engines also aim to maintain a level of diversity to expose users to new and relevant information.

g) **Query Intent Recognition:** Algorithms analyse the intent behind a user's query to deliver diverse results that satisfy different aspects of that intent. For example, a user searching for "apple" might be interested in information about the company, the fruit, or Apple products, and the SERP aims to cover these diverse aspects.

h) **Geographical Diversity:** For location-based queries, algorithms may prioritize results that are geographically relevant. This ensures that users receive information that is applicable to their specific location or region.

i) **SERP Features:** Search engines incorporate a variety of SERP features such as featured snippets, knowledge panels, and related questions. These features add diversity to the presentation of results, offering users different ways to access information.

j) **Feedback Loops:** Search engines use feedback from users, such as clicks, bounce rates, and dwell time, to continuously refine and improve result diversity. If users consistently engage with certain types of content, the algorithms may adjust to reflect those preferences.

Discussion: Ultimately the fight is for getting a top ranking on first page of a **SERP** by two methods, the Organic Ranking, and the Paid Ranking also know as Search Engine Marketing. Organic SERP Ranking is usually achieved through SEO and Paid Ad Ranking through SEM; both combined are called as **'SERP Ranking' or 'SERP Marketing'**. Within SERP Ranking falls the SERP Positioning as told above. The specialized Topicality sites based on **E-E-A-T** *(3E as per this thesis)* parameters get positioning in special features. Here a significant argument can be put to sum-up E-E-A-T into 3E for simplicity of the concept. Search Engine Marketing is a form of digital marketing *(or internet marketing)* that consists of a variety of tactics to promote your business by increasing visibility of your content to searchers, *Moran & Hunt* [2014] - Page 10. Narrowing down the game further to specificity of this thesis, we delve into "The SERP". User interactions play a crucial role in refining the search engine's understanding of relevance and ensuring the

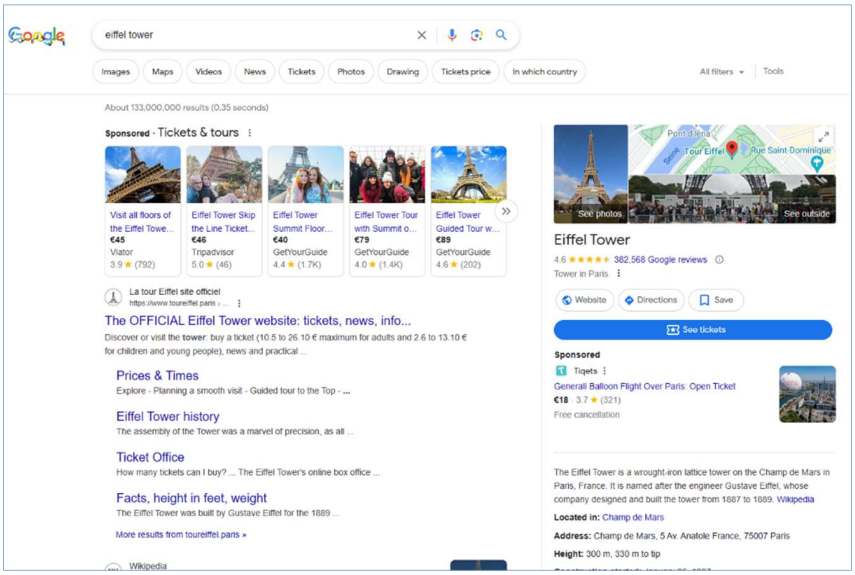
continuous improvement of search result quality. And to achieve and maintain it, SEs continually refine their dynamic algorithms and adapt to evolving user behaviour and information needs. While specific algorithms are proprietary, the overarching goal is to deliver diverse, relevant, and valuable results to users based on their queries. As a result, the factors influencing search rankings can evolve over time. Chart-58 below presents the summary of this section:



Chart-58: Three Stakeholders of SERP

4.3. SERP Interface and Navigation:

SERP Structure and Sections: SERP is a results page of usually 10 snippets of **Organic Results,** 3 to 6 snippets of **Paid** *(Ads)* **Results** and many snippets of **Featured Results** listed in vertical order in one or two columns depending on type and context of a search query on a Search Engine Interface. Refer to pillar-topic 4 of Chapter-1 for an insight of SERP Interface Evolution History. Screenshot-7 below presents an overview of a SERP obtained from Google Search Engine. On the top-left first snippet is of **"Paid-Ads";** 2nd result is an **Organic Webpage;** and on the right side is a Featured Result known as Knowledge Graph.



Screenshot-7: SERP View Dec 2023 | **Source:** Online Google SERP View

According to my intensive lifetime experience spent on SERP Marketing and Optimization and recent-most developments around look and feel of SERP Interface, its Blocks and Navigation Tools can be categorized as per the below Chart-59.

SERP Interface & Navigation				
Standard Organic Results	**Featured Organic Results**		**Layouts & Navigation Tabs & Bars**	
Structured Data Rich Markups	**Zero Click Answer Results**	**On-SERP expandable *(One Click)* Answer Results**	**Universal Search Tabs & Inline Blocks/ Carousels**	**Menus & Buttons**
Deep-links Snippet	Currency Conversion Answer	Paragraph	Images	Text & Voice Search Bar
Extensions Rich Features Snippet	Q&A Block	Bulleted List	Maps	Search Button
Rich Card (mobile)	Recourse Link	Numbered List	Videos	Search Filters
Recipes | Events | FAQs | Movies	Date and Time Answer	Tables	News	Pagination Controls
Sitelinks Search Box	Stocks Index	YouTube Single Video	Books | Scholar	View Options
Product / Service Markup - Start Rating, No. of Reviews, Price, Callouts, Statuaries	Weather Answer	Answer Abstract direct Webpage	Flights	Search Suggestions
	Related Top Brands Snippet	**Media Related Content Results**	Finance	Refinement Tabs
Paid Results	**Related Results**	Top News Stories	Job Sites	Related Searches
Category Ads	Related Search Terms Block	X (Tweeter) Snippet	Places Sites	Tools for Local Search
Text Ads	Similar Topics Block	Related Images Blocks	Products / Shopping	Share and Save Options
Display Ads	Related Products Block	Video Snippets	**Vertical Search Tools**	Advance Features
Shopping Ads	Related Top Brands Result Block	**Mobile Responsiveness**	Any Time Tool	Feedback Options
Video Ads	People Also Ask Results Block	Responsive Design	Any Result Tool	Help and Support
Apps Ads	Related Q&A Block	Voice Search Integration	Safe Search	Personalization Features
GMB Listing Ads	**Knowledge Graphs**	**Local Maps & Itinerary**	Usage Right based Search	
Message Ads	KG - Weather Forecast Snippet	Local Maps Vertical Snippet	**Single Result Layout**	
Position Ads	KG - Task Pane Snippet	Local Maps Horizontal Snippet	Standard Layout	
Top of SERP Ads	KG - Info Pane Snippet	Itinerary Map Results	Schema Extensions	
Bottom of SERP Ads	KG - Knowledge Panel		Featured Layouts	
	KG - Knowledge Card			

Chart-59: SERP Interface Blocks and Navigation Menu

[196]*Jakelabate.com 1* and [197]*Woorank.com 1* present beautiful lists of 54 Rich Result SERP Features, along with [198]*Neil Patel-2,* [199]*SDM-Google.com* and [200]*Madelyn Gardner* [2023] talks of comprehensive understanding of Organic, Paid and Featured results. Keeping these lists in purview and other sources and my lifetime experience, above Chart-59 is designed. Following this chart, we discuss below the various SERP blocks and navigation items:

Standard Organic Results (SORs): The core i.e. main content area of the SERP is occupied by organic search results. These are the web pages deemed most relevant to the user's query based on the search engine's algorithms. Results are often displayed in order of relevance, with the top-ranking pages considered the most pertinent to the user's search query. A single standard organic result snippet is formed of a **Title, URL and Description.** These results can appear on SERP in a block of 10 snippets at one place or as single or grouped in 2 to 10 snippets at 2 or 3 places on SERP. For example, *SAS-1 and SAS-2 in Appendix-1* are the 10 snippets block and individual snippet respectively. With passage of time, advancement of technology and in the zest of serving better user-friendly experience the SORs were enriched with actionable extended features which are usually called as "Rich Features" like site links, reviews, star rating, phone number, location etc. Here to remember that these rich features were added to the root structure of SOR not a totally new feature like "Featured Results", that is a separate category. So based on these rich features following are the Rich SORs:

1. **Sitelinks / Deep-Links Snippet:** An expanded pack of links related to a specific domain. This typically occurs when someone searches for an organization, signifying intent to glean more information or perhaps make a purchase, e.g. Careers or About Us pages below the core domain. Usually, a block of 6 to 8-10 Sitelinks of internal pages of the main site. These links are generated for websites with top influence in their "Topicality" such as *SAS-5.*
2. **Extensions Rich Features:** Additional lines of context based actionable extension such as Ratings, Voting, Prices, Availability, No. of Visitors, Call-out words, Dates, Installs, Size Dimensions etc. For example, refer to *SAS-6.*
3. **Rich Card (for mobile devices):** A cleaner, condensed mobile variation of Rich Snippets in card format.
4. **Sitelinks Search Box:** The Sitelinks Search Box allows users to refine their search within a specific site directly from the SERP. This feature enhances the efficiency of navigation and empower users to explore websites more effectively.

[196] https://www.jakelabate.com/serp-features
[197] https://www.woorank.com/en/edu/seo-guides/what-is-a-googles-serp-feature
[198] https://neilpatel.com/blog/serps-guide/
[199] https://developers.google.com/search/docs/appearance/structured-data/search-gallery
[200] https://www.brafton.com/blog/distribution/a-content-marketers-guide-to-every-2018-google-serp-feature/

5. **Structured Data Rich Results:** These are the rich results, powered by structured data and schema markup, present information in a visually appealing manner. SERPs may showcase visually enhanced rich results, attracting user attention and providing a snapshot of relevant information. For example: Recipes, Reviews, Events and more

Paid-Ads Results (PARs): Paid advertisements are displayed **(Ad Placement)** prominently at the top and bottom of the SERP, labelled as "Ads." These results are often distinguished from organic results through visual cues. Advertisers use **ad extensions** to provide additional information, such as site links, location information, and phone numbers. Text Ads blocks are usually displayed at the top or the bottom of the left column of the SERP. Display Ads can be shown in a horizontal carousel at the top or in a rectangular block in the left-column of SERP. For example, *SAS-3 and SAS-4 in Appendix-1* refer to Text Ads Snippets Block and Shopping Ads Snippets Block respectively. Ads can be categorized based on their type and position (placement) as follows:

Category Ads:

1. **Text Ads:** Simple, keyword-targeted ads consisting of headlines and descriptions shown on search engine results pages.
2. **Display Ads:** Visually rich ads including banners, product carousels, dynamic creatives, and videos shown across websites and apps.
3. **Google Shopping:** A paid search feature that displays product listings with images, prices, and seller info for high-intent shopping queries.
4. **Video Ads:** Promotional content in video format shown on platforms like YouTube or embedded across the web.
5. **Apps Ads:** Ads designed to promote mobile apps through installs, engagement, or re-engagement across Google's networks.
6. **GMB Listing Ads:** Location-based ads that enhance Google My Business profiles with promotional visibility in local search results.
7. **Message Ads:** Click-to-message ads that allow users to initiate direct conversations with businesses via SMS or chat platforms.

Position based Ads:

1. **Top of a SERP Ads:** Through ads based on Quality Scores, several paid ads are featured at the top of SERPs, above organic results. The green "Ad" label informs searchers they're clicking on a paid advertisement, e.g., shopping results, commercial sales, promotional keyword branding.

2. **Bottom of a SERP Ads:** Paid ads featured below organic results are often cheaper, as the four most coveted ad positions (the very top) are more expensive and more competitive. They serve the same purpose as above-the-fold ads but don't contribute to SERP crowding.

Featured Organic Results (FORs): Featured Results quite often used interchangeably with terms Featured Blocks, Featured Snippets or Rich Snippet. These are Organic ones but not the Standard ones and their extendeds. These are designed keeping the searcher behaviour, expectation, and exaggerating 'to the point piece of information' keeping search at mind strategy. They are the natural evolution of the SERP interface in more personalized and signified way to serve the searcher. FORs are with intention of providing the most relevant and direct answer to a search query. As per our synthesis the type of Featured Results varies between 40 to 50 and counting goes on. We don't see all FORs in every SERP, rather their number and type depend on the type of the query and its context. However, these can be categorized as "On SERP Results" and "Off-SERP or Landing Page Results", but contextually usual FORs seen in up and down search can be listed as:

a. **Zero Click Answer Results:** Results with direct answer, instant answer, or concise answer to search query with no need of clicking further are called as the "No Click Answer Results". These Featured Snippets appear at the top of the SERP and provide concise answers to specific queries. These are the short pieces of information that are public facts *(no brands)* and usually doesn't credit a source. It usually carries single word or one line answer to queries asked. For example, "how many meters in a kilometre" or "who is current president of France"? Based on fact type, these results can be of following types:
   1. Currency Conversion Answer
   2. **Instant Answer - Q&A Snippet:** An answer feature addressing queries that may be asking specific questions. This feature has the potential of providing an exact match for the question being asked. It is important to verify that the content is fully accessible without clicking through and that it is relevant to the query; refer to *SAS-8*.
   3. **Recourse Link:** Slight modifications of the original query to account for spelling errors or other mistakes, and a link to return to the original spelling.
   4. Stocks Index
   5. Related Top Brands Snippet
   6. **Instant Answer:** Date & Time Snippet, for example *SAS-7*
   7. **Instant Answer:** Temperature Snippet: Refer to *SAS-9*
   8. **Translation Block:** Refer to *SAS-24*

b. **One Click Answer Results:** Snippets of added context that typically credit a source and is deemed by search engine to be the most contextually relevant result. It is one to 3-4 lines snippet with "click to expand

on SERP itself" feature. Featured Snippets are served in many display formats based on the intent of the search query and the type of information to be featured. These snippets aim to directly address user questions, offering a quick overview without requiring a click-through. Depending on the query, featured snippets may take the form of a paragraphs, lists, tables, or other formats as follows:

1. **Paragraphs:** e.g. FAQs, How tos, Who is, What is
2. **Bulleted lists:** e.g. Rankings, 'best of' lists
3. **Numbered lists:** e.g. Recipes, How tos, DIY Tasks
4. **Tables:** e.g. Prices, Rates, Data
5. **YouTube Single Video:** e.g., a standalone video, a short clip of video or a combination of video and text
6. Answer Abstract directly Website Page

c. **Knowledge Graphs:** It is a tabulated presentation of all key bullet points around an entity searched. Shows up on the right or above organic results, displaying images, information offering details about entities, events or from various sources. They may include images, charts, and maps, enriching the visual experience of the SERP. Search Engines scrapes information from both their data and external sources (e.g. Wikipedia) to feature in Knowledge Graphs usually for searches for people or organizations. These can be of following types:

1. KG - Weather Forecast Snippet: Refer to *SAS-10*
2. **KG - Task Pane Snippet:** Refer to *SAS-11.* It is a compilation of encyclopaedic information about a query provided along the right side of a SERP. The contents and formatting of a task pane change from one topic to another.
3. **KG - Info Pane Snippet:** Refer to *SAS-12.* Nearly identical to Knowledge Graphs, except that search engines pull information only from their databases, for example Google pulls data only from Google Business Profile Listings or Google Maps. It is usually preferred for queries for brands.
4. **Knowledge Card:** Knowledge Cards work similarly to Rich Cards, except these are based on specific data, *Tom J Law* [2019].

d. **Local Maps & Itinerary:** For location-based queries, local packs display a set of business listings relevant to the user's location for example restaurants, hospitals, community establishments, etc. These include details such as addresses, reviews, and contact information. Maps may be integrated into local packs, offering users a visual representation of nearby businesses.

1. **Local Maps Vertical Snippets Result Block:** Refer to *SAS-14*
2. **Local Maps Horizontal Snippets Result Block:** Refer to *SAS-15*
3. **Itinerary Map Result:** Refer to *SAS-16*

e. **Related Results:**

1. **Related Search Terms Block:** Refer to *SAS-17*
2. Similar Topics Block
3. **Related Products Block:** Additional search options that provide the user with information related to the original query. For example, the related searches may appear when a user searches for {Arkansas pharmacy board}. Refer to *SAS-18*
4. **Related Top Brands Result Block:** Refer to *SAS-19*
5. **People Also Ask Results Block:** Additional questions natively displayed in results that are closely related to the original query. For example, refer to *SAS-20*
6. **Related Q&A:** A selection of other queried questions that are similar to the current results. These may add good content to the page, but an irrelevant Q&A doesn't necessarily detract from the quality of the result.

f. **Media Results:** SERPs may include image and video carousels, displaying relevant media content directly on the results page. Users can engage with visual content without navigating to external pages, providing a more immersive experience.

1. **Top News Stories Snippet:** Refer to *SAS-13*
2. **Videos Snippets Result:** Refer to *SAS-21*
3. **Related Images Result Block:** Queries that specifically refer to photos or ones that strongly necessitate relevant imagery (such as storefronts or homes) often return an Image Pack, a horizontal carousel of images, e.g., queries best answered in photo format. Refer to *SAS-22*
4. **Tweets Snippets Results Block:** Refer to *SAS-23*

j. **Vertical Search Filters Menu:** Users may have the option to filter results by date, location, maps, video, news, images, books (Google Scholar), flights, finance, job sites or other criteria, providing a more tailored and relevant set of search results. Searchers may have the option to filter the results using 'Any time', 'Any result' or 'safe search' tabs. As per *Long & Chang* [2014] these auxiliary services, known as verticals, include search engines that focus on a particular domain *(e.g. news, travel, sports),* search engines that focus on a particular type of media *(e.g. images, video, audio),* and APIs to highly targeted information *(e.g. weather forecasts, map directions, or stock prices).* The goal of aggregated search is to provide integrated access to all verticals within a single information context.

k. **Mobile Responsiveness:**

1. **Responsive Design:** SERPs are designed to be responsive, adapting to the screen size and orientation of mobile devices. Mobile-friendly features, such as swipe-able carousels and easy-to-tap buttons, enhance the mobile user experience.

2. **Voice Search Integration:** On mobile devices, voice search is integrated, allowing users to speak their queries for a hands-free search experience.

Single Result Layout and its Ingredients: A single SERP result, whether organic or paid, is composed of several key ingredients, and follows a specific layout. Let's explore the elements that make up a typical single SERP result:

1. **Standard Layout:** Title/Headline, Meta Description, Landing Page URL/Short site links
2. **Schema Extensions:** Deep Links, Rating Bar, Phone Number, Call Outs, Call to action, Author Photo, Site Search Link, Reviews, Publication Date, Event details.
3. **Featured Layouts:** Paragraph, Lists, Tables, Knowledge Graphs *(entities, topics, events etc)*, Local Packs, Images, Videos, Social Media Integrations *(X-Twitter Snippets)*, Swipe-able carousels, Mobile-Responsive Designs.

Search Engine Buttons and Navigations: Search engines typically provide a set of buttons and navigational elements to facilitate user interaction and enhance the search experience. These buttons and navigations are designed to help users refine their queries, access additional information, and navigate between different sections of the search engine interface. Here are common search engine buttons and navigational elements:

1. **Text, Image, and Voice Search Button:** The query input field with microphone icon *(Image and Voice Search)* also know as the central element where searchers enter their search queries. Voice Search allows users to perform searches using voice commands and so with Image Search *(Camera icon)* for Image commands.
2. **Search Bar** *(Execute Search)***:** Writing in this bar and clicking the magnifying lens icon initiates the search based on the query entered in the search bar.
3. **Search Filters:** It offers additional search tools, such as filters for time range, type of content, and usage rights for images and allows searchers to customize search settings, preferences, and advanced search options.
4. **Pagination Controls:** It is located at the bottom of a SERP. Using it searchers can navigate through multiple pages of search results using page numbers and can move backward or forward through the pages of search results.
5. **View Options** *(Grid View and List View)*: This option enables users to switch between grid and list views for search results. Thumbnail Size Controls allow searchers to adjust the size of thumbnail images in the search results.

6. **Search Suggestions:** There are two types of suggestions Dropdown List and Autocomplete are the two types of suggestions, which help searchers to fine-tune their search results. Dropdown List displays suggested queries as users type in the search bar. Autocomplete, automatically completes the query based on popular or relevant searches.
7. **Refinement Tabs:** All button displays all types of content in the search results which gives direct access to specialized search services beyond general search like Images, Videos, News, Maps tabs help in refining search results based on content type.
8. **Related Searches** *(Suggested Queries)*: Offers additional search queries related to the user's original search, aiding in query refinement.
9. **Tools for Local Search:**
    i) **Map View:** Allows users to view search results on a map.
    ii) **Nearby or Location Filters:** Enables local search based on the user's current location or a specified location.
10. **Share and Save Options:** It allows users to share search results through various channels, including email and social media. Save Button enables users to save or bookmark search results for later reference.
11. **Feedback Options:** Feedback Button allows users to provide feedback on the relevance and quality of search results.
12. **Help and Support:** These links provide access to documentation, FAQs, and support resources to assist searchers.
13. **Personalization Features:**
    i) **Sign-in Button:** Allows users to sign into their accounts for personalized search experiences.
    ii) **User Profile Icon:** Access to personalized settings, history, and preferences.

These buttons and navigational elements contribute to a user-friendly search interface, allowing users to tailor their searches, explore different types of content, and access additional features provided by the search engine. They play a crucial role in ensuring a seamless and efficient search experience for users across various devices and platforms.

4.5. Web Search Intelligence:

Search Engine's Search-As-Learning: As talked in the Crawling and Indexing section 4.2 above, every Search Engine hire paid-consultants to annotate their indexed content. This process in technical terms is known as **Search Engine Machine Learning, Search Engine's Automotive (Self) Learning or Search-As-**

Learning (SAL). This concept is based on the philosophy of, 'Searching and learning are co-existing processes. Through SAL SEs assess and learn different cognitive complexities in search patterns and explore information search environments. *Batchakui et al.* [2022] define it as a goal to improve the learning aspect within search engines by associating a layer above traditional search engines to facilitate **content appropriation** during search tasks for a learning context. They further add that search engines apply "Search-as-Learning" by improving **selection algorithms** and the Human Machine Interface ***(HMI)*** to enhance learning during user search tasks. The SAL process involves **the learner, the interface, the information retrieval backend, and the web resources**. The current state of the art in this field is reviewed, and areas of future research are outlined to improve our understanding of search as learning processes, *Hoyer et al.* [2022]. According to them, the field of Search-As-Learning explores information search environments to promote learning during the search process. They present a descriptive backend model of SAL as shown in Chart-60, Appendix-9. Several approaches have been proposed to improve search engines for the learning scenario, such as specific ranking functions, *Pertti Vakkari* [2016]. However, instructional scaffolding, a traditional learning support strategy, has not been studied in the context of search as learning, *Camara et al.* [2021]. Scaffolding techniques within a search system can be effective in helping learners achieve their learning objectives while searching, *Ghosh et al.* [2018]. The incorporation of **scaffolding methods**, such as automatic query expansion, curated subtopics, and real-time feedback, can impact search behaviour and potentially improve learning outcomes. In this row, *Katja Hofmann* [2011] has focused on the development of self-learning search engines that can adapt to user preferences, leading to improved search performance. *Bird & Curran* [2006] has explored the use of search engine projects to drive problem-based learning in computer science courses, enhancing student engagement and analytical skills. [201]*Taibi & Gentile* [2005] has proposed the use of **semantic web technologies** to enhance the search for learning resources, addressing the limitations of keyword-based searches. These studies collectively underscore the potential of search engines to facilitate learning and the need for ongoing research and development in this area. *S. Liaw* [2004] by applying the **theory of planned behaviour** states that the search engines are increasingly being used as learning tools, with factors such as **perceived satisfaction, information retrieval, and sharing experiences influencing** their use. Search-as-Learning as per [202]*Sandra Niemeyer* [2021] involves investigating the conditions necessary for search to facilitate learning, and it aims to make search engines more effective for learning-oriented search tasks. Contrary to above findings here is something underestimating the capacity of Search Engines self-learning, *Collins-Thompson et al.* [2017] noted that traditional web search engines are not inherently designed for complex search tasks that require exploration and learning, despite a significant portion of web searches being

[201] https://www.academia.edu/26890871/A_semantic_search_engine_for_learning_resources
[202] https://blogs.tib.eu/wp/tib/2021/07/21/how-do-we-use-search-engines-for-learning/

complex. So, while there is a growing recognition of the importance of studying and designing search systems to support learning, the integration of SAL into search engines is still an evolving area for research and development. To fulfil this dependency, the ML system of SERP involves the use of advanced algorithms and models that enable them to learn and adapt based on user behaviour, feedback, and evolving patterns in web. Here are key ML methods designed for SEs to learn and evolve automatically:

1. **User Intent Pattern Recognition:** MLAs analyse user queries and interactions to recognize patterns and understand the intent behind different types of searches.
2. **Query Understanding using NLP:** Search engines use NLP to interpret the nuances of user queries, including synonyms, context, and user-specific language.
3. **Relevance Ranking through Click-Through Rates (CTR):** Algorithms learn from user clicks on search results, adjusting the ranking of pages based on the perceived relevance to user queries.
4. **Content Quality Evaluation:** ML models assess the quality of content on web pages, considering factors like relevance, authority, and user engagement.
5. **Personalization based on User History and Preferences:** SEs learn from a user's search history and preferences to personalize future search results, providing more tailored and relevant content.
6. **User Behaviour Analysis based on Dwell Time and Bounce Rate:** Algorithms analyse user behaviour metrics such as dwell time *(time spent on a page)* and bounce rate *(percentage of users who leave a site from landing page)* to gauge content quality and relevance.
7. **Semantic Understanding Context and Meaning:** ML systems enable SEs to understand the context and meaning behind words, improving semantic search capabilities.
8. **Adapting to Trends through Real-Time Learning:** SEs adapt to emerging trends and changes in user behaviour in real-time, continuously updating their algorithms to reflect the most current search patterns.
9. **Local Relevance through Location-Based Learning:** Algorithms consider a user's geographic location, learning to deliver more relevant local results for location-specific queries.
10. **Featured Snippets Optimization learning from Snippet Engagement:** Algorithms learn which types of content are most effective in featured snippets by analysing user engagement with quick-answer formats.
11. **Spam Detection by identifying Low-Quality Content:** ML helps in detecting and penalizing low-quality or spammy content, improving the overall quality of search results.
12. **Algorithmic Updates through Feedback Incorporation:** SEs learn from user feedback and data on search result performance to make algorithmic updates that enhance the overall search experience.
13. **Multimedia Integration by learning from User Preferences:** Algorithms learn to integrate various content types, including text, images, videos, and more, based on user preferences and engagement.

14. **Voice Search Optimization learning using Natural Language:** As voice search becomes more prevalent, ML systems adapt to understand and respond to natural language queries spoken by users.
15. **Continuous Improvement through Iterative Learning:** SEs follow an iterative learning process, where they continuously refine and improve their algorithms based on new data and evolving user needs.

Defining Search Persona from Searcher' Search Pattern: Knowing about what is understood from Web Searcher before feeding to Search Machines by Search Engines helps in defining the **Search Persona**. This is usually imitated through following methods:

i) **Searcher Derivatives:** Although, we have talked about this topic under our thesis variable related to Behaviour in chapter one. But here we intend to present some specific aspects to the general perceiving of the searcher. As per [203]*Adam Hayes* [2024] "Behavioural modelling attempts to explain why an individual makes a decision and the model is then used to help predict future behaviour". As it all depends on the pursuance type of a searcher that what behaviour they will present in their search so in our thesis, we name it as "the psychology under scrutiny". As the word pursuance defines, it is about the interaction between two objects. Here our two objects are the searcher and the search engine. Interestingly and challengingly when we classify this interaction on "Behavioural Model by *Dubey & Jain* [2017], it fits in covert volunteer behaviour class. And as it is covered and invisible expression of behaviour so very typical in predicting and studying. As defined in Chart-61, Appendix-9 model same paper by *Dubey and Jain* hints towards the factors those drive the online searcher behaviour. I find these factors most close and realistic to this study as follows:
   a) Social rules and regulations
   b) Language and understanding
   c) Education and knowledge
   d) Adaptability
   e) Capacity to learn knowledge.
   f) Drive or aim of accomplishing something.

ii) **On-Page Searcher Behaviour Signals:** According to [204]*Dave Davies 1* [2020] of SearchEngineJournal, Google receives the relevancy signals through clicks and post-clicks as below:
   a) SERPs > Click > Quickly Returns > Clicked the next result - Negative relevancy signal to the visited page.

[203] https://www.investopedia.com/terms/b/behavioral-modeling.asp
[204] https://www.searchenginejournal.com/search-engines/user-behavior/

b) SERPs > Click > Returns after time > Clicked the next result - Positive relevancy signal as the user found content of interest and is simply looking for additional information or option.

c) SERPs > Click > Returns after long or short time > refine their query > Search Again - User themselves was not specific as they needed, no good or bad relevancy signal to visited page.

d) SERPs > Click > Returns after long or short time > completely change their query - They found what they were looking for and moved on to a different task. Positive relevancy signal for visited page.

e) Reviews relate to local search provides prominent user behaviour signal.

iii) **Types of User Behaviour Signals that Search Engines don't or hardly use:**

a) Google Analytics installed or not for case of Google Search Engine

b) Social Media Likes, Shares and Comments

iv) **Visual Attention Theory of Searcher:** Taking this signalization one step deeper, it is usually driven by the **visual attention theory**. It gives way to Search Engines to present their SERPs in a List Format and not in matrix view; let's understand it further from some expert views. *Djamasbi and Hall-Phillips* [2014] write in their book "**Eye Tracking** in User Experience Design" about visual search behaviour. They categorize it into goal-directed and exploratory search behaviour, each using a different part of our brain. In **goal-directed search**, we actively search for information following a specific strategy or plan. In **exploratory search**, we just monitor the environment, and we usually do not have a search plan in mind. *Posner & Petersen* [1990] *and Janiszewski* [1998] add further to it that in Exploratory search is our default behaviour because whenever we are not actively involved in looking up information, our visual system keeps screening the environment. In both types of visual search behaviour, our fixations are influenced by how objects are laid out on a web page. Objects on a web page compete to win our attention. The intensity of this competition can be numerically calculated as a function of the number of objects on the viewable screen, their size, and their proximity to each other. The competition for **attention theory** *Janiszewski* [1998] tells us that the race for attention between the objects in our field of view affects how we screen the environment *(exploratory search)* and how we actively search for a piece of information (goal-directed search). Therefore, when we examine search behaviour, we should carefully look at how objects compete for attention *Janiszewski* [1998]. So, the whole game revolves around mapping the searchers goals to what is provided to them in a SERP corresponding to their query. The goal of an online searcher can be as a general visitor, a buyer, a seller, a voter, a researcher etc. depending on at which stage of a buying cycle the searcher is. *Lacoste-Badie et al.* [2013] examined whether the increase in many of information units present on the facing could have a deleterious effect on the attention paid to the brand. Their results show that the attention paid to the brand, measured in terms of number of fixations or fixation duration, is not altered when the brand is surrounded by many information; rather the presence of information additional

preserves the brand while retaining consumer attention. *Lacoste-Badie et al.* [2019] in their Eye-Tracking testing gave a 12-min extract of animated comedy film with 2 commercial breaks in it. In another method, they exposed their audience to eight ads. In both methods, they measured the attention paid to food health warnings and found that only 18% of children audience made more than one eye fixation on health warnings displayed. In another impactful study *Lacoste-Badie et al.* [2018] asked the participants to view a new, on-screen magazine and that their task was to evaluate it. Their study found that 'face presence and gaze directed toward the product' versus 'no face and gaze toward the viewer' have a strong influence on **attention** *(fixation duration, fixation count, and revisit count)* and **memorization** of advertisements, on advertisement and brand evaluation, and on purchase intention. Here in context of SERP interesting would be to observe the user eye movement behaviour on visual content versus text-only content. In her another study regarding visuals in an ad, *Lacoste-Badie* [2014] found that the face present in an advertisement benefit from a privileged attention, although no influence of positive emotional tone of the face (vs. neutral) is observed. In another study related to investigating the influence of front-of-pack (FOP) variations on consumer attention, *Lacoste-Badie et al.* [2020] found that front-of-pack variations catch respondents' attention. There again in their next research on role of visual packaging in influencing the consumer behaviour in stores *Lacoste-Badie et al.* [2021] focused on front of pack (FOP) formats present in images of product in motion i.e. implied motion depicted in static works of art (e. g. the direction the eye takes along an invisible path created by a line, arrow, or pointing finger such as bouncing ball image) found using eye-tracking method that implied motion significantly increases visual attention, consequently generates more frequent choices and implied motion enhances freshness, which translates into greater tastiness, product attractiveness and, ultimately, purchase intention. *Lacoste-Badie et al.* [2022] conducted eye-tracking research about how attention grabbing are the PWLs (Pregnancy Warning Lables) on Wine bottles in France. The results show that almost none of the participants paid attention to the current French PWL. Their research indicates that a larger colourful PWL, with a combined text and pictogram, attracts far more attention, and that participants chose the bottles of wine displaying this type of PWL less frequently. Here in the context of this thesis, this can be interesting to interpret these findings in context of visual content or content blocks in SERPs and how Search Users react to it through their eye-movement and attention falling.

Web Content Annotation by Annotators for Search Engine: Annotation of Search Engine Index through crowdsourcing by Crowdsourced Web Content Annotators enhances the scalability, diversity of perspectives, cost-effectiveness, real-time feedback and adaption for a search engine. This qualitative self-assessment of its Indexed web content by SEs is called as Search Engine Machine Learning **(SEML).** *Chen et al.* [2023] propose

a method to improve learning from crowds by modelling annotator expertise and instance difficulty to predict annotation reliability. *Li et al.* [2023] address the challenge of modelling annotator- and instance-dependent noise transition matrices in crowd-sourcing scenarios and propose a deep neural network-based approach for estimating general noise transition matrices. *He et al.* [2023] proposes large language models, such as **GPT-3.5**, which can serve as excellent crowdsourced annotators by providing them with guidance and demonstrated examples. A two-step approach called **'explain-then-annotate'** is proposed to make LLMs better annotators. First, prompts are created for every demonstrated example to prompt the LLM to provide an explanation for the chosen ground truth answer/label. Then, a few-shot chain-of-thought prompt is constructed using the self-generated explanation to annotate the unlabelled data. Experiments on user input and keyword relevance assessment, **BoolQ, and WiC tasks** show that the annotation results from GPT-3.5 surpass those from crowdsourced annotation for user input and keyword relevance assessment and achieve comparable results for the other two tasks. *Chen et al.* [2013] and *Mendez et al.* [2019] both propose models that incorporate the quality of annotators' contributions in crowdsourcing, with Chen focusing on **pairwise ranking aggregation** and Méndez on a **machine-crowd-expert loop.** *Fang et al.* [2014] introduces a probabilistic model for active learning in crowdsourcing, which transfers knowledge from auxiliary domains to estimate annotators' expertise. *Zhang et al.* [2021] takes a different approach, treating all crowdsourced annotations as gold-standard and applying domain adaptation methods to improve performance. These studies collectively highlight the importance of considering annotator quality and engagement in crowdsourcing, and the potential for knowledge transfer and domain adaptation to enhance the learning process. A **Bayesian framework** has been proposed to enable deep learning models to actively learn from targeted crowds, leveraging the low-rank structure in annotations to infer true labels and reduce the number of required annotations, *Yang et al.* [2018]. Quality control mechanisms, including the use of behavioural features, have been developed to predict annotation quality in crowdsourcing systems, outperforming performance-based methods, *Han et al.* [2016]. Crowdsourcing can be used to acquire large, labelled datasets, although it is important to estimate the reliability of crowd annotators and aggregate individual annotations, [205]*University of Texas* [2017]. This approach involves combining input from multiple annotators to determine the best overall results, enhancing the search engine's ability to understand and process information. Additionally, self-learning AI in e-commerce can extract meaning from product catalogues and customer search queries to match user intent and rank products and services appropriately, thereby improving the search experience, [206]*Kait Spong* [2023]. Kait further adds that it can extract meaning from product catalogues and customer search queries to match user intent and rank products and services appropriately, thereby improving the search experience. Crowdsourcing annotator models

[205] https://phys.org/news/2017-08-future-combine-artificial-intelligence-crowdsourcing.html
[206] https://www.bloomreach.com/en/blog/2022/why-self-learning-ai-is-a-game-changer-in-e-commerce

are often utilized by search engines as part of their self-learning mechanisms. Search Engines do rigorous studies on Searchers Human Behaviour to be proactive in serving its searchers on Customer-Client relation base. Thousands of annotators workday and night in assessing the searcher behaviour and expectations from Search Engines. This job started somewhere in 2010s when the search engines defined dedicated projects regarding it. These models involve gathering labelled data or annotations from a crowd of human annotators to train and improve machine learning algorithms. Here is some common **crowdsourcing annotator models** used in the context of search engines for self-learning:

1. **Relevance Annotation:** Crowdsourcing is employed to gather relevance annotations for search results. Human annotators assess the relevance of search results to specific queries, providing feedback that helps train machine learning models for better result ranking. The relevance of a SERP obtained in return of a query refers to how likely it is to satisfy the query's intent. Content satisfies the query's intent when it provides information that most users (>80%) are attempting to find. A SERP's worth of relevant web results trumps a page of well rendered, but irrelevant, features and content.
2. **Query Intent Annotation:** Annotators help identify and categorize the intent behind user queries. This information is valuable for understanding user needs and improving the search engine's ability to deliver relevant results. Quality annotation includes criteria like Content Lexicality, Topic Depth and Usage Scope.
3. **Named Entity Recognition (NER):** Crowdsourcing annotators may be used to identify, and label named entities in search queries and documents. This aids in improving the search engine's understanding of entities and enhances semantic search capabilities.
4. **Click-Through Rate (CTR) Annotation:** Annotators assess the likelihood of a user clicking on specific search results. This information is used to train models for predicting CTR, enabling search engines to optimize result rankings based on user preferences.
5. **Query Rewriting Annotation:** Crowdsourced annotations help in generating alternative query formulations or rewriting queries. This contributes to the improvement of natural language understanding models and enhances the search engine's ability to handle variations in user language.
6. **User Behavior Annotation:** Annotators assess user behavior data, including clicks, dwell time, and other interactions with search results. This information is crucial for training reinforcement learning models that improve the overall user experience.
7. **Sentiment Analysis Annotation:** Crowdsourcing annotators may assess the sentiment expressed in search queries or document snippets. This sentiment information can be used to improve the understanding of user preferences and intent.

8. **Content Moderation Annotation:** Annotators help identify and flag inappropriate or low-quality content in search results. This contributes to content moderation efforts and helps maintain a high-quality search experience.
9. **Multimodal Annotation:** In cases involving images or videos in search results, annotators may provide labels, tags, or captions to enhance the understanding of visual content. This supports the development of algorithms for visual search.
10. **Spam Detection Annotation:** Annotators identify and label spam or low-quality content in search results. This information is crucial for training models that can automatically detect and filter out undesirable content.
13. **Authority Annotation**: Some SERP results come from a questionable source or domain, possibly affecting the quality of the overall results. Here it becomes necessary to verify if the landing page' information has been peer-reviewed or provided by a credible source or not. Is that result a reputable site? At the same time, searcher's expectations are essential to understand for a search engine. It is not necessary that highly credible or official landing pages always best satisfy the intent. Depending on the query, even content that is gossipy, wrong, or offensive may be relevant to the user. For example, searchers looking for {political satire} or {Hollywood rumours} may not care much about the truthfulness of the content or the trustworthiness of a domain.
14. **Freshness Annotation**: Time context can be important for many newsworthy, product-related, or event queries. It is decided on a seasonal scale. For example, the query {November election} entered in December should return information and analysis about an event in the past, while the same query entered in October should return information and analysis about a future event. Search Engines consider whether a SERP has more up-to-the-minute, breaking-news web results in users' deliberations of page relevance. With passage of time 2010 onward the more and more parameters were added to feed search machines with deeper understanding on how to yield the results in response of a searcher query considering it a Topic. This was the time when the concept of **"Topicality"** was getting popular among Search Engines to prepare their Index Data by "Topic". These factors evolved up as:
15. **Scope Assessment:** Relates to the broadness or narrowness of LP (Landing Page) in relation to the specificity of the query intent. It gazes the width and narrowness of a SERP source on the topic queried and as a result the essence of "**Topicality**" went on developing in priority of Search Engines for aggregating the better SERPs.
16. **Utility:** Refers to the completeness, usefulness, and thoroughness of the content under a LP of a SERP.
17. **Diversity:** This factor is usually weighed for the information intent-based queries as such queries are often seen as multi-intent queries. With further essence of giving specificity to SERPs Search Engines started

focusing on where the searcher is sitting, which language they speak, what are their market trends; like **"Demography"** of an entity. It was seen as the entry of term **"Personalization of SERP";** thus Market, Language and Location were taught to the Search Machines.

18. **Market:** Queries where a user location is explicitly provided or implicitly implied require results from the appropriate market. If one side is more useful than the other for current market or if there are any issues related to market choose this hashtag
19. **Language:** If there are language issues, whether it's the wrong language provided in relation to market or unintelligible language choose this hashtag.
20. **Location:** Location is the most sensitive consideration for Search Engines, for some searcher queries, location provides important context. It is decided based on two considerations:
    a) For a query with an explicit user location i.e. location explicitly stated in the query for example {Starbucks, Paris} query, SERP results must be of Starbucks locations in Paris on top or the near next the maximum. Satisfactory results should be locally relevant (the distance between the user location and the results on the page should be minimal) as well as useful for navigating to the actual physical location (the contact information – address, phone numbers, maps, and directions – should be clear and correct).
    b) **Location mentioned directly in the query**: Location of the user who issued the query (or search engines guess about user location). For example {pizza delivery in Paris}. Though Searcher location is not always relevant for example query {Facebook} {[Perkasie, Pennsylvania] Facebook}
    c) **Discussion:** For some queries, only one is important, and for some, only second is important. While for others, both are important. It all depends on the query type, its words length and sense what it makes. With the increasing diversity in types of devices used for search and to match the understanding and expectation of a searcher, Search Engines then picked up on:
21. **Mobile Friendliness:** Relates to how mobile friendly the SERPs are.
22. **Sitelinkes/Deeplinks:** These are the deep links often appear in first organic result especially in case of navigational intents. But still the topic **relevance and content quality** hold the master key for every machine learning program.
23. **Redundant/duplicative results:** When looking at a complete page of web results, user may see a number of results that look similar, such as:
    a) www.youtube.com
    b) www.youtube.com/index
    c) www.youtube.com//?noredirect=1

    All these URLs link the user to the same page. Consistency of the results or same domain results is desirable, but lack of variety within the results like above example may be undesirable for the user. So

search engine takes care of such risks minutely. *Arora & Govilkar* [2016] state that the inference and analysis of user search goals can have a lot of benefits in improving search engine relevance and user experience which can be achieved by using Content based ranking algorithm, Usage based ranking algorithm and Link based ranking algorithm in unison. A combination of these algorithms enables to filter out the redundant results and persistency problem in SERPs. **Redundancy of results removal can be a greater relief to HISs especially those run searches on urgent basis or on go.**

24. **Ambiguity / Lack of Expected Results:** It often happen in cases of multi-intent queries, where it is difficult to assess that the searcher would be more interest in which type of results and at what rank in SERP. More does not always mean "better" i.e. if any feature does not relate directly to the most likely of possible query intents, then it is bad for the user. A balance of results and context is imperative to be effective.
25. **UI/UX Richness in SERP:** It adds up in the searcher loyalty. A better UI/UX is similar way important for a user as in case of general website, blog or app. They are pleased with beautiful content blocks, well placed and ordered in clear and easy lexical tune.
26. **Contextual Richness:** Richness in terms of contextual interaction and experience is seen as the enhancement of a SERP's relevance via superior page layout combined with supplemental features. Consider this query {directions from Lille to Paris}. One SERP renders a task pane containing pictures and geographical information about Lille or Paris, whereas the other SERP offers several web results (but no task pane or instant answer) yielding driving directions from Lille to Paris. The SERP with the task pane would have richer but irrelevant content in this query context while the SERP with the map results would be more relevant if less rich. The SERP without the task pane but with the driving directions would be the optimal choice.
27. **Ranking Position on SERP:** Results that display at the top carry more weight than the results at the bottom of the page. This means that a poor result displayed at the top of the page impacts the quality of the hit more than a poor result at the bottom of the page. And this makes the "Organic SEO Listing" a hard nut to crack for global to small level SEO experts and still despite we entered in Paid Listing arena why organic listing is still quite important.

Any of the crowdsourcing annotator models above play a significant role in the iterative improvement of search engines by providing valuable labelled data for training and refining machine learning algorithms. The diversity of human perspectives contributes to a more comprehensive understanding of user intent and content relevance.

Search Engines those leverage Self-Learning ML: Several popular search engines leverage machine learning algorithms to learn and adapt to user behaviour, evolving content, and changing search patterns. Here are some examples:

1. **Google:** Google is at the forefront of using machine learning in search. Google's algorithms, including **RankBrain,** are designed to understand user intent, interpret queries, and deliver more relevant results. The big examples are of **Hummingbird and BERT.** These algorithms not only produce SERPs but accumulate the Keywords and categorize them by deciphering each Search Term on many criteria as talked in just above sections. Google Search utilizes machine learning to continuously refine its understanding of context and user preferences.
2. **Bing:** Microsoft's Bing search engine also employs machine learning algorithms to improve search results. Bing uses AI technologies to understand user queries, provide personalized results, and enhance the overall search experience. Bing runs an annotator program named as "**Universal Human Relevance System**"; through which they get their content annotated through crowdsourcing.
3. **Yahoo Search:** Yahoo Search, which is powered by Bing's technology, benefits from machine learning advancements. The underlying algorithms aim to understand user queries, identify intent, and present results that align with user expectations.
4. **Yandex:** Yandex, a popular search engine in Russia, employs machine learning to enhance search relevance. Yandex uses algorithms that consider user behaviour, context, and language nuances to improve the accuracy of search results.
5. **Baidu:** Baidu, the leading search engine in China, integrates machine learning into its search algorithms. Baidu uses AI to understand user intent, personalize results, and adapt to the dynamic nature of the web.
6. **DuckDuckGo:** While DuckDuckGo emphasizes user privacy and does not personalize results based on past user behaviour, it still incorporates machine learning for features like "Instant Answers." Machine learning helps DuckDuckGo provide quick, relevant information for certain queries.

Web Content Annotation Methods: We came to study about this topic in "Off-Search-Request Working" section of "Search Producers" above. We understood there that in what context the SERP Annotation is used. In this section we intend to investigate it in contextual details. The annotation process of search engine results involves the use of explicit and implicit feedback to improve the quality of search results. Based on User Behaviour, there are two main types of annotations: **explicit annotations**, which involve collecting direct user feedback on the relevance of search results, and **implicit annotations**, which are automatically extracted from **search engine query logs**. Explicit annotations can be obtained through **user input**, while implicit annotations are derived from **user behaviour**, such as the queries they submit and the pages they click on. These

annotations are used to enhance the quality of search results and can be integrated into search engines to improve the relevance of the displayed content. Additionally, annotations are also used to define the coverage of search engines and provide greater control over the ranking of search results. The process of annotation is essential for customizing search results based on user preferences and improving the quality, speed, and security of search engines and other computer systems. This helps users avoid clicking on search results where their search term elements are not present, saving time and improving the search process, *Osann-1* [2018] and Osann-3 [2020]. Based on SERP Presented, there are two types of annotations: **annotations indicating missing search term elements** and **annotations indicating endorsed results by the user** or their social network connections, *Liu et al.* [2013]. The impact of search engine results quality annotation on SERP quality is that it improves the relevance and accuracy of search results by providing users with information about the presence or absence of their search term elements on referenced and descendant webpages, *Osann-2* [2018] and *Kim et al.* [2011]. Additionally, annotations based on user endorsements and social network connections help users identify personally relevant results. The annotation process of search engine results involves the **extraction of data units and the assignment of meaningful labels**. This process is crucial for making search results machine processable and is typically done manually, leading to scalability issues, *Liu et al.* [2020] and *Sriramoju* [2014]. To address this, automatic annotation techniques have been proposed, which involve aligning data units into groups and annotating them from different aspects to predict a final label, *Y. P. Jeba* [2013] and *Lu et al.* [2013]. As per Liu and Sriramoju, these techniques have been found to be effective in improving the accuracy and scalability of web search engines. A range of annotation methods have been proposed for improving the relevance and quality of search engine results, *Shao et al.* [2019]. *Rui et al.* [2007] introduced **a search-based method for web image annotation**, using visual and textual consistency measures. Almost every search engine does this activity of **quality control of their information index**. But here we shall be talking about two Search Engines Annotations methods and approaches, which are **Google and Bing** alongside mentioning of Amazon Mechanical Turk ***(MTurk)***. While Google and Bing are public search engines covering all domains; Amazon one is particularly of e-commerce domain. The program Google runs for this purpose in reference to [207]*Google-GSQEG (Google Search Quality Evaluator Guidelines)* is conducted under "Google Search Quality Rater ***(GSQR)***" and Bing runs it under title "Universal Human Relevance Systems ***(UHRS)***". These mechanisms are **Crowdsourcing** based and several thousands of annotators around the globe contribute to this task on paid jobs. Coming to the context of this thesis and these methods, we list out various big and small projects run under these programs. These programs are one of the major bases of Search Engine' "**Search Engine Machine Learning through Crowdsourcing**" ventures. Here to clear that both programs of Google

[207] https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

and Bing use each other's SERP screenshots to compare both and ask annotators to annotate those results through below mentioned methods i.e. projects:

1. **Assessing the User Intent:** This method is used to assess the type of intent behind a query. Types of criteria mentioned in Table-17 below are used in it. These criteria give the idea about what kind of intent that query can be having.

Keywords "modifier" that indicate an intent type			
Informational	**Navigational**	**Transactional**	**Commercial**
How What Who Where Why Guide Tutorial Resource Ideas Tips Learn Examples	Brand Names A product A service From-To	Buy Coupon Order A local store Purchase Cheap Price Pricing Register Download Chat	Best Top Review Product attribute Comparison
Which SERP feature usually means for which Intent type			
Featured Snippet Knowledge Card Videos People also ask	Site Links Tweet Box Knowledge Panel Map Block Image Carousel Video Block Image Block	Ads Shopping results	Ads Featured Snippet
What Searchers Want			
Deep Information with Keywords	Official website *(most likely the homepage)*	Buy page or product catalogue	To make a buying decision
Why pages rank highly			

Text-rich pages	Keywords in title and URL	Keywords in title and text	Trends and reviews
Why Searchers Click			
Keywords in title and snippet with trusted source	Keywords in title and snippet	Keywords in title and snippet with a great offer	Reviews and Social Signals

Table-17: What "Keyword" and "a SERP Feature" indicate "Which" intent?

2. **Identifying Intent Numbers or Combinations:** Search Engines use this method to understand the number or combinations of possible intents behind the searcher query. Table-18 matrix below is used in this method, this shows how the intents are gazed on their likeliness.

Intent Likeliness	Intent Weight	Percentage of public would be expecting it	How searcher will see it	Satisfaction of the intent	Additional Note
Most Likely	Dominant Intent	Above 50%	Perfect	Thoroughly	Usually the official website or webpage
Very Likely	Major Intent	Between 26 to 50%	Excellent	Very well	Other major intents may exist
Likely	Minor Intent	Between 11 to 25%	Good	Moderately	One of several minor intents
Unlikely	A rare intent	Between 1 to 10%	Fair	Weakly	
Very Unlikely	Not an intent or very unrealistic intent	Below 1%	Bad	Poorly	

Table-18: Intent Likeliness versus Satisfaction behind a searcher query

Chart-62, Appendix-9 demonstrates the intent-mapping graphically with more enticing visuality and understanding.

3. **Identifying the Intent Likelihood:** This method is used to gauze the probable weight of intent behind a query. Second and third column of Table-18 maps this probability.
4. **Measuring the overall value of the SERP:**

Parameter	Description
Intent	What type of intent is covered by either side of SERP
Completeness	Are there any Exactly Relevant and Completely Irrelevant to what user is looking for
Positioning	Differences in terms of the ranking order of the results & same placed results on either SERP
Number of Landing Pages	Low or High Number of SERPs on either side
Redundancy	Any Redundancy (duplicate) in results on either side
Captions	Do the results provide snippets or captions which give insight to authority, freshness, relevance, etc.?
Subpages	Does either SERP have a relevant and helpful subpage block underneath the top result?

Table-19: Value of a SERP against a search query

5. **Mapping the Keywords Type to Website Goal**: Below Table-20 demonstrates about possible pairing of website goal and keyword types on three major Intent Types as mentioned in it.

	Priorities for Keyword Types			Search Marketing Approaches
Website Goal	Navigational	Informational	Transactional	
Web Sales	Medium	Medium	High	Organic and Paid
Offline Sales	"	"	High	"
Leads	"	High	Low	"
Market Awareness	High	Low	Low	Emphasis on paid
Information and entertainment	Low	High	Low	Emphasis on organic
Persuasion	Low	High	Low	Organic and Paid

Table-20: Keyword Type mapping to Website Goal | Source: *Moran & Hunt* [2014] - Page 107

6. **Assessing the SERP Content Satisfaction:** Specific Searcher Intent Decides importance of satisfaction factors, for example if the query is navigational and the user is looking to find a specific web page, the page presentation of that web page may not be important. Likewise, a query asking for information on an historical event would have no requirement for very fresh content.
7. **Side-by-Side SERP Quality Evaluation Method:** Project primarily involves **Bing and Google** on **A/B mapping** pattern to assess which SERP yields better satisfaction.
8. **Walking Down Assessment Method:** Assessing the user intent, reviewing the SERP Content, Assessing both content and SERP satisfaction.
9. **Headline Defect Rate Method:** This method is used to annotate the quality and relevancy of the titles used on a webpage against topicality and queried keywords.
10. **News Answer Defect Finding Method - Search Defect Rate:** This method is used to testify the appropriateness of a news intent of SERP against a given search query.
11. **Junk Labelling Method:** This method is used to identify the Junk Pages and Content in SERPs.
12. **Related Suggestions Side by Side Annotation Method:** This method is used to gauze the quality content in "Related Suggestions" Snippet of SERP Features.
13. **Web Page Q & R Evaluation Method:** The QR method is a way to decompose a matrix into two matrices Q and R, where Q is an orthogonal matrix, and R is an upper triangular matrix. It assesses the Quality and Relevance of SERP results.
14. **Web Documents Relevance Annotation Method:** This method evaluates various web documents such as webpages, pdf files, word files etc. for User Intent Relevance.
15. **News and Freshness Addendum Method:** Annotating the various news webpages on Freshness and its durability in the market is done through this method.
16. **Subscribed email donation Method:** This method is basically related to subscription of Search Engines in which we receive periodic updates from Search Engine News and Trends portals. Here people voluntarily *(paid in exchange)* give feedback through donating the emails from own inbox on criteria of freshness, relevance, source after removing the PII (Personally Identifiable Information).
17. **Subscription Notification Accuracy Method:** Like above method but this is done at the Search Engine's end itself. Here the content relevance & quality evaluation received through subscribed email notification is annotated.
18. **Comparative Satisfaction Analysis of SERPs:** In this approach usually Bing and Google SERP Screenshots are compared to assess the results quality and diversity provided against a given search query. We shall talk about it in details in next section.

19. **Map Search Evaluation Method:** This method is used to assess the exactitude of searched location on Google or Bing Map along with the associated information provided in the map.
20. **Resource News Answer Method:** This method verifies the Landing Pages under a Search result and its connection between the searched query and the content provided on this source.
21. **Expert Evaluations:** There two types of experts used in it. The Domain Experts and the In-house Experts. Domain experts i.e. the Subject matter experts assess the quality of search results, especially in areas where specialized knowledge is required. And in-house experts within the search engine company who are well-versed in the domain evaluates the SERP quality internally.
22. **Task-Based Evaluation:** This evaluation includes the scenario-based tasks in which specific search tasks or scenarios are defined for evaluating the effectiveness of search results in completing those tasks. Assessing how well search results align with user intent in such scenarios is associated part of it.
23. **Machine Learning and Predictive Models:** This method is basically for the training data creation by incorporating the ML Algorithms to automatically evaluate the quality of SERPs. This annotated data is then used to train machine learning models to predict the relevance and quality of search results.
24. **SERP Features Analysis:** This method is about evaluating the impact of various SERP features *(featured snippets, knowledge panels etc.)* on user satisfaction and relevance by determining the effectiveness of additional features beyond traditional organic search results.
25. **Binary Relevance Annotation:** Assessors assign binary labels *(relevant or non-relevant)* to each search result, indicating whether it satisfies the user's query. This method is often used for simplicity and efficiency.
26. **Graded Relevance Annotation:** Instead of binary labels, assessors assign graded relevance scores *(e.g., on a scale from 1 to 5)* to indicate the degree of relevance for each search result. This provides a more nuanced understanding of result quality.

The above 26 methods are just exemplary; in total there are many more methods through which the quality and relevance annotation of SERP results are done by search engines through crowdsourcing. This all is done to ultimately achieve following evaluations and accuracies **(purposes)** in SE indexes and search results:

1. **Relevance Assessment:** Relevance assessment involves evaluating the degree to which search results match the user's query. Human assessors review search results and judge their relevance to the given query. This type of annotation helps search engines understand how well their algorithms are returning results that align with user intent.
2. **User Behaviour Assessment:**

i) **CTR Analysis:** User Interaction Data is used in analysing user click-through rates on search results to infer the relevance and satisfaction with the displayed results. Similarly, the performance of results based on their position on the SERP is assessed.

ii) **Dwell Time Analysis:** It includes the assessing of time users spend on a particular page after clicking on a search result, along with the bounce rate percentage of users who navigate away from the page without further interaction.

iii) **Query Reformulation Analysis:** It includes studying how iteratively users reformulate their queries after viewing initial search results and analysing whether users expand or refine their queries for more accurate results.

iv) **User Satisfaction Measurement:** User satisfaction measurement aims to gauge the overall satisfaction of users with the search results. Assessors may consider user feedback, engagement metrics, and other indicators to evaluate the effectiveness of the results in meeting user expectations.

v) **Click Data Analysis:** Assessors analyse actual user click-through data to understand which results users find most relevant. This method leverages real-world user behaviour to assess result quality.

3. **Multimodal Content Evaluation:** It includes the assessing of relevance and quality of search results when users are searching for visual content like videos and images for related search queries.
4. **Content Quality Evaluation:** Content quality evaluation focuses on assessing the overall quality of the content displayed in search results. Assessors may consider factors such as accuracy, comprehensiveness, and credibility of the information provided by the search results.
5. **Sentiment Analysis:** This analysis considers the sentiment signals from social media platforms and analyse the sentiment in user feedback, reviews, or comments related to specific search results.
6. **Query Intent Understanding:** In this method, assessors analyse the search query to determine the user's intent and then evaluate the search results based on how well they fulfil that intent. This type of annotation helps improve the search engine's ability to discern and cater to diverse user intents.
7. **Freshness Assessment:** Freshness assessment involves evaluating the recency and timeliness of search results. Assessors determine how well the results satisfy users looking for the most up-to-date information, especially relevant for queries with a temporal context.
8. **Diversity Evaluation:** Diversity evaluation assesses the variety of results presented in the SERP. Assessors check whether the results cover a range of perspectives, sources, and content types, providing users with a diverse set of information.
9. **Spam Detection:** Spam detection annotation focuses on identifying and flagging low-quality or spammy content in search results. Assessors help train algorithms to recognize and filter out content that violates quality guidelines, ensuring a more reliable and trustworthy search experience.

10. **Entity Recognition:** Entity recognition involves annotating search results to identify and categorize entities *(e.g., people, places, organizations)* mentioned in the content. This annotation helps improve the search engine's understanding of entities and their relationships.
11. **Localization and Language Assessment:** Localization and language assessment involve evaluating how well search results cater to specific languages and geographic regions. Assessors help ensure that results are contextually relevant and linguistically appropriate for diverse user groups.
12. **User Satisfaction Surveys:** Under this method feedback is directly collected from users through surveys to gauge their satisfaction with search results. Users may provide subjective feedback on the usefulness and relevance of the presented information.
13. **Adversarial Evaluation:** Introduce intentionally challenging queries or scenarios to assessors to evaluate how well search engines handle difficult cases. This helps identify potential weaknesses in search algorithms.

These methods are often used in combination to provide comprehensive insights into the quality of SERPs. Search engines continually refine their algorithms based on the results of quality annotation to enhance user experience and deliver more relevant and accurate search results.

Comparative annotation between two Search Engines: This benchmarking analysis is done through comparing the performance of search results against those provided by other search engines or platforms as examples mentioned above of **Amazon, Bing and Google**. It also includes assessing the current SERP quality in comparison to their historical data. A range of studies have explored the differences in search results between various search engines. *Biradar & Kumar* [2006] found significant variations in the number of sites searched and the relevancy of results across different engines. Bar-Ilan et al. [2006] and *Bar-Ilan* [2005] both developed methods for comparing the rankings of search engine results, with the latter finding that major search engines use different ranking algorithms. *Kumar & Pavithra* [2010] compared the searching capabilities of search engines and metasearch engines, finding that while **search engines did not achieve higher precision, metasearch engines (or search aggregator) did not achieve higher recall**. These studies collectively highlight the need for further research into the comparative annotation methods between search results of different search engines. It involves evaluating and comparing the quality of their search results to understand their respective strengths and weaknesses. While the specific details of the annotation process may vary, here's a general overview of how comparative annotation methods can be applied:

1. **Establish Evaluation Criteria:** Define the specific criteria and metrics for comparison. This may include relevance to user queries, accuracy of information, diversity of results, user satisfaction, and other factors relevant to the goals of the comparison.
2. **Select a Set of Queries:** Choose a set of queries that are representative of the types of searches users commonly perform. These queries should cover a diverse range of topics, intents, and complexities to provide a comprehensive evaluation.
3. **Gather Assessors:** Assemble a group of human assessors who are familiar with search engine evaluation and the defined criteria. These assessors should ideally be unbiased and may include individuals with expertise in information retrieval or the specific domain being evaluated.
4. **Blind Assessment:** Conduct a blind assessment where assessors are not aware of which search engine generated the results. This helps ensure unbiased evaluations and avoids potential preconceived notions about the performance of specific search engines.
5. **Evaluate Search Results:** Have assessors review and annotate the search results independently. They should provide judgments on the relevance, accuracy, and other criteria for each search result for both search engines.
6. **Aggregate Assessments:** Aggregate the individual assessments to create a consolidated view of the comparative performance. Statistical analysis may be applied to identify significant differences between the search engines.
7. **Feedback and Iteration:** Collect feedback from assessors regarding the annotation process and refine the evaluation criteria if necessary. Iterative evaluations may be conducted to track improvements or changes in search engine performance over time.
8. **Quantitative Analysis:** Use quantitative metrics, such as precision, recall, and click-through rates, to quantify the performance of each search engine. These metrics provide a numerical basis for comparison.
9. **Qualitative Analysis:** Additionally, perform a qualitative analysis of the annotated results. This may involve reviewing specific examples or cases where one search engine outperformed the other and identifying patterns or areas of improvement.
10. **Benchmarking:** Benchmark the search engines against industry standards or best practices to provide context for the comparative evaluation. This helps in understanding how the search engines perform relative to established benchmarks.

By following these steps, comparative annotation methods enable a systematic and objective assessment of the performance of two search engines. The insights gained from such evaluations can be valuable for search engine providers to identify areas for improvement and for users and businesses to make informed decisions

about which search engine aligns better with their needs. These methods can be used to assess the relevance, accuracy, and comprehensiveness of search results, as well as to identify strengths and weaknesses of each search engine. There are a number of different types of **comparative annotation approaches**, each with its own strengths and weaknesses, which include:

1. **Manual annotation:** This is the most time-consuming and labour-intensive method, but it also provides the most detailed and accurate results. In manual annotation, a human annotator carefully reads and evaluates each search result from each search engine, assigning it a score or rating based on its relevance, accuracy, and comprehensiveness. The all sorts of methods mentioned above which involve human raters fall under this approach.
2. **Semi-automated annotation:** This method combines manual annotation with automated techniques. An automated system initially filters the search results, and then a human annotator reviews and refines the results. This method can be more efficient than manual annotation, but it may not be as accurate.
3. **Automated annotation:** This method uses machine learning algorithms to automatically evaluate search results. These algorithms are trained on a dataset of annotated search results, and they can then be used to evaluate new search results. Automated annotation can be very efficient, but it may not be as accurate as manual annotation. All of the Query parsing, deciphering and indexing techniques talked in various sections of this thesis in above parts fall under this category.

Applications of comparative annotation methods: Comparative annotation methods can be used for a variety of purposes, including:

1. **Evaluating the performance of search engines:** This can be used to identify the strengths and weaknesses of different search engines, and to make recommendations to users.
2. **Improving the ranking algorithms of search engines:** This can be used to identify and correct problems with the ranking algorithms, and to improve the overall quality of search results.
3. **Developing new search technologies:** This can be used to identify new ways to improve search accuracy and relevance.

Challenges in comparative annotation methods: There are several challenges associated with comparative annotation methods, including:

1. **The difficulty of defining and measuring relevance:** Relevance is a subjective term, and there is no single definition that is universally accepted. This makes it difficult to compare and evaluate the relevance of search results.

2. **The difficulty of identifying and measuring biases:** Search engines can be biased in a variety of ways, such as by favouring certain websites or types of content. This can make it difficult to get an accurate assessment of the quality of search results.
3. **The difficulty of automating annotation:** Automated annotation methods are still under development, and they are not always as accurate as manual annotation. This can make it difficult to scale annotation efforts to many search results.

Despite these challenges, comparative annotation methods are an important tool for evaluating search engines and improving search quality. As these methods continue to develop, they are likely to play an increasingly important role in the field of search engine research.

Discussion: It's important to note that the categorization factors for both content creator and web content are interrelated, and search engines use complex algorithms to weigh them and determine the authority of a webpage or author. The combination of these factors allows search engines to categorize web content creators effectively, ensuring that users receive search results that match their intent and provide valuable information. The emphasis on specific factors may vary across different search engines, and algorithms are continuously updated to provide users with the most relevant and trustworthy information. A combined model including all of the above factor, and maybe even more is implemented by Bing as talked in above section and named "**Universal Human Relevance Model**". This discussion is incomplete without bringing to focus back on the content creator. In fact, above study shows that both are important for each other i.e. the Search Engine to Content Creator and Content Creator to Search Engine. However, this topic can arguably include answer to some more questions like, how Good Content Creators are recognized and rewarded or even degraded and penalized by Search Engine but that may take us out of our apparatus of thesis. Moreover, the investigation need in purview of our thesis topic can justifiably be sufficient on this topic as per above findings keeping Chart-62 at center.

Web Search Expressions and Operators: Search Operators are the shorthand special instructions, also called as **Search Commands** to search easily and concretely on Search Engines to get relevant SERP. Search Operators have the power of gauging the content dedicated to a specific topic and **filtering up to 90% of less-focused content.** These are used without putting spaces between the expression and search term. Search operators are special characters or commands that, when combined with a search query, help refine search results. They are a powerful tool for generating targeted, precise results in Google searches or data extraction tools. These operators can be used to narrow down search results, find specific data, conduct competitive

analysis, and enhance SEO efforts. Some examples of search operators include **finding content and link building opportunities, identifying SEO issues, and locating research and statistics to improve content,** [208]Jon Clark [2022], [209]Edwin Toonen [2023] & [210]Matteo Duo [2023]. By mastering Google search operators, users can save time and uncover valuable information that might not be easily accessible through regular searches, [211]Marcin Pastuszek [2023]. Search operators can enhance search results by improving the effectiveness of queries. Using query operators, such as +term or term^10, can lead to improved search results, regardless of specific users. Research conducted on **TREC-7** and **TREC-8** *(Text REtrieval Conference)* test collections with five IR models implemented in the Terrier search engine showed that queries enriched with operators resulted in an improvement in effectiveness of **up to 35.1% over regular queries.** This suggests that end-users would benefit from using operators more often, *Carr & Zamansky* [2018] and Hubert et al. [2011]. This is particularly relevant in the context of web search, where enhanced results, including multimedia objects and interactive elements, have been found to be preferred by users and effective in assessing relevance, Haas et al. [2011]. However, the impact of query operators on search results can vary, with the number of terms, matching terms, and correct operator usage being key factors, Lucas & Topi [2002]. While the use of query operators is often assumed to improve search effectiveness, the actual impact can be minimal, depending on the search engine, Eastman & Jansen [2003]. Based on the resources of [208]Jon Clark [2022], [209]Edwin Toonen [2023] & [210]Matteo Duo [2023], [212]Technolush [2019], [213]Luigi Oppido [2021], [214]Anna Gargioni, [215]Pcmag-GSOs, [216]Condron et al. [2022], [217]Deepawaliseotips [2023], [218]Google-GSOs-1, [219]Google-GSOs-2, [220]Moz-GSOs and [221]Joshua Hardwick [2023] here are some actionable **ways to use search operators** effectively:

1. **Exact Phrase Search ("):** Use double quotation marks to search for an exact phrase. For example, searching for "machine learning" will return results containing the phrase "machine learning" exactly as it is.

[208] https://searchengineland.com/advanced-google-search-operators-388355
[209] https://yoast.com/search-operators/
[210] https://kinsta.com/blog/google-search-operators/
[211] https://seodity.com/blog/google-advanced-search-operators
[212] https://www.technolush.com/blog/advanced-google-search-tips
[213] https://www.wikihow.com/Search-for-Phrases-and-Expressions-in-Search-Engines
[214] https://www.greengeeks.com/blog/top-google-search-secrets-that-will-make-you-look-smarter/
[215] https://www.pcmag.com/encyclopedia/term/search-expression
[216] https://www.spyfu.com/blog/google-search-operators/
[217] https://www.deepawaliseotips.com/mastering-google-search-operators-commands-comprehensive-guide/
[218] https://developers.google.com/search/docs/advanced/debug/search-operators/overview
[219] https://support.google.com/websearch/answer/2466433?hl=en
[220] https://moz.com/learn/seo/search-operators
[221] https://ahrefs.com/blog/google-advanced-search-operators/

2. **Exclude Terms (- or NOT):** Use the minus sign or the word "NOT" before a term to exclude it from search results. For instance, searching for data science -jobs will retrieve results related to data science but exclude those mentioning jobs.
3. **Site Search (site:):** Specify a particular website for search results using the site: operator. For example, site:example.com python tutorials will retrieve Python tutorials specifically from the website example.com.
4. **File Type Search (filetype:):** Use filetype: to search for specific file types. For instance, filetype:pdf artificial intelligence will return PDF documents related to artificial intelligence.
5. **Wildcard Search (*):** Include an asterisk as a wildcard character to represent any word or phrase. For example, searching for machine * algorithms will retrieve results related to machine learning algorithms, deep learning algorithms, etc.
6. **OR Operator (OR or |):** Use the OR operator to broaden your search by including multiple possibilities. For example, data science OR machine learning will return results related to either data science or machine learning.
7. **Intitle Operator (intitle:):** Specify that the search terms must appear in the title of the webpage using intitle:. For instance, intitle:"best practices" data analysis will return pages with "best practices" in the title related to data analysis.
8. **Inurl Operator (inurl:):** Specify that the search terms must appear in the URL of the webpage using inurl:. For example, inurl:research machine learning will retrieve pages with "research" in the URL related to machine learning.
9. **Define Operator (define:):** Use define: to get definitions for terms. For example, define:artificial intelligence will provide definitions for the term "artificial intelligence."
10. **Numeric Range Search (..):** Specify a numeric range using two periods (..). For instance, Python programming language release date 1990..2000 will retrieve results related to the release of Python within the specified years.
11. **Related Operator (related:):** Use related: to find sites related to a specified domain. For example, related:wikipedia.org will retrieve sites related to Wikipedia.
12. **Social Media Search (@ or #):** Include the @ symbol or # symbol to search for content on social media platforms. For example, @username or #hashtag will retrieve content related to the specified username or hashtag.

Basic Search Operators:

1. **Quotation Marks ("):** Use quotes to search for an exact phrase. Example: "artificial intelligence applications"

2. **AND Operator:** Use "AND" to narrow down search results by requiring both terms to be present. Example: artificial intelligence AND machine learning
3. **OR Operator:** Use "OR" to broaden your search, finding results that include either term. Example: renewable energy OR sustainable energy
4. **NOT Operator:** Exclude specific terms from your search using "NOT." Example: electric cars NOT Tesla
5. **Wildcard (*) Operator:** Use an asterisk as a placeholder for any unknown word in a phrase. Example: "benefits of * technology"

Advanced Search Operators:

1. **site Operator:** Search within a specific website or domain. Example: site:wikipedia.org artificial intelligence
2. **filetype Operator:** Specify a particular file type in your search. Example: filetype:pdf climate change report
3. **intitle Operator:** Find pages with a specific word in the title. Example: intitle:"machine learning algorithms"
4. **define Operator:** Get definitions of a term. Example: define:quantum computing
5. **related Operator:** Find sites related to a given domain. Example: related:nytimes.com

Tactics for Advanced Searching:

1. **Numeric Range Search:** Use ".." to search within a numeric range. Example: electric cars 2020..2022
2. **Synonym Searching:** Use the tilde (~) before a word to include synonyms in your search. Example: ~ecofriendly products.
3. **Time-Based Searches:** Use tools to filter results by date or specify a time range for more recent information. Example: climate change news last month
4. **Location-Based Searches:** Include location information for locally relevant results. Example: sustainable agriculture practices California
5. **Combining Operators:** Combine multiple operators for highly specific searches. Example: "renewable energy" AND (policies OR incentives) NOT subsidies

Utilizing Search Engine Features:

1. **Advanced Search Page:** Many search engines provide an advanced search page with built-in options for various operators. Use these for a more user-friendly experience.
2. **Search Suggestions:** Pay attention to search suggestions for commonly used phrases related to your query.

3. **Filters:** Utilize filters provided by search engines, such as image search, news, or specific timeframes.
4. **Preview Snippets:** Check snippet previews to quickly evaluate if a page has relevant information.
5. **Experiment and Iterate:** Don't hesitate to experiment with different combinations of operators and tactics. If your initial search doesn't yield results, iterate and refine.

Actionable ways to use search operators:

1. Find indexation errors.
2. Find non-secure pages (non-https)
3. Find duplicate content issues
4. Find odd files on your domain (that you may have forgotten about)
5. Find guest post opportunities.
6. Find resource page opportunities.
7. Find sites that feature infographics… so you can pith yours
8. Find more link prospects… AND check how relevant they really are.
9. Find social profiles for outreach prospects.
10. Find internal linking opportunities.
11. Find PR opportunities by finding competitor mentions.
12. Find sponsored post opportunities.
13. Find Q+A threads related to your content.
14. Find how often your competitors are publishing new content.
15. Find sites linking to competitors.

SERP Personalization, SEO Challenges and AIRO: Why are we discussing this topic just next to "Query Operators and Processes"? Because "Query types coming from a User Device is in the base of setting-up cookies, preferences and user profile making, which are prime sources of SERP Personalization often termed as "Search User Profiling". So, proceeding to the personalization topic, let's think this way. If a search engine is publishing great info directly in SERPs, then doesn't that **reduce the likelihood of a user clicking on a full-length result**? Yes, and no? If a web page does not earn a featured spot in SERPs, it means someone else has provided information that is more relevant to searchers. It also means that Featured Snippets and the like may appear above all other results, pushing your content further down the page, despite whatever relevancy it may have. This is bad for your SEO, as the former glory of being on Page-1 holds less significance if you don't actually own Position-0 *(*also know as *Featured Snippet)*. Conversely, you can drive higher click-through rates, site traffic and search visibility by winning SERP features, thus improving your SEO. And as others click-

through rates rise, SERP personalization poses challenges for your SEO managers. So, SERP personalization is making it difficult for SEO managers because it involves tailoring SERPs to individual users based on their search and browsing activity, *Ustinovskiy et al.* [2013]. This means that the ranking of SERPs can be personalized using algorithms like **Personalized PageRank** (PPR), *Ronny Lempel* [2017]. This presents a challenge as they need to optimize their websites to rank higher in personalized SERPs. They need to understand the preferences and behaviours of individual users to effectively target their content, *Kaingade et al.* [2015]. Additionally, the **use of user profiles and privacy-preserving search techniques** in personalization further complicates the task for SEO managers, *Timos Sellis* [2011]. SERP personalization makes it challenging due to its impact on search rankings and the need to adapt to personalized search experiences. Personalization, which tailors search results based on factors such as **search history, localization, and browsing history**, can affect the visibility of content and the accuracy of rank tracking, [222]*Masha Maksimava* [2017]. As search engines customize SERPs based on individual data, including location and search history, SEO managers face the difficulty of ensuring their content remains visible and accurately tracked across personalized search experiences, [223]*Nick Barney* [2022]. While personalization can enhance the user experience, SEO managers need to navigate the technical considerations and adapt their strategies to ensure their content remains relevant and visible within personalized search results, [224]*Thais Montjean* [2025]. *Patel et al.* [2015] and *Stamou et al.* [2008] both highlight the difficulty in accurately identifying and ranking user interests, with Jay emphasizing the need for personalized web search and Stamou discussing the use of click history and **topical analysis** to determine user preferences. *Arya et al.* [2014] and *Shafiq et al.* [2015] further underscore the complexity of this task, with Arya reviewing various personalization techniques and Shafiq proposing a method for personalizing web search based on **user preferences** from their social networks. These studies collectively demonstrate the multifaceted nature of SERP personalization and the challenges it poses for SEO managers in understanding and adapting to user preferences. Therefore, the challenge lies in understanding and adapting to the evolving landscape of personalized search experiences to maintain the visibility and effectiveness of SEO efforts. SERP personalization requires SEO managers to adapt their strategies to meet the specific needs and preferences of individual users, making it more challenging to achieve high rankings and visibility in search results. Here are some ways in which personalized search results impact the work of SEO managers:

1. **Individualized Rankings:** SERP personalization means that search results can vary significantly based on individual user profiles, preferences, and past behaviour. This makes it challenging for SEO managers to predict and optimize for specific rankings, as there is no longer a one-size-fits-all approach.

[222] https://www.link-assistant.com/news/personalized-search.html
[223] https://www.techtarget.com/whatis/definition/search-engine-results-page-SERP
[224] https://dotdigital.com/blog/personalization-in-digital-marketing/

2. **Limited Control over User Preferences:** SEO managers have limited control over individual user preferences and behaviours. Personalization factors, such as location, search history, and device usage, can influence rankings. This lack of control makes it difficult to target a broad audience with consistent SEO strategies.
3. **Dynamic Nature and Opacity about Algorithms:** Search engines continuously refine and update their personalization algorithms. SEO managers need to adapt to the dynamic nature of these algorithms, which may undergo changes that affect the visibility and ranking of content. Staying abreast of algorithm updates becomes crucial for effective SEO. Search engine algorithms, especially those incorporating machine learning, can be complex and opaque. SEO managers may not have full visibility into the factors influencing personalized rankings, making it challenging to fine-tune strategies for optimal results.
4. **Impacts on Keyword Strategy:** Personalization influences the interpretation of search queries, and the same keyword may yield different results for different users. SEO managers need to consider a broader range of variations and long-tail keywords to capture diverse user intents, making keyword strategy more complex.
5. **Localized and Contextual Relevance:** Personalization often emphasizes local and contextual relevance. This makes it challenging for SEO managers with a global audience, as search results may vary based on the user's location. Balancing global and local optimization becomes a strategic challenge.
6. **User Engagement Metrics:** SEO managers traditionally rely on metrics like CTR and dwell time to gauge the performance of content. With personalized results, these metrics may be influenced by individual preferences, making it harder to assess the overall success of SEO efforts.
7. **Privacy Concerns:** Increasing privacy concerns have led to the development of privacy-centric algorithms and approaches. While beneficial for user privacy, these measures can limit the availability of user data for SEO analysis and optimization, making it challenging to understand and adapt to user behaviour.
8. **Diverse Intent Matching:** Personalized search aims to match diverse user intents. SEO managers must understand and optimize for various user intents, including informational, navigational, and transactional queries, to ensure visibility across a wide range of search scenarios.
9. **Content Fragmentation:** Personalization can lead to content fragmentation, where different users see different versions of a webpage or content piece. This challenges SEO managers to create content that remains relevant and engaging across various personalized versions.

So, while SERP personalization enhances user experience, it introduces complexities for SEO managers. Navigating the challenges requires a nuanced understanding of the evolving search landscape, continuous

adaptation to algorithmic changes, and a holistic approach to content optimization that considers diverse user preferences and intents.

Relation of AIRO and SEO, and their mutual future: With the advent of G-AI, LLMs, Open-AI, Chat-GPT; the SEO has got a big push in terms of technical advancements and practices. And AIRO has emerged as one of those most promising practices to complement SEO strategies and practices. Simply, AIRO has brought the aura of AI in SEO. So, it is ultimately has started transforming the SERP as well. So as the future of SEO and SERP is going to pass through AIRO, thus it becomes imperative to know about it in this thesis.

Artificial Intelligence Recommendation Optimization *(AIRO)*: AIRO refers to the process of improving and enhancing the user experience of a web product(s) i.e. website, blog, ecommerce platform, mobile app or video channel etc. by using advanced **recommendation systems** of AI. Recommendation systems are Machine Learning Algorithms designed to suggest products, services, or information to users based on various data inputs and user preferences. It involves leveraging user data, historical interactions, and advanced analytics to provide personalized and relevant recommendations to users. Optimization in this context involves refining these algorithms to deliver more relevant and personalized recommendations, thereby enhancing user experience and business outcomes. Research highlights different aspects of recommendation strategies, such as the optimal conditions for offering recommendation services based on commission rates and competition levels, *Zhou et al.* [2023]. Additionally, AI is applied in educational settings to create personalized curriculum recommendations for autistic children, improving cognitive abilities through dynamic learning models, *Hao et Hu* [2022]. In the field of drilling optimization, AI is leveraged to identify optimum drilling parameters by analyzing in-cutter force sensing data, enhancing efficiency and decision-making processes in drilling operations, *Koulidis & Ahmed* [2023]. Moreover, AI-driven recommendation engines are crucial in the oil and gas industry for production optimization, maintenance, and cost reduction, integrating subsurface information to recommend actions for operational excellence and lower unit cost/barrel, *Yuan* [2022]. These examples from diverse business domains reflects the crucial importance of AIRO across domains; thus, for people searching about such diverse topics on SERP. Here's a deeper look into the components and methods involved in this process:

1. **Data Collection:** Gathering user data, which includes Explicit Data such as Ratings, Likes, Dislikes and Reviews; and Implicit Data such as Browsing history, Purchase history, Search queries, Time spent on different items.
2. **Data Processing:** Cleaning and transforming raw data into a usable format.

3. **Algorithm Selection:** Choosing the right type of recommendation algorithm, such as **Collaborative Filtering** for making recommendations based on the preferences of similar users; **Content-Based Filtering** for recommending items like those a user has liked in the past; and **Hybrid Systems** for Combining multiple algorithms to leverage the strengths of each.
4. **Model Training:** Using ML techniques to train the recommendation model on the processed data.
5. **Prediction and Recommendation:** Generating recommendations for users based on the trained model. AI algorithms like collaborative filtering, matrix factorization, and deep learning are employed to analyse user preferences, behaviour patterns, and item characteristics. These models can predict the likelihood of a user engaging with a particular item (product, content, service, etc.), enabling accurate and personalized recommendations, [225]*Pathmonk* [2023].
6. **Content Analysis and Evaluation:** Assessing the performance of the recommendation system using metrics like precision, recall, and F1 score. AI-powered NLP and computer vision algorithms analyse and evaluate the content of items *(text, images, videos, etc.)* to extract relevant features and metadata. This enriched information enhances the accuracy of recommendation models by better understanding item characteristics and their relevance to users, [225]*Pathmonk* [2023] and [226]*Raykova* [2024].
7. **Continuous Learning:** AI recommendation systems continuously learn and adapt to changing user preferences, new data, and evolving trends. They leverage techniques like reinforcement learning and online learning to refine their models and improve recommendation quality over time, [225]*Pathmonk* [2023] and [226]*Raykova* [2024].

AIR Optimization Techniques:

1. **Hyperparameter Tuning:** Adjusting the parameters of machine learning models to improve performance.
2. **Feature Engineering:** Creating new features or modifying existing ones to better capture user preferences and behaviour.
3. **Advanced Algorithms:** Implementing sophisticated models such as **Neural Networks of Deep Learning** particularly those with many layers and can capture complex patterns in data; optimizing recommendations through **Reinforcement Learning** based on long-term user engagement rather than immediate feedback.
4. **A/B Testing:** Running experiments to compare the performance of different recommendation strategies in real-world settings.
5. **User Profiling (Personalization):** Tailoring recommendations to individual user profiles using techniques like clustering, segmentation, and personalized ranking. AI techniques are used to create comprehensive

[225] https://www.linkedin.com/pulse/artificial-intelligence-meets-conversion-optimization-future/
[226] https://uxify.com/blog/post/artificial-intelligence-optimization-website-speed

user profiles based on their browsing history, purchase patterns, demographic data, and other relevant information. These profiles capture user interests, preferences, and contexts, enabling tailored recommendations, [225]*Pathmonk* [2023].

6. **Scalability and Efficiency:** AI algorithms and distributed computing architectures enable recommendation systems to handle large-scale data and provide real-time recommendations efficiently, even for millions of users and items, [226]*Raykova* [2024]. Through tuning and adjustments, AI models are refined to make use of computational resources more effectively for faster and more accurate outcomes, [227]*Ayuya* [2024].

Practical Applications of AIRO:

1. **E-commerce:** Recommending products based on past purchases and browsing behavior.
2. **Streaming Services:** Suggesting movies, TV shows, or music that align with user tastes.
3. **Social Media:** Curating feeds with posts, ads, and content that users are likely to engage with.
4. **Content Platforms:** Offering articles, blogs, or news tailored to user interests.

Benefits of AIRO:

1. **Increased User Engagement:** More relevant recommendations lead to higher user satisfaction and engagement.
2. **Improved Conversion Rates:** Personalized recommendations can drive sales and other desired actions.
3. **Better User Retention:** Consistently providing valuable suggestions keeps users returning to the platform.
4. **Enhanced Competitive Advantage:** Superior recommendation systems can differentiate a service from its competitors.

Mutual relationship of AIRO and SEO: After having an insightful knowledge of AIRO, there is a buzzing question about AIRO and SEO relation. Lots of debates are seen on internet wondering about their coexistence. People ask, is AIRO to replace SEO? Or how both are to sail together. Certainly, AIRO is not going to replace SEO but rather complement it in enhancing digital marketing strategies. As AI can significantly improve website ranking, communication quality, and purchase intent, *Somosi* [2022], by utilizing machine learning for index tuning, resulting in up to a 5x reduction in errors and improved query execution cost predictions, *Ding et al.* [2019] so it is proving a support to SEO. Moreover, AI-based recommender systems address challenges like data sparsity and cold start, leading to more accurate predictions and better user experiences, *Wang et al.* [2019]. AI can also be utilized to determine user intent from search queries, aiding in filtering recommendations and pre-selecting input fields, thereby enhancing user satisfaction and engagement, *Tavernier* [2020]. By

[227] https://www.eweek.com/artificial-intelligence/ai-model-optimization/

incorporating semantic information into entity correlation calculations, AI-based search recommendation methods can improve the accuracy and coverage of recommended entities, *Wei et al.* [2021]. AIRO and SEO serve different purposes. SEO focuses on improving a website's visibility and ranking on search engines like Google for relevant keywords and queries, [228]*Khanal* [2023]. On the other hand, AIRO aims to provide personalized and relevant recommendations to users based on their preferences, behaviour, and context, *Ma et al.* [2022]. While AI recommendations can influence what products or services are suggested to users, it does not directly impact how websites rank on search engines. SEO will still be crucial for businesses to attract organic traffic from search engines and increase their online visibility, [228]*Khanal* [2023]. However, AIRO can complement SEO efforts. By providing more relevant and personalized content recommendations, AI can potentially improve user engagement metrics like time on site, bounce rate, and conversions, [229]*Bodhost* [2024]. These signals can indirectly benefit SEO by indicating a better user experience to search engines. Additionally, as AI language models like ChatGPT become more advanced, they may be used as supplementary tools for SEO tasks like content creation, keyword research, and analysis, [228]*Khanal* [2023]. But they are unlikely to fully automate or replace the strategic and technical aspects of SEO.

So, it is evident from above FLR, that while AIRO can enhance the user experience and support SEO indirectly, it does not negate the need for a comprehensive SEO strategy to improve search engine visibility and organic traffic acquisition, [230]*Berry* [2023]. Both approaches serve different purposes and can work together to drive better overall results. AIRO is a multifaceted process that involves improving the accuracy, relevance, and efficiency of recommendation systems through advanced data processing, machine learning, and continuous refinement techniques. This optimization is crucial for delivering personalized user experiences and achieving business objectives. Therefore, AIRO offers significant benefits, it is more likely to work alongside SEO rather than replace it in the digital marketing landscape.

Discussion: By incorporating these search operators and tactics, search user can develop technical expertise in navigating search engines effectively and obtaining precise information. As per John Mueller, the site: operator doesn't just return URLs that Google has indexed. It also includes URLs that Google just knows, [221]*Joshua Hardwick* [2023]. This allows for a more efficient and targeted approach to information retrieval and give a robust foundation to the use of AIRO technics in our research work, contextually going to appear in upcoming chapters. Mastering search operators and tactics can significantly enhance ability to find specific information

[228] https://www.growandconvert.com/seo/google-rankings-vs-chatgpt-recommendations/
[229] https://www.bodhost.com/blog/ai-powered-webpage-optimization-what-it-is-and-how-to-use-it/
[230] https://www.seo.com/blog/seo-vs-ppc/

efficiently on search engines. Thus, no doubt that searchers equipped with knowledge of search operators have significant edge over others without this backup. Chart-63 below summarizes the facts and findings in above LR.



Chart-63: Web Search Intelligence

4.5. Web Search Market Statistics and Trends:

Based on my personal professional experience of more than a decade on Search Engine Machine Learning projects and as a yesteryear entrepreneur of Digital Marketing Consulting, I can confidently say that all the major search engines worldwide have been heavily investing in studying the web search behaviour to slice the high competition edge for them. Google, Bing, Amazon, Baidu, Yandex, Yahoo etc. are running special intelligence projects, teams and departments to assess their 'Web Search Experience Services' the best way possible since later 2010 augmenting to mass scale year by year and the process has just growth, no decline or

pause till date and time to come. These kind of rigorous commitments and practices of Search Engines have increased their grip on global e-marketing year by year. As per *Maynes & Everdell* [2014] Google search is used at some point in almost every purchase decision process. As a business, you have just over 1 second to capture your customers' attention, so for your content better be relevant. *Aryshandy et al.* [2021] highlights the positive impact of SEO *(the content producer lobby, we talked about them above in Web Content Creator section above)* section on e-marketing outcomes, emphasizing the need for companies to consider this in their strategies. *Egea & Menendez* [2006] underscores the potential of Internet technologies in expanding the reach of global e-marketing, while *Nathan* [2009] and *V. Raja* [2012] both discuss the growing importance of e-marketing in the global context, with the latter specifically focusing on the role of e-commerce in viral marketing. These studies collectively underscore the exponential growth and influence of SERP-marketing in the global e-marketing landscape. Search Engine Marketing (SEM) has a significant impact on various marketing variables, including market share and brand equity, *Bhandari & Bansal* [2018]. The role of SEO in SEM is particularly important, as it can improve the quality of search engine rankings and consumer satisfaction, *Berman & Katona* [2012]. However, the growing use of online affiliate marketing in SEM has raised concerns about content quality and consumer trust, *Bevendorff et al.* [2022]. Despite these challenges, the importance of SEM in the digital world remains high, with SEO analysis and techniques playing a crucial role in increasing website traffic and sales revenue, *Terrance et al.* [2018]. Terrance further added that exponentially augmented grip of SERP marketing on global e-marketing is evident. The interlinkage between Search Engine Marketing Management (SEMM) and SEO plays a vital role in improving the technical aspects of website SEMM and increasing website traffic. The impact of keyword analysis and other SEO-friendly techniques positively affects digital marketing, *John Simon* [2001]. Global electronic commerce, particularly in the service sector, is experiencing exponential growth and opportunities for businesses to expand their markets, *Kaur & Pathak* [2015]. E-commerce and e-marketing have become essential in establishing and maintaining long-term relationships with customers, *Rajamanickam* [2020] and *Ferrera & Kessedjian* [2019]. The development of new technologies and the interconnectedness of the world through the internet have transformed businesses and led to the evolution of e-commerce and global marketing. These changes have had a significant impact on the way businesses market their products and services globally. [231]*Tom J Law* [2019] of Oberlo defines this term in his research article as, "SERP marketing is the process of promoting a business within a search engine's results. SEO and SEM are the two types of SERP marketing. Every second of every day, Google processes 63,000 search queries. This year, we can expect more than 2 trillion searches, [232]*Danny Sullivan* [2016]. Google alone owned 75.4% percent of all search engine market share globally, Chart-71, *Netmarketshare* [2019], at present

[231] https://www.oberlo.com/blog/serp-google-search-engine-results
[232] https://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247

this share is 90.4% as per [233]*Statcounter-live*. The top 5 URLs on the SERPs get 75% of user clicks. Search and email are the top two Internet activities. 70 - 80% of users ignore the paid ads, focusing on the organic results. A study by Outbrain shows that search is the #1 driver of traffic to content sites, beating social media by more than 300%, [234]*Inetsolutions.org-1* [2016]. Let's view some more statistics in a time window segmentation as below:

Key Web Search Market Statistics of 2020s: This section of **"Search Market Data"** is intended to provide the fact-based ground to the importance of this thesis. It assesses the importance of this thesis with recent most data facts of SEM Market also know as SERP Marketing Market, Internet Market, and different types of User Behaviour on SERP Data. The most updated and authentic online sources used in this regard are [235]*Hoque & Munir* [2025], [236]*Dave Chaffey* [2024], [237]*Brian Dean* [2025] and [271]*Semrush Team* [2023]. So, let's know why this thesis is viable and so crucial:

1. SEM is the preferred choice for online advertising among small businesses, with 45% of them investing in this digital marketing strategy, making it the most widely adopted form of online promotion.
2. Daily, at least once, approximately 87% of smartphone users actively search for something on a SE.
3. By 2025, it is projected that nearly 73% of internet users will access internet primarily through mobiles.
4. Google's share was 92.61% in the search engine market in 2023.
5. 2023's top players of Search Engine Market with their Global Market Share have been Google - 92.61%, Bing - 2.79%, Yandex - 1.65%, Yahoo - 1.1%, DuckDuckGo - 0.52%
6. On a monthly basis, Bing handles a total of 12 billion search queries.
7. Approximately 49% of companies receive more than 1,000 average search views of their Google Business Profile listing per month, while 33% of companies receive 1,000 or more Maps views.
8. The average return on ad spend for Google Ads is $2 for every $1 spent.
9. As of March 2023, Bing held approximately 8.23% of the global desktop search market. In contrast, the market leader, Google, dominated with a share of around 85.53%. Yahoo's market share stood at 2.44%.
10. The majority, approximately 85%, of online interactions commence with the utilization of a search engine.
11. Among the search results displayed on the initial page of Google, the top five organic listings collectively attract approximately 67.6% of all clicks from users.

[233] https://gs.statcounter.com/search-engine-market-share
[234] https://www.inetsolutions.org/25-informative-seo-facts-statistics-probably-dont-know/
[235] https://www.agencyhandy.com/search-engine-marketing-statistics/
[236] https://www.smartinsights.com/search-engine-marketing/search-engine-statistics/
[237] https://backlinko.com/google-user-behavior

12. The average ROI for SEM is 250%, meaning that for every 1$ spent on SEM, businesses typically generate a return of 2.5$.
13. Leads generated from search engines typically exhibit a close rate of 14.6%, indicating the percentage of leads that ultimately convert into customers.
14. SEM is the preferred choice for online advertising among small businesses, with 45% of them investing in this digital marketing strategy, making it the most widely adopted form of online promotion.
15. A significant majority, approximately 75% of users, never go beyond the first page of search results when conducting online searches.
16. The percentage of users clicking on organic search results has decreased by 13%, now standing at 47.4%.
17. The click-through rate for search ads has experienced a significant increase of 75%, reaching 3.69%.
18. A considerable majority of Gen-Z and Millennial users, around 62%, prefer visual search technology over other forms of technology.
19. Long-tail keyword searches typically achieve a 3% to 5% higher CTR compared to organic searches.
20. Approximately 70% to 80% of users completely disregard paid advertisements and do not click on them.
21. Compared to social media, search engines generate approximately 300% more traffic to websites.
22. Approximately 87% of smartphone users utilize a search engine at least once every day.
23. In search engine results, the top five websites capture approximately 67% of clicks.
24. 50% of searches consist of four words or more.
25. If a brand appears multiple times in search engine results, around 50% of users are more inclined to click on one of the brand's results.
26. For 61% of marketers, the foremost priorities are optimizing SEO strategies and enhancing their organic presence in online search results.
27. Approximately 75% of users do not click beyond the first page of search engine results.
28. Compared to outbound leads, inbound leads cost approximately 61% less.
29. Approximately 93% of virtual experiences of online interactions and journeys, begin with a search engine.
30. According to 57% of B2B marketers, SEO is marketing effort that drives more leads than any other marketing strategy.
31. Before making a purchase, 88% of consumers engage in some form of online research, whether it's for products in physical stores or online.
32. According to a study, 55% of users are unable to distinguish which links in SERPs are PPC Ads. Additionally, up to 50% of these users have difficulty classifying paid advertisements correctly.
33. Approximately 40% of users abandon a website that requires three seconds or more to load.
34. Approximately 15% of questions asked on Google have never been asked before.

35. A video that is optimized for SEO has up to 52 times higher chances of being featured on Google's first page compared to a complete article.
36. When it comes to organic ranking, videos have the advantage of being 50 times more likely to achieve higher rankings than ordinary text-based results.
37. Around 72% of users who perform a local search visit a store located within a 5-mile radius of their current residence.
38. If a local business has a mobile-friendly website, approximately 61% of mobile users are more inclined to contact or reach out to that business.
39. Companies that incorporate a strong keyword in their business name tend to rank 1.5 spots higher in search results compared to businesses whose names do not include relevant keywords.
40. Approximately 46% of Google searches are conducted with local targets or local intent, meaning users are searching for information, products, or services related to specific locations or nearby businesses.
41. From a share of 18% in 2017, smartphones are projected to account for 44% of the total IP (Internet Protocol) traffic in 2024.
42. Now, voice search accounts for 10% of all searches conducted.
43. Approximately 40.7% of all voice search results are derived from snippets.
44. Every single day, Google handles a staggering 3.5 billion search queries, which translates to a rapid pace of 40,000 searches occurring every second.
45. Google makes approximately 500 to 600 changes to its search engine algorithm on a yearly basis.
46. The first search result receives 18% of organic Google clicks, the second result gets 10%, and the third result captures 7% of the clicks.
47. Approximately 49% of Google searches do not result in a click on any of the search results.
48. Just 0.78% of Google users click on the second page of the search results.
49. For approximately 27.9% of searches on Google, images are displayed in the search results.
50. On Google's SERP, the top results typically receive 3.8 times more backlinks compared to those results that have fewer backlinks.
51. Approximately 55.24% of web pages do not have even a single backlink.
52. The main reason why 91% of all web pages never receive any organic Google traffic is primarily due to their lack of backlinks.
53. Google's advertising services have been utilized by more than 1.3 billion companies, website publishers, and nonprofit organizations globally.
54. Over 70% of shoppers use their phones to buy online.
55. 91% of companies (an all-time high) now use video as a marketing tool.

56. A staggering 98% of people read online reviews for local businesses.
57. Organic web listings get 73% of clicks.
58. Consumers are using search engines 20% more year-over-year.
59. The “Computer and Video Games” category grew web traffic by 183% in 2022.
60. The top traffic source for all websites is organic search.
61. Over two-thirds (68.7%) of all clicks go to the top three organic search results.
62. Snippets have the highest overall CTR at 42.9%.
63. The top organic search result receives an average of 19x more clicks than the top paid search result.
64. Keywords that are roughly 10-15 words long get 1.76x more clicks than single-word terms.
65. Posts that contain target keywords or similar terms in URL have a 45% higher CTR compared to those that don’t.
66. Titles with a positive sentiment have a 4% higher CTR.
67. Users submit 57% of local search queries with a mobile device or tablet.
68. Related Searches is the most common feature in the SERPs for mobile searches.
69. The number of keywords has risen for transactional and commercial searches but dropped for navigational and informational.
70. About 33% of overall traffic to ecommerce sites comes from organic search.
71. Around half of shoppers say they research potential purchases using Google before buying online.
72. In a Google survey, 59% of shoppers said they used Google to research in-store and digital purchases.
73. Over 50% of consumers say they find new brands using Google.
74. B2B companies generate double the revenue from organic search than any other channel.
75. Some 81% of B2B buyers prefer watching videos over reading content when learning how to use a product or service.
76. Organic search and referral traffic are the largest and most effective sources of marketing ROI for brands.
77. Marketers are split on overall sentiment towards the future of AI; 44% foresee a positive impact on their SEO strategy, 51% predict no impact, and 5% see a negative impact.
78. Google is the faraway leader in the desktop search engine market with a share of 86.6%. It is followed by bing (6.7%), Yahoo! (2.71%), DuckDuckGo (0.9%), and Yandex RU (0.64%). Google also leads all mobile internet searches with a market share of 95.03%, followed by Baidu (1.8%), Yahoo! (0.8%), Yandex (0.73%), and DuckDuckGo (0.48%). Google is the search engine leader for tablet users, with a

market share of 90.35%, followed by bing (5.6%), Yahoo! (1.73%), DuckDuckGo (1.06%), and Yandex (0.54%), [238]*Jenny Chang* [2024].

79. As of Aug 2023, Google searchers use one of Google's autocomplete suggestions 23% of the time. People that search for informational and local searches tended to click on an autocomplete suggestion more often than those searching with commercial queries.
80. 50% of Google users click on their results within 9 seconds of searching. And the average amount of time it takes a Google searcher to click on something is 14.6 seconds.
81. Only 9% of Google searchers make it to the bottom of the first page of the search results.
82. 15% of users modify their initial search terms. This suggests that Google tends to serve up highly relevant results. Or that Google users are adept at choosing the right keyword on their first attempt.
83. Only 17% of users bounced back to the search results after clicking on a result. Only 5% of users bounce more than once for the same query.
84. The majority (59%) of Google users visit a single page during their search session. Only 6% need to visit four or more pages to get an answer to their query.
85. 65% of searchers click on a traditional "10 blue link" search result during their search session.
86. Although this varied widely depending on the query, 19% of searchers click on a Google Ad during their search.
87. For local queries, 42% of searchers click on results inside of the Google Maps Pack.
88. 19% of users searching for a product click on a Google Shopping result.
89. On average, only 3% of searchers interact with a "People Also Ask" box. Although this behavior varies widely depending on the query. For example, 13.6% of people searching for supplements clicked on a PAA box.
90. Only .44% of searchers go to the second page of Google's search results.
91. The average search session takes 76 seconds to complete. And half of all search sessions are finished within 53 seconds.

Alongside above bullet points insight of SERP Marketing, below are some graphical trends to visualize the SERP Market Facts memorably. The Chart-64, Appendix-9 displays the Google' Organic CTR trend, which is highly helpful in understanding the value of highest possible ranking in search engines. Chart-65, Appendix-9 explains the main reasons for using the internet. Here interesting to note is the first figure which is of Finding Information, and that related to once of the Intent Types of Searchers, the Informational Intent. We are going to talk about the intents in explicit details in next section below. Chart-66, Appendix-9 identifies the most

[238] https://financesonline.com/google-search-statistics/

common SERP Features on mobile and desktop in 2022 in US market. Chart-67, Appendix-9 is a local consumer review survey of 2023 and explains which sites of apps you have used to evaluate local businesses in the last 12 months.

Key Web Search Market Statistics of 2010s: This data assesses us the connectivity to a consistent habitual trend of growing influence of Search Engine Marketing i.e. the SERP Marketing in overall e-marketing. Chart-68, Appendix-9 is the percentage of all searches happen on any major search property in the USA. Chart-69, Appendix-9 distributes the distribution of search demand across keywords. Chart-70, Appendix-9 shows the Paid vs. Organic CTR difference on mobile versus desktop. Chart-71, Appendix-9 shows the historical distribution of their Search Market Share. A high-level research report of some trends was published by *Alberto Ferreira* [2017], 232*Danny Sullivan* [2016], [239]*Rand Fishkin* [2017] in early 2017, which is counted as among the most vast and considerable study as a base to estimate the searchers behaviour in terms of data and facts. A February 2012 Pew Internet survey finds that 91% of online adults use search engines to find information on the web, up from 84% in June 2004, *Purcess et al.* [2012]. In 2016, Internet advertising revenues in the United States surpassed those of cable television and broadcast television, [240]*IAB-Report* [2017] and [241]*Wiki-OA*.

Key factors behind growing influence of SERP Marketing: Above statistical study of the SERP Market and SEM shows that the grip of SERP marketing on e-marketing has seen exponential growth in last decades which is transforming the landscape of online marketing. This trend is driven by the increasing reliance of businesses on search engines as a primary channel for reaching and engaging with their target audiences. Here are key factors contributing to the exponentially **growing influence of SERP marketing** in the realm of e-marketing:

1. **Search Engine Dominance:** Google continues to dominate the search engine market with a significant share, making it the go-to platform for users seeking information and solutions. While Google is the leader, other search engines like Bing, Yahoo, Baidu, Yandex also contribute to the overall impact of SERP marketing.
2. **User Behaviour Shift (Search as a Starting with Search is a Folk Custom now):** A growing number of consumers start their online journey with a search query. Whether its product research, information gathering, or solving a problem, users often turn to search engines first as folk custom now. The increasing use of mobile devices further amplifies the importance of search, as users conduct searches on the go.

[239] https://moz.com/blog/state-of-searcher-behavior-revealed
[240] https://www.iab.com/wp-content/uploads/2016/04/IAB_Internet_Advertising_Revenue_Report_FY_2016.pdf
[241] http://en.wikipedia.org/wiki/Online_advertising#Delivery_methods

3. **Mobile-First Approach leveraging Local Searches:** The prevalence of mobile devices has led to a mobile-first approach in web design and marketing strategies. Mobile-friendly websites are prioritized by search engines, impacting SERP rankings. Mobile users often engage in local searches, influencing businesses to optimize for local SEO and enhance their visibility on local SERPs.
4. **SERP Features and Rich Snippets: Featured snippets** provide concise answers to user queries directly on the SERP, increasing visibility and authority for the source. Information presented in **knowledge graphs** enhances the user experience by providing context and related details.
5. **Paid Advertising Opportunities:** Paid advertising on SERPs, such as Google Ads, Bing Ads allows businesses to prominently display their products or services to users with specific intent. E-commerce businesses benefit from the visibility provided by shopping ads directly on the SERP.
6. **Local SEO and Google Business Profile:** Local businesses leverage **Google Business Profile** to enhance their visibility on local SERPs, especially for users searching for nearby products or services. Maps and local business information integrated into SERPs contribute to increased engagement with local businesses.
7. **User Intent Understanding with Semantic Search:** Search engines, particularly Google, have improved their understanding of user intent through **semantic search and natural language processing**. Marketers optimize content for **long-tail keywords** to align with specific user queries and capture niche audiences.
8. **Evolving Algorithms and Updates:** Ongoing algorithm updates, such as Google's BERT, Core Web Vitals and Self-learning of Search Engines through their “Search Intelligence” methods explained in above parts of this thesis aim to improve the understanding of context and nuances in user queries, enhancing the relevance of search results. Emphasis on user experience metrics in ranking algorithms encourages website optimization for better SERP performance.
9. **Multichannel Integration:** SERP marketing is integrated with other digital marketing channels, creating a **cohesive cross-channel online presence**. Social media, content marketing, and email campaigns often intersect with SERP strategies.
10. **User Trust and Credibility:** High SERP rankings are often associated with **credibility and authority signals** in the eyes of users. Businesses strive to secure top positions for increased trust. **Reviews and ratings** displayed on the SERP contribute to users' decision-making process and impact the reputation of businesses.

Discussion: The grip of SERP marketing on e-marketing is fuelled by the central role search engines play in users' online experiences. Businesses recognize the need to optimize their digital strategies for SERP visibility to effectively reach and engage their target audiences on this increasingly competitive online landscape. As search engines continue to evolve, the influence of SERP marketing is likely to remain a pivotal aspect of

comprehensive e-marketing strategies. The "Search Intelligence" programs talked above, and similar other inventions and updates search engines keep on experimenting, all together confirm that Search Engines have long life in coming era of GenZ *(we shall talk about it in next section of this chapter).*

4.6. SERP in Generation Zoomer (GenZ) and Generation Alpha:

After having this much foreground of our research context, it is reasonable here to see its importance not only from past backdrops but how it will be seen by current and future generations; and that is how we can establish the futuristic importance of our research work in this thesis with a NOVEL approach and psyche. As per [242]*Wiki-GenZ*, GenZ born between 1997 to 2012 and Generation Alpha born between early 2010s to mid-2020s. In the context of GenZ and ML, the evolution of SERP is influenced by the unique characteristics, behaviours, and expectations of this generation. Gen Z, born roughly between the mid-1990s and early 2010s, is known for being tech-savvy, digitally connected, and having distinct preferences in how they consume information. The evolution of search engines to meet the needs of Generation Z is a complex process, involving advanced models for **information retrieval**, *Jouis et al.* [2012], **natural language user interfaces, accurate search results, and interactive user interfaces**, *Yang et al.* [2000], and the development of more intelligent search engines that provide insights and knowledge, *Wei-Ying Ma* [2005]. These advancements are further enhanced by the extraction of data records from web pages using a **vision-based approach** and the exploration of **complementary properties between the surface and deep web**, *Ma et al.* [2004]. These developments collectively aim to provide more relevant, accurate, and intelligent search results, catering to the preferences and expectations of Generation Z. Keeping these expectations in purview, the next generation of search engines aims to provide highly personalized search results that satisfy individual information needs, *Li Zheng* [2014]. They are designed to identify relevant information in specific domains or topics and provide users with up-to-date and focused insights, *Serrano et al.* [2017]. These search engines also utilize domain **taxonomy** to generate actionable information and enhance user searching experiences, *Widyantoro* [2007]. Additionally, search engines are incorporating **collaborative, community-based ontology** creation environments to make the system scalable, *Steven Miller* [2016]. The next generation of search engines also focuses on data-driven innovations and the utilization of all available data to make better decisions, *Devarakonda et al.* [2011]. They aim to integrate and retrieve scientific data from various sources, ensuring accessibility and retrievability as technology evolves. Generation Z's search behaviour is showing a notable shift, with an increasing number of individuals turning to platforms like **TikTok** for search purposes. While Google remains a dominant force in

[242] https://en.wikipedia.org/wiki/Generation_Z

search, TikTok is gaining traction as a search engine among Gen Z. According to a survey, **51% of Gen Z women prefer TikTok over Google for search**, with 74% of Gen Z using TikTok search overall. The reasons cited include the **video format** of the results, more **relatable answers**, and **personalized content**, [243]*Danny Goodwin* [2023] and [244]*Matt G. Southern* [2024]. Additionally, a study by Adobe revealed that 40% of consumers, including Gen Z, are using **TikTok to search for various information, recipes, music**, and more, [244]*Matt G. Southern* [2024]. Despite these trends, it's important to note that TikTok is not positioned to replace Google as a search engine entirely. However, its unique format and personalized content delivery are resonating with users, particularly in the Gen Z demographic, [245]*Neal Sivadas* [2023] and [244]*Matt G. Southern* [2024]. This shift in search behaviour underscores the evolving landscape of search engines and the need for businesses to consider the growing influence of platforms like TikTok in their digital marketing strategies. Generation Z's evolving search behaviour is prompting search engines to adapt in several ways:

1. **TikTok's Influence:** Gen Z's increasing use of TikTok for search and product discovery is reshaping the search landscape. TikTok's **video format and personalized content** are resonating with this demographic, challenging traditional search engines like Google, [246]*Kitty Bates* [2002].
2. **Long-Tail Keywords:** Gen Z's preference for **long-tail queries** is impacting SEO. They use more specific, nuanced searches, influencing the choice of keywords and content optimization strategies, [247]*House-of-communication* [2023].
3. **Voice Search and Local SEO:** Gen Z's reliance on **mobile devices and voice assistants** is driving the need for mobile-optimized content and local SEO strategies to target this demographic effectively, [247]*House-of-communication* [2023].
4. **Content Prioritization:** Search engines, including Google, are adjusting their algorithms to prioritize original, helpful content over irrelevant information, aiming to meet the needs of Gen Z's efficient search behaviour, [246]*Kitty Bates* [2002].
5. **Diversification of Search Platforms:** The emergence of TikTok and other platforms as search engines is diversifying the search landscape, prompting businesses to consider alternative platforms for reaching Gen Z, [246]*Kitty Bates* [2002].

In response to these shifts, search engines are likely to continue evolving to meet the unique search behaviours and preferences of Generation Z. Search engines are catering to the mood and trends of the next generation by

[243] https://searchengineland.com/gen-z-tiktok-google-search-survey-431345
[244] https://www.searchenginejournal.com/tiktok-gains-traction-as-a-search-engine-among-gen-z-study/505633/
[245] https://www.linkedin.com/pulse/why-gen-z-uses-tiktok-search-engine-you-should-too-neal-sivadas
[246] https://mrs.digital/blog/tiktok-gen-z-search-engine/
[247] https://www.house-of-communication.com/uk/en/newsroom/2023/09/how-to-target-gen-z-via-search.html

customizing the user's experience based on their activities on social media sites. By analysing a user's posts and other **activities on social media**, search engines can infer the user's emotional state, current interests, likes/dislikes, and aspirations, *Hebenthal et al. [*2010]. To adapt to the changing interests of users, crawlers are being designed to track the **current pattern points** and download web pages that meet the user's present needs and current trends, *Gupta & Dixit* [2021]. Additionally, search engines are evolving to include **context-driven information supply**, where relevant information is provided without the need for an explicit query, and social search, where information is obtained by asking members of a web milieu rather than constructing elaborate queries, *Nilsson & Alanko* [2018]. These advancements in search engine technology allow for a more personalized and efficient search experience that caters to the mood and trends of the next generation, *Petrova & Trunin* [2020] and *Andrei Broder* [2007]. Search engines are adapting to user behaviour by integrating results from different collections and guiding users to relevant content, *Lewandowski* [2008]. The theory of planned behaviour suggests that individual satisfaction with search engines, their use as an information retrieval tool, and sharing search experience and information influence their intention to use search engines, *S. Liaw* [2004]. The Google Generation, or Generation Z, behaves differently on the internet, with **less confidence in searching, more cut-and-paste search statements**, and **poorer working memories and multitasking abilities**, *Nicholas et al.* [2011]. Here's how machine learning is shaping SERP for Generation Z:

1. **Enhanced Personalization:** Machine learning algorithms analyse user behaviour, search history, and preferences to deliver highly personalized SERP experiences for Gen Z individuals. SERPs may adapt based on past interactions, providing content that aligns with the user's interests and preferences.
2. **Conversational Search using NLP:** Gen Z tends to use more conversational and natural language in searches. NLP-driven algorithms enable search engines to better understand and respond to conversational queries. With the rise of voice-activated devices, machine learning enhances **voice search** capabilities, allowing for more accurate interpretation of spoken queries.
3. **Visual Search Integration:** Machine learning facilitates **image recognition** and **visual search** capabilities on SERPs. Gen Z, being visually oriented, benefits from the ability to search using images or by interacting with visual elements directly on the SERP. Visual search enhances **product discovery**, allowing users to find items similar to those in images, aligning with Gen Z's preference for visual content.
4. **Instant Gratification:** Machine learning algorithms prioritize speed and relevance, delivering **instant results** to cater to Gen Z's preference for quick information retrieval. SERPs may prominently **feature snippets** providing concise answers to common queries, aligning with Gen Z's desire for immediate and concise information.
5. **Interactive Rich Media Integration:** Machine learning enables the integration of **rich media**, such as interactive elements and videos, directly on the SERP. This aligns with Gen Z's preference for engaging

and interactive content. For e-commerce queries, machine learning may facilitate **virtual try-on experiences** directly on the SERP, enhancing the shopping journey for Gen Z.

6. **Privacy and Transparency:** Gen Z places a high value on privacy. Machine learning models are designed to respect user privacy, with features like **federated learning** that allow personalization without compromising individual data. Search engines may prioritize **transparency in how algorithms work**, building trust with Gen Z users who are concerned about how their data is used.
7. **Continuous Learning and Adaptation:** Machine learning models incorporate **user feedback** from Gen Z individuals to continuously adapt and improve the relevance of search results. Algorithms **learn from evolving user behaviours**, ensuring that the SERP stays attuned to the changing preferences and trends among Gen Z.
8. **Social Integration:** Machine learning algorithms may **integrate relevant social media content** directly into the SERP. This reflects Gen Z's strong presence on social platforms and their preference for socially curated information. SERPs may feature **recommendations from influencers** or social media personalities based on predictions of user interests.
9. **Cross-Channel Unified Experience:** Machine learning contributes to a unified digital experience by integrating SERP strategies with other channels, such as **social media, email, and content marketing**, catering to Gen Z's preference for seamless cross-channel interactions.
10. **Ethical AI Practices:** Machine learning models in SERPs strive to **mitigate biases**, ensuring fair and unbiased representation of information. This aligns with Gen Z's expectations for ethical AI practices. SERPs may prioritize **diverse and inclusive content**, reflecting Gen Z's emphasis on representation and inclusivity.

Discussion: Our study finds that the Search Engines are already acting pro in understanding and coping up with the expectations of Gen-Z. As Generation Z continues to shape the digital landscape, machine learning in SERPs evolves to meet their behaviours, and expectations. The integration of advanced technologies ensures that SERPs remain dynamic, personalized, and relevant, providing an optimal search experience for the next and next to next generations of users. Chart-72 below summarizes the facts and findings in above research.



Chart-72: The Searcher: Search Expressions and Operators

Unit-2: Research Framework

Synopsis: While Unit-1 gives realisation of a foreground coverage of **'Domain Level'** *(as per Table-4 of Introduction section)* diverse general literature around the SERP, which was necessary to dig in a broader and deeper understanding and knowledge of SERP in context of our thesis topic; Unit-2 aims to be more specific *(in manner of filtering using a V-shape funnel - refer to Chart-95) about* the **'Thematic'** of our thesis as mentioned in Table-4. This unit does not only present a critical review but sophisticated synthesis of the Literature primarily focusezd on the title of this thesis, which enable us to contretely reach the Research Scope for this thesis, thus ultimately to draw the research model to test. So, this unit is comprised of four chapters:

Chapter-5: Synthesized Literature Review: It gives emphasis to Literature of highest relevancy to further derive the raw-hypotheses and then out of them to choose the considerable hypotheses set of this research. Literature covered in Unit-1 stimulates this Chapter by providing the required foreground.

Chapter-6: Research Scope: Based on the considerable hypotheses figured-out in Chapter-5; this chapter sets the scope of each hypothesis derivative. So, you shall see the research scope validation, the most relevant 5 or 6 validated hypotheses chosen out of total considered. In the last part, this chapter talks about the statements of research motivation and goal(s), justification of our thematic research question and finally the research questionnaire to test the hypotheses set.

Chapter-7: Research Model, Methodology, Method, Metrics and Modalities (R5Ms): Based on the LR and Research Scope, five **'R5Ms Testable Frameworks'** are figured out in this chapter. All possible literary relevant research methodologies, methods, and models pertaining to our Research Model, the Research Problem and qualitative questionnaire are studied here. Then based on the contemporary models, resource's availability, thesis suitability, and internal and external dependencies the most compatible R5Ms Framework is finalized to derive the research test. A robust Research Model is designed to test and do the statistical calculations.

Chapter-8: Bibliographic Data Analysis: This chapter is comprised of a systematic analysis of five types of Literature Review viz-a-viz **FLR** *(Foreground Literature Review),* **SLR** *(Synthesized Literature Review),* **R5MLR** *(Research Methodology, Method, Model, Metrics and Modalities Literature Review),* **RPCLR** *(Research Problem Contemporary Literature Review)* and **ALR** *(Arbitrary Literature Review)* done for this thesis.

Chapter-5

Synthesized Literature Review

Synthesized here means the study and analysis of the literature as a vector to ML relation format, which should entice the reader's interest by fulfilling the purpose of reading or referring to this thesis. Just like we see in a 'Vector and Machine Learning Relation', the 'Foreground Literature Review (FLR)' can be seen as the Machine Learning Field and 'Synthesized Literature Review (SLR)' the specific Vector Set(s) of that ML field. So, in the context of this thesis 'FLR is a broader foreground of SERP and 'SLR is particularly for the theme of this thesis as a subset within FLR. FLR aspects shall be used in references in this chapter wherever applicable. In a nutshell, where FLR represents 'Domain of this thesis', SLR intends to elaborate its 'Thematic'.

5.1. Knowledge Gaps and Further Scopes - From overall SERP Perspective:

This section aims to figure out the hypothesis derivatives for this thesis based on two stakeholders the web searcher and the search engine. Web Searcher for the direct behavioural characteristics generation and search engine for assisting these characteristics as a technical facilitator.

Derivatives draining from the direct behavioural characteristics of a web searcher:

A) Domain Knowledge Pursuance and Perceived Relevance *(Obtaining SERP Granularity and falling into Filter Bubble)*: *Schultheib & Lewandowski* [2021]-1 find that users who have little knowledge of the result types are more likely to trust and favour Google than users with greater knowledge. *Schultheib & Lewandowski* [2021]-2 establish in their research that users with a low level of knowledge on search advertising are more likely to click on ads than users with a high level of knowledge. Also, the users with little knowledge show less willingness to scroll down to organic results. Chart-73, Appendix-9 by *Schultheib* [2023] states in his SEM (Paid Ads) research model that the overall Knowledge Gain from a SERP navigation is influenced by the quality of information and position of the result in SERP, which a user with Search Engine Literacy can take better benefit of. Then domain knowledge derives the user in their selection and information retrieval behaviour. While *Schultheib* is prominently focused on paid results, this thesis aims to cover organic and paid both. [248]*Tegze* [2023] states that overtime grown cognitive understanding about search

[248] https://newsletter.fullstackrecruiter.net/p/how-filter-bubble-affecting-your-search

process and task decision approach has the tendency to land you in **filter bubble** if not pursued on changes and updates. Knowing about the habitude of French HISB with changes and updates and tendency to escape the filter bubble would be interesting for this thesis. Following one-liners are drawn from above synthesis of Domain Knowledge around SERP:

1. Less knowledge more dependency on Google.
2. Less knowledge, less willingness to scroll organic results.
3. Cognitive understanding without pursuing updates lands you in filter bubble.
4. Higher domain knowledge and SE literacy means careful reading.
5. Perceived relevance capacity comes from domain knowledge and SE Literacy.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Studying the habitude of French HISB with changes and updates, their cognitive tendency to escape the filter bubble, how knowledge level of domain differentiates different HISBs.
2. **Viability:** Impact of technology change on new generation and their foresightedness about future Healthcare Information user.

Based on above synthesis, Table-21 below presents a matrix view of Gap Types versus Future Scopes on aspect of Domain Knowledge Pursuance and Perceived Relevance in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	Yes	Organic Results Surfing	*Schultheib* [2023]	Studying the habitude of French HISB with changes and updates, their cognitive tendency to escape the filter bubble, how knowledge level of domain differentiates	Impact of technology change on new generation and their foresightedness about future Healthcare Information user.
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	Young Age, technology and knowledge	[248]*Tegze* [2023]		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	No	N/A	N/A		

Empirical (Evaluation Void)	No	N/A	N/A	different HISBs.	
Theoretical (Application Void)	No	N/A	N/A		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	Yes	Knowing the filter bubble in HISB	[248]*Tegze* [2023]		

Table-21: Knowledge Gaps and Future Scopes - Domain Knowledge Pursuance and Perceived Relevance

B) Biasing in Searcher's Behaviour *(Biographical, Social, Cultural and Economic Status impact on SERP Navigation Behaviour)*: Chin et al. [2015] while found in their research that older adults do more clicking exploitation on websites opened from SERP and less exploration on SERP compared to younger adults, they express the need of further research in examining the engagement versus exploration of results in different contexts and age groups. *Tennant et al.* [2015] state that being younger, more educated with higher Socio-Economic Status means high probability of using internet more for HISB. *Kopeinik et al.* [2023] try to articulate the sources of biasing behaviour originated from the search user's end, specifically due to stereotypical gender representations by users in formulating online search queries. **Cognitive, Social Norms** *(norms and patterns of discrimination)* and **Algorithmic** biasing are the three major source of biasing their work is focused on. This shows the need of evaluating the cognitive biasing in SERP clicking behaviour. Their research gives the scope to how users can be informed about bias-mitigating query strategies during their interactions with search engines challenge. *Bitirim & Ertugrul* [2020] based on three demographic characteristics gender, age, and program studying *(i.e. female, male, under-20, 20-24, above-24, English computer engineering, Turkish computer engineering, software engineering)* conducted research on selected SERP features with 12 close-ended questions found that most of respondents have rarely or more click for tab, spelling suggestion, related query suggestion, instant search suggestion, video result, image result and pagination, except Advertising and amount of clicking results. Between 78% to 85.4% respondents click 10 or fewer results. This shall be interesting to study the same audience set in French HISB. *Karanam et al.* [2017] found that young users make better click behaviour on difficult task than older who make lesser clicks and their explorations recommend working on Latent Semantic Analysis (LSA) similarity value of search results with query, which can give augmented relevancy of results and query for each other. From here, an efficient **hypothesis** can be derived saying "Number of search tasks in a given search journey, articulates the task

difficulty level and user behaviour, intent depth and information urgency". Following oneliners are drawn from above synthesis of Biasing in Searcher' Behaviour around SERP:

1. Old age users do more clicking exploitation than exploration, while younger users do more exploration than exploitation.
2. Being younger, more educated and high Socio-Economic Status tends to high probability of using internet for HIS.
3. Cognitive, Social Norms and Algorithmic biasing render biasing in Search Behaviour.
4. Student age searcher's shows very varied behaviour in clicking specific SERP features.
5. Youngers tender better clicking in difficult search tasks than elders.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Knowing about exploration vs. exploitation clicking, task level browsing behaviour and social-SES-cognitive biasing in French HISB.
2. **Viability:** Finding into the 'escaping from search filter bubble' tendency of French HIS.

Based on above synthesis, Table-22 below presents a matrix view of Gap Types versus Future Scopes on aspect of Biasing in Searcher's Behaviour in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	Yes	Exploration vs. Exploitation finding in French HISB	*Chin et al.* [2015], Bitirim & Ertugrul [2020]	Knowing about exploration vs. exploitation clicking, task level browsing behaviour and social-SES-cognitive biasing in French HISB.	Finding into the 'escaping from search filter bubble' tendency of French HIS.
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	SES, Cognitive and Social Norms behaviour in French HISB	*Tennant et al.* [2015], *Kopeinik et al.* [2023]		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	Yes	Difficult vs. Easy Search Tasks testing in French HISB	*Karanam et al.* [2017]		
Empirical (Evaluation	No	N/A	N/A		

Void)					
Theoretical (Application Void)	No	N/A	N/A		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	No	N/A	N/A		

Table-22: Knowledge Gaps and Future Scopes - Biasing in Searcher's Behaviour

C) Search Query Scope and SERP Utility: *Yamin & Ramayah* [2013] divides the query in two types as **breadth query and depth query**. Breadth query strategy is a broad usage of query. The query formulated is general, wide and not specific to the domain. Depth query strategy is a narrow usage of query. The query is narrowed into the domain, and the use of keyword is more specific towards the search task. It is assumed in their finding that the users switched their strategy from broad to narrow to achieve their goal. Going by the flow, *Yamin et al.* [2014] state that the series of query stored in the transaction log of a search engine represent the users' behaviour *(selecting, changing, and modifying the query)* during the searching. They further state that formulating a search query is a challenging task for most of users because they are required to express their information need to search system. [249]*Sikora et al.* [2005] state that the breadth of the search structure representations did not vary considerably among search topics, the depth varied substantially. *Liu et al.* [2019] state that according to the cognitive viewpoint of information retrieval (IR) research, a search task can be conceptualized as a sequence of information seeking intentions which both motivate and are influenced by search behaviours. To develop a more comprehensive, multi-level *(i.e., task level, intention level, and behaviour level)* understanding of web search, the authors analysed intention and search behaviour data collected from 693 query segments generated by 40 participants and found that, 1) Task features significantly affected the frequency of occurrence of most of the information seeking intentions, and these effects gradually faded away as search sessions proceeded; 2) The presence of a variety of intentions in both current and subsequent query segments were connected with and detectable by different subsets of behavioural measures. Based on their findings, it might be obvious to say that the breadth search query search number is moderately higher compared to depth search query. In our thesis this connects to the Search skills and their intangible behaviour. It shall be interesting to know, what type of search queries are typed more in HIS users in France. *Tung et al.* [2019] work on three user behaviour factors of search efforts, content-triggers and application context in naturalistic information search tasks establishes four findings as **'search activity is integrally associated with user' creative processes'**. Searching within intellectual and creative tasks is more time-

[249] https://humanlinkllc.com/trends-in-web-searching-task-specific-characteristics/

intensive, while search activity occurring as a part of daily routine tasks is associated with more frequent searching within a search task. Third, searching is more often induced from utility applications in tasks demanding a degree of intellectual effort. Fourth, users' leisure information-seeking activity is occurring inherently within social media services or comes from social communication platforms. The kind of tasks and applications distributed, studied in Chart-74, Appendix-9 can be a complementary reference to data collection model of this thesis. By referring to *Marchionini* [2006] for its two broad categories of search tasks as **look-up search and exploratory search**; *Medlar & Glowacka* [2020] assert in their research that users performing exploratory search and lookup search exhibit different behaviours, enabling them to be distinguished using a classifier trained on easy to measure interaction variables. These two types of search categories motivate us to set same kind of approach to check its standing in our context audience market as iterated above. In another similar kind of user behaviour modelling by *Hendahewa & Shah* [2015] propose a simulation framework for evaluating exploratory search based on implicit features and user search action sequences extracted from the transactional log data to model different aspects of exploratory search namely uncertainty, creativity, exploration, and knowledge discovery. Their research shows that there is a substantial improvement in the user's performance in reaching the task specific goals met by the proposed recommendation strategy. This strategy can be tested in healthcare environment as people seek good recommendations with ultra serious attention. *Russo & Russo* [2020] puts an offer of empirical research around the intrinsic limitations of SEs and cognitive biasing of searchers around health topics. They advocate to reveal the possibilities of "**filter bubble**" a user can surround due to semantics used by a search engine and the possible biased use of that SERP under the influence of knowledge and presumptions a searcher may have in its cognitive set up. This study indeed proclaims a highly interesting problem if solved could straight so many search issues. Ours aim is nearby to this especially studying the cognitive biasing due to searcher' previous knowledge and searching pervasiveness. *Lewandowski & Kammerer* [2020] present a narrative literature review of **eye-tracking** studies examining three factors as contextual factors, resource factors and individual factors. The most important finding related to contextual factors is that task type does not seem to have a systematic effect on users' viewing behaviour, or that it is at least difficult to systematically investigate such task effects *(this thesis aims to fill this gap with possible efforts)*. However, contrary to it, the findings of *Sendurur & Yildirim* [2015] demonstrated that search patterns can change according to types of tasks, especially the number of page visits and correct hits, the way keywords used, and task completion success differed across task type. Although the statistical analysis did not generate significant relationships, qualitative analysis demonstrated that time spent, scan type, number of keywords, and reading styles differed for all tasks. As per *Hienert et al.* [2018] users show different behaviour for different task types, e.g., for fact finding vs. information gathering tasks. Daniel et al. emphasize the role of the task topic - in and of itself - over parameters such as the search results and read content pages, dwell times,

session variables and subjective measures such as task difficulty or task success. Both references verify the importance of giving more attention to the task topic as an important measuring factor for knowing user behaviour. Regarding resource factors, the most crucial overall finding still is that when search results are presented in the form of lists, search results at the top of the list receive most attention. Concerning effects of individual factors on users' viewing behaviour on SERPs, both age and prior topic knowledge seem to be related to a more systematic and more careful reading of the SERPs. So, we can lead from here that prior domain/topic knowledge and Search Results positioning are of highest impact in influencing user search behaviour. *Singer et al.* [2012] investigated the characteristics of complex information needs in the context of web search engines, the possible simple measures to define these complexities, predictability of search task success and the different behaviours of successful and unsuccessful searchers. The results confirm that complex tasks show significantly different characteristics than simple tasks. Successful search behaviour can be distinguished from unsuccessful search behaviour. They conclude that complexity can be expressed by the effort *(in terms of time, sessions, queries, and browser tabs) needed* to carry out a search task. We can derive one hypothesis from here that how health information seekers behave on being successful versus unsuccessful in their search. Following oneliners are drawn from above synthesis of Search Query Scope and SERP Utility:

1. Old age users do more clicking exploitation than exploration, while younger users do more exploration than exploitation.
2. Breadth query is a general broad usage non-specific to domain search strategy.
3. Depth query is a narrow specific use of keywords search strategy.
4. The breadth search query search number is moderately higher than depth search query.
5. A searcher can be surrounded in a 'search filter bubble' due to semantics used by SEs, search bias due to own knowledge influences and cognitive presumptions.
6. Search activity is integrally associated with user' creative processes.
7. Search efforts, content-triggers and application context are the search tasks in a naturalistic information seeking.
8. Searching within intellectual and creative tasks is more time intensive.
9. Search activity occurring as a part of daily routine tasks is associated with more frequent searching within a search task.
10. Searching is more often induced from utility applications in tasks demanding a degree of intellectual effort.
11. Users' leisure information-seeking activity is occurring inherently within social media services or platforms.
12. Searcher performing Look-up search and Exploratory search exhibit different behaviours.
13. Different aspects of exploratory search are uncertainty, creativity, exploration, and knowledge discovery.

14. Contextual scenario does not seem to have a systematic effect on users' viewing behaviour.
15. In resource scenario search results presented in list top receive most attention.
16. In individuality scenario both age and prior topic knowledge seem to be related to a more systematic and more careful reading of SERP.
17. Complex tasks show significantly different characteristics than simple tasks.
18. Search Task complexity can be expressed by the efforts in terms of time, sessions, queries, and browser tabs used to carry out that search.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Is it depth or breadth queries typed more in HISB, how searcher's creativity play role in concrete HIS, studying contextual and cognitive biasing in French HISB, how searchers react in successful versus unsuccessful search task.
2. **Viability:** Searcher creativity level escapes them from filter bubble, so knowing creativity in search can help in profound B2B as well as B2C strategies.

Based on above synthesis, Table-23 below presents a matrix view of Gap Types versus Future Scopes on aspect of Search Query Scope and SERP Utility in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	Is it depth or breadth queries typed more in HISB, how searcher's creativity play role in concrete HIS, studying contextual and cognitive biasing in French HISB,	Searcher creativity level escapes them from filter bubble, so knowing creativity in search can help in profound B2B as well as B2C strategies.
Knowledge (Information / Industry / Upgradation / Technology / Data)	No	N/A	N/A		
Practical (Mechanism Conflict)	Yes	Task Types and difficulty levels in HISB	*Lewandowski & Kammerer* [2020]		
Methodological (Better Alternate Method)	No	N/A	N/A		
Empirical (Evaluation Void)	No	N/A	N/A		

Theoretical (Application Void)	Yes	How filter bubble function in HISB	*Russo & Russo* [2020]	how searchers react in successful versus unsuccessful search task.	
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	Yes	Searcher' creativity in HISB	*Tung et al.* [2019]		

Table-23: Knowledge Gaps and Future Scopes - Search Query Scope and SERP Utility

D) Result Ranking and Search Behaviour: Joakim [2020] while studying the impact of **ML Algorithms** (Word2Vec) **training** output, load and accuracy with Search Engine's **relevance score** compared to standard TF-IDF search engine and about its comparison to collaborative filtering; he found that both ML and Filtering impact measurably for smaller depth size of 3 to 5. Talking about the future research around ML Algorithms, he expressed the scope in experimenting with diverse domains, using multiple data sources, measuring execution, training and build times for training ML Algorithms. Ullah et al. [2018] studied 6 Page Ranking Algorithms comparatively and express the need of discussing the shortcoming of these algorithms. The studied algorithms were 'Particle swarm optimization', 'Distance rank algorithm', 'Generic Algorithm', 'Ontology page ranking', 'Weighted PageRank' and 'Reranking' algorithms. Ullah et al. state that one cannot rank the different ranking algorithms in terms of performance as different approaches of page ranking algorithms are suitable for different applications, and we need different algorithms for offline page ranking and online page ranking. So here it can be interesting to research about the specific ranking algorithms for Healthcare Results Ranking. *Pawitra* [2015] investigated the effect of the number of search results and their ordering on search satisfaction and behaviour. The participants' behaviour revealed that they were influenced by what they believed to be true about the ordering of returns in such a way that they would select options displayed early on the search results list when they perceived the ranking was reliable. From here we drive a probable variable to state that the ordering of results is treated proportionally to user's belief or to say user's prior knowledge and perceived perception of that subject. It shall be interested to investigate this fact in HIS in France. Following oneliners are drawn from above synthesis of Result Ranking and Search Behaviour around SERP Utility:

1. ML and Filtering algorithms training impacts SE's relevance score for smaller depths of 3 to 5.
2. One cannot rank the different ranking algorithms in terms of performance as different approaches of page ranking algorithms are suitable for different applications.
3. We need different algorithms for offline and online page ranking.

4. Users would select options displayed early on the search results list when they perceived the ranking was reliable.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** What are the specific ranking algorithms and their functioning for Healthcare Results Ranking. How French HISB happens around ordering of results is treated proportionally to user's prior knowledge and perceived perception of that subject.
2. **Viability:** These finding can help the scientists and strategists in articulating more robust search algorithms.

Based on above synthesis, Table-24 below presents a matrix view of Gap Types versus Future Scopes on aspect of Result Ranking and Search Behaviour in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	What are the specific ranking algorithms and their functioning for Healthcare Results Ranking. How French HISB happens around ordering of results is treated proportionally to user's prior knowledge and perceived perception of that subject.	These finding can help the scientists and strategists in articulating more robust search algorithms.
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	Prior knowledge and perceived perception	*Pawitra* [2015]		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	Yes	Algorithmic improvising in HISB	*Joakim* [2020]		
Empirical (Evaluation Void)	No	N/A	N/A		
Theoretical (Application Void)	Yes	Different ranking for different algorithms. Different algorithms for online and offline content in HISB.	*Ullah et al.* [2018]		
Population (Audience)	No	N/A	N/A		

Conceptual (Ambiguity)	No	N/A	N/A		

Table-24: Knowledge Gaps and Future Scopes - Result Ranking and Search Behaviour

E) User Gain Expectancy *(Web Searcher's browsing behaviour of reaching to desired result and stopping the SERP navigation)*: *Di Caprio et al.* [2022] analyse the effects that the position of the alternatives ranked *(different results)* by a search engine and the relative impatience of users have on their information retrieval behaviour. They observed it through their two mimicking impatience algorithms; the first version assumes that users proceed sequentially through the ranking until they find an alternative satisficing their expectations. Once they find a satisficing alternative, they continue retrieving information until they observe an alternative that violates their expectations. The second version increases users' impatience, who stop retrieving information as soon as an alternative does not satisfy their expectations - even if it is the top-ranked one. According to them future research should aim at validating the combinatorial optimality of the behaviour observed, particularly when accounting for the evaluation of different subsets of alternatives within those composing the initial page of search results provided by an engine. Here we see scope of testing this aspect in context of the theme of this thesis. From the findings of *Ahn et al.* [2018] a new research scope evolves for strengthening the understanding about the user search attention capturing variations. SERP diversity captures more user attention, that attention gets renewed on new alternatives countering, after that attention decays rapidly. *Linus Philip* [2015] revealed that visual hierarchy, competition for attention and banner blindness, are three important factors that may influence transactional search behaviour. He further found that the way visual hierarchy is organised, is not directly related to transactional search behaviour. Next, he found that the banner blindness phenomenon only exists to a certain extent and search behaviour is only slightly influenced by this. Overall, no major differences were found in the transactional search behaviour for two different search engine interface designs. Instead, Linus found a lot of similarities, although the approaches for presenting the SERPs varied. This leads us to the conclusion that fundamentals of searcher's behaviours are responsible for how users view transactional search engine interfaces. Similar experiment can be tried for navigational and informational user behaviour. Adding to foray of Linus, *Azzopardi et al.* [2018] propose a new measure based on Information Foraging Theory, which can account for the heterogeneity of elements *(web results, Ads, specialized answers),* through different costs, and which naturally motivates the development of a user stopping model that adapts behaviour depending on the **rate of gain** with potential to deliver better results that fit to observed behaviours such as stopping rank, time on SERP and inferred utility i.e. rank, time, gain respectively. They define the SERP surfing behaviour on two lines, one as 1-1-1-1 and second as 2-1-2-1 while traversing through core and right rail of SERPs. It entices the idea to study the heterogeneity of French Health SERPs and user behaviour on

them about their ranking, stopping rank, time and/or gain rate. *Zanjani & Khadivi* [2014] states that new objects such as user, ad or query reduce precision of targeted advertising, this finding forages the idea of checking similar traits around over all SERP in a different audience set like this thesis. *Schultheib & Lewandowski* [2021]-2 demonstrate that Ads on top in mobiles get more visual attention than desktop, this can be attributed to larger aperture and space to show multiple items in above-the-fold screen on larger screens like desktop. In context of our thesis, we can relate it to behaviour about the organic results as well. *Arguello* [2014] work in Search Task difficulty assessment through predictive models reveals about the features generated from scroll and mouse-movement events on the SERP in three contexts as user interaction, user interest and user prior knowledge of task domain. His study found that the most predictive features were different for whole-session vs. first-round prediction, that mouseover features were effective for first-round prediction, and that level of interest and prior knowledge features did not improve performance. Interesting would be to study these factors for predicting the search task difficulty in French HISB. *Eldad & Elad* [2021] in their research on self-stated autism bounded search user's SERP investigative behaviour found that the differences in scrolled distance were larger for popular queries be it textual query or image query. This gives scope to a new possible hypothesis to test for health queries about how people scroll on popular health queries and medium or least heard health queries. *Wolfram* [2013] states that the browsing habits of users are largely uniform, regardless of the investment in individual query formulation, or the number of query topics undertaken per user session. Internet users on their search habits *(Logan & Driscoll-Eagan, 1998; Spink, Bateman & Jansen, 1998)* indicate that users stop searching for a variety of reasons, including **frustration or satisfaction** with the results. Wolfram further states that based on the uniformity of browsing habits, users **expect quick retrieval** and are not willing to engage in additional browsing to satisfy their information needs, or it is possible that users are coming up with sufficient numbers of relevant hits in the first few pages, making additional browsing unnecessary. This study asserts the impatient search behaviour and expectations of the searcher. A kind of hypothesis derivable from this study can state the expectations of the quality results in quicker time possibility flaunts searchers to jump on top results. By large human tendency is to expect best from their search habits bubble, majority of times squeezed to few search tactics set. Ultimately to find how much patience and pursuance HIS adopt in comparison of other domain searchers. Following oneliners are drawn from above synthesis of User Gain Expectancy around SERP Utility:

1. Users proceed sequentially on SERP until a result *(alternative)* satisficing their expectation is found.
2. Users continue retrieving information until they observe a result *(alternative)* violates their expectations.
3. A result not satisfying user expectation, even if top-ranked; it increases user' impatience to stop IR.
4. SERP diversity captures more user attention, that attention gets renewed on new alternatives countering, after that, attention decays rapidly.

5. Visual hierarchy, competition for attention and banner blindness influence transactional search behaviour.
6. Visual hierarchy is not directly related to transactional search behaviour.
7. Banner blindness phenomenon only exists to a certain extent and slightly influences the search behaviour.
8. There are no major differences in transactional search behaviour across search engines.
9. Heterogeneity of SERP elements alters user behaviour in terms of stopping rank, time on SERP and user gain rate.
10. How new SERP objects entry reduces precision of ad targeting.
11. Ads on top in mobiles get more visual attention than desktop due to small aperture of screen.
12. In Search Task difficulty prediction, the most predictive features are usually different for whole-session vs. first-round prediction.
13. In Search Task difficulty prediction, the mouseover features are effective for first-round.
14. In Search Task difficulty prediction, the level of interest and prior knowledge features don't improve performance.
15. Differences in scrolled distance were larger for popular queries be it textual query or image query.
16. Browsing habits of users are largely uniform, regardless of the investment in individual query formulation, or the number of query topics undertaken per user session.
17. Users stop searching for a variety of reasons, including frustration or satisfaction with the results.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** How French HISB reacts to results violating user expectations for scrolling further or stopping IR. Fundamentals of searcher's behaviours are responsible for how users view SE interfaces. How heterogeneity of SERP impacts the stopping rank, time spent and user gain rate. Knowing Search Task Difficulty Prediction for French HISB. French HIS scrolling for popular, medium, least heard queries. Patience and pursuance adopted in French HISB.
2. **Viability:** Knowing the User Gain Expectancy opens up the gates to explore into real time user search pattern.

Based on above synthesis, Table-25 below presents a matrix view of Gap Types versus Future Scopes on aspect of User Gain Expectancy in context of SERP:

Relevant Knowledge Gaps				Viable Future Scope	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**

Evidence (Contradiction or Counter)	Yes	Entry of new objects in SERP and HISB change	*Zanjani & Khadivi* [2014]	How French HISB reacts to results violating user expectations for scrolling further or stopping IR. Fundamentals of searcher's behaviours are responsible for how users view SE interfaces. How heterogeneity of SERP impacts the stopping rank, time spent and user gain rate. Knowing Search Task Difficulty Prediction for French HISB. French HIS scrolling for popular, medium, least heard queries. Patience and pursuance adopted in French HISB.	Knowing the User Gain Expectancy opens up the gates to explore into real time user search pattern.
Knowledge (Information / Industry / Upgradation / Technology / Data)	No	N/A	N/A		
Practical (Mechanism Conflict)	Yes	Information Foraging Theory and Stopping IR in French HISB	*Di Caprio et al.* [2022], *Linus, Azzopardi et al.* [2018], *(Logan & Driscoll-Eagan, 1998; Spink, Bateman & Jansen, 1998)*		
Methodological (Better Alternate Method)	Yes	Search Attention Renewal in French HISB	*Ahn et al.* [2018]		
Empirical (Evaluation Void)	No	N/A	N/A		
Theoretical (Application Void)	Yes	Scrolling distances for popular, medium and least heard queries in French HISB	*Eldad & Elad* [2021]		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	Yes	Banner blindness and results hierarchy in French HISB Desktop vs. Mobile attention on top results	*Linus Philip* [2015] *Zanjani & Khadivi* [2014]		

Table-25: Knowledge Gaps and Future Scopes - User Gain Expectancy

F) Query Drifting and User Interest Decay: Xu et al. [2023] in their research based on PGMs for defining their method Counter Factual Explanation with Editing ***(CFE2)*** tested the effectiveness of producing

SERP by editing the original query on its counter facts and achieved significant improvements. They propose to further test their work for the end use-case of query suggestion/reformulation. *Kiseleva et al.* [2015] studied detrimental cases behind failing SERPs due to temporal drift and dynamic nature of the content using their conceptual behavioural model and proposed to **predict query drift** beyond current SERP. They define four types of drifts as sudden, incremental, gradual and reoccurring as shown in Chart-75, Chart-76 and Chart-77 in Appendix-9. Sudden drift occurs on for example breaking news queries and varies from few hours to few days on timeline, having search abandonment rate highest. Incremental drift occurs on less popular tail types of queries, volume of queries remains almost stagnant, search abandonment is frequent initially. Gradual drift can be seen as a combination of Sudden and Incremental both. Such changes occur on secondary news and events around a public entity. Reoccurring change has a regular type of behaviour, like movie release events. If query reformulation grows over time, it is taken as a positive change while if it decreases then it is counted as negative. They expressed about the intricate dependencies of their model on continually changing ranking, personalization, customization, and localization, where our research has the potential to fill the gap by knowing situation of all four query drift types in context of French HISB. *Yi Song* [2019] studies the calculation and method of user interest similarity to discover user groups with similar interest. **User interest is not constant for a long time.** User short-term interest accumulated to a certain extent may evolve into long-term interest, which requires us to learn and update user interest model. If not updated, user interest drift may occur. This research gives a probable scope to testing user interests change, and in what situations how a user interest sustains for longer in French HISB, is it service quality or self-updating on interest. *Zerhoudi et al.* [2021] use Markov approach in which they utilise the context discovery process to model user-type specific behaviour and search intent type by capturing the user's query change in a search session. They found that a search session consists of a sequence of queries where the query count may change depending on users' intent shift. They explored two different techniques as possible contexts for the domain: (1) classifying users according to their navigation behaviour into two categories: exploratory and lookup, and (2) grouping users by learning a mixture of intent-level query change characteristics into five categories *(i.e., specification, generalisation, word substitution, word repetition and new topics)*. From here one possible hypothesis for this thesis can be studying the 'intent shifts in HISB' through query change behaviour. *Kazutoshi* [2016] discussed techniques and methodologies for estimating search intents and for evaluating search outcomes in the context of complex search tasks. He focused on searchers' gazing behaviour on their browsed web pages and assumed that terms drawing a high degree of a searcher's attention are likely to be relevant to his/her search intent. He also found that the following behavioural features were most effective for the subsequent query reformulation type prediction: query term length, attention area in web pages, and clicks on search results, based on which he further classified problem with five query reformulation types *(i.e., Generalization, Specialization, Parallel*

Move, Mission Change, and Error Correction). Here he ascribed to further research possible about re-ranking search results containing more attention drawing terms to higher positions. *Shi et al.* [2014] state that SE users usually have a higher probability of leaving when loyalty is found to be continuously decreasing through their work on loyalty and defection behaviour of search user across multiple search engines through their LoyaltyTracker *{consisted of flow metaphor (flow-view) analytics, density map and word cloud}* with which they proposed an interactive visualization to track user loyalty and switching behaviour towards multiple search engines from the vast amount of user log data by integrating it with density map and word cloud to study the patterns in flow-view, see the Chart-78, Appendix-9. They demonstrate another crucial use of their invention in tracking and analysing the dynamics of user engagement in e-learning courses from web log data. Healthcare being YMYL *(Your Money Your Life)* domain of search, Loyalty in medical products and services is expected very high by customers, so defection of search can be directly related to it and can be included in test of this thesis. *Chen et al.* [2021] conducted a field study to analyse fine-grained user reformulation behaviours including reformulation type, entry, reason (*Chart-79, Appendix-9*), and the inspiration source *(Chart-80, Appendix-9)* with various search intents. Their results demonstrate that query reformulation behaviour in web search varies with the type of search tasks. They further found that the current query suggestion/related query recommendations provided by search engines do not offer enough help for users in complex search tasks. They further state that user intent and domain expertise will influence their search process. For instance, a broad intent may cause a search process with more exploration. A clear intent, on the contrary, can lead to more exploitation. Testing about what kind of health queries lead to exploitation vs exploration can be of interest to this thesis. In what ratio health query intents are broad vs. clear can be another question to answer in this thesis. *Zhitomirsky et al.* [2016] investigate 'wisdom of crows' about whether and how the user preferences change over time in accordance with the ranking of search results according to their relevance to the user needs in aggregated consensus judgments or individual user judgements. Their results show that consensus ranking calculated by aggregating the individual user rankings and relevance judgments resulted in substantially fewer changes compared to the averaged individual user rankings. In other finding quite low similarity was found between the search engines' and consensus users' rankings. This research may have practical implications for personalisation of SERP in context of the theme of this thesis, to state that as users' preferences change over time, therefore the ranking of a search engine should adapt to this. *Granka et al.* [2004] did an eye-tracking study about how does rank influence the amount of attention a link receives and how do users explore the list on SERP. They found that time fixated on result at rank one and two are almost equal. After the second link the fixation time drops off sharply. The abstracts ranked 6 to 10 receive approximately equal attention. This can be explained by the fact that typically only the first 5-6 links were visible without scrolling. Once the user has started scrolling, rank becomes less of an influence for attention. A sharp drop occurs after link 10, as ten

results are displayed per page. For the second question about exploring, they researched about how many results above and below the selected document users scan on average. Here they found that, users who selected the lower ranked documents viewed proportionately more abstracts overall. The number of links viewed below a click is low beyond rank 1, indicating that users do tend to scan the list from top to bottom. According to *Nettleton & Gonzalez-Caro* [2012] in their Eye-Tracking SERP analysis about successful and unsuccessful search found that users who are successful formulate fewer queries per session and visits a smaller number of documents than unsuccessful users. Measuring cognitive efforts through fixation frequency and fixation duration is more related with the strategy of search of the user than with the difficulty of the search task. About the scanning behaviour, the search strategies of the successful users are more focused and systematic than of the unsuccessful. *Strzelecki* [2020] skimmed through 54 research papers from 2004 to 2019 particularly on Eye-tracking methodologies used in evaluation of different aspects of searcher's SERP behaviour. So far none of these papers was from French Market; however, Lacoste-badie et al. have conducted several Eye-Tracking research but all other than SERP domain. Following oneliners are drawn from above synthesis of Query Drifting and User Interest Decay around SERP Utility:

1. Editing original query on counter facts can yield improved SERP.
2. Sudden Query Drift occurs on breaking news queries varying from few hours to few days, they have highest search abandonment rate.
3. Incremental Query Drift occurs on less popular tail types of queries, volume of queries remains almost stagnant, search abandonment is frequent initially.
4. Gradual Query Drift can be seen as a combination of Sudden and Incremental both, occurring on secondary news and events around a public entity.
5. Reoccurring Query Drift has a regular type of behaviour, like movie release events.
6. If query reformulation grows over time, it is taken as a positive change while if it decreases then it is counted as negative.
7. User interest is not constant for a long time.
8. User interest duration depends on user's appetite for learning and self-updating.
9. If a user interest is not updated, it becomes prone to interest drift.
10. Terms drawing a high degree of a searcher’s attention are likely to be relevant to his/her search intent.
11. Five query reformulation types are Generalization, Specialization, Parallel Move, Mission Change, and Error Correction.
12. Query reformulation prediction factors can be length of the searched query, attention area in web pages, clicks on search results.
13. SE users usually have a higher probability of leaving when loyalty is found to be continuously decreasing.

14. Query reformulation behaviour in web search varies with the type of search tasks.
15. Query suggestion/related query recommendations provided by search engines do not offer enough help for users in complex search tasks.
16. User intent and domain expertise influence the search process.
17. A broad intent may cause a search process with more exploration.
18. A clear intent can lead to more exploitation.
19. Consensus ranking resulted in substantially fewer changes compared to averaged individual user rankings.
20. There is quite low similarity between the search engines' and consensus users' rankings.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Query Drifting in French HISB. How French HISB sustains interest and falls into drift in interest. Intent shifts in HISB through query change process. Re-ranking results containing more attention drawing terms to higher positions. Loyalty and defection in French HISB. Query suggestions use rate in HISB. Broad intent vs. Clear intent in HISB. Queries leading to exploitation vs. exploration in HISB.
2. **Viability:** In user's query shift behaviour, re-ranking higher with attention drawing terms in SEO.

Based on above synthesis, Table-26 below presents a matrix view of Gap Types versus Future Scopes on aspect of Query Drifting and User Interest Decay in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	Query Drifting in French HISB. How French HISB sustains interest and falls into drift in interest. Intent shifts in HISB through query change process. Re-ranking results containing more attention drawing terms to higher positions. Loyalty and defection in	In user's query shift behaviour, re-ranking higher with attention drawing terms in SEO.
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	Query suggestions significance in HIS	*Chen et al.* [2021]		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	No	N/A	N/A		
Empirical (Evaluation Void)	Yes	Query drifting in HIS	*Kiseleva et al* [2015]		

Theoretical (Application Void)	No	N/A	N/A	French HSIB. Query suggestions use rate in HSIB. Broad intent vs. Clear intent in HSIB. Queries leading to exploitation vs. exploration in HSIB.	
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	Yes	User interest decay in health SERP	*Yi Song* [2019]		

Table-26: Knowledge Gaps and Future Scopes - Query Drifting and User Interest Decay

G) Clicking Intent Intensity: *Ashkan & Clarke* [2012] in their work on development and evaluation of an ontology for commercial search studied the characteristics of query categories *(location of ad, rank of ad and query category)* in reference to CTR on Ads links, along with the query intent impact. Their finding said that the query-based features, along with the content of SERPs, are effective in detecting query intent. Factors such as query category, the rank of an ad, and the total number of ads displayed on a result page relate to the context of the ad, rather than its content. This finding gets us to assuming a possible hypothesis stating that the intensity of intent is directly proportional to likeliness of clicking a result conditioned to relevancy and quality of that result. *Shen & Martin* [2010] give a measurement of **'loyal users in search engine'** based on the length of their queries as short, medium or long; and whether their clicking is related to high CTR or previous search experiences. Subjectivity of the user behaviour through granular analysis talks about their knowledge of their subject. They define seven user categories regarding overall access times (OAT) as one off (of), infrequent (inf), occasional (occ), intermittent (int), frequent (freq), very frequent (vf) and loyal (loy). As per their findings overwhelmingly the users in all categories click top- and high-ranking results. When submitting short length queries, frequent and very frequent users are more likely to click top ranking results while loyal users are more interested in high-ranking results. When submitting medium length queries, it is interesting to note that loyal users are most likely to click high ranking results than top ranking results. On relation of user knowledge, CTR and query length, they observe that loyal users when submitting medium length queries are very likely to achieve high CTR *(with confidence value 0.74)* by clicking high ranking results. They are also less likely to be affected when medium length queries failed to return satisfactory answers. From these findings, we can derive a hypothesis that SERP personalisation can be achieved by knowing the search user's loyalty index. *Zhang & Moffat* [2006] in their research on query log of search.msn.com find some crucial relations between user search behaviour, queries, sessions and CTRs. Their findings say that long queries have a smaller CTR. User will sometimes take long jumps between consecutive clicks and are also almost as likely to move backward as forward. Users dislike going beyond the first results page. Users can make quick decisions about pages they

have clicked on. From this study, interesting would be to know if the CTR behaviour vs query length of 2006 has changed in 2024 considering updated technological upbringing and generation shift of users, that can derive a new hypothesis calling upon with the statement that the technology and generation change also changes the user search behaviour traits in comparison of previous generation. *Bai et al.* [2017] study on Yahoo Web Search query log data shows that the users of a fast search system are more sensitive to delays than the users of a slow search system. The study further finds that users are more likely to notice the **response latency delays** beyond a certain latency threshold, their search experience potentially being affected. This analysis demonstrates the significant change in click behaviour as the response latency increases, users are more likely to perform clicks on the result page that is served with lower latency. The user study revealed that up to a certain point *(500ms)* the added response time delays are not noticeable by the users. However, after a certain point *(1000ms),* the users could feel the added delays with very high likelihood. Additional analyses indicated that the degree to which response latency affects the click behaviour varies with the user, query, and context attributes. *Su et al.* [2018] investigated the relation between user intents, search behaviour and perceived satisfaction in the context of three types of product search as Target Finding (TF), Decision Making (DM) and Exploration (EP). TF users tend to issue few specific queries and browse only top ranked results; DM users tend to issue short queries, browse deep, and click more results; and EP users issue many diverse queries, browse deep, but do not click often. It shall be quite interesting to find the similar line set of Intent Types in HISB on French Health audience set. Following oneliners are drawn from above synthesis of Clicking Intent Intensity around SERP Utility:

1. Information displayed on SERP is contextual to query rather than its content.
2. Query-based features along with the content of SERPs are effective in detecting query intent.
3. 7 types of SE Loyals are one off, infrequent, occasional, intermittent, frequent, very frequent and loyal.
4. Frequent, very frequent and loyal users give highest CTR on medium length queries.
5. All 7 types of SE Loyals click top- and high-ranking results.
6. Long queries have a smaller CTR.
7. Users take long jumps between consecutive clicks and are likely to move backward as forward.
8. Users dislike going beyond the first results page.
9. Users can make quick decisions about pages they have clicked on.
10. Searchers of a fast search system are more sensitive to delays than the users of a slow search system.
11. Up to 500ms response time delays in SERPs are not noticeable, after 1000ms users could feel the added delays with very high likelihood.
12. Searchers are more likely to perform clicks on SERP that is served with lower latency.
13. Target Finding users tend to issue few specific queries and browse only top ranked results.

14. Decision Making users tend to issue short queries, browse deep, and click more results.
15. Exploration users issue many diverse queries, browse deep, but do not click often.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** SERP personalization relation to loyalty index in French HISB. Intent intensity relation to clicking likeliness in HISB. Testing the temporal behavioural change in HIS. SERP response latency in HIS. How different searchers issue queries.
2. **Viability:** Knowing the French HISB query issuing pattern can help businesses in creating content accordingly.

Based on above synthesis, Table-27 below presents a matrix view of Gap Types versus Future Scopes on aspect of Clicking Intent Intensity in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	Yes	SE loyalty in French HISB	*Shen & Martin* [2010]	SERP personalization relation to loyalty index in French HISB. Intent intensity relation to clicking likeliness in HISB. Testing the temporal behavioural change in HIS. SERP response latency in HIS. How different	Knowing the French HISB query issuing pattern can help businesses in creating content accordingly.
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	Clicking Intent likeliness and SERP display latency	*Zhang & Moffat* [2006] *(MSN Study); Bai et al.* [2017] *(Yahoo Study)*		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	No	N/A	N/A		
Empirical (Evaluation Void)	No	N/A	N/A		
Theoretical (Application Void)	No	N/A	N/A		
Population (Audience)	Yes	Temporal behaviour change	*Zhang & Moffat* [2006] *(MSN Study)*		
Conceptual (Ambiguity)	No	N/A	N/A		

				searchers issue queries.	

Table-27: Knowledge Gaps and Future Scopes - Clicking Intent Intensity

H) SERP Personalization Impact *(on Search Behaviour)*: *Salehi et al.* [2020] states in their research that personalisation of search results does not increase users' satisfaction with their search results. However, it does reduce the time spent by users *(students in their study)* in locating information they judged to be satisfactory answers to their questions. Health being related to user's life, so they are super sensitive and critical about HIS comparative to other domains; so, it would be interesting to test how SERP Personalization alter their satisfaction quotient in context of the theme of thesis. A probable hypothesis can be counter and critical to what Salehi et al. say and that is "SERP personalization doesn't reduce time spent on SERP in HIS". Following oneliners are drawn from above synthesis of SERP Personalization Impact around SERP Utility:

1. Personalisation of search results does not increase users' satisfaction with their search results.
2. SERP Personalisation reduces time spent by users in locating information they need.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Does SERP personalization reduces time spent on SERP in HIS. SERP personalization can be critical for HISB.
2. **Viability:** Personalization of Health SERP, necessary or over exaggerated thing.

Based on above synthesis, Table-28 below presents a matrix view of Gap Types versus Future Scopes on aspect of SERP Personalization Impact in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	Does SERP personalization reduces time spent on SERP in HIS. SERP personalization can be critical	Personalization of Health SERP, necessary or over exaggerated thing.
Knowledge (Information / Industry / Upgradation / Technology / Data)	No	N/A	N/A		
Practical (Mechanism Conflict)	No	N/A	N/A		

Methodological (Better Alternate Method)	No	N/A	N/A	for HISB.	
Empirical (Evaluation Void)	No	N/A	N/A		
Theoretical (Application Void)	Yes	Criticality of Health SERP personalization	*Salehi et al.* [2020].		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	No	N/A	N/A		

Table-28: Knowledge Gaps and Future Scopes - SERP Personalization Impact

Derivatives draining from the technical facilitations to SERP by the Search Engines:

I) Semantic Clustering and IR: Gong & Cosma [2023] researched on combining Visual Semantic Embedding ***(VSE)*** network and Vision Transformers with Relation-focused ***(VITR)*** learning with GPT-3.5 turbo large language mode and defined a novel cross-model search engine named BOON, enabling image-to-text, text-to-image retrieval tasks. They found improvements in search engine accuracy by this association. They hint on use of BOON with Bing and Google for experimenting more around ITT and TTI retrieval more relevantly. *Ngo & Nguyen* [2004] discuss an algorithm for web search results clustering based on Tolerance Rough Set that uses clustering techniques for grouping similar documents together to facilitate presentation of results in more compact form and enable thematic browsing of the results set. They further expressed the scope in improving the representations, which have positive effects on clustering quality. Philipp & White [2017] investigated potential strategies to mine queries and searcher histories for clues that could help search engines choose the most appropriate information to present in response to exploratory medical queries. They raised the need for detailed study of the impact of SERP composition on users' future behaviour. *Nakamuraa et al.* [2018] while proposing a search engine using modern neural network techniques, express that we believe deep learning centric systems are going to be more and more important in the coming years, and their work is one significant step towards exploring and understanding that world. However, they equally accept that neural NLP models lack in intuition interpretability and debuggability, important aspects of commercial solutions. *Bae et al.* [2020] work on machine-human interactive clustering finds that one should consider **interactive clustering** as a discipline that combines machine learning/data mining and Human-Computer Interaction. Talking of the room for future work, they state that there is a need for developing solutions where the machine would initiate quality improving operations, for example, indicate to the users' specific clusters that require feedback in a form of a

query, and the users would provide information based on such requests. This challenge is noticed, and as the demand for user involvement grows, many researchers mention user study as an area in need of future work. *Zamani et al.* [2020] worked on better information retrieval using large scale search log analysis based on more clarifying user questions arising from their queries with possible multiple search intents. They used expert annotation method prepare the training data as solution to this issue, like crowdsourcing used by leading SEs. This gives scope to collection of users' feedback for improvements in IR, using user histories is another scope in obtaining more coherent results. Borisov et al. [2016] express the need of extending their work on distributed representations for user behaviour modelling beyond the work they did on **relevance prediction** both in ranking documents and setting documents relevancy to search query. They recommend experimenting about user actions like clicking on ads, zooming mobile results, reforming the query, audio queries, image search and mouse movements. Kumar [2022] concludes in his research that some people believe Web 3.0 is where computers will generate new information, rather than humans. Others feel Web 3.0 is all about the Semantic Web and personalization. In our model we want to connect **semantics with predictability and personalization** both. So, taking further up on his observation, we feel that the SERP is a joint output of Web 3.0 activities like Indexing, Schema, Semantics, Blogging, Chatting, Commenting and Social Media. Rashid & Wasif [2016] research about the viability of semantic search engine ***(SSE)*** over conventional ones, they talk about technology used for SSEs, semantic searching about SSEs with examples. While they find that SSEs yield unique search experiences rich in results relevance and efficiency; they expressed the need of further research on use of SSEs, their features, advantages and developing **SSE technology** capable of meeting the challenges efficiently and their compatibility with global standards. *Chen & Décary* [2018] while working on fine-tuning of search describe further scope of working on achieving true machine text understanding using NLP and Cognition frame for better relevance, usability, and learnability. They express the need to incorporate a refine-grained semantic analysis method developed for analysing lexical elements and semantic structure of a sentence. Mostafa & Gwizdka [2016] use three **Neuro-Physiological Evidence methods Eye Tracking, EEG, and fMRI** for Improving Search. In their future recommendations they emphasize on need of standardising search tasks, their evaluation, combinations, and transparent research. *Serrano & Gelenbe* [2018] proposed an Intelligent Search Assistant ***(ISA)*** that uses a Random Neural Network ***(RNN)*** to record, reorder and present results based on perceived relevance to user needs, outperforming commercial search tools. Their ISA adapts and learns from user previous relevance measurements *(based on Gradient Descent for quality and Reinforcement Learning for higher learning rate)* significantly increasing its quality and improvement within the first iteration. As per *Priyadharshini et al.* [2018] when it comes to IR, almost all the previous papers work at the level of query or session *(query or session trails)*. Their research focused on the prediction of user behaviour on the task level by implementing automated clustering and query suggestion which clusters the

given user queries into tasks from web search log. The task clustering efficiently produces the task for the given set of queries under the user identification. So is for the query suggestion which gives efficient output for the given list of words. Following oneliners are drawn from above synthesis of Semantic Clustering and IR around SERP Utility:

1. Three Neuro-Physiological Evidence methods for improving search in standardising search tasks are Eye Tracking, EEG, and fMRI.
2. Number of search tasks hints the task difficulty, user behaviour, intent depth and information urgency.
3. Neural NLP models lack in intuition interpretability and debuggability, important aspects of commercial solutions.
4. Interactive clustering is a discipline that combines machine learning/data mining and Human-Computer Interaction.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Using Neuro-Physiological Evidence methods Eye Tracking in French HISB for improving search in standardising search tasks, their evaluation, combinations, and transparency.
2. **Viability:** Knowing the SERP browsing approaches of Gen-Z.

Based on above synthesis, Table-29 below presents a matrix view of Gap Types versus Future Scopes on aspect of Semantic Clustering and IR in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	Using Neuro-Physiological Evidence methods Eye Tracking in French HISB for improving search in standardising search tasks, their evaluation,	Knowing the SERP browsing approaches of Gen-Z.
Knowledge (Information / Industry / Upgradation / Technology / Data)	No	N/A	N/A		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	Yes	Eye Tracking in Health SERP on French HISB Search Log Analysis for	Mostafa & Gwizdka [2016], Zamani		

		French HISB	et al. [2020], Priyadharshini et al. [2018]	combinations, and transparency.	
Empirical (Evaluation Void)	No	N/A	N/A		
Theoretical (Application Void)	No	N/A	N/A		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	No	N/A	N/A		

Table-29: Knowledge Gaps and Future Scopes - Semantic Clustering and IR

5.2. Knowledge Gaps and Further Scopes - From Health SERP Perspective:

Pre-Web HIS (Health Information Seeking) **Outcomes and Methods:** Our aim starts with developing the understanding of Health Information Seeking Behaviour ***(HISB)*** pre-web-traditional-era then traversing through initial web-era of Web 1.0 & 2.0 and reaching to the ultra-web era of Web 3.0; bigger aperture of which is covered in Literature Review Unit above. Here we shall try to gauze the knowledge gaps by figuring out the scopes, this thesis can fulfil particularly around the thematic question of this thesis. *Liu & Zou* [2023] study extended the literature by demonstrating the psychological effect of **discrete outbreak events** such as of COVID-19 on health information search. They found that discrete outbreak events could increase people's search for typical symptoms. Their study shows that as severe as the health issue like Covid is, it results into bigger stresses and health concerns and when you are nearer to outbreak area it manifolds the psychological stress. Taking a cue from this study about the highest urgency of health information search, a gazing scale can be defined from highest urgency to normal and least urgent health issues. For understanding the kind of information seeking, as per *Lambert & Loiselle* [2007] several studies reviewed measured the influence of HISB on individuals' health-related search outcomes. Commonly, outcomes or consequences of seeking health information include:

1. **Cognitive outcomes**, such as increasing knowledge *(Andreassen et al., 2005; Muha et al., 1998),* informed decision making *(Davison et al., 2002; Muha et al., 1998; Warner & Procaccino, 2004),* increasing perception of control *(Echlin & Rees, 2002),* and coping *(Edgar, Remmer, Rosberger, & Fournier, 2000).*
2. **Behavioural outcomes**, including discussing information obtained with health care professional *(Andreassen et al., 2005; Czaja et al., 2003; Muha et al., 1998),* increased self-care abilities and adherence

to treatment *(N. Gray et al., 2005),* and change in health behaviour *(Shi et al., 2004; Szwajcer et al., 2005; Warner & Procaccino, 2004).*

3. **Physical outcomes**, such as increase physical quality of life *(Ransom et al., 2005).*
4. **Affective outcomes**, including decrease anxiety, fear, and distress *(Brereton & Nolan, 2002; Huber & Cruz, 2000)*, increase hope *(Huber & Cruz, 2000)* and empowerment *(N. Gray et al., 2005).*

Throughout the literature, two main dimensions of HISB emerge:

1. **The information dimension:** The information dimension emphasizes the characteristics of the information sought, particularly in terms of **type and amount**. The type refers to the **content and diversity** of the search. The amount refers to **how much information** (details) about a given topic one seeks, **underlining the depth of the search**. Individuals have been found to vary greatly along this dimension:
   a. Some might search a lot of health-related information on a wide array of topics *(Clark, 2005; Echlin & Rees, 2002; Leydon et al., 2000; Szwajcer et al., 2005).*
   b. Some might choose to seek little or no health-related information *(Case et al., 2005; Echlin & Rees, 2002; Longo, 2005; Szwajcer et al., 2005).*
   c. Some individuals might seek health-related information only on a particular issue but avoid other types of health-related information *(Friis, Elverdam, & Schmidt, 2003; Leydon et al., 2000; Loiselle, 1995).*
   d. Some might seek general information, whereas others prefer details or specifics *(Ford, Wilson, Foster, Ellis, & Spink, 2002).*
2. **The method dimension:** The method dimension of HISB focuses on the discretionary actions individual use to obtain health related information and sources of information used. Discrete information seeking activities or strategies include:
   a. Direct and indirect questioning *(e.g., Borgers et al., 1993; Brashers, Goldsmith, & Hsieh, 2002; Feltwell & Rees, 2004; Johnson, 1997).*
   b. Asking for clarifications *(e.g., Beisecker & Beisecker, 1990).*
   c. Discussing and exchanging information with others *(e.g., Beisecker & Beisecker, 1990; Friis et al., 2003; Matthews, Sellergren, Manfredi, & Williams, 2002).*
   d. Reading *(e.g., Brereton & Nolan, 2002; Feltwell & Rees, 2004).*
   e. Observing *(e.g., Brereton & Nolan, 2002).*
   f. Use of a third party *(e.g., Johnson, 1997).*
   g. Browsing *(e.g., Johnson, 1997).*
   h. Listening *(e.g., Brereton & Nolan, 2002).*

Most often, individuals seek health-related information, at any given time, from a combination of personal *(e.g., self, friends, family)* and impersonal *(e.g., book, Internet)* sources. The use of multiple sources might reflect individuals' desire to acquire as much information as possible *(J. Brown et al., 2002; Shuyler & Knight, 2003)* and/or to validate *(R. Gray et al., 1998; Muha, Smith, Baum, Maat, & Ward, 1998)* or complement information received from a prior source *(Brereton & Nolan, 2002; Fleming, Goodman, Geraghty, West, & Lancaster, 2002; Muha et al., 1998). Johnson (1997)* referred to the sources of information an individual consults to obtain information as their information field. Most individuals indicate a preference for health professionals when seeking medical facts *(Andreassen et al., 2005; J. Brown et al., 2002; Johnson, 1997; Loiselle, Semenic, Côté, Lapointe, & Gendron, 2001; Warner & Procaccino, 2004).* This might reflect individuals' belief that professionals can provide unbiased, reliable information that is in their best interest *(J. Brown et al., 2002; Gollop, 1997; James, James, Davies, Harvey, & Tweddle, 1999).* Individuals also report a preference for other sources of information, such as friends or others experiencing the same health issue, when seeking psychosocial information *(Beresford & Sloper, 2003; Dunne, 2002)* or the Internet when seeking sensitive information *(anonymity provided) (N. Gray et al., 2005). Rusova et al.* [2022] collected 32428 responses in Czech Republic around how and why people search for health information. Around 66% people search on Search Engines with around 49% searching it for self while 20% do it for their kins, rest sought it for studies or research. This finding again asserts the strong relativity between search engines and HISB. *Schoenherr* [2022] states that a recent survey found that 81% of US residents surveyed indicated that they relied significantly on their own research when making major life decisions. In contrast, only 31% indicating that they relied on expert advice. Schoenherr presents in a **Sociotechnical** approach that requires interaction and integration of humans and technology in a systematic manner as shown below in Chart-81, Appendix-9. The macro-social level of Chart-81 includes IT infrastructure of a society, Search engine algorithms, censorship of information and ethical AIS policies. And Search engine algorithms are what research of this thesis comprised of. Schoenherr has recently suggested that a therapeutic alliance will necessarily develop between users and AI systems that they interact with. Consequently, developers of **search engines** must account for this trust and the concomitant distribution of responsibility in their design. We intend to contribute to this regard on any proportional stance. Following oneliners are drawn from above synthesis of Pre-Web HIS Outcomes and Methods around SERP Utility:

1. As severe the health issue like Covid found as would be the psychological stress and health concerns and it manifolds in outbreak areas.
2. 4 HISB outcome types: Cognitive, Behavioural, Physical and Affective.
3. 2 dimensions of HISB: Information and Method.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Defining a health urgency severity scale in French HISB.
2. **Viability:** Knowing the health severity points can help in mitigating bigger mass risks.

Based on above synthesis, Table-30 below presents a matrix view of Gap Types versus Future Scopes on aspect of Pre-Web HIS Outcomes and Methods in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	Defining a health urgency severity scale in French HISB.	Knowing the health severity points can help in mitigating bigger mass risks.
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	Health urgency severity scale in French HISB	*Liu & Zou [2023], Rusova et al.* [2022]		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	No	N/A	N/A		
Empirical (Evaluation Void)	No	N/A	N/A		
Theoretical (Application Void)	No	N/A	N/A		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	No	N/A	N/A		

Table-30: Knowledge Gaps and Future Scopes - Pre-Web HIS Outcomes and Methods

HIS in Web Era: Moving from pre-web knowledge on HISB, we enter in post-web era. As per *Jacobs et al.* [2017] one in three US adults use the internet to diagnose or learn about a health concern. Jacobs et al. research on whether web is reducing or creating disparities in health information availability and use for making health decisions indicated that there is a possibility that while the web is an easily available source of health information, it could also create inequalities in health information accessibility. This research also found that there are **digital disparities in health information seeking** especially with regards to age and Socio-Economic

Status ***(SES)*** *(Lorence & Park, 2007; Massey, 2016; Rains, 2008).* The study results also highlight that being **younger, more educated,** and having a **higher SES** were predictors of internet use for health information seeking *(Tennant et al., 2015).* Being **older, having low internet skill, lower SES and being Hispanic** were determinants of using a health care provider or traditional media, such as print and magazines, as source of health information. **Digital disparity** which is another reason widening the gap of information access *(Hargittai & Hinnant, 2008; Lorence & Park, 2007; Massey, 2016).* Chart-82, Appendix-9 shows that Internet has become the most preferred HIS platform post web invention. And that is how this aspect entices this thesis to include this topic in its title. *Lambert & Loiselle* [2007]; *Zimmerman & Shaw* [2020] take it to Information and Communication Technologies ***(ICTs)*** contributing to the maturity of the concept of HISB. As per *(Deng & Liu, 2017; Roettl et al., 2016)* the accessibility of online health information has changed the way people engage in the health decision making process. The inclusion of online communities and health information seeking online indicates the impact ICTs have had on the conceptual understanding of HISB *(Baumann et al., 2017; Gonzalez et al., 2016).* Technological advances such as virtual and augmented reality give many evidence and inefficient scales of requirement of further research in HISB through ICTs. According to review of *Jia-2 et al.* [2021], 83% of the health information consumers searched health information using a general search engine, such as Google and Yahoo, while 15% of the online seekers tended to search on specific health information websites. Their review shows the need of studying HISB in specific demographic groups, different subject groups, geographic groups, economic groups, and their comparison. Future studies can conduct comparative behavioural similarities and differences. *Kerstin Horch* [2021] also produces similar kind of behavioural trends in her KomPaS *(Communication and Patient Safety in German Language*) study done in Germany stating internet is the first choice with 26% of all of the media types available to search or obtain health information and people in the low socioeconomic status group search the Internet for health information less often than people in the high or medium status groups. *Mirzaei et al.* [2021] in their research review list out the most central predictors of HISB being age, education, gender, health condition, and financial income. Specifically, they have been able to develop a sociodemographic group of predictors from **5 predictors** *(age, gender, education, race, and health literacy)* to **16 predictors** *(caregiver, employment, household, language, sexual orientation, finances, societal engagement, and location of residence).* *Chang & Huang* [2020] mentions of four complementary characteristics of health information sources i.e. **access to medical expertise, tailorability, anonymity, and convenience** are crucial in individuals' processes of acquiring health information. They further expressed the need of exploring more of methodological variables that exhibit consistency in conceptualization and operationalization regarding present results to ultimately deal with the **statistical biases**. We aim to answer some of these biases through our variable model. *Kim et al.* [2023] find in their research that people who seek health information from social media tend to have more experiences of

encountering Human Papillomavirus (HPV) vaccine misinformation, while no significant relation between the two. This study also found that people with above mentioned encounter are more likely to use professional health websites for same purposes. However, there was no relationship between exposure to misinformation on social media and the use of search engines for health information seeking, which leaves the scope for another research. From the perspective of this thesis, this is evident from above finding that **people still trust the Search Engines more than social media.** Following oneliners are drawn from above synthesis of HIS in Web Era around SERP Utility:

1. Web may create inequalities in health information accessibility.
2. There are digital disparities in HIS especially with regards to age and Socio-Economic Status.
3. Being younger, more educated, and having higher SES are predictors of internet use for HIS. Health
4. Being older, having low internet skill, lower SES and Hispanic are determinants of using a health care provider or traditional media, such as print and magazines, as source of health information.
5. Digital disparity widens the gap of information access.
6. Internet has become the most preferred HIS platform post web invention.
7. 83% of searchers use SEs like Google and Yahoo for HIS, while 15% use specific health websites.
8. People in low socioeconomic status group search the Internet for health information less often than people in the high or medium status groups.
9. Most central predictors of HISB are age, education, gender, health condition, and financial income.
10. Four complementary characteristics of health information sources are access to medical expertise, tailorability, anonymity, and convenience.
11. People who seek health information from social media tend to have more experiences of encountering misinformation.
12. People with misinformation encounter on social media are more likely to use professional health websites for same purposes.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Digital disparities and inequalities of SERP in French HIS.
2. **Viability:** Knowing SERP disparities can help SERP content creators to create smart content.

Based on above synthesis, Table-31 below presents a matrix view of Gap Types versus Future Scopes on aspect of HIS in Web Era in context of SERP:

Relevant Knowledge Gaps				Viable Future Scope	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	Yes	Disparities in French Health SERP	*Jacobs et al.* [2017]	Digital disparities and inequalities of SERP in French HIS.	Knowing SERP disparities can help SERP content creators to create smart content.
Knowledge (Information / Industry / Upgradation / Technology / Data)	No	N/A	N/A		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	No	N/A	N/A		
Empirical (Evaluation Void)	No	N/A	N/A		
Theoretical (Application Void)	No	N/A	N/A		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	No	N/A	N/A		

Table-31: Knowledge Gaps and Future Scopes - HIS in Web Era

Web HISB Characteristics: *Pang et al.* [2014] put a HISB framework *(borrowed from Alzougool et al. 2013, see Chart-83, Appendix-9)* in their research about 'Exploratory HISB'. They state that individuals who need to handle health issues of themselves or their loved ones are more likely to use the exploratory approach to search. Other factors such as contradictory information, curiosity and task clarity also lead to exploratory health search. During the search process, an exploratory seeker has no specific targets but keeps learning new search keywords and information. Moreover, exploratory health search alternates with other approaches for searching, depending on the uncertainty level of the seeker. They state that very few research has been done in this regard. This fact gives a base to one of our hypotheses that "Life/health care or threat urgence of correct information leads people to do exploratory searches more". As per Chart-83 such kind of searching behaviour falls under recognized and demanded information seeking need. *Song et al.* [2022] state in their research that misinformation density in health SERPs affects people's **search activities** *(e.g., how they formulate queries and click results),* their **correctness of answering** factual questions and their perceived **goodness of the search systems**. As their audience were 60 students in China, so they recommend expanding the audience number and

market. *Neeragatti et al.* [2023] combines Digital Health Information Searching ***(DHIS)*** and Unified Theory of Acceptance and Use of Technology ***(UTAUT)*** to evaluate the impact of later the former. In their findings about search mediums for Health information, Google/another browser stands at 50.70%, YouTube at 27.7%, Facebook 7.3%, Health Apps 12.9% and others 1.40%. They tested these results with their **intent variables** model shown in Chart-84, Appendix-9. As the future scope in their research, they emphasize to try the test on different demographics. Selection of determinants can be increased towards usage of DHI. The study also suggests the extension with usage behaviour of DHI and social media influence as determinants. In their recommendation the **behaviour aspect can be relevant to this thesis**. *Weimann-Saks* [2022] states that differences will be found in **people's attitudes towards the disease**, depending on whether knowledge was obtained from internet sources or from a credible source. The attitudes of the former about the disease will be more negative *(the disease seen as more serious)* than among those of people that obtained their knowledge from a credible source. The study groups they chose were very limited in size and different motivations; both criteria should be tried with different contexts. This study is somehow counter to curiosity and assumption of HISB as positive, interesting can be to see where the SERP stands in this comparison between Internet and Credible Source. *Schoenherr & Thomson* [2020] demonstrate that there are strong correlations between information search behaviour and mobility around the time of **public health announcements** suggesting that, directly or indirectly, health communication was associated with changes in behaviour. They relate this behaviour to Health Behaviour Model adapted from Chart-85, Appendix-9 and *Rosenstock* [1974] and public announcements and instructions issued by governments and politicians during Covid-19 phase. We also intend to study the effect of similar announcements found in SERPs. *Paunic et al.* [2023] in their critical analysis of clinical trial websites' search engines, revealed **pitfalls in the clinical trial search system**, such as dysfunctional search filters, inconsistent results across different websites, outdated trial information, and most importantly, lack of reproducibility as well as a lack of thoroughness of search results in Canadian market. This signals a similar scope of verifying in another audience set. *Liu & Xu* [2020] finds in their research that information search models based on Information retrieval algorithms produce a relatively ideal performance on relevance, readability, and credibility. They state about a challenge regarding improving healthcare information search for laypersons still eminent despite various IR technologies trials using NLP and ML; it shows scope to refine these technical aspects to yield SERPs evaluable by a layperson. *Chiu et al.* [2022] research reveal about the **HISB of laypersons**. It has been indicated that patients use online medical information to integrate with advice from friends, family, and physicians to confidently make their medical decisions. They further state that supporting laypersons in exposing to online health information through complex and simple search approaches, can in turn induce them to consult medical experts about such information. Following oneliners are drawn from above synthesis of Web HISB Characteristics around SERP Utility:

1. Individuals who handle health issues of themselves or their loved ones are more likely to use the exploratory approach to search.
2. Contradictory information, curiosity and task clarity lead to exploratory health search.
3. An exploratory SERP seeker has no specific targets but keeps learning new search keywords and information.
4. Misinformation density in health SERPs affects people's search activities, their correctness of answering factual questions and their perceived goodness of the search systems.
5. In health information search mediums in Google/another browser stands at 50.70%, YouTube at 27.7%, Facebook 7.3%, Health Apps 12.9% and others 1.40%.
6. Depending on whether knowledge was obtained from internet or credible source, attitude of the former about the disease will be more negative compared to later.
7. Public health announcements put direct impact on information search behaviour.
8. Pitfalls such as dysfunctional search filters, inconsistent results, outdated information, lack of reproducibility and thoroughness of SERPs distracts HIS.
9. Patients use online medical information to integrate with advice from friends, family, and physicians in order to confidently make their medical decisions.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** French HISB with exploratory intent. SERP credibility in HISB. Public announcements impact on French HISB.
2. **Viability:** Exploring new characteristics of HISB.

Based on above synthesis, Table-32 below presents a matrix view of Gap Types versus Future Scopes on aspect of Web HISB Characteristics in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	French HISB with exploratory intent. SERP credibility in	Exploring new characteristics of HISB.
Knowledge (Information / Industry / Upgradation / Technology / Data)	No	N/A	N/A		

Practical (Mechanism Conflict)	No	N/A	N/A	HISB. Public announcements impact on French HISB.	
Methodological (Better Alternate Method)	No	N/A	N/A		
Empirical (Evaluation Void)	Yes	Exploratory intent and HISB	*Pang et al. [2014]*		
Theoretical (Application Void)	No	N/A	N/A		
Population (Audience)	Yes	Public announcement impact on French HISB	*Schoenherr & Thomson* [2020]		
Conceptual (Ambiguity)	No	N/A	N/A		

Table-32: Knowledge Gaps and Future Scopes - Web HISB Characteristics

DHL (Digital Health Literacy) **Impact on HISB:** *Htay et al.* [2022] examined in their research the association of DHL with information seeking behaviour and satisfaction regarding Covid-19. Regarding the HISB most respondents used search engines and social media. They state that DHL enables people to critically evaluate the information in hand. People should be encouraged to utilize credible sources such as websites of public bodies, news portals *(e.g., newspapers, national TV stations)*, websites of doctors/ insurance companies, and official health portals for COVID-19 related information searching. *Ni et al.* [2024] describes **eight common sources of information** in their research related to HISB, out of which the **internet** was the most frequently mentioned source of information, while **face-to-face communication with healthcare professionals** was the preferred source. Associated factors were categorized into six categories as **demographic characteristics, psychological aspects, role-related or interpersonal traits, environmental aspects, source related characteristics, and disease-related factors.** Moreover, the results showed **five types of HISB** among people with inflammatory bowel disease (IBD), including **active searching, ongoing searching, passive attention, passive searching, and avoid seeking**. Notably, active searching, especially social information seeking, appeared to be the predominant common type. They expressed the necessity for repeated surveys to explore evolving HISB trends in evolving digital era. *Ghezzi et al.* [2019] in their research on privacy ethics and quality biasing of SERPs information for commercial or political purpose state that SEs have a major role in spreading low-quality health information such as that of anti-vaccine websites. Their results show that "alternative" search engines *(Duckduckgo, Ecosia, Qwant, Swisscows and Mojeek)* may return more anti-vaccine pages (10-53%) than Google.com (0%). The Chart-86, Appendix-9 shows the percentage of vaccine-negative webpages in different SERPs, out of it, they found Google to be hardly bothered about being

ethical, while the alternative search engines are better fulfilling this responsibility. Our thesis finds this information in relation to the "low quality SERPs usually displayed by Search Engine can probably take Searchers into wrong direction".

Khamsum et al. [2014] state that users' cognitive styles influence their information searching strategies, query reformulation behaviour, web navigational styles and information processing approaches. A probable hypothesis scope in context to this thesis from this fact can be **finding the relation between cognition capacity of a HIS and their SERP surfing behaviour**. *Yan Zhang* [2013]'s research on consumer health information searching indicated that **health information searching is as much social as it is private,** as participants not only relied on search engines, but also consulted their close social ties and healthcare providers, stating it as an exploratory search type. According to Zhang health searchers perform various cognitive activities, including building mental models, testing hypotheses, comparing, and validating information, and taking mental notes, to process the information that they found. Chart-87, Appendix-9 is the Zhang's model of health information search using which he states that people use Google the most as first information source, then comes the family & friends, at third number comes the medical websites and doctors. *Alfano et al.* [2020] did research on possibility of facilitating the search of online health information in context of language complexity based on specific user requirements. Their research proposes the use of **schema structured data** in combination with the evaluation of the term familiarity index, which has led to a mapping between the language complexity user requirements and different types of audiences yielding more satisfying results. They emphasize the further need of more comprehensive set of tests need to be undertaken for evaluating more effectively the relation between language complexity levels and the different audience types. It gives ground to another possible hypothesis as "The capacity of understanding complex languages in HIR *(Health Infromation Retrieval)* is directly proportional to education and cognition level of a HIS *(Health Information Searching)*". This gives scope to our "lingual command" variable. *Tuliao et al.* [2022] researched about 'search strings' made by searchers on what information **mental health seekers** look for in a service provider through SERPs. Their content analysis indicated seven search string categories: **location of nearby services, symptoms, types of services/specialty, asking for advice, questions about resources, questions about whether they have a problem, and looking for the "best."** Potential clients generally search for information associated with geographical proximity of available services, information about the psychological problem such as symptoms, and information about the type of treatment and the available treatment resources. This finding is very helpful in describing the cognitive parameters of this thesis, which can be used in proposing solutions for reducing the cognitive burdens involved in mental health searches. *Zhang-2 et al.* [2021] categorized various online **channels** into three categories for their use in health information seeking in China as **Searching, Browsing and**

Scanning channels. In context of this thesis, their research found the use of online searching channels which include Search Engine, Encyclopedias were affect driven and characterized by a stronger need for health knowledge. Their results suggest that health communication practitioners and scholars may consider measuring the use of internet, new media, or online media more precisely instead of simply asking the public about the frequency of online channel use or internet use in the acquisition of health information. This is an attentive point for us in designing our research framework. *Kim et al.* [2019] in their research on use of mobile devices and applications for health information search in Croatia found that students with less prior medical knowledge tend to consider the video feature of applications more useful. Students rely heavily on searching health topics in search engines on mobile devices, rather than browsing or using applications. Here the video trend is usable in the Video SERP feature of our study and relying heavily on search engine usage in mobile search as well affirms the viability and scope of this research and should be tested in French market as well. A possible hypothesis this finding sets for this research can be, 'users with less prior knowledge or least knowledge, prefer more of video sources in comparison of other sources. Just like we studied above on many occasions about effects of Covid-19, *Li-2 et al.* [2023] studied the affect in health information search behaviour due to heatwave a nature disaster in China. They observed increased population behaviours of seeking information on respiratory diseases, urological diseases and heatstroke associated with the heatwave intensity in 2022. The heatwave duration in 2022 was also associated with an increase in seeking information on cardiovascular diseases and diabetes, mental health, and heatstroke. From this study we can pull out a hypothesis about HISB that HISB is affected by incidental pandemics like Covid Pandemic, Heatwave and yearly routine seasons. *Renhe et al.* [2022] particularly worked on **search-and-go behaviour**, an end-to-end deep learning model to predict whether a user will check-in the searched place or not; using user search query data and trajectory data from Yahoo Japan. The results based on standard metrics show that their model achieved a high estimation accuracy and outperforms the other methods significantly under multiple search scenarios. There can be similar research around the probability if someone reaching to doctor, medical or pharmacy visit after doing first search on internet. Following oneliners are drawn from above synthesis of DHL Impact on HISB around SERP Utility:

1. DHL enables people to critically evaluate the information in hand.
2. 8 common sources of HIS, internet, face-to-face communication with healthcare professionals, demographic characteristics, psychological aspects, role-related or interpersonal traits, environmental aspects, source related characteristics, and disease-related factors.
3. 5 types of HISB include active searching, ongoing searching, passive attention, passive searching, and avoid seeking.
4. SEs have a major role in spreading low-quality health information.

5. Users' cognitive styles influence their information searching strategies, query reformulation behaviour, web navigational styles and information processing approaches.
6. Health information searching is as much social as it is private.
7. 7 search string categories in HIS are location of nearby services, symptoms, types of services/specialty, asking for advice, questions about resources, questions about whether they have a problem, and looking for the "best."
8. Three types of HIS channels are Searching, Browsing and Scanning.
9. Students with less prior medical knowledge tend to consider the video feature of applications more useful.
10. Students do HIS heavily in SEs on mobile devices, rather than browsing or using applications.

Interesting would be to test these facts for our defined audience set i.e. French HISB for:

1. **Scope:** Identifying low-quality health information in SERP and its impact on French HISB. Ethical commitment of SEs in Health SERP. Language complexity and HIR. Less knowledge users jump for videos instead of other source in HIS. Climatic, pandemic pitfall for HISB. Finding pitfall ratio to physical medical services after HISB.
2. **Viability:** Studying DHL impact on HISB gives future scope for producing better SERP and Web Content.

Based on above synthesis, Table-33 below presents a matrix view of Gap Types versus Future Scopes on aspect of DHL Impact on HISB in context of SERP:

Relevant Knowledge Gaps				**Viable Future Scope**	
Gap Type	**Applicable**	**Gap**	**References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	Yes	Knowledge level vs. French HISB	*Khamsum et al.* [2014], *Kim et al.* [2019]	Identifying low-quality health information in SERP and its impact on French HISB. Ethical commitment of SEs in Health SERP. Language complexity and HIR.	Studying DHL impact on HISB gives future scope for producing better SERP and Web Content.
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	Climatic, pandemic pitfall for HISB	*Li-2 et al.* [2023]		
Practical (Mechanism Conflict)	No	N/A	N/A		
Methodological (Better Alternate Method)	No	N/A	N/A		

Empirical (Evaluation Void)	Yes	Ethical commitment of SEs in Health SERP.	*Htay et al.* [2022], *Ni et al.* [2024] , *Ghezzi et al.* [2019]	Less knowledge users jump for videos instead of other source in HIS. Climatic, pandemic pitfall for HISB. Finding pitfall ratio to physical medical services after HISB.	
Theoretical (Application Void)	Yes	Medical establishment pitfall after online HIS in French HISB	*Renhe et al.* [2022]		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	No	N/A	N/A		

Table-33: Knowledge Gaps and Future Scopes - DHL Impact on HISB

5.3. 54 Research Hypotheses Derivatives based on SLR: 54 probable hypothesis derivatives are derived in Table-34 below based on 61 Literature References from above SLR, Research Gaps and Future Scope Tables, and Appendix-5:

54 Research Hypotheses Derivatives based on SLR							
Sr.	**Citation Key**	**Possible Derivative Hypothesis Statement**	**One-word pick**	**Behavioural Aspect**	**Stakeholder**	**Link to thesis variable**	**Hypothesis derived**
1	Karanam et al. [2017]	Young users clicks more on difficult tasks than old users and explore more.	Generation Change	B - Biasing in Searcher's Behaviour	Search User	Persuasion - Intangible Behaviour - Search Skills - Experience	**H-1:** Exploratory search behaviour dominates the Health Search than the Exploiting one.
2	Chin et al. [2015]	Elders exploit /engage / click more on website / LP opened while less exploration of SERP, youngers do its opposite.	Generation Change	B - Biasing in Searcher's Behaviour	Search User	Persuasion - Intangible Behaviour - Search Skills - Experience	
3	Di Caprio et al. [2022]	Combinatorial optimality of search results means user continue to IR until strikes an alternative violating his/her experience.	User Gain Rate	E - User Gain Expectancy	Search User	Personalization - Tangible Behaviour - User Satisfaction	**H-2:** Health being YMYL sensitive category field, user gain rate is foraged with highest tendency, which is influenced by ranking, diversity, personalization and quality of
4	Shi et al. [2014]	Percieved Loyalty of using domain specific results is bounded with level of achieving task goals and public belief pattern towards such results.	User Gain Rate	F - Query Drifting and User Interest Decay	Search User	Persuasion - Intangible Behaviour - Search Skills - Experience Nuances & Context	
5	Bitirim & Ertugrul [2020]	Their study on SERP feature clicking with three demographic charateristic gender, age, and program studying found that most of respondents have rarely or more click for tab, spelling suggestion, related query	User Gain Rate	B - Biasing in Searcher's Behaviour	Search User	Personalization - Tangible Behaviour - User Satisfaction	

		suggestion, instant search suggestion, video result, image result and pagination, except Advertising and amount of clicking results.					SERP
6	Azzopardi et al. [2018]	Result rank, time on SERP and inferred utility derives users to click.	User Gain Rate	E - User Gain Expectancy	Search User	Personalization - Tangible Behaviour - User Satisfaction	
7	Singer et al. [2012]	The HISs behaves differently on being successful versus unsuccessful in their search task. The satisfaction level indicators reveal these differences.	User Gain Rate	C - Search Query Scope and SERP Utility	Search User	Understanding - Art of Phrasing - Semantic Understanding - Neuro Understanding	
8	Shen & Martin [2010]	SERP personalisation can be achieved by knowing the search user's loyalty index.	User Gain Rate	G - Clicking Intent Intesity	Search User	Persuasion - Intangible Behaviour - Search Skills - Experience Nuances & Context	
9	Bae et al. [2020]	Quality and parameters of user feedback clustering through human and machine interaction, provides more quality SERP and interaction experience.	User Gain Rate	I - Semantic Clustering and IR	Search User	SERP Personalization - Personalization Factors	

10	Ahn et al. [2018]	SERP Diversity captures more user attention, attention gets renews on new alternatives countering, after is attention decays rapidly.	User Gain Rate	E - User Gain Expectancy	Search User	Persuasion - Sources & Algorithms - Intangible behaviour	
11	Wolfram [2013]	The expectations of the quality results in quicker time possibility flaunts searchers to jump on top results.	User Gain Rate	E - User Gain Expectancy	Search User	Contextual personalization - Search user satisfaction	
12	Bai et al. [2017]	Users are more likely to perform clicks on the result page that is served with lower latency.	User Gain Rate	G - Clicking Intent Intesity	Search User	Semantic Matching - User Expectations	
13	Arora & Govilkar [2016]	Removal of redundancy in results can be a greater relief to HISs especially those run searches on urgent basis or on go.	User Gain Rate	In LR	Search User	Results Mapping & Sorting - Semantic Matching - Semantic Algorithms	
14	Salehi et al. [2020]	Personalization of SERP doesn't increase user satisfaction, however, reduces their time spent in locating information to satisfy their question.	User Gain Rate	H - SERP Personalization Impact	Search User	SERP Pesonalization - Search User Satisfaction	
15	Tuliao et al. [2022]	Understanding health search strings and providing content accordingly can reduce cognitive burdens.	User Gain Rate	DHL Impact on HISB	Search User	Art of Phrasing - Understanding - Cognitive Capabilities	

16	Russo & Russo [2020]	Semantics of SERP, Cognitive biases of searchers due to their previous knowledge and presumptions can limit their IR output.	Knowledge & Experience Level	C - Search Query Scope and SERP Utility	Search User	Cognition - Topicality - Context - Topical Ontology	**H-3:** User durability and adaptability with a Search Engine is directly proportional to Information Quality and Update.
17	Kim et al. [2019]	Users with less prior knowledge or least knowledge, prefer more of video sources in comparison of more knowledgeable ones.	Knowledge & Experience Level	DHL Impact on HISB	Search User	Cognition - Topic knowledge and Cognitive Evaluation	
18	Lewandowski & Kammerer [2020]	Age, prior domain/topic knowledge and Search Results positioning are of highest impact in influencing user search behaviour.	Knowledge & Experience Level	C - Search Query Scope and SERP Utility	Search User	Understanding - Art of Phrasing - Semantic Understanding - Neuro Understanding	
19	Schultheib & Lewandowski [2021]-1 & 2	Users with little knowledge of result types are more likely to trust and favour Google than users with greater knowledge; and so is for the case of Ads; and users with little knowledge show less willingness to scroll down to organic results.	Knowledge & Experience Level	A - Domain Knowledge Pursuance and Perceived Relevance	Search User	Cognition - Topic knowledge and Cognitive Evaluation	
20	[248]Tegze [2023]	Overtime grown cognitive understanding about search process and task decision approach has the tendency to land you in filter bubble If not pursued on changes and updates.	Knowledge & Experience Level	A - Domain Knowledge Pursuance and Perceived Relevance	Search User	Understanding - Art of Phrasing - Semantic Understanding - cognitive capability	

21	Yamin & Ramayah [2013]	HIS queries are more breadth search queries due to lack of domain knowledge. Lack of domain knowledge drives searchers to type more breadth search queries while domain knowledge rich users adopt to depth search queries behaviour.	Knowledge & Experience Level	C - Search Query Scope and SERP Utility	Search User	Understanding - Art of Phrasing - Neural and Semantic Understanding	
22	Yamin et al. [2014]				Search User		
23	Pawitra [2015]	The ordering of results is treated proportionally to user's belief or to say user's prior knowledge and perceived perception of that subject.	Knowledge & Experience Level	D- Result Ranking and Search Behaviour	Search User	Predictability - User Expectations - Semantic Approaches	
24	Jonsson & Stahl [2015]	Previous experience and the trustworthiness of a result, strongly affected if a user would choose to click on that result.	Knowledge & Experience Level	In LR	Search User	Predictability - Semantic Matching & Approaches	
25	Ashkan & Clarke [2012]	Intensity of intent is directly proportional to likeliness of clicking on a particular search result conditioned to relevancy and quality of that result.	Intent & Relevancy	G - Clicking Intent Intesity	Search User	Intent Precision - Ontological approaches	**H-4a:** Urgency, Precautionary and Courtesy are the HIS Intents **H-4b:** HISB is more about informational intent comparative to
26	Priyadharshini et al. [2018]	Number of search tasks in each search journey, articulates the task difficulty level and user behaviour, intent depth and information urgency.	Intent & Relevancy	B - Biasing in Searcher's Behaviour	Search User	Intent inferring - Query and Content Deciphering	
27	Zerhoudi et al. [2021]	Query count and navigation behaviour in a search task, depends on search intent type and intent shift.	Intent & Relevancy	F - Query Drifting and User Interest	Search User	Intent and Content Clicking	

				Decay			navigational or transactional.
28	Su et al. [2018]	Finding HIS Intent Types just like there are three Intent types for products search in ecommerce. Urgency, Precautionary, Courtesy Intent Types of HIS.	Intent & Relevancy	G - Clicking Intent Intesity	Search User	Intent Precision - NLP Components	
29	Zhang-2 et al. [2021]	Search Engines are more used for Searching Health Information instead of Browsing or Scanning results, meaning that HISB is more informational intent instead of Navigational intent.	Intent & Relevancy	DHL Impact on HISB	Search User	Intent precision - Ontological approaches	
30	Zamani et al. [2020]	Annotations regarding more query intent clarity, results better information retrieval as SERP.	Intent & Relevancy	I - Semantic Clustering and IR	Search User	SERP Personalization - Personalization Factors	
31	Khamsum et al. [2014]	Users' cognitive styles influence their information searching strategies, query reformulation behavior, web navigational styles and information processing approaches.	Cognitivity & Understanding	DHL Impact on HISB	Search User	Cognitive Evaluation of topicality	**H-5:** For lehmen, health sector is one among least learned but highly sensitive on top of head; it increases their
32	Lyounghee et al. [2023]	People continue to trust the Search Engines more than social media pertaining to maintain their natural tendency of scientific temperament.	Cognitivity & Understanding	HIS in Web Era	Search User	Cognition - Topical Knowledge Enrichment	

33	Arguello [2014]	Level of interest and prior knowledge of domain eases the search task diffculty or pushes to find the interest.	Cognitivity & Understanding	E - User Gain Expectancy	Search User	Persuasion - Intangible Behaviour - Search Skills - Experience	reliability on SERPs more.
34	Tennant et al. [2015]	Being younger, more educated with higher Socio-Economic Status means high probability of using internet more for HISB.	Cognitivity & Understanding	B - Biasing in Searcher's Behaviour	Search User	Intent Precision - NLP Components	
35	Alfano et al. [2020]	The education and cognition level of searcher proportionally define their capacity of understanding complex languages in HIR.	Cognitivity & Understanding	DHL Impact on HISB	Search User	Neural Understanding - Linguistic Knowledge	
36	Jiang et al. [2014]	Longer search sessions means deep search user searching more alternatives. Longer Search Session means user has higher cognitive capacity or SERPs lack required quality.	Cognitivity & Understanding	In LR	Search User	Semantic understanding - Cognitive capabilities	
37	Schoenherr [2022]	Creating systems for self learning sooth the nature of people self-decision making upto 81% of total.	Cognitivity & Understanding	Pre-Web HIS Outcomes and Methods	Search User	Art of Phrasing - Neural Understanding - Cognitive Capabilities	
38	Chen et al. [2021]	SEs assistance in saving users' search efforts in complex or task-driven sessions is limited and SE' have big scope of personalizing their query recommendation	SERP Personalization	G - Query Drifting and User Interest Decay	Search Engine	Semantic Matching - User Expectations	**H-6:** Probability of accepting and rejecting

		mechanism.					Health SERP personalization is an equal and early possibility
39	Hendahewa & Shah [2015]	SE' recommendation strategy can substantially improve user's performance in reaching the task specific goals.	SERP Personalization	C - Search Query Scope and SERP Utility	Search Engine	SERP Personalization - Personalization Factors - Tangible Behaviour	
40	Kazutoshi [2016]	Re-ranking search results containing more attractive terms to higher positions provides more satisfied search experience.	SERP Personalization	G - Query Drifting and User Interest Decay	Search Engine	Cognition - Search Context - Topical ontology	
41	Linus Philip [2015]	Visual hierarchy (VH) is proportionally related to navigational and informational search intents, trasactional intent has no relation with VH.	SERP Personalization	E - User Gain Expectancy	Search Engine	Intents Precision - Language Ontology - Intent inferring through NLU	
42	Yi Song [2019]	User interest may drift if the information served is not updated one, thus shorten their interest life.	Freshness & Updates	F - Query Drifting and User Interest Decay	Search Engine	Cognition - Topicality - Context - Multifaceted nature of topicality	**H-7:** Information Quality and User gain rate keeps the users loyal to SERP and Search Content
43	Zanjani & Khadivi [2014]	New objects such as user, ad, feature or query reduce precision of SERPs, it can be handled through using similar old audience bracket.	Freshness & Updates	E - User Gain Expectancy	Search Engine	Cognition - Topicality - Context - Multifaceted nature of topicality	

44	Zhang & Moffat [2006]	The technology and generation change also changes the user search behaviour traits in comparison of previous generation.	Freshness & Updates	G - Clicking Intent Intesity	Search Engine	Understanding - Art of Phrasing - Semantic Understanding - cognitive capability	
45	Zhitomirsky et al. [2016]	Users' search preferences change over time and therefore the ranking of a search engine should adapt to this.	Freshness & Updates	F - Query Drifting and User Interest Decay	Search Engine	Persuasion - Search Skills - Nuances & Context	
46	Liu & Xu [2020]	Requirement of continued refining technical aspects to yield SERPs for fitting bigger audience size of laypersons will never end, complexity of search tasks derive them to consult the network more. So SERP simplicty in HIS field is directly proportional to search satisfaction. That is why "HIS is as much social as it is private".	Freshness & Updates	Web HISB Characteristics	Search Engine	Semantic Personalization - SERP Algorithms	**H-8:** HISB is as much social as it is private
47	Chiu et al. [2022]			Web HISB Characteristics	Search Engine		
48	Yan Zhang [2013]			DHL Impact on HISB	Search Engine		
49	Htay et al. [2022]	DHL enables people to critically evaluate the information in hand. Internet and face-to-face interactions are the most preferred sources of HIS.	Cognitivity & Understanding	DHL Impact on HISB	Search User	Neural Understanding - Cognitive Capabilities	
50	Ni et al. [2024]			DHL Impact on HISB	Search User		
51	Ghezzi et al. [2019]	Low quality SERPs usually displayed by Search Engine can probably take Searchers into wrong direction depending on	System Discrepancies	DHL Impact on HISB	Search Engine	Persuasion - Search Skills - Sources & Algorithms	

		exposure and experience of search user.					
52	Nakamuraa et al. [2018]	Neural NLP models lack in intuition interpretability and debuggability, important aspects of commercial solutions.	System Discrepancies	B - Semantic Clustering and IR	Search Engine	Intent Precision - NLP Components	**H-9:** Algorithmic, technical and display systems behind SERP presentation can get easy excuse from users but then users may shift somewhere else
53	Jacobs et al. [2017]	Web can create disparities and inequalities in health information accessibility.	System Discrepancies	HIS in Web Era	Search Engine	Picking Contextual Results - Personalization Algorithms	
54	Paunic et al. [2023]	Pitfalls such as dysfunctional search filters, inconsistent results, outdated information, lack of reproducibility and thoroughness of SERPs distracts HIS.	System Discrepancies	Web HISB Characteristics	Search Engine	Semantic Matching - Semantic Algorithms	
55	Schoenherr & Thomson [2020]	Correlations between information search behaviour and mobility around the time of public announcements bring HISB change.	Life threat & influence	Web HISB Characteristics	Search User	Cognitive Accuracy - Cognitive Evaluation Factors	**H-10:** Social-political-governmental influence signals generate feel of urgency and influence regarding health topics, which turn the
56	Rosenstock [1974]			Web HISB Characteristics	Search User		
57	Eldad & Elad [2021]	Popular queries SERPs are scrolled to more depth and distance, can be attributed to the interest arose from familiarity with the query topic due to its popularity.	Life threat & influence	E - User Gain Expectancy	Search User	Predictability - Semantic Matching & Approaches	
58	Pang et al. [2014]	Life/health threat urgence of correct information leads people to do exploratory	Life threat & influence	Web HISB Characteristics	Search User	Intent Type - Ontological	

59	Song et al. [2022]	searches more.		Web HISB Characteristics	Search User	Approaches	users in precautionary intent on priority
60	Neeragatti et al. [2023]			Web HISB Characteristics	Search User		
61	Li-2 et al. [2023]	HISB is affected by incidental pandemics like Covid Pandemic, Heatwave and yearly routine seasons.	Life threat & influence	DHL Impact on HISB	Search User	Intent Precision - NLP Components	

Table-34: Research Hypothesis Formation

5.4. Eight Research Hypothesis Derivatives based on FLR: Below Table-35 below presents the 8 testable research gaps or scopes derived from "Discussion" sections of Chapter-2, Chapter-3 and Chapter-4 of Foreground Literature Review Unit and Appendix-5:

Relevant Knowledge Gaps				Viable Future Scope	
Gap Type	**Applicable**	**Gap**	**Thesis Chapter References**	**Scope**	**Viability**
Evidence (Contradiction or Counter)	No	N/A	N/A	Following aspects can be researched and updated: Who learn from whom and in what extent i.e. SE and User; SERP display delay and adjustable quotient of user; EEAT to 3E; Entertainment Intent,	A three-dimensional value addition to all three stakeholders of SERP
Knowledge (Information / Industry / Upgradation / Technology / Data)	Yes	Who learns from whom, SE from User or User from SE and in what ratio. Percentage of users reformulate health queries	*2.1. Art of phrasing of Search User and Search Engine* *2.2. Natural Language Ontology*		
Practical (Mechanism Conflict)	Yes	For quality and relevancy user may adjust with delay in SERP display, but upto what extent and frequency.	*2.5. SERP Personalization*		

Methodological (Better Alternate Method)	No	N/A	N/A	Mapigation Intent, Chat-GPT as another SERP Feature, MetaSE adaptability, Advantage of Search Operators knowledge	
Empirical (Evaluation Void)	Yes	EEAT to 3E, Entertainment and Mapigation/Destination as major Search Intent Categories	*2.6. Search Skills* *4.2. Three Stakeholders of SERP*		
Theoretical (Application Void)	Yes	Acceptance level of Chat-GPT as another SERP feature Adaptability for MetaSE	*3.5. G-AI and LLMs* *3.6. MetaSearch Engine and SERP*		
Population (Audience)	No	N/A	N/A		
Conceptual (Ambiguity)	Yes	Searchers equipped with search operator knowledge have significant edge over others	*4.4. Web Search Intelligence*		

Table-35: Research Hypothesis Derivatives based on FLR

Chapter-6

Research Scope

6.1. Research Scope Validation – 44 Research Audience and Scope Selection:

We have tried to accomplish our FLR and SLR on standards mentioned by *Amanda Bolderston* [2008] as "*A good Literature Review can be an invaluable tool to the practitioner, providing a succinct summary and analysis of the pertinent information in a given area*" and by *Hamilton & Clare* [2004] as "*One of the most common reasons for a literature review is to use the literature to enhance and develop an argument in a thesis or report*". The Framework Model for Literature Review by *Brocke et al.* [2009] *(shown in Chart-88, Appendix-9)* is taken in due consideration to meet the research standard. So, this way, abiding by the true spirit of research, we talk of research scope validation now:

Carving out a scope for any scientific or academic research is a tedious task. Just like any other thesis or research paper, by now in this thesis, we have reached a stage of justifiable literature review and its synthesis in hand *(in fact its major finding is done in Synthesized Literature Review Chapter above, here it is about its validation on the standards and formats in practice)* to carve-out the scope of this thesis and its could be contribution to scientific community and general world both. For this validation we have used eight types of research gaps. We have prepared a matrix table used in Chapter-5 from Tables series number Table-21 to Table-33 through which the positioning of our research questions and formation of hypotheses is tried. Below is a descriptive list of types of research gaps taken into consideration for the FLR and SLR of this thesis:

1. **Evidence Gap** (Contradictory evidence gap): Results from studies allow for conclusions in their own right but are contradictory when examined from a more abstract point of view, *(Jacobs, 2011; Maller-Bloch & Kranz, 2014; Miles, 2017),* [250]*Every Author* [2023].
2. **Knowledge Gap** (Knowledge Void Gap): Desired research findings do not exist, *(Jacobs, 2011; Miller-Bloch & Kranz, 2014: Miles, 2017),* [250]*Every Author* [2023]. Changing Industrial Information and Technology Gap, *Kavuri & Milne* [2009]. A knowledge gap is the difference between what is known and what needs to be known about a specific topic. A knowledge gap can occur for various reasons, including insufficient or incomplete information, contradictory information, limited research, and outdated information, *Mueller-Bloch & Kranz* [2015], *Brocke et al.* [2009] & [251]*Tutors India* [2023].

[250] https://everyauthorpub.com/every-author-blog/f/7-types-of-research-gaps-in-literature-review

[251] https://www.tutorsindia.com/blog/what-are-the-different-types-of-research-gaps-in-the-literature-review/

3. **Practical-Knowledge Gap** (Action-Knowledge Conflict Gap): Professional behaviour or practices deviate from research findings or are not covered by research *(Jacobs, 2011: Miller-Bloch & Kranz, 2014; Miles, 2017),* [250]*Every Author* [2023]. Practical gap arises when there is a discrepancy between theory and practice, where there is a necessity to conduct further research to address this gap and improve the application of the theoretical findings into practice. A practical gap is also called Practical-Knowledge Gap or Action-Knowledge-Conflict Gap, *Kavuri & Milne* [2009], *Mueller-Bloch & Kranz* [2015], *Brocke et al.* [2009] & [251]*Tutors India* [2023].
4. **Methodological Gap:** A variation of research methods is necessary to generate new insights or avoid distorted findings, *(Jacobs, 2011; Muller-Bloch & Kranz, 2014; Miles, 2017),* [250]*Every Author* [2023]. A methodological gap, or method and research design gap, is a gap in the research procedures and techniques employed in previous studies that may affect the precision and reliability of the findings, *Kavuri & Milne* [2009], *Mueller-Bloch & Kranz* [2015], *Brocke et al.* [2009] & [251]*Tutors India* [2023].
5. **Empirical Gap** (Evaluation Void Gap): The empirical gap refers to the difference or discrepancy between observed data or evidence and the values or ideals that are desired or expected. It is a concept that is relevant in various fields of study, including economics, law, and technology. Under this category the research findings or propositions should be evaluated, or empirically verified, *(Jacobs, 2011; Miller-Bloch & Emipielcal Gaps Kranz, 2014; Miles, 2017),* [250]*Every Author* [2023]. The term "empirical gap" refers to the missing or undiscovered components of a phenomenon that can be addressed through empirical research. When the available data or empirical research is unable to answer specific research questions, a gap results, known as an empirical gap or empirical void gap, *Kavuri & Milne* [2009], *Mueller-Bloch & Kranz* [2015], *Brocke et al.* [2009] & [251]*Tutors India* [2023].
6. **Theoretical Gap** (Theory Application Void Gap): Theory should be applied to certain research issues to 'generate new insights. Wherever is lack of theory thus a gap exists, *(Milller- Bloch & Kranz, 2014, Jacobs, 2011; Miller-Bloch & Kranz, 2014; Miles, 2017),* [250]*Every Author* [2023]. It is a gap in existing theories and models in the field that has not been effectively addressed, *Kavuri & Milne* [2009], *Mueller-Bloch & Kranz* [2015], *Brocke et al.* [2009] & [251]*Tutors India* [2023].
7. **Population Gap:** Research regarding the population that is not adequately represented or under-researched in the evidence base or prior research (e.g., gender, race/ethnicity, age and etic), *(Robinson, et al, 2011),* [250]*Every Author* [2023].
8. **Conceptual Gap:** Conceptual gaps occur when there is inconsistency or ambiguity in a concept's definition, interpretation, or application, resulting in confusion and misunderstanding. Conceptual gaps lead to inconsistent or unclear definitions and interpretations, *Kavuri & Milne* [2009], *Mueller-Bloch & Kranz* [2015], *Brocke et al.* [2009] & [251]*Tutors India* [2023].

44 Research Audience and Scope Selection: In the purview of Knowledge Gaps and Scopes figured out through the Tables series number Table-21 to Table-33, we have prepared 143 testable facts in Appendix-5. Appendix-5 is used in figuring out the 62 Hypothesis Derivatives shown in Table-34 and Table-35 in Chapter-5. Table-36 below displays the total types of scopes and their numbers:

Research Audience and Scope Selection			
Gap Type	**No.**	**Gaps**	
Evidence (Contradiction or Counter)	6	Knowledge level vs. French HISB Disparities in French Health SERP SE loyalty in French HISB	Exploration vs. Exploitation finding in French HISB. Organic Results Surfing. Entry of new objects in SERP and HISB change.
Knowledge (Information / Industry / Upgradation / Technology / Data)	9	Who learns from whom, SE from User or User from SE and in what ratio. Percentage of users reformulate health queries Climatic, pandemic pitfall for HISB SES, Cognitive & Social Norms behaviour in French HISB	Health urgency severity scale in French HISB Clicking Intent likeliness and SERP display latency Query suggestions significance in HIS Prior knowledge and perceived perception Young Age, technology and knowledge
Practical (Mechanism Conflict)	3	For quality and relevancy user may adjust with delay in SERP display, but upto what extent and frequency.	Task Types and difficulty levels in HISB Information Foraging Theory & Stopping IR in French HISB
Methodological (Better Alternate Method)	5	Eye Tracking in Health SERP on French HISB Search Log Analysis for French HISB Search Attention Renewal in French HISB	Algorithmic improvising in HISB Difficult vs. Easy Search Tasks testing in French HISB
Empirical (Evaluation Void)	5	EEAT to 3E Entertainment and Mapigation/Destination as major Search Intent Categories	Ethical commitment of SEs in Health SERP. Exploratory intent and HISB

			Query drifting in HIS
Theoretical (Application Void)	8	Acceptance level of Chat-GPT as another SERP feature Adaptability for MetaSE Medical establishment pitfall after online HIS in French HISB How filter bubble function in HISB	Criticality of Health SERP personalization Scrolling distances for popular, medium and least heard queries in French HISB Different ranking for different algorithms. Different algorithms for online and offline content in HISB.
Population (Audience)	2	Public announcement impact on French HISB	Temporal behaviour change
Conceptual (Ambiguity)	6	User interest decay in health SERP Searchers equipped with search operators' knowledge have significant edge over others Knowing the filter bubble in HISB	Banner blindness and results hierarchy in French HISB Desktop vs. Mobile attention on top results Searcher' creativity in HISB
Total	44		

Table-36: Number of probable Research Gaps and audience selected

Specific description about the research audience selected: Our thesis is focused on Healthcare Information Search Behaviour in French Market. In terms of demographics of audience, we aim to first study our thesis on graduate and post-graduate students of Nord-France. Another audience category shall be the healthcare experts, scientists, and teachers / professors. As another expansion, this research may include local patients in hospitals through face-to-face interviews; we shall mention about it in our research data analysis post our surveys and data collections.

6.2. Hypotheses Formation - 10 Potentials:

Based on FLR and SLR done above, summarizing their finding on research gaps and scopes in Table-34 in Chapter-5, here below in Table-37 are the 10 research hypotheses developed to test through this thesis:

Hypotheses derived from citations referred		
Sr.	**Citation Key**	**Potential Hypothesis Derivatives**
1	Karanam et al. [2017], Chin et al. [2015]	**H-1:** Exploratory search behaviour dominates the Health Search than the Exploiting one.
2	Di Caprio et al. [2022], Shi et al. [2014], Bitirim & Ertugrul [2020], Azzopardi et al. [2018], Singer et al. [2012], Shen & Martin [2010], Bae et al. [2020], Ahn et al. [2018], Wolfram [2013], Bai et al. [2017], Arora & Govilkar [2016], Salehi et al. [2020], Tuliao et al. [2022]	**H-2:** Health being YMYL sensitive category field, user gain rate is foraged with highest tendency, which is influenced by ranking, diversity, personalization and quality of SERP
3	Russo & Russo [2020], Kim et al. [2019], Lewandowski & Kammerer [2020], Schultheib & Lewandowski [2021]-2, [248]Tegze [2023], Yamin & Ramayah [2013], Yamin et al. [2014], Pawitra [2015], Jonsson & Stahl [2015]	**H-3:** Information Quality and User gain rate keeps the users loyal to SERP and Search Content
4	Ashkan & Clarke [2012], Priyadharshini et al. [2018], Zerhoudi et al. [2021], Su et al. [2018], Zhang et al. [2021], Zamani et al. [2020]	**H-4a:** Urgency, Precautionary and Courtesy are the HIS Intents **H-4b:** HISB is more about informational intent comparative to navigational or transactional.
5	Khamsum et al. [2014], Kim et al. [2023], Arguello [2014], Tennant et al. [2015], Alfano et al. [2020], Jiang et al. [2014], Schoenherr [2022]	**H-5:** For layman, health sector is one among least learned but highly sensitive on top of head; it increases their reliability on SERPs more.
6	Chen et al. [2021], Hendahewa & Shah [2015], Kazutoshi [2016], Linus Philip [2015]	**H-6:** Probability of accepting and rejecting Health SERP personalization is an equal and early possibility
7	Yi Song [2019], Zanjani & Khadivi [2014], Zhang & Moffat [2006], Zhitomirsky et al. [2016]]	**H-3:** Information Quality and User gain rate keeps the users loyal to SERP and Search Content
8	Liu & Xu [2020], Chiu et al. [2022], Yan Zhang [2013], Htay et al. [2022], Ni et al. [2024], Ghezzi et al. [2019]	**H-7:** HISB is as much social as it is private
9	Nakamuraa et al. [2018], Jacobs et al. [2017], Paunic et al. [2023]	**H-8:** Algorithmic, technical and display systems behind SERP presentation can get easy excuse from users but then users may shift somewhere else
10	Schoenherr & Thomson [2020], Rosenstock [1974], Eldad & Elad [2021], Pang et al. [2014], Song et al. [2022], Neeragatti et al. [2023], Li-2 et	**H-9:** Social-political-governmental influence signals generate feel of urgency and influence regarding health topics, which turn the users in precautionary intent on

	al. [2023]	priority

Table-37: Hypotheses derived from citations referred

Elaboration: Like any other standard research, the major research idea, question and its possible variables conceived are shaped with required gradients like hypothesises set, psychological model and research framework. By doing the related Literature Review and Synthesis of Literature Review above, we figured out the requisite aperture of our research question, its variables and then figuring out the possible hypothesises around it. These gradients can be described in necessary explanations as below:

1. **Research Variables:** As a recap, referring to "Table-9: Six variables of this thesis" and "Diagram-8: Psychological model of SERP Behaviour", both form the ground required for psychological and technical temperament of this thesis. In the light of the FLR done in Unit-1 and the SLR done in Chapter-5 of Unit-2, we justify our thesis question and its associates in next section.
2. **Research Question and its associates:** Beside the entitled question of this thesis, what other supportive or associating questions we aim to answer *(obviously by connecting with FLR and SLR)* can be as follows:
   a. Do users explore more or exploit more the SERPs for healthcare issues?
   b. With what attention level the first and first three results are treated on Desktop view and Mobile view.
   c. What gain rates keep the users stick to SERPs?
   d. What are different Health Search Intents?
   e. What kind of intent the HISB is out of three famous search Intents?
   f. How the influencing aura of a Search Engine takes the Search Users in its influence, probably leading them to the infamous Search Filter Bubble?
   g. Why people accept or reject Health SERPs in early stages?
   h. What personalization of SERP brings in its users the most?
   i. How social is health search?
   j. How private is health search?
   k. What impact the technical and algorithmic system discrepancies put on its users?
   l. Why government and political influences health information seeking/searching *(HIS)* the most?
3. **Hypotheses Finalization:** Taking from **62 Raw-Hypotheses** derived in Table-34 from FLR and SLR, we further aggregated them into **10 probable** ones in the Table-37, as the most Potential Hypotheses Derivatives as below:

H-1: Exploratory search behaviour dominates the Health Search than the Exploiting one. Optimizing the SERPs from exploratory point of view can yield better turn-out of Gen-Z.

H-2: Health being YMYL sensitive category field, user gain rate is foraged with highest tendency, which is influenced by ranking, diversity, personalization, and quality of SERP.

H-3: Information Quality and User gain rate keeps the users loyal to SERP and Search Content

H-4a: Urgency, Precautionary and Courtesy are the HIS Intents

H-4b: HISB is more about informational intent comparative to navigational or transactional.

H-5: For layman, health sector is one among least learned but highly sensitive on top of head; it increases their reliability on SERPs more.

H-6: Probability of accepting and rejecting Health SERP personalization is an equal and early possibility.

H-7: HISB is as much social as it is private.

H-8: Algorithmic, technical and display systems behind SERP presentation can get easy excuse from users but then users may shift somewhere else.

H-9: Social-political-governmental influence signals generate feel of urgency and influence regarding health topics, which turn the users in precautionary intent on priority.

Hypotheses Matching to Psychological Model Variables:

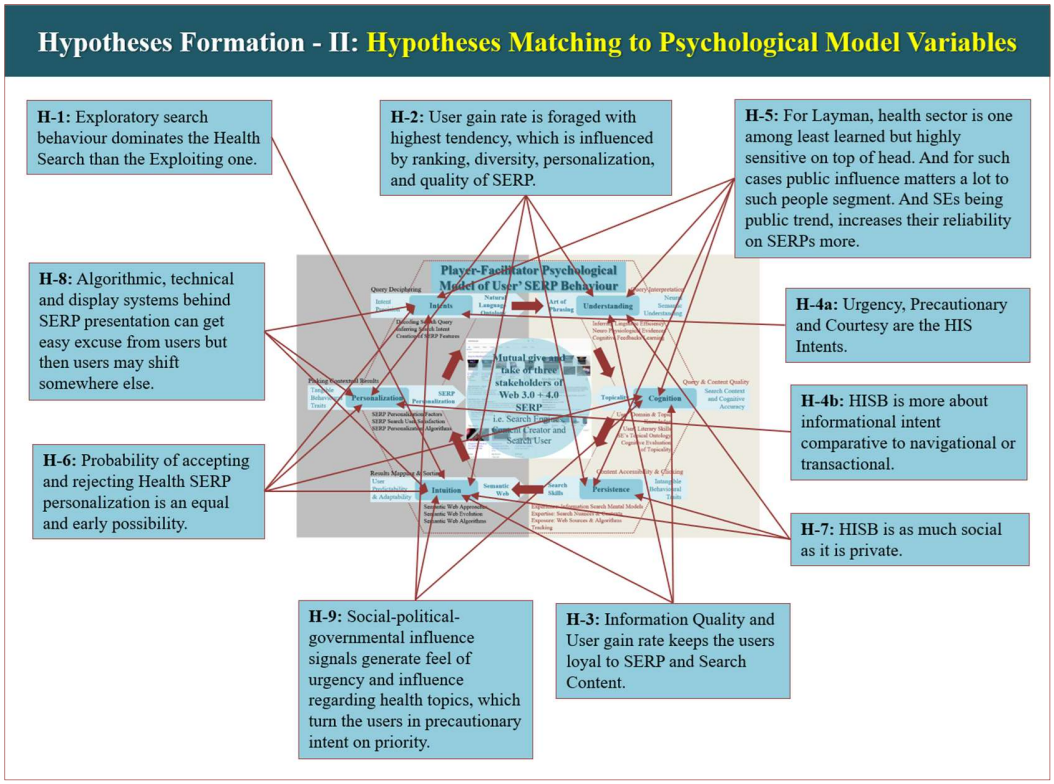


Chart-89: Hypotheses Matching to Psychological Model Variables

After a careful review of these 10, only 6 are to be retained for final testing in this research, which is further elaborated in "Research Goals" section next. These are to be selected based on the suitability to our thesis theme, resources available for testing and research model compatibility.

6.3. Research Motivations and Goals:

Research Motivation: Thesis writing is a service to mankind. Service to facilitate the humanity, to serve the locality and to adding value to the civilization. This thesis is a foraged contribution to the digital marketing needs of the world. Its specificity lays in contributing to the sector of Generative Artificial Intelligence based Digital Health Marketing through utilization of SERPs on any Search Engine worldwide. Our utmost motivation goes for attaining a renowned name in serving the Digital Healthcare Information facilitation services globally.

Research Goals: We see the goal of this research in universal and specific intent categories:

Universal category goals: These are basically the conceptual goals, where we intend to justifiably establish two new Search Intent Categories using different psychological aspects of the literature review, these goals are:

1. **Exploring and Explaining the SERP Navigation Behaviour and Search Intent Classification:** To get the answers of our variables and hypotheses developed from LR referring to Table-37 and Chart-89 we find it feasible to develop a Search Intent Classification Supervised Machine Learning Model on real human data tested against AI-Simulated Data. Its model-method and results are discussed in Chapter-9.
2. **Preparing a Psycho-Cognito Profile Method:** Talking about Search Intents is incomplete without figuring out the Psycho-Cognito profile of targeted audience. It helps to relate the results to viable reasons. So, by this exercise we want to gauze the scale of Understanding, Congintive Capacity, Semantic Relevance, Proficiency and Personalization factors of our target audience i.e. the test participants and as shown in Diagram-8 in Chapter-2.
3. **Positioning the "Mapigation Intent" category as fifth major intent category:** Beside above two goals, we want to draw ways to further possibilities of defining two new 'Search Intent Types'; one is talked here in this point and another in next point. It comes from the gaps, experts and influencers often leave unaddressed, but it entices us every time and it is regarding the types of Web Search Intent. Here we want

to put our argument that the four major types of Intent are not sufficient and justified to include all mind-states of searching on internet. Just like when it is talked about the definition of navigational intent, it is iterated in a traditional fashion "having a specific site or brand name in mind to search or navigate". Throughout our literature review, we hardly find anyone talking about the "mapigation" in navigation or information intent category *(there is no sense of talking about mapigation in transactional or commercial intent categories).* The ground we give here to our argument is "state of mind", "type of action needed" and "type of information collected". When people search for addresses and location what is their 'mind of state' at that moment; this is found almost unaddressed tildate or otherwise considered under 'Navigational' intent only; but **there should be difference stated between searching a Digital Property Location and a Phyiscal Address Location**. So, it can also be interesting to identify the **"mapigation intent"** as fifth major intent type after information, navigation, commercial and transaction.

4. **Identifying "Entertainment Intent" category as sixth major intent category:** It comes from another similar gap as for the mapigation intent. And it is about the "entertainment" category of intent. We feel that this deserves the sixth categorization because it is a time spending activity just like hanging out. The person of this category has totally different "state of mind" in comparison of rest of the four major ones. Person is in no hurry, having lot of time to spend, having intention of relaxing, refreshing his/her mental and psychic state, which means a purpose of driving a mood change, mood recharge or mood refreshing. So, it is not a task to do rather an activity to enjoy watching and relaxing or celebrating; thus, it doesn't fit as such in present four standard categories. So, it should be named as "Entertainment" or "Watching" or "Mind-relaxing" Intent".

Specific Category Goals: These are basically the hypothesis testing goals. As we have shortlisted 10 hypotheses in Table-37 from our SLR in previous chapter. So, based on these hypotheses these goals can be defined as follows.

1. Online HIS is more exploiting or exploratory? - *based on H1.*
2. How the user gain rate derives online HIS mind around a SERP - *based on H2.*
3. Third goal can be set on figuring out the health specific search intent types, which as per our SLR can be Urgency Intent, Precautionary Intent and Courtesy Intent - *based on H4a.*
4. In the fourth testing a comparative gazing of traditional intent types can be elaborated in context of HIS in French Market – *based on H4b.*
5. In the fifth goal such factors can be exploited which make people to search privately or socially - *based on H7.*

6. Not making it bigger and complex the last and sixth goal can be finding about acceptance and rejection extent of Health SERP personalization - *based on H6.*

6.4. Evolution of Research Question:

As evident from the title of this thesis, the central question at focus has been, "Psychological factors behind French Web Searcher's Health Information Searching Behaviour on SERP, in purview of G-AI". In short "French HISB on SERP". Below Table-38 shows about the transformation in thesis title along with the progress of this thesis.

Research impact on Thesis Title			
Year	**Moment**	**Title**	**Comparative Impact**
2019-20	Thesis Proposal	Study of the psychological factors behind the SERP Navigation Behaviour	Domain Level Generic
2023-24	FLR and Research Framework Review	Psychological factors behind French Web Searcher's Health Information Searching Behaviour on SERP, *in purview of G-AI.*	Contextually Domain Specific
		Short Title: French HISB on SERP	Crisp Keywords Set

Table-38: Research impact on Thesis Title

First, in general our thesis question is well assessed by the vast Literature Review *(FLR, SLR, R5MLR, RMLR, ALR)* sections, which should be well realized from complete study above. On top of this, this Chapter-6 in specific justifies the viability of our thesis question through our mechanism of hypotheses selection and pairing a particular goal to each. Which all in unison provide evident support to our research question.

Chapter-7

R5Ms *(Research Model, Methodology, Method, Metrics and Modalities)* Framework

7.1. Five R5Ms Testable Frameworks:

As a research custom, after having FLR, SLR, Central Psychological Model and Potential Hypotheses Set defined in previous chapters; just like each research proceeds toward designing its testable research model; this chapter intends to first figure out the **five probable Research Methodologies (RMs)** relevant to this thesis, then developing a practical and technical viability knowledge base of these five RMs in purview of relevant Literature Review **(R5MLR).** Based on this foregrounded understanding of what practically viable, we conceptualize the problem of this research in this chapter. The research model defining shall be done with a minutely and specifically concised Literature Study **(R5MLR)** in parallel. This way once the foolproof testable model is obtained, we shall be proceeding towards the testing of that model in Chapter-9.

Based on our knowledge and understanding guts developed through above rigorous research, we see defining of a research problem as a **Topic Modelling with 5Ms gradients**; first M being the Model itself, once the Model is defined its compatible Methodology to be carved-out; then comes the Method of testing the Methodology of the Model with relevant measure units of results called as the Metrics and finally the Modalities to be used in performing the complete research testing. So, this way we name it as **R5Ms set of a research work**.

While Model, Metrics and Modalities are easily separably identifiable terms; the Methodology and Methods most often keep the biggest confusion for researchers due to their terminology and adjacent conceptual nature. And that is why these are most often used replaceable. So, to set a clear understanding about both; we start with understanding the difference between Methodology and Method from Table-39, Appendix-9. So, the R5Ms can be defined in one line for each of 5Ms as below:

1. **Model:** Model is the centric psychological or technical set of testable dimensions.
2. **Methodology:** Methodology is the set of Method, Metrics and Modalities to test a Model.
3. **Method:** A step-by-step process or prompt of testing the Model.
4. **Metrics:** Pair of a dimension and its mathematical value i.e. Metric.
5. **Modalities:** Set of equipments, resources and environment required to perform the test.

As per the FLR, SLR and Hypothesis driven, we can assume following possible methodologies for testing of the hypotheses of our research model:

1. Analysing a Search Engine's Log Data on parameters defined to test in our hypotheses set.
2. Filming the searcher's navigation of French HISB on SERP and then analysing this human-generated real-data against AI-Simulated Data.
3. Eye Tracking Method to observe the eye movement of SERP users on defined search tasks and then analysing that data.
4. Doing an online anonymous field survey on a defined questionnaire.
5. Interviewing industry experts and professionals.

Table-40 below presents a comparative assessment of possibility of all five test methods:

Comparative assessment of five Research Methodologies

Sr.	Methodology	Quality Probability	Research Modalities Required	Modalities Availability
1	Search Engine Log Data Analysis	Highest possible	**Data Source:** French HISB Log Database of a Search Engine like Google, Bing etc. **Analytical Tool**: SPSS, Excel **Training requirement:** Yes **Finance:** Depending on SE's terms & conditions	Partially available, for the rest it requires huge budget and resources
2	**Analysis of Filming of French HISB on SERP**	Highest possible	**Data Source:** Text Transcripts Data retrieved from Filmed Videos, AI-Simulation Data and YouTube API Data **Analytical Tool**: Python, Google Colab GPU, VSCode, LLMs, Excel **Training requirement:** Yes **Finance:** Not as such	Complete set available in MTD-LUMEN Lab
3	Eye-Tracking Observation of French HISB on SERP	Highest possible	**Data Source:** Eye-Tracking data collection through live observation tool and system. **Analytical Tool**: SPSS, Excel **Training requirement:** Yes	Complete set available in MTD-LUMEN Lab

			Finance: Not as such	
4	Online anonymous field survey	Midway	**Data Source:** To be collected through online survey filling. **Analytical Tool**: SPSS, Excel **Training requirement:** Yes **Finance:** Not as such	All modalities manageable.
5	Interviewing industry experts and professionals	Midway	**Data Source:** Face to face or video interviews of professionals, professors, consultants etc. **Analytical Tool**: SPSS, Excel **Training requirement:** Yes **Finance:** Depending on case to case of person to be interviewed.	All modalities manageable, finance to be figured out.

Table-40: Comparative assessment of five feasible research methods

Based on above analysis, **2nd Method is most favourable in purview of suitability of Method, Metrics and Modalities available!**

In next section, we shall investigate relevant literary references about technical aspects and their feasibility to adapt in 2nd Method of Filming French HISB as picked from Table-40 to define its proper R5Ms set.

7.2. Relevant R5Ms Literature Review (R5LMR):

Before writing what kind of relevant R5Ms should be studied, let's understand the kind of data we shall require collecting as per our selected research methodology of 'Filming SERP Navigation Behaviour' and generating 'AI-Simulation-Data' and 'YouTube API Data'. Here we start with a brief of the methodology and related 5Ms Literature Review. Model testing and results collection is described in Chapter-9. [252]*Adita Joshi* [2022] cites some most important and frequently used research methodologies such as descriptive research, analytical research, fundamental research, applied research, conceptual research and empirical research; all these methods shall have their stake in our data analysis. As our research intend to reveal facts around present current state of affair of SERP Search Behaviour so it falls under **descriptive research**. And as User Search Behaviour always

[252] https://researcher.life/blog/article/choose-research-methodology/

is fresh thus it reflects current market mood, which can be applied to get better results in future, and it falls under **applied research**. As our research intends to bring some new concepts, recommendations and implications to give base to new concepts or add-up to existing concepts of SERP Marketing and SERP Search Behaviour thus it falls under **conceptual research** as well. These three types of conceptual findings are discussed after having the results in hand and assessed in Chapter-10. Adding to further type of a research, this research work is exploratory in nature that intends to understand the state of mind of search users related to field of psychology *(as* [253]*Uttkarsha Bhosale* [2023] *helps in understanding it)* and it falls under **exploratory research** category. There is lot of appropriations and data adjustments require in understanding the search user patterns so it makes it an **analytical research** category as well. Which type of research constituted how much percentage, it is dicussed in Chapter-10 and Chapter-11. In addition, *our 'Knowledge Gaps and Scopes' synthesized through the Tables series number Table-30 to Table-43 and Table-44: all in unison presents our research work as an amalgamation of Q&Q, Descriptive, Applied, Exploratory, Analytical and Conceptual Research.*

To further assess our research methodology with appropriate research methods and metrics, we traversed through a peculiar list of vibrant user behaviour models invented and cited remarkably over the past two decades. Chart-90, Appendix-9 presents the chronological list of all such major "**user search behaviour evaluation models and metrics**". First, we investigate each of these models & metrics one by one and then we discuss the most relevant ones in reasonable details to set a symmetry to our research needs. Before the blossoming of resources in the Web and IR *(Information Retrieval)* evaluation metrics were traditionally developed for set retrieval. These metrics, such as **precision, recall, and F-score/measure,** are not connected with user models, *Zhang-1 et al.* [2021]. In the field of "user search behaviour evaluation", there are two segments of evaluation, the online evaluation *(e.g. A/B Testing, Interleaved comparisons, and their variants)* and offline evaluation *(also called test collection evaluation or batch evaluation), Zhang-1 et al.* [2021]. Here *Moffat et al.* [2012] cite that despite being repeatable and efficient, offline evaluation approaches cannot directly take the perception of users to the systems into consideration. To address this issue the modern evaluation metrics *(such as listed in Chart-90)* provide a simulation of users of search systems to bridge the gap between offline evaluation and realistic user behaviour. Let's try to do a comprehensive overview of these developments of user search behaviour modelling and position us.

[253] https://www.enago.com/academy/choose-best-research-methodology/

As a custom a research framework or method is to be defined to formulate evaluation metrics with corresponding user models. In this regard, *Carterette* [2011] categorizes three underlying models to compose evaluation metrics:

1. **Browsing Model:** It describes how a user interacts with the SERP.
2. **Document Utility Model:** It describes how a user derives utility from individual relevant results in the SERP.
3. **Utility Accumulation Model:** It describes how a user accumulates utility while browsing.

Different from Carterette, *Moffat et al.* [2013] explained the metrics with three different but interrelated ways and established the **C/W/L framework**:

1. **Conditional (C):** Means the probabilities of a user continuing to read past each rank.
2. **Weights (W):** The weight given by the user on each document rank.
3. **Leaving (L):** Probabilities of a user leaving (L) at a given rank.

R5Ms around Information Retrieval (IR) Needs: One of the most important aspects of our thesis in terms of research methodologies is the "Information Retrieval". We may need to peep into some small to big size information data bases such as "log data", "general user behaviour data". Work of *Jarvelin & Kekalainen* [2002] on modelling around IR state the "**recall and precision based on binary relevance judgment**" methods segment as a traditional evaluation method. While talking about the computing of "**cumulative gain-based measurements"** in reference to TREC-7 *(Text REtrieval Conference),* they talk of "relevance score of retrieved documents along the ranked result list", "A discount factor application to the relevance scores in order to devaluate late-retrieved documents" and "Computing the relative-to-the-ideal performance of IR techniques". These novel measures are defined and discussed, and their use is demonstrated in a case study using TREC data by Kalervo & Jaana under the methods termed as **Direct Cumulated Gain, Discounted Cumulative Gain (DCG), Normalized Cumulative Gain (NCG)**. This method is highly helpful in search behaviours mannerism analysis. Not going into much description of models, we plan an easy write-up of 37 most relevant Search User Behaviour Metrics/Models based on 41 research papers in Table-41 below:

Sr.	Citation Key	Evaluation Metric	Evaluation Method	Methodology

1	Jarvelin & Kekalainen [2002]	DCG: Direct & Discounted Cumulative Gain	Handle the position bias and graded relevance assessments	A gain and the discount graded relevance judgments model that computes the cumulative gain users obtain from the retrieved results and equally a log-based discount function to reduce the gain of documents by their ranks with an assumption that documents at greater ranks will be less likely examined due to time, effort, or previous information.
2	Moffat & Zobel [2008]	RBP: Rank biased precision	A user stopping model, the parameter p is the probability of examining the next result.	The users always examine the first result and keep making only one decision, to view the next result with fixed persistence (probability) p or not, until they decide to stop.
3	Robertson [2008]	AP: Average Precision		Different users stop scanning the result list at different relevant documents and the stopping probability distribution is assumed to be uniform across all relevant documents.
4	Craswell et al. [2008]	CUM: Cascade User Model		Click > Skip Above model explaining the position effect, which assumes a linear traversal through the ranking, and that documents below a clicked result are not examined.
5	Sakai & Robertson [2008]	NCU: Normalized Cumulative Utility	This model examines the family of NCU metrics in terms of rank correlation and discriminative power, i.e., the proportion of statistically significantly different system pairs given a confidence level.	NCU maintains high correlation with AP and shows the highest discriminative power, i.e., the proportion of statistically significantly different system pairs given a confidence level, by utilising graded relevance in a novel way.
6	Dupret & Piwowarski [2008]	UBM: User Browsing Model	Predicted user behaviour is compared with the conclusions of an **eye-tracking**	The browsing model assumes that all users behave the same way independently from the query.

			experiment.	
7	Clarke et al. [2008]	α-nDCG: Normalized Discounted Cumulative Gain		Trade-off between relevance and diversity metric for diversified search evaluation.
8	Agrawal et al. [2009]	nDCG-IA: Normalized Discounted Cumulative Gain - Intent Aware		On per-intent relevance assessments, intent-aware approach applies to traditional evaluation metrics for diversified search evaluation.
9	Chapelle et al. [2009]	ERR: Expected Reciprocal Rank		Graded relevance-based model of the expected reciprocal length of time that the user will take to find a relevant document. The probability Ri that users stop after examining a document depends on how they are satisfied with this document, which is directly affected by the relevance ri of the document.
10	Chapelle & Zhang [2009]	DBN: Dynamic Bayesian Network	Position bias urls appearing in lower positions are less likely to be clicked even if they are relevant.	This metric proposes a Dynamic Bayesian Network which aims at providing with unbiased estimation of the relevance from the click logs in predicting both click-through rate and relevance.
11	Guo et al. [2009]	DCM: Dependent Click Model	Based on the dependencies between multiple clicks.	It generalizes the cascade model to multiple clicks by including a set of position-dependent parameters to model probabilities that an average user returns to the search result page and resumes the examination after a click.
12	Li et al. [2009]	GA: Good Abandonments		A good abandonment as an abandoned query for which the user's information need was successfully addressed by the search results page, with no need to click on a result or renew the query.

13	Yilmaz et al. [2010]	EBU: Expected Browsing Utility	It accounts click decisions of users. EBU also takes the snippets into account along with documents.	Based on distinguishing between the clicked documents and only scanned documents from previous logs. This model can be defined as a sequence of actions and decisions of users.
14	Chapelle et al. [2011]	CBM-IA: Cascade Based Metric - Intent Aware	It has better correlation with abandonment rate as in ERR-IA.	It assumes the equal importance of all the intents. CBM is submodular; DCG is modular; and AP is super-modular.
15	Carterette [2011]	CF: Conceptual Framework		Conceptual framework is a connection between DCG *(probability that the document d_k is viewed by user)* and RBP *(probability that user stops at rank k)*; which means if users stop at ranks which are greater or equal to k, the document dk is viewed by users.
16	Smucker & Clarke-1-2-3 [2012]	TBG: Time-Biased Gain	The accumulative gain discounted over the expected time T(k) to reach rank k.	This model uses the time to measure the effort users make to examine snippets and further read documents. Shorter documents are expected to take less time than longer documents.
17	Moffat et al. [2012]	INSQ: Inverse Square	It evaluates the Continuation Probability and leaving *(or stopping)* probability	It is based on the influence of users' intention to find useful information on their stopping decision. The assumption is that the probability of a user continuing search having reached rank i is positively related to T, the anticipated number of relevant documents.
18	Moffat et al. [2013], Moffat et al. [2019], Moffat & Alfan [2020] & Wicaksono [2020]	C/W/L Framework - in line with INSQ	C/W/L = The **Continuation** Probability, **Weight** of a document contributing to the metric and **Leaving** Probability.	It is based on the influence of users' intention to find useful information on their stopping decision. The assumption is that the probability of a user continuing search having reached rank i is positively related to T, the anticipated number of relevant documents. Metrics with user models that reflect typical behaviour also tend to be the metrics that correlate well with user satisfaction ratings.
19	Chuklin et al. [2013]	CMBM: Click model-based	This method transforms a click	The main constituent of this model-based metric is a click model a probabilistic model aimed at

		metrics	model into a metric and summarized series of click model-based metrics	predicting user clicks.
20	Sakai & Dou [2013]	U-measure		A unified U-measure framework based on trail-texts to directly handle the issue of nonlinear traversals. A trail-text tt is concatenated by n strings s1s2…sn, where each string is a snippet or full text of document in traditional search scenario.
21	Ferrante et al. [2014]	MP: Markov Precision	It is assumed that users could start from an arbitrary document and move forward or backward to any other documents after spending a random time on the document.	The main contributions of MP are the novel idea to inject user models into metrics with Markov Chains and the unified framework to bridge the gap between rank-based and time-based metrics via using different versions of Markov Chains.
22	Giannakouris et al. [2014]	CSMR: Cosine Similarity & MapReduce		It is a method for pairwise text similarity on massive datasets, using the Cosine Similarity metric and the **tf-idf** normalization method.
23	Xenophon et al. [2014]	RRP: Reciprocal Rank using webpage Popularity		It assumes that the probability of relevance of a web document is a function of a combination of two factors: its relevance grade and its popularity grade.
24	Moffat et al. [2015] & Moffat et al. [2016]	INST: Information Storage & Retrieval	Useful effectiveness metrics must be goal sensitive	Like INSQ with difference that it takes the assumption that the probability of users stopping search increases as they get closer to their goal of finding T relevant documents into account.
25	Chuklin et al.-2 [2015]	RCM: Random Click Model	Simplest click model with one parameter only.	Every document has the same probability of being clicked and this probability is a model parameter. Being single parameter model, it is the safest model from overfitting.

26	Evangelopoulos et al. [2016]	IRDP - Information Retrieval using document popularity	It uses Cosine Similarity on MapReduce model	An evaluation metric for information retrieval systems which employs two relevance factors; a relevance judgement grade and a popularity grade that represents users' vote for a document.
27	Jiang & Allan [2016]	AdapEM: Adaptive Efforts Metrics	Like TBG, it considers different effort for different documents and modified existing metrics to Adaptive Effort metrics. In contrast to TBG and U-measure, which estimate effort based on document length, the adaptive effort metrics only consider the influence of relevance on the effort.	It uses an adaptive effort e(i) while retaining the other parts of existing metrics in two ways as: first to fix the ratio of effort between relevant and non-relevant documents and let the effort on relevant documents be one unit. Another approach is to estimate effort based on relevance from search logs like TBG.
28	Chuklin & Rjike [2016]	CAS: Click, Attention & Satisfaction		The CAS model is an extension of click mode-based metrics by incorporating the effects of "good abandonments" and non-linear SERP layouts. Attention model uses the user's attention on SERP elements determined by feature vectors of the elements. Then based on PBM model, users click, and non-click is examined as click model. Finally, the satisfaction model determines user satisfaction in examining SERP elements and clicked documents.
29	Luo et al. [2017]	HBG: Height-Biased Gain	A document ranking (height) based model. Its central idea is to cumulate discounted gain over height, rather than time or clicks.	It assumes that users first examine the snippet of the first result with height (upper ranking). Then they decide to click this result or not based on the Snippet Relevance and Click Necessity.

30	Zhang et al. [2017]	BPM: Bejewelled Player Model	Cumulative gain and effort model	Inspired by a game Bejewelled, it is based on imposing constraints with users' expectations for both **gain and effort** to determine when users stop.
31	Jiang & Allan [2017]	AdapPM: Adaptive Persistence Metrics	It is adaptive version of RBP & TBG. A persistence parameter of a metric as a linear model based on the relevance of results on the SERP.	Users' persistence should be adaptive on various SERPs with different qualities; thus, this model is based on adaptive persistence versions of existing metrics of RBP & TBG.
32	Azzopardi et al. [2018]	IFT: Information foraging theory-based measure	Based on Goal Sensitive rule and Rate Sensitive rule models	IFT, *Pirolli & Card* [1999] models how people search for information with instinctive foraging mechanisms that similarly adopted to find food through evolution; is made base to suggest two rules thath affect the user stopping model. First, the Goal Sensitive rule which assumes that users' continuation probability decreases as gain accumulates. Second, assumes the Rate Sensitive rule of *Eric Charnov* [1976] that users' continuation probability decreases as the rate of gain (average gain obtained with a unit effort) decreases.
33	Thomas et al. [2018]	CAM: Card-Aware Metrics	It determines the usability probability of SERP Cards i.e. Instant Answers, Maps, Image Frames etc.	This model shares the idea of CAS model and assumes that users can obtain gain from the card or the full document of a result. Given each card-document pair, users first examine the card, and then have four possible decisions: (1) stop after the card with probability continue past the card with probability card click on the full document with probability (2) stop after the document with probability or (3) continue past the document to the next card with probability (4) continue past the card with probability, not click with probability and continue to the next card.

34	Azzopardi et al. [2020]	DDM: Data Driven Metrics	This method investigated the influences of four factors: Position *(rank of the result),* Focus *(how focused the query on navigational intents, measured by click ratio),* Relevance *(relevance label of the result),* and Type *(item type of the result, such as web, ads, video etc.).*	The model moves in the footsteps of C/W/L Framework and Adaptive Persistence but extends its aperture towards maximum likelihood. So DDM proposes a more general data-driven approach to infer the continuation probability with Maximum Likelihood Estimate
35	Ferrante & Ferro [2020]	P@H: Precision at stopping time	Three browsing methods: DFBM: Deterministic Forward Browsing Model SFBM: Stochastic Forward Browsing Model RWBM: Random Walk Browsing Model	This model is in lines with forward and backward transitions of Markov Precision, differentiating by focusing on stopping time precisely. It proposes three major groups of browsing model as: DFBM (assumes users always sequentially examine each result until the last one), SFBM (User may move forward to the next result with probability pi or stop) and RWBM (Users may move forward probability pi or backward probability qi or stop probability 1 - pi-qi.
36	Diaz et al. [2020]	EE-D: Expected Exposure Disparity, EE-R: Expected Exposure Relevance	It computes the squared error between the target exposure and system exposure.	It is based on the principle of Equal Expected Exposure which argues that given a fixed information need, no item should receive expected exposure than any other item of the same relevance grade.
37	Breuer et al. [2021]	Repro_Eval	A Python package for search evaluation.	This toolkit focuses on reactive reproducibility studies of system-oriented IR experiments.

Table-41: R3M for 37 Metrics for Search User Behaviour Measurement

Based on the study of above Metrics/Models, Table-42, Table-43 and Table-44 represent a comparative understanding of different model-based evaluation metrics against different user search behaviour dimensions. These dimensions are categorized in three evaluation streams as 'User behaviour parameters', 'User decision assumptions' and 'User Utility Function'. Table-25 depicts four major categories of User behaviour as 'Interacting Order on SERP', 'Snippet/Card Interaction', 'Document Interaction' and 'Stopping Interaction'. Wherever "x" is marked in a cell, it means that corresponding parameter is counted or considered in that metric/model. These markings are based on the assumptions derived from Table-41 above; so, for an easy understanding, annotate the below information keeping both in purview.

Comparison of model-based evaluation metrics - I

			User Behaviour Parameters											
			Interacting Order on SERP			Snippet/Card Interaction				Document Interaction			Stopping Interaction	
Sr.	**Citation Key**	**Metric**	Forward	Backward	Random	Examining	Gain	Examining Efforts	Clicking	Examining	Gain	Examining Efforts	Leaving Snippet	Leaving Document
1	Jarvelin & Kekalainen [2002]	DCG	x							x	x			x
2	Moffat & Zobel [2008]	RBP	x							x	x			x
3	Robertson [2008]	AP											x	x
4	Craswell et al. [2008]	CUM							x					
5	Sakai & Robertson [2008]	NCU												
6	Dupret & Piwowarski [2008]	UBM	x	x		x				x				
7	Clarke et al. [2008]	nDCG												
8	Agrawal et al. [2009]	nDCG-IA												
9	Chapelle et al. [2009]	ERR	x							x	x	x		x
10	Chapelle & Zhang	DBN												

	[2009]													
11	Guo et al. [2009]	DCM				x				x				
12	Li et al. [2009]	GA											x	x
13	Yilmaz et al. [2010]	EBU	x			x			x		x		x	x
14	Chapelle et al. [2011]	CBM-IA											x	x
15	Carterette [2011]	CF	x							x	x	x		x
16	Smucker & Clarke-1-2-3 [2012]	TBG	x			x		x	x	x	x	x	x	x
17	Moffat et al. [2012]	INSQ	x							x	x			x
18	Moffat et al. [2013] & [2019]	C/W/L	x							x	x	x		x
19	Chuklin et al. [2013]	CMBM	x			x			x	x	x	x	x	x
20	Sakai & Dou [2013]	U-measure	x	x	x	x	x	x	x	x	x	x	x	x
21	Ferrante et al. [2014]	MP	x	x	x					x	x	x		x
22	Giannakouris et al. [2014]	CSMR						x				x		
23	Xenophon et al. [2014]	RRP												
24	Moffat et al. [2015] & [2016]	INST	x							x	x			x
25	Chuklin et al.-2 [2015]	RCM							x					
26	Evangelopoulos et al. [2016]	IRDP												
27	Jiang & Allan [2016]	AdapEM	x			x		x	x	x	x	x	x	x
28	Chuklin &	CAS	x			x	x		x	x	x		x	x

	Rjike [2016]													
29	Luo et al. [2017]	HBG	x			x	x	x	x	x	x	x	x	x
30	Zhang et al. [2017]	BPM	x							x	x	x		x
31	Jiang & Allan [2017]	AdapPM	x			x		x	x	x	x	x	x	x
32	Azzopardi et al. [2018]	IFT	x							x	x	x		x
33	Thomas et al. [2018]	CAM	x			x	x		x	x	x		x	x
34	Azzopardi et al. [2020]	DDM	x							x	x	x		x
35	Ferrante & Ferro [2020]	P@H	x	x						x	x	x		x
36	Diaz et al. [2020]	EED, EER	x	x										
37	Breuer et al. [2021]	Repro_Eval												

Table-42: Comparison of model-based evaluation metrics-I

Table-43 depicts three major categories of User decision assumptions as 'Factors Influencing Stopping Decision', 'Factors Influencing Clicking Decision' and 'Intent Relevance'. Wherever "x" is marked in a cell, it means that corresponding parameter is counted or considered in that metric/model. These markings are based on the assumptions derived from Table-41 above; so, for an easy understanding, annotate the below information keeping both in purview.

Comparison of model-based evaluation metrics - II												
			User Decision Assumptions									
			Factors Influencing Stopping Decision						Factors Influencing Clicking Decision			Intent Relevance
Sr.	**Citation Key**	**Metric**	Position	Relevance	Cumulative Gain	Cumulative Effort	User Expectations	Others	Doc Relevance	Snippet Relevance	Click Necessity	Assessment
1	Jarvelin &	DCG	x									

	Kekalainen [2002]											
2	Moffat & Zobel [2008]	RBP										
3	Robertson [2008]	AP										
4	Craswell et al. [2008]	CUM	x									
5	Sakai & Robertson [2008]	NCU	x	x								
6	Dupret & Piwowarski [2008]	UBM										
7	Clarke et al. [2008]	nDCG		x								
8	Agrawal et al. [2009]	nDCG-IA										x
9	Chapelle et al. [2009]	ERR		x								
10	Chapelle & Zhang [2009]	DBN	x	x								
11	Guo et al. [2009]	DCM	x									
12	Li et al. [2009]	GA										
13	Yilmaz et al. [2010]	EBU		x					x			
14	Chapelle et al. [2011]	CBM-IA										x
15	Carterette [2011]	CF	x	x								
16	Smucker & Clarke-1-2-3 [2012]	TBG				x			x			
17	Moffat et al. [2012]	INSQ	x				x					
18	Moffat et al. [2013] & [2019]	C/W/L	x	x	x		x					
19	Chuklin et al. [2013]	CMBM	x	x					x			
20	Sakai & Dou [2013]	U-measure				x			x			
21	Ferrante et al. [2014]	MP	x									

22	Giannakouris et al. [2014]	CSMR										
23	Xenophon et al. [2014]	RRP		x					x	x		
24	Moffat et al. [2015] & [2016]	INST	x		x		x					
25	Chuklin et al.-2 [2015]	RCM									x	
26	Evangelopoulos et al. [2016]	IRDP		x					x	x		
27	Jiang & Allan [2016]	AdapEM	x	x		x			x			
28	Chuklin & Rjike [2016]	CAS	x					x	x			
29	Luo et al. [2017]	HBG				x			x		x	
30	Zhang et al. [2017]	BPM			x	x	x					
31	Jiang & Allan [2017]	AdapPM	x	x			x		x			
32	Azzopardi et al. [2018]	IFT			x	x	x	x				
33	Thomas et al. [2018]	CAM	x	x	x		x	x		x		
34	Azzopardi et al. [2020]	DDM	x	x				x				
35	Ferrante & Ferro [2020]	P@H	x	x								
36	Diaz et al. [2020]	EED, EER		x					x	x		
37	Breuer et al. [2021]	Repro_Eval							x	x		

Table-43: Comparison of model-based evaluation metrics-II

Table-44 depicts User Utility Functions based on Gain and Efforts with further emphasis on Cumulative, Reciprocal and Expected Gain and Effort. Wherever "x" is marked in a cell, it means that corresponding parameter is counted or considered in that metric/model. These markings are based on the assumptions derived from Table-41 above; so, for an easy understanding, annotate the below information keeping both in purview.

Comparison of model-based evaluation metrics - III			
			User Utility Functions

Sr.	Citation Key	Metric	Gain	Cumulative Gain	Reciprocal Cumulative Effort	Expected Average Gain	Expected Cumulative Gain	Others
1	Jarvelin & Kekalainen [2002]	DCG		x				
2	Moffat & Zobel [2008]	RBP	x					
3	Robertson [2008]	AP						
4	Craswell et al. [2008]	CUM						
5	Sakai & Robertson [2008]	NCU						
6	Dupret & Piwowarski [2008]	UBM						
7	Clarke et al. [2008]	nDCG						
8	Agrawal et al. [2009]	nDCG-IA						
9	Chapelle et al. [2009]	ERR			x			
10	Chapelle & Zhang [2009]	DBN						
11	Guo et al. [2009]	DCM						
12	Li et al. [2009]	GA						
13	Yilmaz et al. [2010]	EBU		x				
14	Chapelle et al. [2011]	CBM-IA						
15	Carterette [2011]	CF	x	x	x	x		
16	Smucker & Clarke-1-2-3 [2012]	TBG		x				
17	Moffat et al. [2012]	INSQ	x					
18	Moffat et al. [2013] & [2019]	C/W/L	x	x	x			
19	Chuklin et al. [2013]	CMBM	x		x			
20	Sakai & Dou [2013]	U-measure		x				
21	Ferrante et al. [2014]	MP				x		
22	Giannakouris et al. [2014]	CSMR						
23	Xenophon et al. [2014]	RRP						
24	Moffat et al. [2015] & [2016]	INST					x	
25	Chuklin et al.-2 [2015]	RCM						
26	Evangelopoulos et al. [2016]	IRDP						
27	Jiang & Allan [2016]	AdapEM				x	x	
28	Chuklin & Rjike [2016]	CAS		x				
29	Luo et al. [2017]	HBG		x				

30	Zhang et al. [2017]	BPM		x	x	x		
31	Jiang & Allan [2017]	AdapPM	x	x	x			
32	Azzopardi et al. [2018]	IFT	x	x				
33	Thomas et al. [2018]	CAM	x					
34	Azzopardi et al. [2020]	DDM	x	x	x			
35	Ferrante & Ferro [2020]	P@H				x	x	x
36	Diaz et al. [2020]	EED, EER						
37	Breuer et al. [2021]	Repro_Eval						

Table-44: Comparison of model-based evaluation metrics-III

For defining Search Tasks around the finalized research methodology: The above demonstrated study of 37 relevant research models furnishes a vast ground as reference point for defining the search tasks for our research methodology. Based on R5MLR learned above, we can figure out the most suitable research method and model for us as follows:

1. **To test H1:** *Eye tracking method, Log file data analysis, Filming of SERP Navigation Behaviour* ***(FSNB)***
2. **To test H2:** *Eye tracking method, Filming of SERP Navigation Behaviour*
3. **To test H4a:** *Landing Page Analysis method, Filming of SERP Navigation Behaviour*
4. **To test H4b:** *Purpose diagnosis and information access norms, Govt. laws regulations, Filming of SERP Navigation Behaviour*
5. **To test H7:** *Demographic, cultural, and ethical norms depiction set*
6. **To test H6:** *Aesthetic, user-friendliness, usability measuring method, Filming of SERP Navigation Behaviour*

7.3. Conceptualization of the Research Statement (Research Problem and Model to test)

According to [254]*Badenhorst* [2022] Problem Statement succinctly states the key problem in one or two sentences in one or two lines *(also called as the rationale of the problem);* rest of the problems shall fall in its background. Through a problem statement, Knowledge Gap(s) in previous research(es) is/are next point to be addressed. Based on FLR, SLR and pertinent research methodologies reviewed in previous chapters and above we can define multiple research statement as below:

[254] https://www.youtube.com/watch?v=KNgPbPM2tW8

1. **Problem-1:** No data or research exists about **Exploiting vs. Exploratory** Health Information Search Behaviour in French Audience with their demographics. Knowing this fact can help business strategists and content creators in formulating impeccable content pieces for French SERP surfers.
2. **Problem-2:** There is least or no information or research till date regarding the SERP **User Gain Rate**, due to which French Content Creators always struggle in assessing their data from psychological satisfaction angle; the web analytics data are not efficient enough in calculating the user gain. However, only big giants of Search Engines like Google may know these facts from their Log Data but not the content creators. This research shall be helpful in providing them with French Audience SERP surfing Gain Rates.
3. **Problem-3:** There is no study found in the French Market regarding **Search Intents** specific to Health Search Users. This thesis can perhaps be the first in proposing 6 Intent Types specifically related to Health Search named as Urgency Intent, Precautionary Intent, Reassurance Intent, Diagnostic Intent, Treatment Intent and Courtesy Intent. Our Search prompt Tasks are scripted to identify each of these Intents.
4. **Problem-4:** There are four types of **standard intents** as Informational, Navigational, Transactional and Commercial; while there is sizeable research found in other markets in the world that French Health market lacks in spotting their propositions. So, this thesis aims to propose these intents with the intensity in SERP searches.
5. **Problem-5:** There are no research or data found regarding how the **privacy concerns** drive Health Search Behaviour; privacy concerns specific to cultural norms. Cultural norms are something which make the Psycho-Cognito acumen of its follwoers. So, it can be interesting to get such a profiling through this research.
6. **Problem-6:** How it works regarding **personalization of SERP** content in Health Searches in French market. We find no study at this aspect and audience set, so our research can have the possibility of addressing this aspect as well.

Thus, we want to test and find solutions to above problems based on logics and evidence.

Context of the research study: Just like any other research requires a ground and realisable data experiment context in terms of environment, time, institution and audience set; the selected audience ground set is the French Market Health Information Searchers on Google; particularly the students of 20 to 30 years of age at University of Lille, France.

7.4. R5Ms Components of the Research Problem to Test

Conceptualization Framework: The problems above and the context conceived further needs a Framework to caricate the problem testing phenomenon, which for this, we need to think as a SERP User mindset and first of all script the possible variation of 'Health-Search-Prompts/Queries' which are depicted in Chat-91 below.



Chart-91: SERP Search Path Mind-Map

Next thing comes for figuring out is the kind of search tasks a SERP user usually acquires to accomplish their research goal. This goal achieving task is formed based on various psychological characteristics as depicted in

Diagram-8. To figure out the all dependencies and requirements of achieving the User's Search Behaviour characteristics, we prepare a R5Ms Framework as in Chart-92:

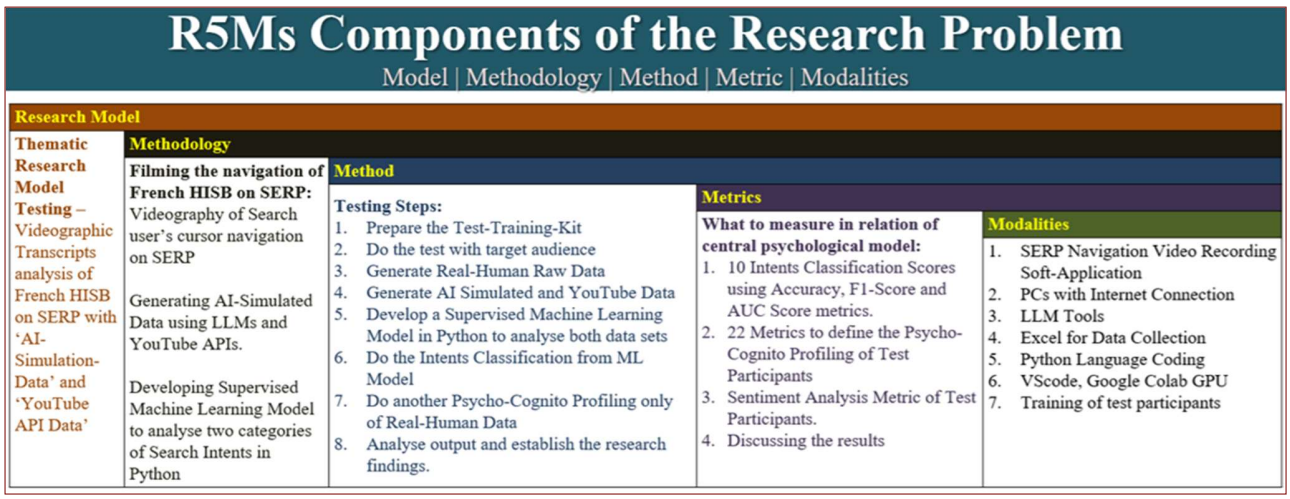


Chart-92: R5Ms Framework of the Research Problem

7.5. Research Problem Contemporary Literature Review 1 (RPCLR-1):

Referring to Table-40 above an assessment of choosing one of the five possible methodologies, "Filming of SERP Navigation Behaviour" was chosen to proceed with practical testing. After having its complete set of R5Ms figured out technically as per Chart-92 above, here we intend to review the relevant literature around similar R5Ms especially around collecting navigation data through Filming of SERP Navigation Behaviour.

Filming of SERP Navigation Behaviour (FSNB) Methods: *Crockford & Agius* [2006] while researching about video files database searching methods cite four types of methods as Circular incremental (55%), Circular decremental (10%), Educated guess (29%) and Random selection (6%); and regarding video watching browsing behaviour they found Straight viewing (21%), Speed switching (46%), Inaccurate shuttle determination (13%), Accurate shuttle determination (7%), Half and refine approaches (13%). However, this finding is not as such directly related to SERP navigation behaviour; but and as screen behaviour is a common aspect of both studies so these findings give a certain level confidence and idea related to this thesis question.

Talking about the comparative importance of Filming of SERP Navigation method over other *Nettleton & Gonzalez-Caro* [2012] state that the web log data has several important shortcomings. The main one being that we don't really know what the user is doing - is s/he looking at the screen or doing something else? While if we

use Filming of SERP Navigation method, it has the full potential of answering to what user does on SERP screen. **So, this point can be seen as an answer to a why that Filming of user behaviour was chosen instead of analysing a user log data.**

The decision to use video recording for data collection should be informed by the purpose and aims of the study as well as the unit of behaviour being measured, *Caldwell & Atwal* [2005]. On the other hand, video recording provides only a window of what happens in real time and may lack important contextual data, *Latvala et al.* [2000]. Participant reactivity is a persistent issue in any observational research, but the use of cameras means the presence of an additional "eye", *Gross* [1991]. The "Hawthorne effect" is a classic example of reactivity that can be decreased by exposing participants to longer periods of observation to acclimate them to the presence of the observer. This can be accomplished by conducting a sensitizing session or by analysing data taped later in a predetermined "middle" segment of the session, *Harrison Elder* [1999]. A reasonable representation of the behaviour of interest can be obtained by sampling either by time or by event, *Casey* [2006]. Researchers need to carefully define the temporal and spatial boundaries of the phenomena of interest.

Video recording has been under-utilized as a data collection tool because of confidentiality and privacy issues. However, it has many benefits, and recent studies using video recording methods have introduced new research areas and approaches such as performance evaluation and analysing clinician-patient interactions. Video techniques have benefits such as multi-channel recording and video coding and comparing "unmanned" video recording with the traditional observation method in primary care research, *Asan & Montague* [2014]. *Onur & Enid* presents a comparatist list of Pros and Cons of using Filming (Video Recording) method for observing the behaviour as in Table-45, Appendix-9.

Filming of SERP Navigation Behaviour (FSNB) Metrics: To understand the FSNB Testing metrics, this section furls the related Literature about their outputs and mutual relationship. *Fubel et al.* [2023] in their assessment of the impact of 24 different SERP features on organic CTR findings underscore that SERP features have a tangible impact on CTR, thereby holding their own against other result page characteristics. Their findings not only confirm the widely accepted view that the ranking position of a SERP greatly influences its CTR but also reveal that other SERP characteristics - such as the presence of specific SERP features, the nature of the searched keyword - can considerably affect CTR. Chart-94, Appendix-9 from them shows that out of all SERP features, Job Search, FAQ, Knowledge Panel, People also ask and Tweet have strong positive correlation with CTR. In the FSNB experiment of this thesis, it can be of high importance to know about these traits in French HISB on SERP.

Filming of SERP Navigation Behaviour (FSNB) Modalities: Once the behavioural measurement criteria *(Method and Metrics)* is established, the next step is to select appropriate measurement instruments i.e. Modalities. Behavioural measures can be as simple as frequency counts of the occurrence of a target behaviour or a continuous measure of the duration of that behaviour, *Roter et al.* [2000]. The first decision in selection of observational measures is whether the level of measurement requires a microanalytic or macroanalytic approach. Microanalytic approaches can produce fine discrimination between units of analysis and a variety of information, but are very time consuming, resource-intensive, and produce large datasets. *Janice* [1988] recommended microanalysis for single-subject case designs or small-sample studies. Macroanalysis, or analysis of an entire interactional process, using global measures may be more appropriate for studies evaluating the effectiveness of an intervention. Observational tools designed for use with video recording allow researchers to use manualized procedures and decision rules to assign codes to nonverbal behaviours, verbal expressions, or vocal utterances. For example, the Facial Action Coding System (FACS) is a microanalytic system for rating facial movements and emotional facial expressions, *Emily et al.* [2015]. Across all studies, training of video coders avoid fatigue when coding tapes and to take frequent breaks to counteract drift. Coders are also asked to avoid coding tapes on days when they are emotionally upset or ill, as their own emotional state can influence the behavioural observations they make. *D'Angeli et al.* [2022] in their research studied about 14 Videorecording tools, out of which Scala di Valutazione delle Interazioni Alimentari **(SVIA)** was used by *Cimino & Cerniglia, 2022 and Cimino, Almenara, & Cerniglia, 2022* in their online observation studies on Covid subject. However, not as such directly related except the online recording, but to keep in mind; this tool can be tested on usability of our requirements.

Asan & Montague [2014] have enlisted 12 Video Analysis Computer Programs utilized in several studies such as Observer/Noldus, MacSHAPA, A.C.T., Vanna, VINA, Tagging Software, Computer assisted time and event recorder (CATER), The ALFA (Activity Log Aggregation) toolkit, Atlas.ti, QSR NVivo and HyperResearch; **Atlas.ti and HyperResearch** sounds nearly compatible to the Filming of SERP Navigation Behaviour.

The kind of navigation behaviour we want to analyse in our thesis falls under nonverbal behaviour methods, for which the Observational Quantitative *(Mast, Hall, Klockner, Choi -2008),* Nonverbal Accommodation Analysis System **(NAAS)** *(D'Agostino & Bylund, 2010),* Correlation Analysis *(Collins, Schrimmer, Diamond & Burke, 2010)* and Lag-Sequential Analysis *(Connor, Fletcher & Slamon 2009)* can be in relevance to our methodology.

Research Prompts Scripting and Test-Sets defining: In order to develop our test-sets to provide to our test-participants we searched about the possible use cases of search on Google by French Health Information Searchers. As per our own learning and experience in healthcare research, we state that French people often search for a variety of topics related to healthcare, especially on Google. Some of the most common searches include:

1. **Health Conditions and Symptoms:** French users frequently search for information about specific diseases, symptoms, and medical conditions such as diabetes, hypertension, migraines, and mental health issues like anxiety and depression.
2. **COVID-19:** Since the pandemic, there has been a high volume of searches related to COVID-19, including symptoms, vaccine information, testing locations, and updates on government guidelines.
3. **Diet and Nutrition:** Topics related to healthy eating, weight loss, and specific diets *(like keto or Mediterranean)* are popular. Searches for the nutritional content of food and advice on managing chronic conditions through diet are common.
4. **Medications and Treatments:** Many people search for information about prescription drugs, over-the-counter medications, side effects, and alternative treatments, including homeopathy and herbal remedies.
5. **Healthcare System Information:** French users often look up details about how to navigate the healthcare system, including how to find a doctor, information about health insurance (like "Mutuelle"), and hospital services.
6. **Pregnancy and Childcare:** Pregnancy-related queries, information about childbirth, breastfeeding, and childcare are commonly searched, reflecting a strong interest in maternal and child health.
7. **Mental Health:** There's a growing interest in mental health, with searches focusing on therapy, stress management, and conditions such as depression and anxiety.
8. **Fitness and Wellness:** Searches for exercise routines, wellness tips, and information on maintaining a healthy lifestyle are popular, especially post-pandemic.
9. **Public Health Campaigns:** People often search for information on national health campaigns, such as those related to vaccination, smoking cessation, or cancer screening.
10. **Healthcare News and Policy:** Searches for the latest news related to healthcare, such as changes in healthcare policy, public health issues, or medical advancements, are also common.

These trends reflect a combination of interest in both personal health management and broader public health concerns. Based on [255]*Chloé R* [2020] French people frequently search for following information and topics about healthcare on Google:

1. **Coronavirus/COVID-19:** In 2020, searches related to coronavirus outpaced all other health topics, generating over 26 million searches in France.
2. **Flu:** Before the COVID-19 pandemic, flu was one of the top health concerns for French people searching online.
3. **Common health issues:** Other popular searches include topics like gastroenteritis, urinary tract infections, and migraine.
4. **Diet-related queries:** Searches for topics like the "ketogenic diet" have seen significant increases.
5. **Medications:** The top 10 Google searches for drugs in France include painkillers, antibiotics, and contraceptives.
6. **Specific diseases:** Diseases like lupus have seen increased searches, often linked to celebrity discussions of health issues.

According to *Ducrot et al.* [2021] and *Beck et al.* [2014] key trends in Online Health Information Seeking have been as follows:

1. There was rapid growth in internet use for health information from 2010 to 2014, followed by a decrease in 2017.
2. Young adults (aged 15-30) are particularly likely to use the internet for health information.
3. Trust in online health information has been decreasing, highlighting the need for reliable sources.

To assess the top three traffic grabbing Public Healthcare websites of France, as per [256]*SEMrush* [2024] these are ameli.fr, doctolib.fr and sante.fr. To assess how importantly Google Search would be contributing to Telemedicine, the largest trend emerged after Covid-19, [257]*Demoule* [2021] research reveals that the interest in telemedicine has grown significantly, especially during the COVID-19 pandemic in figures by Nov. 2020, 86% of French people knew what telemedicine was, and 92% were familiar with teleconsultation.

[255] https://www.alioze.com/en/top-health-searches-google-france/
[256] https://www.semrush.com/trending-websites/fr/healthcare
[257] https://www.apizee.com/telemedicine-the-state-of-play-in-france.php

These trends show that French people are increasingly turning to online sources for health information, with a focus on current health issues, common ailments, and emerging healthcare technologies like telemedicine. We shall be taking this assessment as base for defining our four test sets in Chapter-9.

7.6. Probable Research Biases

After having a thorough learning of the concept of biasing in research like ours, we found following possibilities of research biasing that can take place in our methodology. Based on biases mentioned in Table-46, we have discussed about our intuitions about it in Chapter-10.

Sr.	Bias Type	Occurrence Probability in Variables of my thesis
1	Hostile Attribution Bias	Exploring-Exploiting-Engaging Vars.
2	Vividness Bias	Intent Types, User Gain Rate
3	Optimism Bias	Privacy and Personalization Concerns
4	Perception Bias (false consensus effect)	User Gain Rate, Privacy
5	Positive Illusion	Personalization
6	Recall Bias	Partial impact on all variables
7	Response Bias	Audience participation interest
8	Observer Bias (detection bias or ascertainment bias)	Audience participation interest
9	Bandwagon Effect	Privacy and Cultural Habitude
10	Observer-expectancy effect (Pygmaion or Rosenthal effect)	Audience participation interest
11	Actor–observer bias	Privacy and Cultural Habitude
12	Availability heuristic (Availability bias)	Exploring-Exploiting-Engaging Vars.
13	Confirmation bias	SERP Personalization and Persistency
14	The halo effect	SERP Personalization and Persistency
15	The Baader–Meinhof phenomenon	Exploring-Exploiting-Engaging Vars.
16	Sampling bias or ascertainment bias	SERP Personalization and Persistency
17	Attrition bias	Audience participation interest
18	Volunteer or self-selection bias (also a sampling	Audience participation interest

	bias type)	
19	Undercoverage bias (also a sampling bias type)	In data analytics
20	Pre-screening or advertising bias	User Gain Rate, SERP Personalization and Persistence
21	Healthy user bias	Audience participation interest
22	Social desirability bias	Privacy and Cultural Habitude

Table-46: Probable Research Biases of this thesis

Chapter-8

Bibliographic Data Analysis

This chapter presents a complete analysis of Literature Review done at four place of this thesis; first the Unit-1 of Foreground Literature Review (FLR), second the Chapter-5 of Synthesized Literature Review, third the Pillar-Topic-2 of Chapter-7 of Literature Review regarding the Research Metrics, Methods, Methodologies, Models and Modalities (R5Ms) relevant to this thesis and the fourth the Pillar-Topic-3 of Chapter-7 dedicated to the Literature Review around the final Research Model of this thesis i.e. the Eye Tracking Model in context of SERP. This analysis is distributed as and in following sections:

8.1. Qualitative and Quantitative (Q&Q) Analysis of Bibliography:

Quantitative Segmentation and Analysis of Bibliography: Referring to Table-4 in Introduction section, here below in Table-47 presents the segmentation of bibliographic references distribution on criteria of Source Type, Doc Type, Utility Type, Time Period, Pillar Topic and Topic, R5Ms Facts and Annexures.

Q&Q Analysis of Bibliography: Quantitative Segmentation

Analysis as per Source Type, Doc Type, Topic Type, Time Period, Utility Type, R5Ms Facts & Annexures

Source Type	No.	%age
Scientific Journals	427	37.98%
Company Researches	156	13.87%
Institutional Researches	89	7.91%
Publishers	143	12.72%
Blogs/Online Articles	267	22.50%
Local References	56	4.98%

Annexures	No.
Abbreviations	199
Data Visuals	195
Keyword Categories	29
Oneliner Facts-list	143
SERP Features	79

Time Period	No.
From 1929 to 1950	7
From 1951 to 1975	12
From 1976 to 2000	67
From 2001 to 2005	89
From 2006 to 2010	134
From 2011 to 2015	201
From 2016 to 2020	305
From 2021 to 2025	309
Total	**1124**
Oldest paper of year	1929
No year references	101
Last 10 years	55%

Utility Type	No.
Scanned	1704
Considered	1178
Included	1124
Articles	789
Books	134
Unique Sources	364
Zotero Docs	1418
LR Folder of PDFs	1503

Doc Type	No.
Articles	789
Reports	160
Patents	67
Books	134
Chapters	78
Videos	12

R5Ms Facts	No.
Research Models	41
Research Metrics	37
Hypothesis Derivatives	62
Research Scopes	44
Specific SLR Biblio	255

Table-47: Quantitative Analysis of Bibliography

Qualitative Segmentation and Analysis of Bibliography: This analysis categorizes the academic papers, articles, books and other document types used in this thesis on Source Types, Source Brand with total number of times a Source Type is used in retrieval of a relevant document.

Source	Total references	Source Types	Source	Total references	Source Types
researchgate.net	130	Scientific Journals	onlinelibrary.wiley.com	5	Scientific Journals
dl-acm-org	85	Scientific Journals	tandfonline.com	5	Scientific Journals
typeset.io	74	Scientific Journals	bmcmedinformdecismak	5	Scientific Journals
arxiv.org	53	Scientific Journals	blog.google	4	Blog
link-springer-com	41	Scientific Journals	citeseerx.ist.psu.edu	4	Institutional Research
ieeexplore	36	Scientific Journals	marketbrew.ai	4	Blog
sciencedirect.com	33	Scientific Journals	moz.com	4	Blog
patents.google.com	32	Company Research	patentimages.storage.googleapis.com	4	Company Research
doi.org	28	Scientific Journals	algolia.com	4	Company Research
semanticscholar.org	23	Scientific Journals	ryenwhite.com	4	Blog
wikipedia.org	20	Blog	core.ac.uk	3	Scientific Journals
academia.edu	16	Scientific Journals	diva-portal.org	3	Scientific Journals
searchenginejournal.com	13	Blog	ebscohost-com	3	Scientific Journals

linkedin.com	12	Blog	faculty.ist.psu.edu	3	Institutional Research
ncbi.nlm.nih.gov	12	Scientific Journals	forbes.com	3	Company Research
youtube.com	12	Video	frontiersin.org	3	Scientific Journals
searchengineland.com	8	Blog	ibm.com	3	Company Research
developers.google.com	6	Company Research	iopscience.iop.org	3	Scientific Journals
dl.acm.org	6	Scientific Journals	journals.sagepub.com	3	Scientific Journals
google.com	6	Company Research	jstor-org	3	Scientific Journals
onlinelibrary-wiley-com	6	Scientific Journals	kopp-online-marketing.com	3	Blog
pubmed.ncbi.nlm.nih.gov	6	Scientific Journals	neilpatel.com	3	Blog
aclanthology.org	6	Company Research	people.eng.unimelb.edu.au	3	Institutional Research
jmir.org	6	Scientific Journals	support.google.com	3	Company Research
link.springer.com	5	Scientific Journals	techtarget.com	3	Blog
microsoft.com	5	Company Research	**Sources with 2 references**	64	N/A
Sources with 1 reference	292	N/A	**Grand Total**	**1124**	

Table-48: Qualitative Analysis of Bibliography

VOSviewer Maps and Connected Papers Qualitative Analysis: In Appendix-6, there are two scientific tools used to assess the overall and specific papers quality used in this thesis in terms of their Citation Score evaluation.

VOSviewer is a powerful software tool designed for constructing and visualizing bibliometric networks, such as co-authorship, citation, and keyword co-occurrence maps. These visualizations, known as VOSviewer maps, enable researchers to explore the structure and dynamics of scientific literature by identifying clusters, trends, and influential publications within a research field. Using data from sources like Web of Science or Scopus, VOSviewer applies advanced text mining and clustering techniques to generate intuitive, color-coded maps that reveal relationships among authors, journals, or terms. These maps are widely used in literature reviews, research evaluations, and science mapping to support data-driven insights into scholarly communication.

Like VOSviewer, **Connected Papers** maps are also another kind of visual literature graph designed to help researchers explore academic papers and understand the structure of a research field. By inputting a seed paper, the tool generates a network of related works based on co-citation and bibliographic coupling, even if the papers do not directly cite each other. Each node in the map represents a paper, with its size and position reflecting citation count and thematic proximity. This intuitive visualization enables users to identify foundational studies, recent developments, and potential research gaps, making it a valuable resource for conducting comprehensive literature reviews and building a thesis bibliography.

8.2. Research Thesis Shaping-up Funnel:

As evident from the name, this Thesis Funnel reflects the complete research work done in systematic stages to finally bring down to the final research model of this thesis. Taking on the arrowheads in the left column of Chart-95 below, '**Source Types'** refers to kind of resources furled; '**Docs Types**' refers to number of different resources furled; '**Publishing Period**' mentions the different time zones in which the said resources were published; '**Reference Sources'** are the number of different research databases from where the resources were retrieved; '**Bibliography Size'** refers to number of total and unique documents reviewed and included in this thesis respectively; '**Inclusion in SLR'** indicates the number of resources included in SLR; **'Hypotheses'** tells about the number of Raw Hypotheses derived from FLR and SLR Combined; **'Framework'** is the final number of hypotheses retained for framing the '**Variables**' of the central '**Psychological Research Model'** of this thesis. Then comes the next arrowhead '**R5MLR**' referring to the various research models analysed in purview of the central model. Next in this FUNNEL is the '**Technically Testable Research Model'** proposed as **Analysis of Filming of French HISB on SERP** after having shortlisted the five probables. Its details are talked in the 3rd section of Chapter-7.

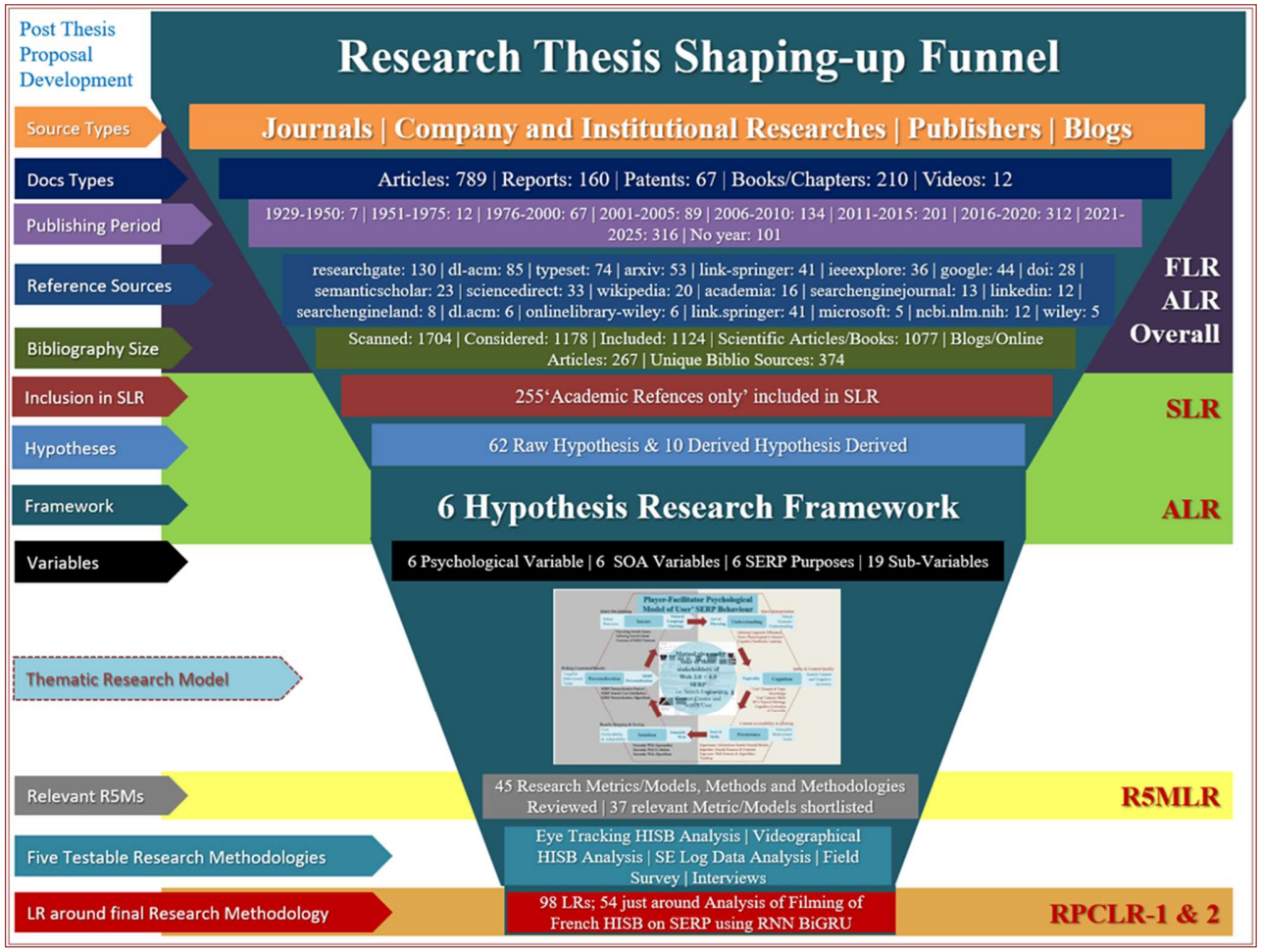


Chart-95: Research Thesis Shaping-up Funnel

8.3. PRISMA Validation of FLR, SLR, R5MLR, RPCLR and ALR:

The complete bibliographic statistics talked in section 8.1 and 8.2 above are also presented as the standard PRISMA Validation Model by *Moher et al.* [2009]. The Chart-96 represents the statistics on the parameters defined by Moher et al. Hypothetically; the complete literature review and synthesis can be seen in three PRISMA Set-ups.

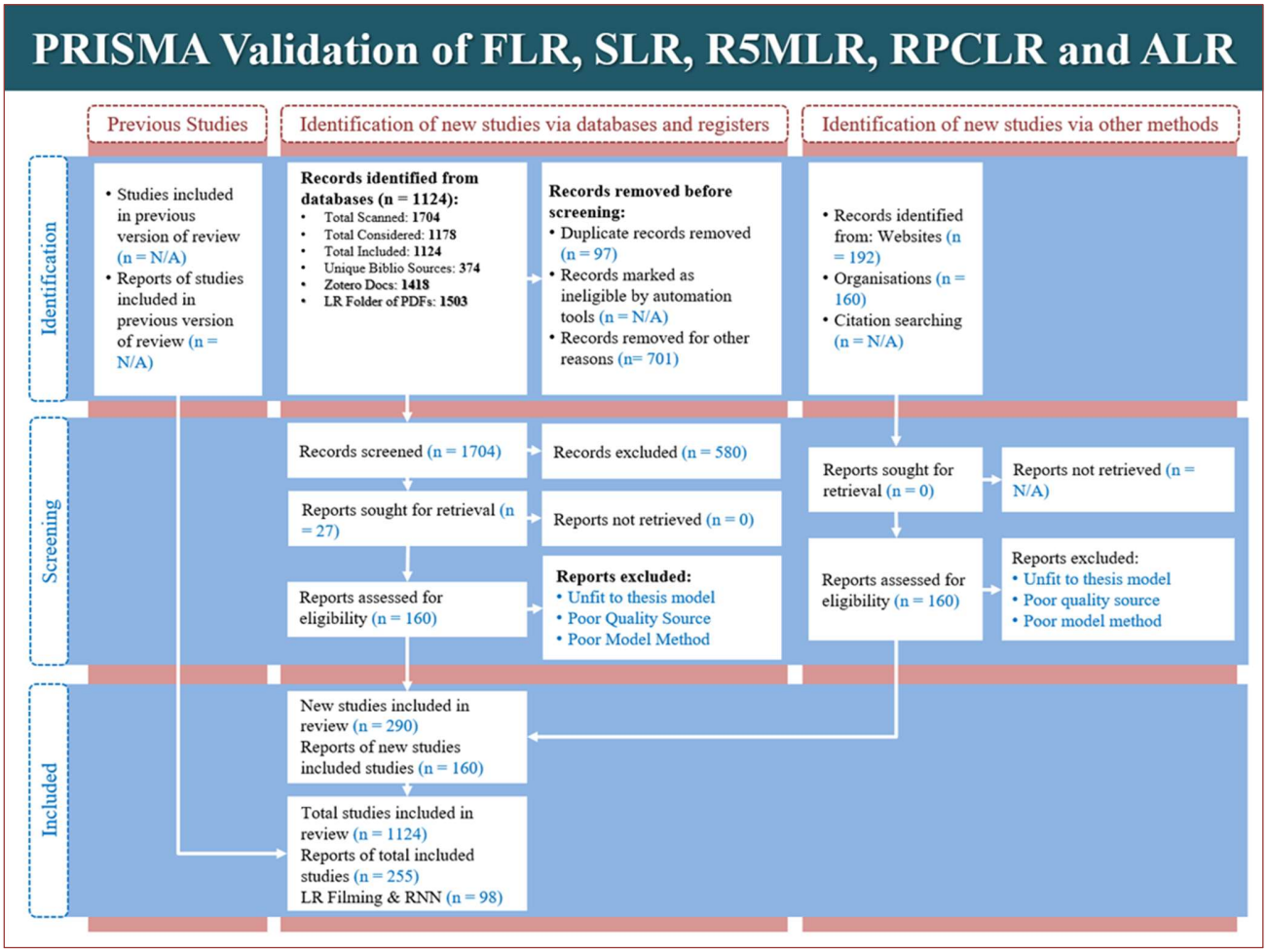


Chart-96: PRISMA Validation of FLR, SLR, R5MLR, RPCLR & ALR

Unit-3: Research Test Ontology

Synopsis: This unit is comprised of research test ontology, variables and data metrics mapping, test pathways scripting, research data collection methods, calculations, analysis, conclusion, discussion of the results, and future recommendations; along with a talk about the limitations of this thesis and future scope. This way, this unit is comprised of two chapters:

Chapter-9: Test-Participants Profiling, Data Taxonomy, Research Model and Results Validation: This chapter presents a synchronized analogy of ontology of the research model test, Research Method adapted to Test our research model, Results Data and Statistics, Results Analysis and Results Validation.

Chapter-10: Thesis Wrap-up *(Research Discussion, Conclusion, Limitations, Future Scope)*: This chapter elaborates and conclude the findings of our Research Model Test and overall findings of our research work. At the end Limitations along with Future Scopes and Recommendations are talked to tender move-forward about the subject matter of this thesis.

Chapter-9

Test-Participants Profiling, Data Taxonomy, Research Model and Results Validation

Taking Variables and Hypotheses into Research Test: Before moving to Research Model Testing and Analysis, this is a moment where syncing the things developed till this stage should enthuse a consistent understanding of key components of this thesis in a sequential manner. The Chart-97 below depicts all those components in sequentially.

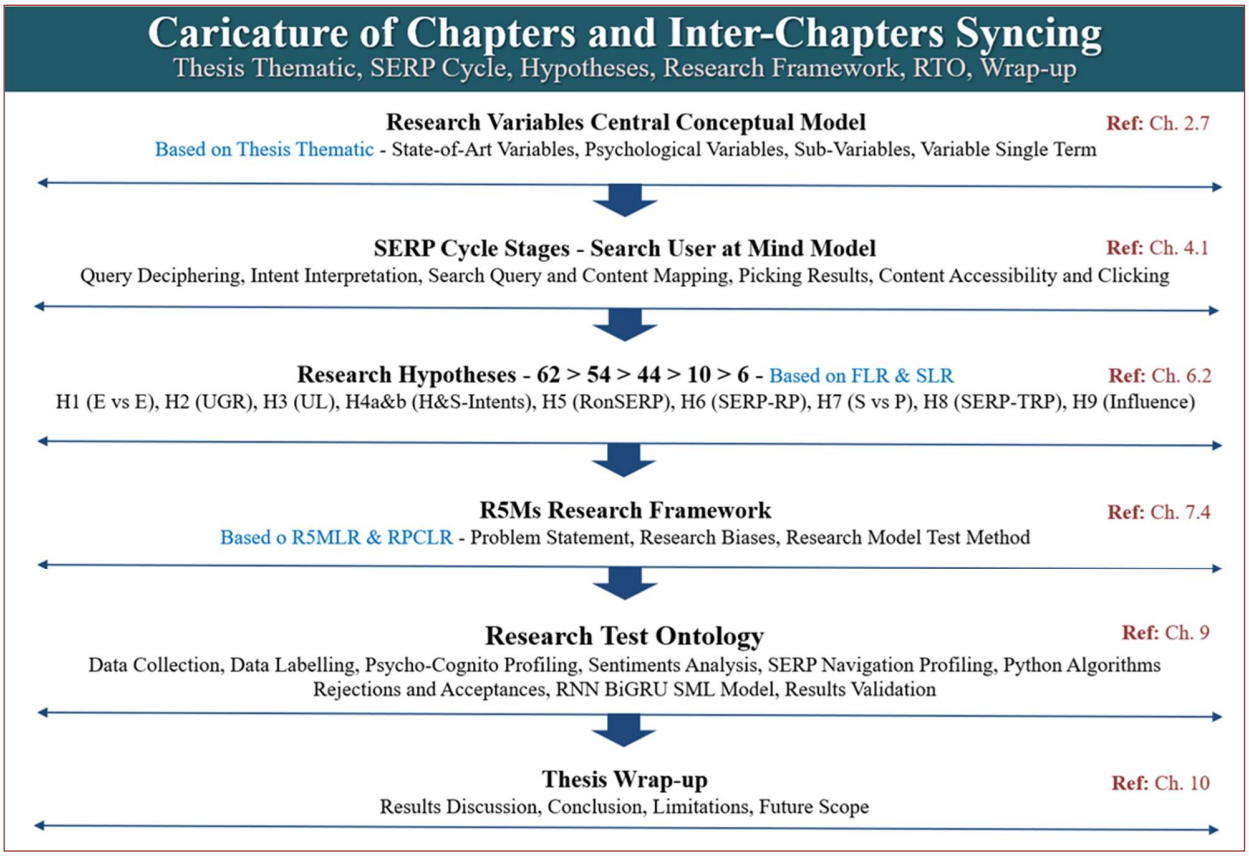


Chart-97: Caricature of Chapters and their Inter-Chapters Syncing

9.1. Psycho-Cognito Profiling and Hypotheses Evaluation:

However, 'Data Taxonomy' is discussed in Section 9.2, and here we are starting directly with 'Psycho-Cognito profiling' of our test-participants, because it is based on post-test Google Data survey analysis findings as well.

'Data Taxonomy' section follows the 'RNN BiGRU Intents Classification Modelling' and as prime concentration of the data is with this modelling so to talk in flow 'Data Taxonomy section is placed after this section'.

Psycho-Cognito Profiling: To analyze the six factors as per our central Psychological model, Diagram-8 viz-a-viz Cognitive Level, Understanding Level, Sentiment Involvement Level, Intuitive Relevance *(Semantic)* Level and Personalization Concerns *(here to note that Intent is not included, because we think that Intent is a next level but immediate outcome of Psycho-Cognito Profiling, so despite putting Intent together with all others in Diagram-8, its calculations are still independent of the whole model)* of all test participants fully or partially; a Python Algorithm was coded to analyse 627 video test transcripts *(inserted into Python code through an Excel file)*, on 22 different Psychological and Cognitive statistical metrics *{Type Token Ratio* ***(TTR),*** *Avg Sentence Length, Parse Tree Depth, Lexical Density, Abstractness Score, Ambiguity Score, Cognitive Verbs, Reasoning Connectors, Abstract Words, Comprehension Difficulty, Cognitive Load Score, Cohesion Ratio, Sentiment Polarity, Sentiment Subjectivity, Self-Reference, Empathy Terms, Certainty Terms, Tentative Terms, Flesch Score, Gunning Fog Score, SMOG, Scolarius Level}* to know Pycho-Cognito level of each participant individually and as 'combined average of all of them' *(Python file names were as below)*:

i) **cognitive.analysis2.py**
ii) **linguistic.analysis4.py**
iii) **sentiment-analysis2.py**

After a rigorous scrutiny of possible flaws, inconsistencies, missing statistical rigor, dubious data, and interceptive overreach, we found a strong linguistic analysis with moderate psychological interpretations. Only 7 out of 22 metrics considered above were reliable enough for our test, as rest of them need more rigorous insights from the transcripts; which we have left for another time for another research.

Validated Metrics List:

1. **Type Token Ratio (TTR)** – Linguistic Complexity: 0.467
2. **Average Sentence Length -** Linguistic Complexity*:* 16.80 words
3. **Lexical Density** - Linguistic Complexity*:* 0.581
4. **Flesch Reading Ease** - Readability*:* 62.88
5. **Sentiment & Tone Polarity*:*** 0.151
6. **Sentiment & Tone Subjectivity*:*** 0.439
7. **Self-Reference Score -** Content Focus*:* 20.42

The test-participants profile shows **moderate linguistic sophistication** with a **TTR** of 0.467 indicating balanced vocabulary diversity, neither repetitive nor overly complex. **Average sentence length** of 16.80 words suggests structured, deliberate expression typical of educated users seeking specific information. The **high lexical density** (0.581) confirms information-packed queries rather than conversational language. The **Flesch readability score** of 62.88 places content at 8th-grade level, accessible but substantive. **Sentiment analysis** reveals slight positivity (0.151) with **balanced objectivity/subjectivity** (0.439), suggesting factual inquiry with personal investment. The extremely **high self-reference** score (20.42) dominates the profile, indicating **strongly personal-context queries**, likely health*, (another same set topics are legal, or financial)* concerns where the user's specific situation drives search behaviour. Test-participants behavioural pattern constructs well-formed, content-dense queries about personal situations, maintaining slight optimism while seeking specific, actionable information. Language suggests functional literacy with domain-specific needs rather than casual browsing.

Cognitive Level Assessment: This assessment reveals an upper-intermediate to advanced processing capacity, supported by a TTR of 0.467 and an average sentence length of 16.8, indicating the ability to manage complex information structures. A lexical density of 0.581 reflects efficient handling of information-rich content, while a Flesch score of 62.88 suggests comfort with 8th-grade complexity, though structured queries point to a higher actual capability. The understanding capability is domain-specific, with a balanced subjectivity score of 0.439 showing integration of factual and personal context, a high self-reference rate of 20.42 with moderate complexity indicating sophisticated personal knowledge mapping, and a positive sentiment score of 0.151 reflecting confidence in evaluating information quality.

Semantic Analysis & Intuitive Relevance Capabilities: This analysis highlights contextually sophisticated semantic processing, demonstrated by high lexical density and structured sentences that enable compression of complex semantic relationships. A self-referential focus reveals the ability for analogical mapping, linking general information to personal situations, while a moderate TTR reflects semantic flexibility in expressing similar concepts through varied language. Intuitive relevance capabilities are highly developed, with strong self-reference scoring indicating automatic filtering through a personal relevance lens, balanced objective/subjective processing suggesting intuitive source quality assessment, and consistent positive sentiment showing confidence in relevance judgments.

Proficiency Skills Assessment: This proficiency skills assessment identifies an advanced-intermediate level of information processing, with a lexical density of 0.581 indicating efficient extraction and synthesis of information, sentence complexity reflecting comfort with multi-layered argumentation, and readability preferences supporting optimal cognitive load management. Communication proficiency is functionally sophisticated, demonstrated by balanced TTR enabling expressive flexibility without excessive complexity, sentiment control indicating emotional regulation during information seeking, and self-contextualization revealing advanced metacognitive awareness.

Personalization Sensitivity: This analysis indicates extremely high personalization dependency, with a self-reference score of 20.42 reflecting hyper-personalized information processing. Every query is filtered through a personal relevance framework, suggesting a strong preference for content that connects directly to individual context. There is a likely struggle with abstract or general information that lacks personal connection, and the perceived value of information is directly tied to its personal applicability.

Limitations of Psycho-Cognito Profiling*:*

- Over-personalization may **narrow search** scope unnecessarily.
- May miss valuable general information due to **self-reference bias**.
- Potential filter bubble creation through **excessive personalization**.

Categorical Definition and Description of 7 Validated Metrics:

Lexical and Structural Complexity Measure:

1. **TTR (Type Token Ratio):** TTR is a measure of vocabulary richness i.e. lexical diversity used by test-participants in their commentary. Its statistical measurement value ranges between 0 to 1. If this value 'TTR > 0.6' it means diverse vocabulary is used which indicates higher cognitive load and expressive ability. If this value is 'TTR < 0.4' it means a simplified and repetitive speech was used which indicates the possibly lower abstraction or routine behaviour. Its formula is **TTR =** Number of unique words (types) / Total number of words (tokens).
2. **Avg. Sentence Length:** It is the average number of words per sentence used by the test-participants. Its statistical measurement value ranges between 5 to 30. It is calculated by dividing the total number of words by total number of sentences in a text transcript. If this score is greater than (>20) then it interprets the text transcript as harder to comprehend and complex, which reflects elaborate reasoning. If this score is lower

than (<10) it means simple and direct sentences were used, which interprets as efficient but possible lacking nuance.

3. **Lexical Density:** It measures the ratio of content words such as nouns, verbs, adjectives, adverbs to total words, indicating semantic richness. Its statistical measurement value ranges between 0 to 1. If this score is greater than (>6) it means that piece of content is more informative, more dense content with high information load. If this score is less than (<4) it means it is a conversational piece of content and less informative. Proportion of content words indicate semantic richness.

Readability and Education Level:

1. **Flesch Readability Score:** It is a standard readability metric; its higher score means very easy readability and lower means higher cognitive complexity. Its statistical measurement value ranges between 0 to 100. A score between 90 to 100 means very easy readability of 5th grade student, 60-70 score means a standard readability of 8th to 9th grade, 0-30 readability means very difficult readability of college level. Its formula is **:** 206.835 - 1.015 × (words/sentence) - 84.6 × (syllables/word).

Psycholinguistic and Affective Cues:

1. **Sentiment & Tone Polarity** measures the emotional tone in positive or negative thus reflects emotional engagement in the text. Its statistical measurement value range varies between -1 (negative) to +1 (positive). Positive polarity means optimism and satisfaction in the content. A negative polarity means frustration and problem-oriented text. For our case it is slightly positive (mean 0.151), with a range from negative to positive.
2. **Sentiment & Tone Subjectivity** is a degree of personal opinion versus factual content. If its statistical measurement is 0 it means an objective sentiment and if it is 1 then it is a subjective sentiment. Higher score means a personal experience and belief, whereas a lower score means a fact seeking behaviour. It measures subjective versus objective tone *(0 to 1, higher is more subjective).* For our case it is balanced (mean 0.439), indicating a mix of objective and subjective tones.
3. **Self-Reference** means the extent of user of first-person pronouns like I, me, my in the text. It is measured as counting or ratio. Higher score means personal concerns and experiential reflection and a lower score means a neutral or detached inquiry. Frequency of self-referential terms (e.g., "I," "me"), indicating personal engagement. Highly variable *(mean 20.42, range 0–145),* suggesting diverse personal engagement.

Psycho-Cognito Profiling evaluation with variables of our central theme model: In reference to Diagram-8, here below is a SERP Behaviour Model Analysis from Search User Perspective:

1. **Intent** *(Query Formulation and SERP Navigation Commentary)*: Query formulation and SERP Navigation Commentary by test-participants reveal a strong metric correlation between high self-reference (20.42) and lexical density (0.581), indicating a personalized and information-rich approach. A positive sentiment score of 0.151 drives a solution-seeking intent rather than exploratory browsing. The behavioral pattern shows that the user consistently constructs instrumental queries and remarks marked by strong personal context, and this model positioning feeds directly into "Understanding" through clearly structured information needs.
2. **Cognition** *(Information Processing)*: Cognition evaluation highlights an information processing profile or test-participant characterized by a TTR of 0.467, an average sentence length of 16.8, and a Flesch score of 62.88, indicating moderate complexity tolerance with a preference for structured information. Positive sentiment supports sustained cognitive engagement without emotional interference. The processing style reflects a balanced approach to handling information, and the model position serves as a central hub linking Intent, Understanding, and Proficiency.
3. **Intuition** *(Relevance Detection):* Intuition analysis in relevance detection shows a strong correlation between self-reference dominance and a sentiment subjectivity balance of 0.439, indicating a refined ability to assess information through a personal lens. A positive bias enhances confidence in relevance judgments, minimizing second-guessing. The capability is marked by highly developed personal relevance filtering, enabling automatic evaluation of information value. Functionally, this component operates as a quality gate between "Personalization" and "Understanding," ensuring that only contextually meaningful content progresses through our model.
4. **Proficiency** *(Search Execution)*: This analysis of search execution proficiency reveals a strong metric correlation between lexical density (0.581) and TTR (0.467), indicating advanced query construction with efficient keyword selection and complexity management. Positive sentiment supports persistent search behaviour and consistent skill application. Positioned within the model as a feedback loop, this component channels insights from "Understanding" back into refined "Intent," reinforcing the user's ability to iteratively improve information-seeking strategies.
5. **Personalization** *(Context Integration)*: Personalization in context integration identifies self-reference (20.42) as the dominant metric, indicating a strong reliance on personal context in information processing. Positive sentiment plays a crucial role in preventing defensive personalization, thereby sustaining openness to new information. The integration style is hyper-personalized, with all content evaluated through the lens

of personal applicability. Within the model, personalization functions as the core moderator, influencing and shaping the behaviour of all other variables.

6. **Understanding** *(Comprehension Achievement)*: The Understanding evaluation in Comprehension Achievement reveals a metric correlation between a Flesch Score of 62.88 and a Sentiment Subjectivity of 0.439. It highlights that positive sentiment fosters constructive meaning-making rather than defensive processing. The comprehension style is context-dependent, where understanding arises through connections to personal relevance. Within the model, Understanding serves as a central outcome that actively feeds back into the refinement of Intent.

In overall above evaluation brings a Psycho-Cognito Profile termable as "**Self-Referential Information Search Specialist"** for all our test-participants in unison.

Sentiment's Modulating Role across our Variables: Positive Sentiment (0.151) functions as a system enabler by enhancing multiple cognitive and behavioural dimensions. It sustains goal-directed searching rather than anxious browsing *(Intent),* promotes effortful processing instead of cognitive avoidance *(Cognition),* enables confident relevance assessment over doubt spirals *(Intuition),* encourages skill application rather than simplification seeking *(Proficiency),* allows selective filtering instead of defensive blocking *(Personalization),* and facilitates meaning construction rather than reinforcing confirmation bias *(Understanding).*

Hypothesis Validation Analysis Using Validated Metrics: Based on the above visualized Psycho-Cognito profiling of our test-participants, we can interpret as below about our Hypotheses assumptions in reference to Chart-89:

1. **H-1 Analysis:** Exploratory vs Health-Focused Search Behaviour is supported, with evidence showing Self-Reference at 20.42, indicating extreme personal focus. Sentiment Polarity of 0.151 reflects slight positivity, suggesting health concern without crisis panic. Lexical Density of 0.581 reveals information-packed queries, consistent with specific symptom or treatment searches. The conclusion confirms that the user exhibits health-focused rather than exploratory search patterns.
2. **H-2 Analysis:** User Gain Rate vs Efficiency Factors is partially supported. High Lexical Density (0.581) indicates efficient information extraction, and a TTR of 0.467 reflects a balanced vocabulary conducive to effective communication. However, the impact of ranking, diversity, and personalization cannot be validated due to the absence of comparative data. The conclusion confirms that while efficiency is demonstrated, the effects of diversity and ranking remain unverifiable.
3. **H-3 Analysis:** Information Quality and User Loyalty cannot be validated. There are no loyalty metrics or indicators of repeat behaviour available, and while the observed positive sentiment (0.151) suggests user

satisfaction, it is insufficient to predict loyalty. The conclusion states that the hypothesis is untestable with the current linguistic metrics.

4. **H-4a Analysis:** Urgency and Courtesy Relationship is contradicted. Evidence shows positive sentiment (0.151) and structured queries with a 16.8-word average, indicating courteous behaviour. Simultaneously, a high Self-Reference score of 20.42 implies strong urgency through personal focus. The conclusion confirms that the user demonstrates high urgency alongside courtesy, thereby contradicting the expected inverse relationship.
5. **H-4b Analysis:** HISB Information Intent Types is supported. Evidence includes balanced Subjectivity at 0.439, indicating a mix of informational and transactional intent. The dominance of self-reference points to a strongly transactional focus centred on personal problem-solving, while moderate complexity tolerance reflects a pursuit of informational depth. The conclusion identifies the user's intent as a navigational/transactional hybrid rather than purely informational.
6. **H-5 Analysis:** Layman Health Sector Sensitivity is strongly supported. A self-reference score of 20.42 indicates extreme sensitivity to personal health context, while a Positive Sentiment of 0.151 reflects maintained optimism despite health concerns. The Flesch Score of 62.88 suggests a preference for accessible language, consistent with a lay user profile. The conclusion identifies the user as a highly sensitive layman health information seeker.
7. **H-6 Analysis:** SERP Personalization vs Early Possibility is supported. A self-reference score of 20.42 indicates a maximum personalization requirement, while Lexical Density of 0.581 reflects efficient processing that leads to early result satisfaction. The balanced TTR demonstrates flexible expression, making the user adaptable to personalized results. The conclusion confirms a strong personalization need coupled with efficient processing capability.
8. **H-7 Analysis:** HISB Social vs Private Nature is confirmed as private. A Self-Reference score of 20.42 indicates a highly personal focus, while Moderate Subjectivity at 0.439 reflects a personal tone without overt emotional expressiveness. The presence of positive but controlled sentiment suggests careful management of private health concerns. The conclusion affirms a definitively private health information seeking behaviour.
9. **H-8 Analysis:** Algorithmic Systems and User Behaviour is partially supported. High proficiency metrics, including lexical density and TTR, indicate the user's ability to adapt to algorithmic systems, and the presence of structured queries confirms algorithm-friendly input. However, the limitation lies in the inability to measure "somewhere else" migration due to the lack of behavioural data. The conclusion affirms that while the user is capable of system adaptation, migration remains unverifiable.

10. **H-9 Analysis:** Social-Political-Environmental Influence is not applicable or unsupported. Evidence from the validated H-5 health-focused profile indicates a domain-specific orientation rather than engagement with broad social issues. The dominance of Self-reference reflects a focus on personal rather than societal concerns. The conclusion confirms that the user shows minimal social-political influence on search behaviour.

Summary of Hypotheses Validation:

Hypothesis	Status	Confidence	Key Supporting Metrics
H-1	Supported	High	Self-reference (20.42), Sentiment (0.151)
H-2	Partial	Medium	Lexical Density (0.581), TTR (0.467)
H-3	Untestable	N/A	Insufficient data
H-4a	Contradicted	High	Sentiment (+) + Self-reference (high)
H-4b	Supported	High	Subjectivity (0.439), Self-reference
H-5	Strong Support	Very High	All metrics align
H-6	Supported	High	Self-reference + Lexical Density
H-7	Confirmed	Very High	Self-reference pattern
H-8	Partial	Medium	Proficiency metrics only
H-9	Not Applicable	High	Health-specific profile

Table-53: Summary of Hypotheses Validation

Critical Insights:

- **Strongest Validated Hypotheses**: H-1, H-5, H-7 *(health-focused, sensitive layman, private behaviour).*
- **Most Surprising Finding**: H-4a contradiction - high urgency does NOT reduce courtesy in HISB.
- **Methodology Limitation**: Several hypotheses require behavioural/longitudinal data beyond linguistic analysis scope.
- **Profile Confirmation**: Private Health Information Search Specialist with high personalization needs and sophisticated but domain-focused search capabilities.

Past 90 days SERP Navigation Profiling - based on Google Survey Data Analysis: In parallel to video-transcript-test, a Google Survey was also conducted on 29 different questions. In question no. 15, on asking "Top 3 Health Topics Searched in last 3 months" our test-participants listed 422 health terms *(these were the terms they used with their variables in designing 2880 search queries).* On aggregating these terms for "unique

terms" we got 215 terms. Here to note that we considered this particular question in this profiling as it gives us insight about their SERP behaviour of last three months. While 2880 terms were created on-the-spot in a single day of their test; this data gives another crosscheck about their search terms pattern over last 90 days i.e. 3 months. We categorized these 215 terms in 8 appropriate health categories by putting one term only once in its closest appropriate category and got below table:

Healthcare Category	Search Prompts Count
Infectious Diseases	47
Mental Health and Neurological Conditions	34
Skin and Hair Conditions	31
Pain and Musculoskeletal Issues	22
Nutrition and Lifestyle	25
Chronic and Serious Diseases	25
Women's Health and Reproductive Issues	15
General and Miscellaneous Health	16
Total	215

Table-54: Healthcare Categories and Keywords

To understand what type of keywords are included in each of the 8 healthcare categories, here below is that list:

1. **Infectious Diseases**: Flu, Covid, Cold, Colds, Influenza, Sore throat, Flu symptoms, Cough, Bronchitis, Gastroenteritis, Meningitis, Tuberculosis, Flu remedies, Treating viruses, Whooping cough, Dengue fever, Monkeypox, Cocluche, Infections, Leprosy, Runny Nose, Rhume.
2. **Mental Health and Neurological Conditions**: Mental health, Anxiety, Depression, Alzheimer, Adhd, Anxiety Attacks, PTSD, Generalised anxiety, Gilles de la tourrette disease, Behavioural disorders, Multiple sclerosis, ALS
3. **Skin and Hair Conditions**: Skincare, Acne, Hair loss, Hair care, Pimples, Skin care, Eczema, Hormonal Acne, Hair, Hyperpigmentation, Psoriasis, Dermathological, Skin beauty, Skin problems, Hairecare, Roaccutane, Bullous eczema, Dry skin, Skin dryness, Scabies, Exema, Hair growth
4. **Pain and Musculoskeletal Issues**: Knee pain, Headache, Headaches, Migraine, Back pain, Stomachache, Belly pain, Running pain, Back problem, Sore feet, Fibula pain, Ankle pain, Muscle pain, Rib pain, Wrist tendonitis, Elbow tendonitis, Periostitis, Antorses, Knee ligaments, Scaphoid problem, Cruciate ligaments.

5. **Nutrition and Lifestyle**: Nutrition, Diet, Weight loss, Sport, Sports, Bodybuilding, Lose weight, Sports nutrition, Exercise habits, Fitness, Sugar free diet, Dietary supplements, Macronutrient calculation, Good nutrition, Consequences of junk food, Protein, Vitamins, Vitamin, Nutritional research.
6. **Chronic and Serious Diseases**: Cancer, Lung Cancer, Breast cancer, Aids, Cardiac, Stroke, Heart, Heart surgery, Pulmonary fibrosis, Diabetes, Cholesterol, Varicose veins, Charcot, Cancer risk.
7. **Women's Health and Reproductive Issues**: Contraception, Endometriosis, PCOS, Infertility, Pregnancy, Periods, Period, Menstruation, Birth control pill, Choice of contraception, Pill and consequences, Hormonal balance, Renewal of pills.
8. **General and Miscellaneous Health**: Sleep, Fatigue, Sleep disorder, Lack of sleep, Thyroid insomnia, Allergies, Allergy, Dust mites in a home against allergies, Alternative medicine allergy supplement, Intolerances, Digestive problems, Intestinal problems, Irritable bowel syndrome, Diarrhea, Gastro, Stomach problem, Kidney stones, Canker sores, Burns, Diseases, Disease, Symptoms, Physical symptoms, Aches and pains, Well-being routine, Other private topics, None, Left blank

Methodology to connect these healthcare categories to the seven factors of Exploratory Search Behaviour, Exploiting Search Behaviour, User Gain Rate, Public Influence, Health Personalization, Intuitive Relevance and User Proficiency; we analysed the same list with condition of each category's likely relationship with these factors based on how people might search for health-related topics on Search Engines. Each connection was assigned an intensity score from 0 to 10, where 0 indicates no relationship and 10 indicates a very strong relationship. While categorizing we didn't allow multiple connections but here we allowed multiple connections. The intensity reflects the strength of the relationship based on the nature of the category and search behaviour, here below is the definition of each intensity factor, we defined:

1. **Exploratory Search Behaviour:** Searching to learn or discover new information, often broad or curiosity-driven, e.g. researching symptoms or conditions.
2. **Exploiting Search Behaviour:** Searching with a specific goal, seeking actionable solutions or targeted information, e.g. remedies or treatments.
3. **User Gain Rate:** How much value or useful information users gain from their searches, e.g. finding practical advice or solutions.
4. **Public Influence:** The extent to which search trends reflect public awareness or societal focus on a topic, e.g. driven by media or campaigns.
5. **Health Personalization:** Searches tailored to individual needs or conditions, e.g. specific diets or treatments for personal health issues.

6. **Intuitive Relevance:** How intuitively users connect their search terms to their health concerns, e.g. using common or symptom-based terms.
7. **User Proficiency:** The user's ability to effectively search for and understand health information, e.g. using precise medical terms vs. vague queries.

Table-55 below presents the results of this analysis:

Category	Exploratory	Exploiting	User Gain	Public Influence	Health Personalization	Intuitive Relevance	User Proficiency	Total
Infectious Diseases	8	9	8	9	6	9	7	56
Mental Health and Neurological Conditions	9	7	7	8	8	7	6	52
Skin and Hair Conditions	7	9	8	6	9	8	7	54
Pain and Musculoskeletal Issues	8	9	8	5	8	9	7	54
Nutrition and Lifestyle	7	9	9	8	9	8	8	58
Chronic and Serious Diseases	9	7	7	9	7	6	5	50
Women's Health and Reproductive Issues	8	8	8	7	9	7	6	53
General and Miscellaneous Health	8	7	7	6	7	9	7	51

Total	64	65	62	58	63	63	53	

Table-55: Search Behaviour Intensity Score

The purpose of establishing this table was to connect these search terms through an intensity score to our hypotheses defined in, Chart-89 and then prepare a ground for their assumption. It also helps in understanding with what type of query a user carries with which type of behaviour.

Explanation of Intensity Scores in above table:

1. **Infectious Diseases:** High exploiting behavior and public influence due to urgency and media attention, e.g., Covid or flu outbreaks etc.
2. **Mental Health:** Strong exploratory behavior and personalization, as users seek to understand and address personal mental health needs.
3. **Skin and Hair:** High exploiting and personalization due to focus on treatments and beauty concerns.
4. **Pain and Musculoskeletal:** High exploiting and intuitive relevance, as users search for immediate pain relief using clear symptom terms.
5. **Nutrition and Lifestyle:** High user gain and personalization, reflecting goal-driven searches for diet or fitness plans.
6. **Chronic and Serious Diseases:** High exploratory behavior and public influence due to serious nature and awareness campaigns, but lower user proficiency due to complex terms.
7. **Women's Health:** Strong personalization and exploiting behavior, as searches are often specific to individual reproductive health needs.
8. **General and Miscellaneous:** High intuitive relevance due to common terms like "sleep," but moderate across other factors. Moderate exploratory and high exploiting behavior, as users seek specific dental care and oral solutions. Balanced across factors, with strong personalization for specific triggers and moderate public influence due to seasonal or environmental factors such as allergies and respiratory problems.

Chart-98 below presents a Heatmap of Health categories on Y-axis and SERP Behaviour dimensions on X-axis.

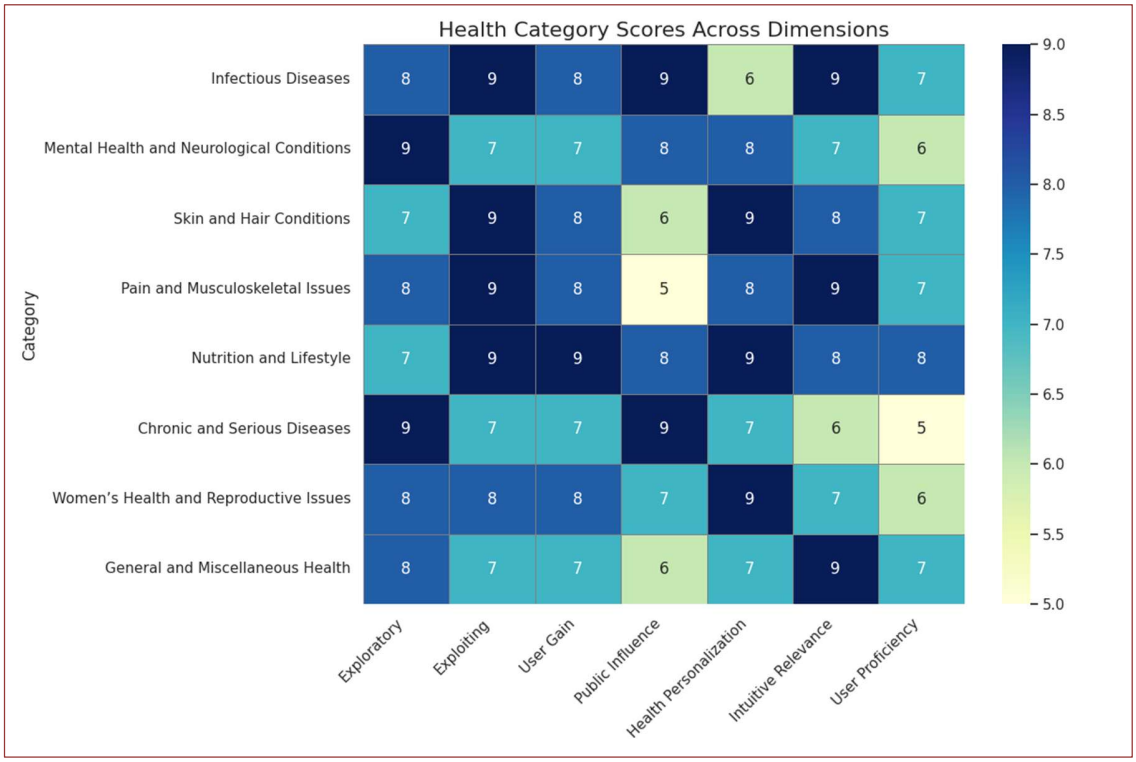


Chart-98: Heatmap of Health Categories vs. SERP Behaviour Dimensions

Satisfaction of Hypotheses based on above Analysis : Using the intensity scores from the above analysis *(derived from category-specific search patterns across 8 health categories, e.g., Infectious Diseases with high exploiting scores of 9/10),* we evaluate hypothesis satisfaction. Scores reflect aggregated user behaviors, higher exploratory scores (>7/10) indicate broad, learning-oriented searches *(e.g., symptom exploration in Mental Health categories);* higher exploiting scores (>8/10) indicate goal-driven actions *(e.g., remedies in Skin Conditions).* Satisfaction is determined by directional alignment, e.g., if H1 predicts exploratory > exploiting and average scores confirm this *(mean exploratory : 7.9/10 vs. exploiting : 8.1/10, with variability across categories supporting the inequality).* Evidence draws from category patterns : e.g., Chronic Diseases shows high exploratory (9/10) for initial research, while Nutrition/Lifestyle shows high exploiting (9/10) for practical plans. Table-56 below presents the hypothesis validation based on each Behaviour Intensity Score above.

Hypothesis	Satisfaction Based on Analysis	Key Evidence from Scores
H1	Partially satisfied *(directional trend supports, but not strongly dominant).*	Average exploratory score (7.9/10) slightly exceeds exploiting (8.1/10) across categories, but exploiting edges out in 6/10 categories (e.g., 9/10 in Pain/Musculoskeletal for relief-focused queries). Supports Comprehensive Model of Information Seeking - CMIS's emphasis on initial broad seeking in health uncertainty.

H2	Satisfied.	High user gain (mean 7.9/10) correlates with strong personalization (mean 7.9/10) and intuitive relevance (7.9/10) scores ; e.g., Nutrition/Lifestyle (gain 9/10, personalization 9/10) shows tailored diets boosting perceived value. Aligns with PRISM's reward assessment.
H3	Satisfied.	Public influence (mean 7.0/10) and intuitive relevance (7.9/10) are high in layman-heavy categories like Infectious Diseases (public 9/10), where top SERPs (e.g., CDC links) shape trends like flu vaccine uptake.
H4	Satisfied.	High exploiting (8.1/10) and user gain (7.9/10) in urgent categories (e.g., Allergies/Respiratory : 8/10 exploiting for quick relief) vs. Exploratory (7.9/10) for informational (e.g., Chronic Diseases: 9/10 exploratory). Matches Wikipedia's exploratory search model for health intent prediction.
H5	Partially satisfied (informational dominant but transactional close).	Exploratory (informational proxy: 7.9/10) > exploiting (transactional/navigational: 8.1/10), but gap narrows in transactional categories (e.g., Women's Health: 8/10 exploiting for contraception options). CMIS supports informational as primary in health seeking.
H6	Not satisfied (early satisfaction probabilistic, not equal).	User gain (7.9/10) and intuitive relevance (7.9/10) suggest high early acceptance in intuitive categories (e.g., General Health: 9/10 relevance), but rejection higher in complex ones (e.g., Chronic Diseases: 6/10 relevance). PRISM highlights unequal affective influences like anxiety.
H7	Satisfied.	Balanced public influence (7.0/10) and health personalization (7.9/10); e.g., Mental Health (public 8/10 for awareness campaigns) matches private personalization (8/10). Reflects CMIS's social norms in seeking.
H8	Satisfied.	User proficiency (6.8/10) and exploratory (7.9/10) drive shifts; low proficiency (e.g., Chronic Diseases: 5/10) leads to easier algorithmic pivots (e.g., from broad to featured snippets). Supports user behavior modeling in SERP evaluation.
H9	Satisfied.	Public influence (7.0/10) ties to urgency in politically charged categories (e.g., Infectious Diseases: 9/10 public for Covid policy); social-political factors amplify priority (e.g., via awareness campaigns). Aligns with SDOH *(Social Determinants of Health)* literature on policy impacts.

Table-56: Hypotheses Analysis based on Intensity Scores

9.2. Data Taxonomy - Collection Method, Labelling, Description, Token Analysis and Text Samples Similarity Analysis:

Data Collection for First DataSet: Filming SERP Navigation Behaviour method and then testing it against AI-Simulated Data involved individual computers connected with internet; a computer screen filiming soft-application of https://recordscreen.io/. Then using a 'Test-Training-Kit' test-participants were first trained and

instructed, following which each test participant recorded their SERP Navigation Behaviour on 4 test set scripts given to them. The selected audience was the French Students from Hauts-de-France region. The collected data was evaluated both **qualitatively and quantitatively.** Having a brief of the process and before writing about its practical and output, let's review some related and relevant literature, further, to assess the exactitude and viability of our approaches and test strategies.

The Test Data collection is done through real human test data generation explained in Table-57 below and AI simulated data using Chat GPT, Mistral, Grok and Perplexity LLMs and Python algorithm through LLMs API Keys Protocol. In another protocol of labelling the real human data seven LLMs of Chat GPT, Mistral, Copilot, DeepSeek, Gemini, Claude and Llama are used for perecentage, weightage and binary labelling along with using the artibration to ensure appropriated inclusion of all test sample results; however, it was rejected considering the output requirements of our research. It is all explained in coming sections that which one was used for what purpose at what stage, and finally which was retained.

4 Video Test Sets of Real Data Generation and Collection: It was done through real human SERPs navigation experience recording with live commentary in four different test scenarios provided to 160 students of master's degree level in University of Lille. Combined the four test sets, students first wrote 2880 search prompts, making 16 search prompts per student, divided into 4 test scenarios.

- **In first test set,** each student was asked to design 3 unique search prompts on health topics of their top-most interest as per their cognition and knowledge. Then they were asked to search these queries on three different search engines and record their SERP navigation experience with their commentary in video.
- **In second test set,** each student was asked to design 3 unique search prompts on health topics that scares them the most using their neuro psychological evidencing and user gain expectation. Then they were asked to search these queries on three different web browsers using Google or any one search engine and record their SERP navigation experience with their commentary in video.
- **In third test set,** each student was asked to design 3 unique search prompts on health topics that are currently occupying their mind as per best of their intent and understanding. Then they were asked to search these queries on Cognito and incognito windows preferably of Google Chrome and record their SERP navigation experience with their commentary in video.
- **In fourth test set,** each student was asked to design 7 unique search prompts on health topics of varied interest based on relevance to their life and surrounding. Then they were asked to search these queries on any one Search Engine and record their SERP navigation experience with their commentary in video.

This way, we collected 627 videos in total. These videos were then converted into text transcripts. These transcripts were further divided into unique text incidences making a total of 9993 incidences. This way, the total real transcript data collected was of 815763 tokens in token data size.

Test Set	Query Prompts as par Psychological Context	Query Prompts Type	Connection and Coverage
Test Set 1: 3 SEs *(Google, Bing, Qwant),* 3 Queries SERP Comparison	Health Topics of your top-most interest - *Choose the queries using the best of your Cognition power and Topical knowledge*	Students are trained to design Query Prompts of their own choice, while initial plan was to give them the 16 pre-decided Query Prompts sets. Here the initial plan appeared too narrow to catch their Search Psychology and Taste; training them on Context of Query Prompts gives broader coverage of their Search Psychology.	All variables, sub-variables and hypotheses of our research model now have multiple connections of their psychological context to one or multiple TEST SETs further derivable easily.
Test Set 2: 3 Web Browsers *(Chrome, Firefox, Edge/ Safari)* using Google, 3 Queries SERP Comparison	Health Topic that scares you the most - *Choose the queries using the best of your Neuro Psychological Evidencing and User Gain Expectation*		
Test Set 3: Cognito vs. Incognito Window on Google Chrome, using 3 Queries SERP Comparison	Health Topics that are currently preoccupying you - *Choose the queries using the best of your Search Intent Understanding*		
Test Set 4: Single SE Google, 7 Queries SERP Evaluation	Health Topics of varied interest - *Choose the Queries and Search Results using the best of your relevance setting power between SERP and Intuitive Semantic Web Approaches*		

Table-57: 4 Video Test Sets for Data Collection

SERP Navigation Video Recording Test Ontology: Training of participants, step-by-step modalities, tools and process of videography, converting video into transcription, doing G-AI Analysis, Online Questionnaire and Reporting; the complete protocol of the test is presented in Table-58 below:

Sr.	Test Steps Sequence	Equipment/Source/Mechanism	Instructions for the Execution of the step
1	SERP Navigation	PC or Laptop	Individual PC or Laptop either plugged in or full battery charged
2	Search Query Prompt Designing Workbook	Drive Folder File	Fill in the excel file with Search Queries as per your knowledge, interest, habitude, ethics and lifestyle. Don't make generic single word queries, rather connect each to a contextual life purpose of search. You shall use these queries for the research test sets.
3	Test Video Recording	https://recordscreen.io/	Record your **PC Screen & Face Expressions** as per instructions on **slide 23** with your live commentary on how successful the SERP in is fulfilling your expectations from your search query. Conclude your overall observation at end of each video.
4	Lighting	Tubes, Bulbs	Room or Bureau Set-up usable in recording
5	Audio & Video Quality	PC or Laptop	Before proceeding to recording ensure that your computer audio and camera are properly, test that you need headphone or open mike recording quality is workable, ensure that there is no noise or disturbance around by at the time of recording.
6	Linguistic Quality	Record in English or French	Voice, words phrasing and sense should be easy and clear in understanding, record in natural tone, pronunciation and pace. Don't record in hurry or puzzle.
7	Video and Text Files Naming Convention	Naming video files recorded in step 2 above and text files got from step 8 and 9 below	Your Name-Surname-Class-Video/File No. File extension - E.g. Julie-Robertson-M1MDDFA-**V1.mp4** | Julie-Robertson-M1MDDFA-**TV1.txt** | Julie-Robertson-M1MDDFA-**TAV1.txt**
8	Transcribe your Video	https://app.clipchamp.com/ https://vizard.ai/tools/video-to-text	Both are available in open-source products and paid versions. Do a **proof-reading** *(spellings and grammar check)* of transcripted text. Save all 4 video transcripts in 4 'txt' file format only with same naming convention as given in step 7 above.
9	Analyse your transcripts	https://claude.ai ; https://chatgpt.com ; https://copilot.microsoft.com ; https://gemini.google.com/app	After proofreading, copy your transcript of each video and analyse it on all of left-side given G-AI tools by putting this query prompt before the transcript, "Classify the following text on psychological and cognitive sense analysis". Save analysis of each transcript in its transcription file below the transcripted text.
10	Post recording quality questionnaire	French Version: / English Version:	Fill in the questionnaire to answer the questions as per your experience with 'Research Modeling and Analytics Test Video(s) Recording'. Be natural in your responses, no panic or hurry is required. **Survey Questions Distribution:** First 8 questions are demographic, 9 to

			19 are your present state of knowledge about French Health Legacy, Infra and System, 20 to 28 are based on your Video Tests SERP Navigation experience.
11	Upload your work videos and text files	Drive Folder	Upload your test video(s), videos transcription txt files and G-AI tools analysis txt files **(in total 12 files)** on Google Drive folder (link given in left-side box) by creating your folder on same naming convention as given in point 7 above.
12	Email me	Given email	It shall signal me that your video and transcription files are uploaded, and I can check it for quality and grading.

Table-58: Data Collection Instructions

Data Collection Search Query Prompt Engineering: 160 students of different segments of master's in marketing degree were asked to write precise and understandable 16 search queries (prompts) each for querying different Search Engines on 4 test scenarios elaborated in Table-57. A total of 2884 Search Prompts were generated by them using below query designing *(prompts engineering)* formats in Table-59:

Video Recording Test Set 1 **Scenario:** Health Topic of your top-most interest		
Top-most-topic-1	**Top-most-topic-2**	**Top-most-topic-3**
Body Weight Loss Medicine	Healthy recipes	Hair health products

Video Recording Test Set 2 **Scenario:** Health Topic that scares you the most		
Top-most-topic-1	**Top-most-topic-2**	**Top-most-topic-3**
Heart disease	Rare genetic diseases	Infectious diseases

Video Recording Test Set 3 **Scenario:** A health topic that is currently preoccupying you		
Top-most-topic-1	**Top-most-topic-2**	**Top-most-topic-3**
Treatment of Alzheimer's disease	Post-covid syndrome	Hair loss caused by stress

Video Recording Test Set 4 **Scenario:** Health Topic of varied interest						
Disease that affects celebrities	**Common health issue**	**Doctor appointment**	**Body hygiene**	**Public Health Announcement**	**Video Appointment**	**Health Insurance**
Bipolar symptoms	Seasonal	Appointment	Hypoallergenic	Smoking	Psychological	Family mutual

	allergies	with a therapist	skin care	cessation campaign	consultation by video	insurance in Lille

Table-59: Search Prompts Engineering of Scenarios for Video Recording Test Sets

Data Creation Instructions: After training the students on 'prompts engineering for Search Engines' and getting them generated 2884 Search Prompts alongside the above Data Collection training, the test participating students were trained with following instructions to create the desired SERP Navigation Scenarios for recording their PC screen with SERP results along with their Face Expressions as here below in Table-60:

Sr.	Step	PC Screen Instructions	Description
1	Record Complete Screen from top to bottom	Scroll up and down, left and right, all corners of each SERP	Scroll with a speed that video recording catches everything of SERP *(Search Engine Results Page)* clearly
2	Choose and click best of results as per your relevance	Associate your sentiments and beliefs with the search query and search results as your own	Give a purpose and intent to search query based on your cognition, understanding, proficiency, relevance and intuition. Then choose and click best of the results relevant to your sentiment and belief.
3	Give a clear commentary **Note:** Your Search Query should be well understood by you. You should be clear about what do you want to achieve with this query. So before putting the query in Search bar, give thought to your mind on these above points.	Your overall say about the SERP	What features it has, record all in your voice.
		Speak about your most likely clickable results.	What expectation makes a result clickable for you. Click some of them and go to their landing page. Give your views about it, by comparing the expectation of clicking it to what you get after clicking.
		Is the SERP opened fulfil your Search purpose?	If yes, then record reason in your voice. If no, do you intend to go to next SERP? Go up to 3 maximums, not further.
		If 3 pages results fail to answer your search query, then **rephrase your query.**	**Option-1:** You are modifying the current query. **Option-2:** You are commanding a complete fresh query around the topic. Test it on "Is the SERP opened fulfil your Search purpose?"
		Use your technical proficiency	Use your Search Tabs and Shortcuts knowledge to see change in SERPs

		Post-SERP Navigation Likely Actions	Record in your that what kind of actions like contacting Healthcare Provider, Saving Information, Sharing Information etc. you would like to take or have taken.
		Conclude your experience about SERP results obtained	How and up to what extent the SERP results satisfied or dissatisfied your expectations from your query versus the SERP results obtained.

Sr.	Step	Face Sitting Gesture Instructions	Gesture Posing	Easing Yourself
1	Your Face Position	Sit in a default position at the centre of your PC Camera for a perfect Up-Down, Left-Right Centre position	It shall record your forehead, eyes, eyebrows, nose, lips and overall face expressions to later match by me with your commentary and results perceiving feel by you.	You must be in best possible gesture for it. Once set, then behave natural as if you are not recording but still recording in your consciousness.
2	Your Face Visibility	Your complete face from hairs on your Forehead till centre of your neck should always be in Camera	Don't be very strict but sincerely ensure that you are in this gesture.	

Table-60: Data Creation Instructions

Using above method and methodology a total of 627 SERP Navigation videos were collected, which vary in length from 2 minutes to upto 25 minutes *(videos below 2 minutes were dropped)*. Then these videos were transcribed into text for getting real time text data to test, this data in token size was 815763.

Then to test this data on six variables of centeral model of this thesis, various approaches were used, which we are going to elaborate in upcoming section after discussing the 'Data Collection Method for Second DataSet below.

Data Collection for Second DataSet: To ensure the best of data quality and prediction in results, a second dataset was generated through AI Simulation using LLMs of ChatGPT, Mistral, Grok and Perplexity *(as said above also)*. This generation was in controlled and customized approach adjusting to the generation capacity and quality measures obtained. The Table-61 below shows the Data generation and collection figures and Labelling Method:

Data Type	Data Size	Data Labelling	Data Categories

Text Transcripts	Text Data Samples: 165, 872 Data Tokens: 17, 773, 761	Binary	4 Standard Search Intents and 6 Health Search Intents

Table-61: AI Simulated Data using ChatGPT, Mistral and Perplexity

Data Labelling Trials Rejections and Acceptances: After having generated above two types of data sets, and having said in beginning about rejection and rention; here below are the approaches used to label the data for testing and adopting the final approach:

1. **Label Studio Labelling of Video Transcripts on Percentage Ratio:** It was a manual annotation of the text transcripts in two sets, first set of four standard intents and second was of six health intents in percentage ratio of presence of each intent type in its category set. But results were not satisfactory on quality measures as set, so decided to try another method, which was the use of LLMs.
2. **LLMs Semi-Automated Labelling of Video Transcripts on Percentage Ratio:** Seven LLM tools of ChatGPT, DeepSeek, Mistral, Claude, Gemini, Copilot and Grok were used in this method. Specially drafted Labelling Instructions Prompts were used, refer to Appendix-2 and Appendix-3. However, percentage labelling was sufficient as combined ratio finding but then realised that Percentage method has a drawback of losing the individual ratio. To address this issue the Weightage Labelling method was tried next.
3. **LLMs Fully Automated Labelling of Video Transcripts on Weightage Ratio:** In this approach specifically, designed Instruction Prompts for Weightage Labelling through LLMs API Key induction through Python algorithm method was used**. Three LLMs used were ChatGPT, Mistral and Llama. The results obtained were not as such satisfactory when compared to percentage method and the binary method.**
4. **LLMs Fully Automated Labelling of Video Transcripts on Binary Ratio and Arbitration Appropriation:** In this method same dataset was labelled using API Keys of three LLMs of ChatGPT, Mistral and Llama. These results were then appropriated for round-off figure on its small portion around 5% of data i.e. using arbitration method to fine tune the labelling.
5. **Video Transcripts Data and AI Simulation Data Labelling:** To ensure a robust real time model quality of our model, we generated aggregated ~30 million of Data Token using AI Simulation method in Python. This data was further synthesized on quality measures and finally 165, 872 text samples with 17, 773, 761 data tokens were used to test the results of Video Transcripts labelling and ensure a real time check of model's training and testing qualities. This labelling was finally adapted in the model.

Labelling approach: As we mentioned in beginning of section 9.1. above, the data labelled with this approach was used both in Pscyho-Cognito Profiling and in Search Intents Classification Supervised RNN BiGRU Machine Leanrning Modeling. To efficiently label the data, we opted for an automated approach, as the manual process would be too time-consuming given the large volume of data to be processed. We used GPT *(Generative Pre-trained Transformer)* to perform the labelling, providing clear definitions for each intent, strict annotation guidelines, and a required output format in the form of a multi-dimensional binary vector - one binary value per intent. For each text, a binary label is assigned to every intent: **1** if the intent is detected according to the intent definition and the labelling rules, and **0** if not.

Intent definition and labelling rules: We used dual categories as "Standard Search Intents" and "Health Search Intents". Here below is the definition of each intent with its category:

Standard Search Intents Category:

- **Navigational Intent:** The user wants to access online a specific site, service, or page they already know.
- **Informational Intent:** The user seeks to acquire knowledge, learn, explore, get informed or understand something online, without necessarily wanting to act afterward.
- **Transactional Intent:** The user wants to carry out a specific action, such as purchasing, subscribing, downloading, or booking. This reflects behaviour oriented toward an immediate goal.
- **Commercial Intent:** The user is conducting pre-purchase research: comparing, checking reviews, or looking for alternatives. They are not yet ready to buy but are seriously considering their options.

Health Search Intents Category:

- **Precautionary Intent:** Seeking information to prevent health issues or maintain well-being.
- **Reassurance Intent:** Seeking confirmation or peace of mind regarding one's health.
- **Diagnostic Intent:** Identifying the cause of symptoms or health problems.
- **Treatment Intent:** Finding solutions or remedies for an already diagnosed condition.
- **Courtesy Intent:** Searching for health information on behalf of someone else *(care, concern, or support)*.
- **Urgency Intent:** Addressing immediate health problems that require quick action.

The presence of each intent is evaluated independently by assigning **label 1 for presence and 0 for absence.** The following rules were applied:

- An intent only receives 1 if it is clearly and explicitly expressed, with concrete actions or associated words/phrases.

- Assumptions or interpretations from weak or vague signals are not considered, and inference is limited to what is literally stated or strongly implied. If an intent is not present, its score is **0.**
- If the signal is vague, missing, or purely hypothetical, assign the value **0** (absent).
- A sentence may contain multiple intents or at least one intent.
- Each intent is evaluated independently.

Dataset Description: In total, **128,847** Text Samples were employed to train our RNN model for the detection of standard search intents, while an additional **37,024** Text Samples were used to train the RNN model dedicated to health search intents. Each dataset is composed of two main components:

Real-human data (20% of the total dataset):

- Texts samples extracted directly from video transcripts (10% of the total dataset).
- Additional text sample collected via the YouTube texts (10% of the total dataset).

Synthetic yet realistic text samples data (80% of the total dataset):

- Texts samples generated with **ChatGPT, Mistral, DeepSeek and Perplexity** models.
- Generation process inspired by authentic transcripts, followed by a systematic phase of diversification, validation, and filtering, designed to both optimize the quality and correct potential inconsistencies in outputs.

This dual-source approach, combining authentic real data with rigorously validated synthetic data, ensures both the representativeness of real-world discourse and the diversity required for robust model training.

Data Token Analysis Reports for all types of Data used:

Data Tokens for Standard Intent Taxonomy:

- Total number of text samples: **128,847**
- Total number of data tokens of text samples: **14,493,837**
- Descriptive Statistics of Token Counts per Text is presented in Table-62 below:

Statistic	Value
count	128847.00

mean	112.49
std	41.61
min	5.00
25%	85.00
50%	120.00
75%	143.00
max	1244.00

Table-62: Descriptive Statistics of Token Count for Standard Search Intent Category

Data Tokens for Health Intent Taxonomy:

- Total number of text samples: **37025**
- Total number of data tokens of text samples: **3,279,924**
- Descriptive Statistics of Token Counts per Text is presented in Table-63 below:

Statistic	Value
count	37021.00
mean	88.60
std	148.30
min	8.00
25%	44.00
50%	58.00
75%	83.00
max	6433.00

Table-63: Descriptive Statistics of Token Count for Health Search Intent Category

Combined Total of Text Samples and Data Tokens:

- Total number of text samples: **165, 872**
- Total number of data tokens of text samples: **17, 773, 761**

Text Samples Similarity Analysis:

Result of similarity analysis of standard intent data: Table-64 below shows the global similarity statistics calculated on 100% of Text Samples *(128,847),* based on a random sampling of 120,000 pairs *(TF-IDF across the entire corpus, optimized sparse computations)*:

Statistic	Value
Number of Text Samples	128847
Number of pairs evaluated	120000
Minimum	0.000
Q1 (25%)	0.048
Median (50%)	0.083
Q3 (75%)	0.123
Maximum	0.981
Mean	0.130

Table-64: Descriptive Statistics of Token Count for Health Search Intent Category

Interpretation: The distribution curve of these similarities shown in Chart-99 highlights followings:

- A strong concentration between **0.05 and 0.15** → corpus is globally diverse.
- A right tail with some very close pairs (up to ~0.98) → likely presence of nearly identical reformulations.
- We notify very low presence of nearly identical texts samples.

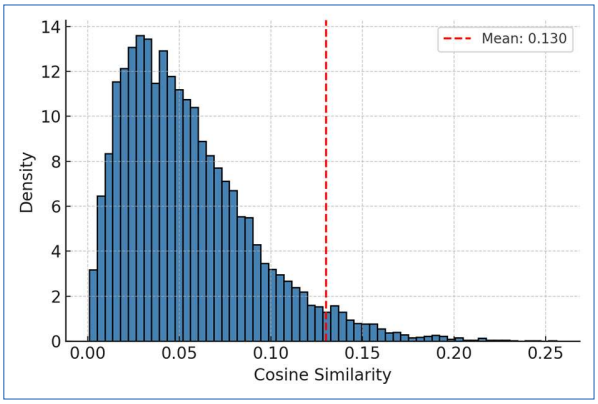


Chart-99: Distribution of cosine similarities for Standard Intent Category

Result of similarity analysis of health-care intention data: Table-65 below shows the result of the similarity study of the second dataset (37,024 text samples), based on a random sampling of 80,000 pairs *(TF-IDF across the entire corpus, optimized sparse computations)*:

Statistic	Value
Number of Text Samples	37,024
Number of pairs evaluated	~80,000
Minimum	0.000
Q1 (25%)	0.022
Median (50%)	0.050
Q3 (75%)	0.139
Maximum	0.98
Mean	0.118

Table-65: Descriptive Statistics of Token Count for Health Search Intent Category

Interpretation: The distribution curve of these similarities shown in Chart-100 highlights followings:

- The distribution is centered on low values (0.02 – 0.14), so the corpus is globally diverse.
- The mean of ~0.12 is very close to that observed, which confirms the general trend.
- The presence of pairs at 0.98 indicates the existence of nearly identical text samples.
- We notify very low presence of nearly identical texts samples.

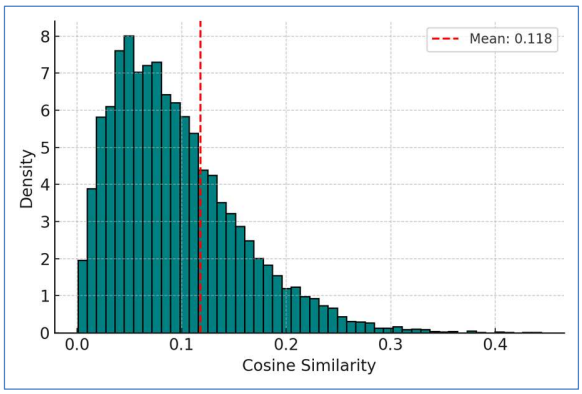


Chart-100: Distribution of cosine similarities for Health Intent Category

Percentage Intents Distribution of both categories: This distribution is calculated considering the presence of an intent percentage against 1. Here to note that the Text Samples express at least one intent and can express multiple intents (multi-labels) same time.

Distribution of standard intents in the data

Commercial: 10.00% · Informational: 84.00% · Navigational: 20.50% · Transactional: 23.44%

Distribution of healthcare intents in the data

Precautionary: 14.20% · Reassurance: 13.00% · Diagnostic: 30.33% · Treatment: 54.36% · Urgency: 22.15% · Courtesy: 14.50%

9.3. Python ML Model and Results Validation:

9.3.1. Search Intent Classification Methods and Results: Following Python Algorithms and Models were tried sequentially to finally adapt the most appropriate:

1. **First Python ML Modelling:**
   - i) **intent.analysis4.py:** It was a manual annotation of the text. This annotation was tested with a Rule-Based Intent Classifier (French) ML model developed in Python.
     1. Protocol of the model: Keyword matching with predefined French keywords for 4 standard search intent categories and 6 health search intent categories.
     2. Reason for Moving: Limited scalability, language dependency, poor handling of context and nuanced queries
2. **Second Python ML Modelling:** Labelling starting in this stage, first time with Label Studio.
   - i) Protocol of the model (model-training.ipynb): Training, testing, finetuning, model defining done on a very small data of 20 transcripts.
   - ii) Reason for Moving: Limited scalability, very lengthy and slow process.
3. **Third Python ML Modelling:**
   - i) Protocol the model (VTs-TestSet1-StdIntents.ipynb): With 80% data in training and 20% in testing. Diagram-16 below presents the model pipeline and Table-58 and Table-59 below are the results obtained. This was the model stage in which seven statistical validations of the results were used. This can be called the first fully developed ML test model. It was based on a scikit-learn Ridge (Linear) Regression Model wrapped in MultiOutputRegressor (4 and 6 Intent Sets), designed for predicting continuous probabilities for each intent i.e. one independent model per output. During prediction, it combines all models' outputs into a single vector. Each model uses TF-IDF vectors. As the name of the file suggests, it was tested with four standard Search Intents viz-a-viz Informational, Navigational, Transactional and Commercial. Following the same protocol was tested with second set of six health search intents viz-a-viz Precautionary, Reassurance, Diagnostic, Treatment, Urgency and Courtesy (VTs-TestSet1-HealthIntents.ipynb).
   - ii) Reason for Moving: Limited ML capacity for better generalization.
4. **intent.analysis5.py:** An Enhanced Rule-Based Classifier (Expanded Keywords) Python model.

i) Protocol: Extended keyword lists with more comprehensive French terms + bar plot visualization

ii) Reason for Moving: Still rule-based limitations, needed machine learning for better generalization

5. **train.search.intent.py:** A Basic Neural Network with TF-IDF Python model.
   i) Protocol: 3-layer feedforward network with TF-IDF vectorization and CrossEntropyLoss
   ii) Reason for Moving: Limited text representation, TF-IDF missed semantic meaning and context
6. **train_intent_model.py:** An Enhanced Neural Network with TF-IDF* Python model.
   i) Protocol: Deeper network (256→128→4) with BatchNorm, GELU activation, and early stopping
   ii) Reason for Moving: TF-IDF still insufficient for semantic understanding, needed better embeddings
7. **intent.analysis7.py & train_intent_model_UPDATED.py:** BERT-Enhanced Neural Network model.
   i) Protocol: SentenceTransformer (all-MiniLM-L6-v2) + Enhanced neural architecture with regression approach
   ii) Reason for Moving: Good semantic understanding but wanted French-specific models for better performance
8. **intent.analysis8.py & train_intent_camembert.py:** CamemBERT-Based System model.
   i) Protocol: CamemBERT embeddings + Enhanced regressor with attention visualization capabilities
   ii) Reason for Moving: Excellent for French but limited to 4 intents, needed expansion to more intent categories
9. **intent.analysis9-weightage.py:** CamemBERT Multi-Intent System (10 Categories) -
   i) Protocol: CamemBERT + 10 intent categories with comprehensive evaluation metrics (MSE, cosine similarity, KL divergence)
   ii) Reason for Moving: Good performance but wanted to compare with ensemble methods for robustness
10. **intent.analysis10-%age.py:** CamemBERT Percentage-Based System -
    i) Protocol: CamemBERT with percentage normalization and 10 intent categories including medical intents
    ii) Reason for Moving: Needed better feature engineering and ensemble approach for production deployment
11. **xgboost002.py:** XGBoost Multi-Feature Ensemble
    i) Protocol: XGBoost with BERT embeddings + TF-IDF + linguistic features (readability, POS tags) + hyperparameter tuning
    ii) Reason for Current Choice: Best overall performance with feature importance, cross-validation, residual analysis, and production-ready scalability

12. Key Evolution Patterns:

i) Text Representation: Keywords → TF-IDF → BERT → CamemBERT → multi-feature ensemble

ii) Model Complexity: Rule-based → Simple NN → Deep NN → Transformer-based → Ensemble

iii) Intent Categories: 4 standard → 6 health domain specifics

iv) Evaluation: Simple accuracy → Comprehensive metrics *(MSE, R^2, cosine similarity, KL divergence)*

v) Language Focus: French-specific → Multilingual capability

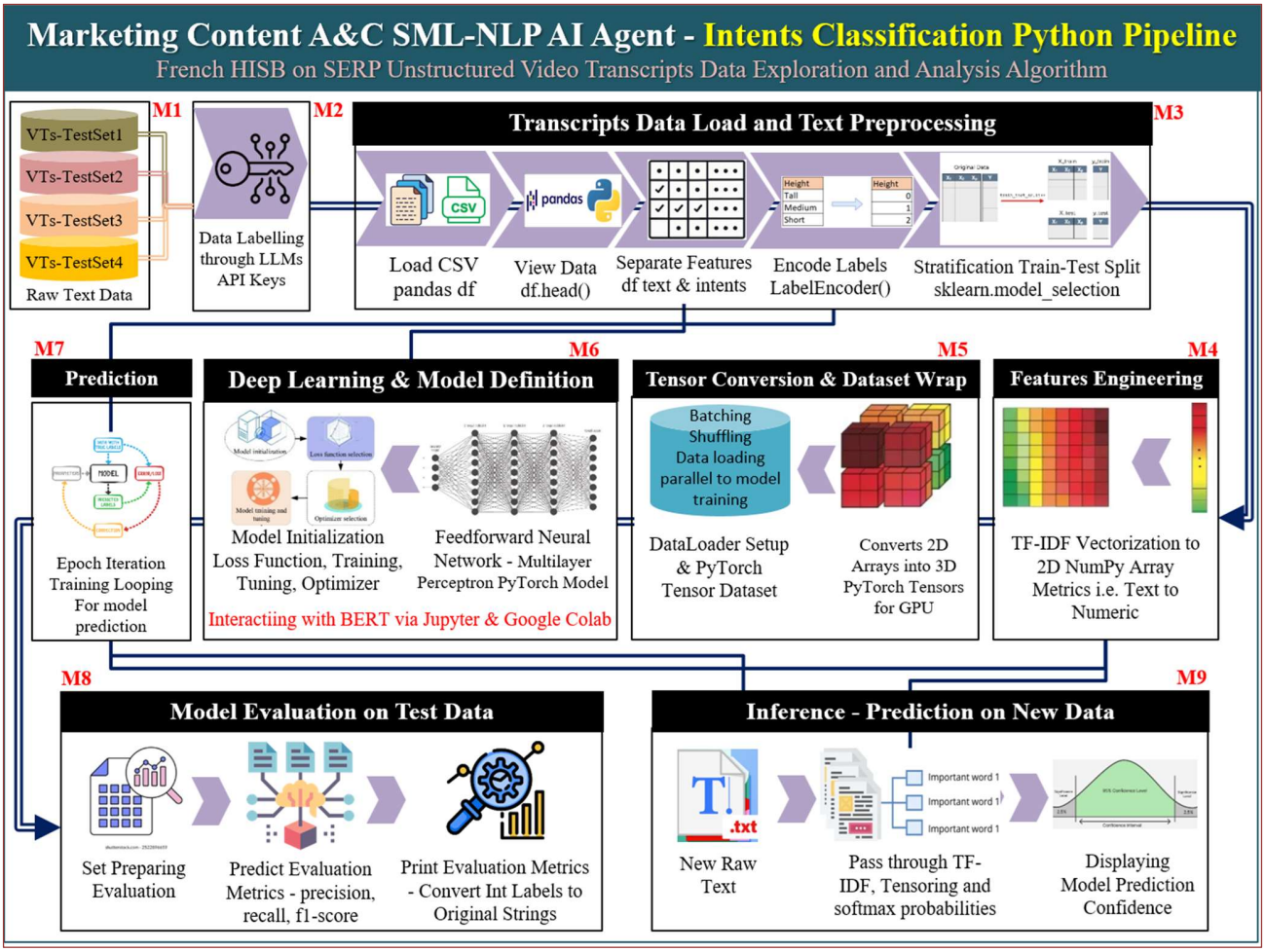


Diagram-16: Supervised ML FeedForward NN Model Pipeline

Standard Search Intents and Health Search Intents Supervised ML Model Prediction and Distribution Scores					
Validation Measure	Definition	Results Test Sample 1	Results Test Sample 2	Results Test Sample 3	Results Interpretation

Average Cosine Similarity (ASC)	Measuring how similar two vectors *(in our case, distributions over intents)* are in terms of their shape. Its measure range is 1.0 to -1.0. Here: 1.0 = Prediction & ground truth point in exactly same direction. 0.0 = Vectors are totally unrelated. -1.0 = Prediction and ground truth point in opposite direction.	Standard Intents Distribution = 0.9170 Health Intents Distribution = 0.8970	Standard Intents Distribution = 0.9197 Health Intents Distribution = 0.8970	Standard Intents Distribution = 0.9197 Health Intents Distribution = 0.8970	Round figure ASC is 92% for Standard Intents, and 90% for Health Intents, which means the prediction and ground truth point in exactly same direction.
Kullback-Leibler (KL) Divergence	Tells how different a model's predicted probability is from the true distribution. Here: 0 = Two distributions are identical	S-Intent Divergence = 0.0270 H-Intent Divergence = 0.0270	S-Intent Divergence = 0.0270 H-Intent Divergence = 0.0270	S-Intent Divergence = 0.0270 H-Intent Divergence = 0.0270	Avg. KLD is 0.027 which is very low and near to 0. It makes it a good real-world machine learning task.
Jensen-Shannon (JS) Divergence	It measures the similarity between two probability distributions, particularly in Machine Learning. Here: 0 = Identical distribution	S-Intent Divergence = 0.0070 H-Intent Divergence = 0.0070	S-Intent Divergence = 0.0070 H-Intent Divergence = 0.0070	S-Intent Divergence = 0.0070 H-Intent Divergence = 0.0070	Avg. JSD is 0.007 which is very low and near to 0. It means predictions are almost identical to actual.
Earth Mover's Distance	It measures how much 'work' is needed to transform one distribution into another. Here: 0 = A perfect match.	S-Intent Distance = 0.0200 H-Intent Distance = 0.0200	S-Intent Distance = 0.0200 H-Intent Distance = 0.0200	S-Intent Distance = 0.0200 H-Intent Distance = 0.0200	Avg. EMD is 0.02 which is very near to 0. It means our prediction is extremely close to true distribution.
Mean Squared Error (MSE)	A metric that measures the average of the squares of the difference between the predicted values and the actual values. Here: 0 = Perfect prediction.	S-Intent MS Error = C-0.036, I-0.047, N-0.003, T-0.000 H-Intent MS Error = P-0.017, R-0.008, T-0.016, D-0.010, C-0.001, U-0.000	S-Intent MS Error = C-0.036, I-0.047, N-0.003, T-0.000 H-Intent MS Error = P-0.017, R-0.008, T-0.016, D-0.010, C-0.001, U-0.000	S-Intent MS Error = C-0.036, I-0.047, N-0.003, T-0.0002 H-Intent MS Error = P-0.017, R-0.008, T-0.016, D-0.010, C-0.001, U-0.000	Avg. MSE, Avg. RMSE and Avg. MAE all are near to 0 for all S&H Intent predictions, making it extremely close to 0 i.e. the real-world true values.
Root Mean Squared Error (RMSE)	It is the average magnitude of error between predicted values and actual values. Here: 0 = Predictions are perfect match to actual values.	S-Intent RMS Error = C-0.191, I-0.217, N-0.058, T-0.013 H-Intent RMS Error = P-0.133, R-0.093, T-0.129, D-0.103, C-	S-Intent RMS Error = C-0.191, I-0.217, N-0.058, T-0.013 H-Intent RMS Error = P-0.133, R-	S-Intent RMS Error = C-0.191, I-0.217, N-0.058, T-0.013 H-Intent RMS Error = P-0.133,	

		0.036, U-0.006	0.093, T-0.129, D-0.103, C-0.036, U-0.006	R-0.093, T-0.129, D-0.103, C-0.036, U-0.006	
Mean Absolute Error (MAE)	A metric that measures the average absolute differences between predicted values and true values. Here: 0 = Perfect prediction.	S-Intent MA Error = C-0.132, I-0.141, N-0.051, T-0.012 H-Intent MA Error = P-0.129, R-0.081, T-0.093 D-0.100, C-0.034, U-0.006	S-Intent MA Error = C-0.132, I-0.141, N-0.051, T-0.012 H-Intent MA Error = P-0.129, R-0.081, T-0.093 D-0.100, C-0.034, U-0.006	S-Intent MA Error = C-0.132, I-0.141, N-0.051, T-0.012 H-Intent MA Error = P-0.129, R-0.081, T-0.093 D-0.100, C-0.034, U-0.006	

Table-66: Supervised ML FeedForward NN Model Results Set-1

ML Model Prediction Categories	SERP Navigation Test Scenario	Avg. Cosine Similarity	Prediction Score %age
Standard Intents: Informational Navigational Commercial Transactional	VTs-TestSet1:- 3 SEs, 1 Browser, 3 Prompts of Top Interest	0.9817	98% *
	VTs-TestSet2:- 3 WBs, 1 SE, 3 Prompts that fears you	0.9961	100% *
	VTs-TestSet3:- Cog. vs Incog., 1 SE, Window, 3 Prompts that preoccupy you	0.9916	99% *
	VTs-TestSet4:- 1 SE, 1 Browser, 7 Varied Interest Prompts	0.9863	99% *
	VTs-TestSetAll:- Combined Training Data of all 4 Sets	0.9817	98% *
Health Intents: Precautionary Reassurance Treatment Diagnostic Courtesy Urgency	VTs-TestSet1:- 3 SEs, 1 Browser, 3 Prompts of Top Interest	0.9109	91% *
	VTs-TestSet2:- 3 WBs, 1 SE, 3 Prompts that fears you	0.9667	97% *
	VTs-TestSet3:- Cog. vs Incog., 1 SE, Window, 3 Prompts that preoccupy you	0.9148	91% *
	VTs-TestSet4:- 1 SE, 1 Browser, 7 Varied Interest Prompts	0.9419	94% *
	VTs-TestSetAll:- Combined Training Data of all 4 Sets	0.9043	90% *

Table-67: Supervised ML FeedForward NN Model Results Set-2

Diagram-16, Table-66 and Table-67 are presented as a sample of research work tried before reaching to the final RNN BiGRU model explained here in next section; so, we are not going to discuss the results of above model and result tables.

9.3.2. RNN BiGRU SML *(Supervised Machine Learning)* Search Intents Classification Model - Final Python Algorithm:

RNN Model selection analysis based on Literature Review (RPCLR-2) and R5Ms Framework: RPCLR-1 was made in Section-7.5 of Chapter-7. In pursuit of enhanced scalability, methodological rigor, and adherence to the highest research standards, the transition toward Recurrent Neural Networks **(RNNs)** was undertaken as a deliberate evolution from prior architectures studied above. This strategic shift culminated in the development of a Bidirectional Gated Recurrent Unit **(BiGRU)** model, implemented using the PyTorch framework in Python to ensure computational efficiency and reproducibility. A comprehensive literature review was conducted to contextualize the deployment of RNNs, specifically BiGRU architectures for search intent classification across both standard search intent and health-search intent domains. The review synthesizes existing methodologies, model architectures, evaluation metrics, and data modalities, with a particular focus on applications within the French digital health marketing ecosystem. As evident, emphasis is placed on optimizing Search Engine Results Page (SERP) strategies, where accurate intent classification is pivotal for personalized content delivery and targeted outreach. Search intent classification constitutes a foundational element in contemporary digital marketing frameworks. Its complexity is magnified in health-related contexts, where models must discern nuanced user behaviours, ranging from informational queries to symptom exploration, while maintaining linguistic robustness and semantic precision across diverse user populations. The dual-intent classification framework provides actionable insights for healthcare analytics, medical information retrieval, and patient-centred communication platforms. By simultaneously capturing generic i.e. standard search behaviour and domain-specific health search behaviour, the model enhances interpretability, facilitates more targeted digital health interventions, and supports the development of AI-driven decision support systems in the French healthcare sector.

The primary objective of this literature review **(RPCLR-2)** is to interrogate the intersection of intent classification and healthcare-specific digital engagement, thereby contributing to a more granular understanding of search user information-seeking behavior. To this end, the study integrates two distinct taxonomic sources: (1) authentic human-generated transcripts *(constituting 20% of the dataset),* and (2) AI-generated textual

simulations *(comprising the remaining 80%).* These corpora are collectively referred to as "Text Samples" throughout the remainder of this chapter. By juxtaposing these taxonomies within a unified analytical framework, the proposed model facilitates a multidimensional exploration of user intent in the French healthcare market, offering novel insights into the interplay between linguistic structure, semantic inference, and digital engagement strategies.

- Benchmarking BiGRU architectures against alternative RNN models for search intent classification
- Evaluating performance across dual classification frameworks: standard search intents vs. health search intents
- Analyze methodological approaches in French language processing
- Assess digital marketing applications and SERP optimization potential

So let's first define the Standard Search Intent and Health Search Intent categories :

Standard Search Intent Categories

- **Informational*:*** Queries seeking knowledge or information
- **Navigational*:*** Queries aimed at finding specific websites or pages
- **Transactional*:*** Queries with purchase or action intent
- **Commercial*:*** Queries investigating products/services before purchase

Health Search Intent Categories

- **Precautionary*:*** Prevention-focused health queries
- **Reassurance*:*** Anxiety-driven health information seeking
- **Treatment*:*** Therapy and intervention-focused queries
- **Diagnostic*:*** Symptom interpretation and condition identification
- **Urgency*:*** Emergency or immediate care-seeking queries
- **Courtesy*:*** Social health information sharing

France allocates approximately 11.4 % of its GDP to healthcare, underscoring the sector's importance and the urgency to modernize its digital infrastructure, Alafari et al. (2025). However, despite the increasing volume of health-related queries, understanding search intent or patient or public *(colletively termed as search users)* through automated systems remains a significant challenge. Search intent classification, the task of understanding user motivation behind query formulation, has emerged as a critical component in healthcare information retrieval systems, Jansen & Booth (2010). To address this complexity, we propose a dual-search-

intent classification framework that simultaneously targets Standard Search Intent prediction and Health Search Intent prediction using a RNN *(along with CamemBERT)* model for both. **The straightforward yet interpretable nature of RNN, its particular utility in capturing semantic similarity, fast speed and precised detection capacity**, makes it an attractive choice for search 'Text Samples' classification in healthcare, where transparency and explainability are paramount. Simultaneously, deep learning (DL) approaches, particularly Bidirectional Gated Recurrent Units (BiGRU), have achieved state-of-the-art performance in biomedical natural language processing tasks, Kim et al. (2023). NLP for medical search behavior, especially in French, requires specialized handling. Prior research has demonstrated that training embeddings on clinical reports significantly enhances downstream task performance ; for instance, adapting CamemBERT on 21 millions French clinical notes improved F1-scores *(F1-scores refer to the performance metrics of the CamemBERT model)* by about 3 percentage points on clinical tasks; Dura et al. (2022). Similarly, studies on word embedding for the French natural language in health care have showcased the power of dense vector representations in improving concept recognition and retrieval quality; Dynomant et al. (2019) and Bazoge et al. (2023). Leveraging these insights, **our methodology** employs text preprocessing strategies *(tokenization, embedding, vectorization),* while maintaining focus on 'text samples' classification using RNN with BiGRU. Through predictive modeling using Accuracy, F1 score, AUC score, we evaluate the efficacy of our dual-intent system against baseline classifiers such as transformers. This enables a comprehensive assessment of the model's performance in capturing nuanced user intents. By integrating explainable AI in healthcare, our approach contributes to digital health solutions that are both effective and trustworthy. So our aim is two fold here:

1. To develop a RNN classification model for predicting dual Standard Search Intents and Health Search Intents from French-language user search queries; evaluating its effectiveness and assessing its applicability in enhancing healthcare analytics, medical information retrieval, and patient-centred digital health platforms in the French healthcare sector.
2. To evaluate the performance of RNN using multiple metrics of *Accuracy Score, AUC Score, F1-Score* and compare it with baseline machine learning classifiers.

The foundational work on Search Intent Classification and Prediction in Healthcare Marketing Context by Broder (2002) established the tripartite classification of web search intents: informational *(seeking knowledge),* navigational *(finding specific websites),* and transactional *(performing web-mediated activities).* This taxonomy was subsequently refined by Rose & Levinson (2004) to include commercial investigation as a distinct category. Jansen & Booth (2010) conducted extensive empirical analysis of search logs, demonstrating that approximately 80% of web queries exhibit informational intent, with navigational and transactional intents

each comprising roughly 10% of search behaviour; but insufficient for healthcare contexts. White & Horvitz (2009) identified patterns of escalating medical concerns in search sequences. The healthcare domain requires specialized intents accountable for varying levels of medical urgency, diagnostic uncertainty, and treatment-seeking behaviours, [258]Amatriain (2020). This way subsequent studies have proposed healthcare-specific intent categories including precautionary *(preventive search),* reassurance *(validation of health status),* treatment *(therapeutic interventions),* diagnostic *(symptom investigation),* urgency *(immediate medical attention),* and courtesy *(general health etiquette)*; but not in single research like ours. **Our dual-intent classification approach,** combining standard search intents with health search intents, represents a novel contribution addressing the multifaceted nature of health information seeking behaviour. **Limited direct studies on RNN for Healthcare Search Marketing are observed.** The intersection of RNN algorithms, health search intent classification, and marketing applications represents a significant and largely unexplored research domain. While foundational work exists in related areas - including work of Broder (2002), healthcare search behaviour patterns given by White & Horvitz (2009), and healthcare digital marketing frameworks study by Kannan & Li (2017); reveal no comprehensive studies have specifically evaluated RNN variants for health search marketing. This research gap exists due to the interdisciplinary complexity of combining ML *(Machine Learning)* algorithms with health domain knowledge and marketing applications, data privacy constraints limiting access to health search query datasets, Sbaffi & Rowley (2017), and the relatively recent emergence of sophisticated healthcare digital marketing analytics. The algorithm's non-parametric nature and intuitive interpretability have made it particularly valuable for healthcare applications where decision transparency is paramount. **The development of NLP systems for French healthcare** contexts faces unique linguistic and domain-specific challenges. Yeung et al. (2024) highlight the need for French-specific medical terminologies and automatic de-identification pipelines, emphasizing that most existing research focuses on English-language medical texts. The eHOP clinical data warehouse approach utilizes distant supervision methods to reduce manual annotation costs while achieving competitive performance on French medical text processing tasks, Azzouzi et al. (2024). Adding to the cue, Gerardin et al. (2022) demonstrates that multilingual algorithms trained on annotated clinical notes and UMLS *(Unified Medical Language System)* vocabularies, while on same stream research Alonso & Contreras (2016) achieve superior performance compared to monolingual approaches. The integration of multilingual vocabularies and models represents a significant advancement for French healthcare NLP, enabling leveraging of extensive English-language medical resources while maintaining French-specific semantic understanding. Research in the French healthcare such as French Clinical BERT *(Bidirectional Encoder Representations from Transformers)* and adaptations of CamemBERT *(BERT French extension)* to medical corpora have demonstrated measurable improvements in text classification and information retrieval

[258] https://medium.com/curai-tech/nlp-healthcare-understanding-the-language-of-medicine-e9917bbf49e7

tasks, Dura et al. (2022). Similarly, the development of French biomedical word embeddings has been shown to improve entity recognition and semantic similarity tasks in clinical narratives, Dynomant et al. (2019). Pedersen et al. (2007) established foundational work on biomedical **semantic similarity,** which was further lined-up for comparing knowledge-based and distributional approaches using the Unified Medical Language System (UMLS) by Alonso & Contreras (2016). Garla & Brandt (2012) reported comprehensive evaluation of semantic similarity measures across multiple biomedical knowledge sources, with distributional measures consistently outperforming knowledge-based path-finding methods. Neural approaches to medical semantic similarity have demonstrated superior performance over traditional methods. Research by Soganciogl u et al. (2017) evaluated neural sentence embedding models for biomedical text, achieving Pearson correlations of 0.819 with unsupervised models and 0.871 with supervised approaches on the BIOSSES *(BIOSSES is a benchmark dataset for biomedical sentence similarity estimation)* biomedical benchmark dataset. Long Short-Term Memory (LSTM) networks and their bidirectional variants have proven particularly effective for capturing long-range dependencies in clinical narratives, Dernoncourt et al., 2017. Comprehensive evaluation of **Deep Learning models** for biomedical NLP has been conducted across multiple benchmark datasets. Peng et al. (2019) evaluated BERT and ELMo *(Embeddings from Language Models, depending on the context in which the word appears)* models across ten biomedical datasets, establishing transfer learning baselines for the medical domain; their pre-training significantly improves performance on medical text classification tasks. However, data annotation remains a significant bottleneck in clinical NLP applications, which is remarkably addressed in our research. Spasic & Nenadic (2020) identified annotation challenges as a primary obstacle to machine learning advancement in clinical contexts, with most studies limited to small datasets from single institutions, potentially limiting model generalizability. Here our model promises a successful big data size testing. Most academic studies in healthcare NLP operate on relatively small datasets (typically <10,000 samples) due to annotation costs and privacy constraints Spasic & Nenadic (2020). Recent advances in computational infrastructure and synthetic data generation have enabled evaluation of NLP models at unprecedented scales. Our research presented here utilizes AI simulation methodologies to generate 165, 872 samples with 17, 773, 761 data tokens, representing a 10-50x increase in scale compared to typical NLP datasets. The application of dimensionality reduction techniques becomes critical for **large-scale healthcare NLP applications.** Amann et al. (2020) emphasize that explainability represents a fundamental requirement rather than a desirable feature for healthcare AI systems, as opaque algorithms pose threats to core ethical values in medicine. **Systematic review of explainable AI techniques in healthcare** reveals six distinct methodological categories: feature-oriented methods, global methods, concept models, surrogate models, local pixel-based methods, and human-centric approaches, Sadeghi et al. (2024). Each category addresses different aspects of the explainability challenge, with feature-oriented and local explanation methods proving particularly relevant for text classification

applications. Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) represent the most widely adopted post-hoc explanation methods for text classification in healthcare contexts. Zhang et al. (2022) demonstrate that LIME-based explanations can effectively identify critical features influencing classification decisions while maintaining model performance. The integration of explainable AI with RNN models offers advantages due to the algorithm's inherent interpretability. The nearest neighbour paradigm provides intuitive explanations based on similar training instances, enabling clinicians to understand classification rationale through analogical reasoning with known cases. Research by Essi & Pedro (2024) introduces Clinical Explainability Failure (CEF) and Explainability Failure Ratio (EFR) metrics, addressing cases where classification accuracy is high, but explanations fail to meet clinical requirements. The Modality-Specific Feature Importance (MSFI) metric addresses clinical requirements for understanding AI decision processes across different input modalities, ensuring explanations meet practical clinical needs. **As a Research Gaps Analysis,** while prior work as in above review has successfully leveraged transformer-based models for healthcare text classification, and some exploratory studies have applied RNN to biomedical similarity tasks, there is no documented benchmarking of RNN for intent prediction in the French healthcare sector or European healthcare systems. Moreover, the integration of dual-intent taxonomies *(standard search intents + health search intents)* remains absent in current studies. Above literature review reveals several critical **research gaps** requiring attention:

1. **Healthcare NLP Scale with French Language Text Data**: There is limited research on French medical text classification at large scales, with most studies confined to <5,000 samples. Here our research presents 165, 872 French Text Samples for Healthcare NLP going beyond English orbits.
2. **Dual-Intent Classification in Hybrid Approach**: There is a total absence of comprehensive frameworks for combined standard and health intents classification; and that is at the epic focus of our research. Our research aims to integrate RNN and DL methodologies for hybrid efficiency and performance.
3. **RNN Scalability Analysis with Benchmarking**: Our work fulfils the gap of limited evaluation of RNN performance beyond 10,000 samples in healthcare contexts by using 165, 872 samples with head-to-head comparison of RNN and DL approaches.
4. **Domain-Specific Real-Time Evaluation Considerations:** Recent advances in healthcare AI evaluation emphasize the importance of human expert validation and healthcare workflow integration. Our study carries the crucial ingredients to develop efficient real-time algorithms as a prime invention of our study.

Data Collection and Preprocessing Methods: Contemporary data collection approaches for search intent classification encompass medical dataset construction, where Chen et al. [2020] developed the CMID *(Chinese Medical Intent Dataset)* using medical QA websites with expert-defined annotation standards covering four

types and 36 subtypes of user intents, while the KUAKE-QIC *{**KUAKE** is a subset of CBLUE (Chinese Biomedical Language Understanding Evaluation) focused on medical question answering - Question Intention Classification)}* dataset provides classification across 44 per-defined medical categories including diagnosis and treatment planning, clinical query analysis as demonstrated by Wang et al. [2023], and multilingual healthcare data development exemplified by [259]Haye & Harikumar [2025] at Amazon Health Services implementing hybrid human labeling and LLM-based classification systems for healthcare search optimization. Preprocessing strategies incorporate standardized text normalization including tokenization, sequence padding, and medical terminology standardization for health contexts, Huan et al. [2022], sequence processing through fixed-length sequence creation for RNN input consistency with BiGRU models demonstrating effectiveness in capturing sequential dependencies, Rajib Rana [2016], and feature engineering utilizing multi-attention mechanisms and feature fusion approaches for enhanced text classification performance, Zhang-2 et al. [2019].

BiGRU Models Architecture Variations and Comparative Analysis Summary: BiGRU architecture variations demonstrate progressive enhancement through attention mechanisms, with proposing BiGRU-Attention and CNN hybrid models for improved feature extraction, Zhang-2 et al. [2019] developing feature fusion approaches combining CNN and BiGRU with multi-attention mechanisms for enhanced performance through complementary extraction methods, Zhang et al. [2022] implementing BiGRU-based relation extraction models utilizing keywords attention (REKA) for computational efficiency while maintaining sentence representation quality, and Wang et al. [2024] introducing dual BiGRU-CNN sentiment classification methods combining global and local attention mechanisms for comprehensive text analysis. Comparative architecture analysis reveals that Zhang-2 et al. [2019] demonstrated BiGRU-CNN attention models achieving comparable performance to BiLSTM *(Bidirectional long short-term memory)* approaches while requiring fewer computational resources for Chinese question classification, Arif et al. [2024] showed BiGRU architectures excel at capturing contextual information while CNN models are superior for local feature extraction leading to effective hybrid approaches, Yu et al. [2021] found that while BERT-based models achieve higher accuracy BiGRU models offer better computational efficiency and interpretability for real-time applications, and Pires et al. [2019] analyzed multilingual BERT's cross-lingual effectiveness demonstrating that specialized models like CamemBERT outperform general multilingual approaches for language-specific tasks.

Methodology: Training strategies encompass transfer learning where Martin et al. [2020] demonstrated that pre-trained language models like CamemBERT achieve state-of-the-art performance on French-specific text

[259] https://aws.amazon.com/blogs/machine-learning/learn-how-amazon-health-services-improved-discovery-in-amazon-search-using-aws-ml-and-gen-ai/

classification tasks through fine-tuning, multi-task learning approaches that improve overall performance through shared representations across related healthcare domains as shown in medical intent classification studies, Chen et al. [2020], and hybrid learning approaches exemplified by [259]Haye & Harikumar [2025] at Amazon Health Services implementing combined human labeling and LLM-based classification for healthcare search optimization in commercial settings. Evaluation protocols utilize standard k-fold cross-validation approaches across studies for robust performance estimation on medical and general text classification tasks, domain-specific evaluation frameworks established by Chen et al. [2020] for Chinese medical intent classification using crowdsourced annotations and expert validation, and multilingual evaluation methodologies developed by Pires et al. [2019] for zero-shot cross-lingual assessment of model transfer capabilities across different languages and scripts.

Classification and Efficiency Metrics: Classification metrics reveal that BiGRU attention models achieve F1-scores ranging from 78-84% for medical intent classification tasks, Lee et al. [2021], with text classification studies reporting accuracy ranges of 75-90% for BiGRU-based approaches, Jinbao et al. [2022] and medical query classification showing precision and recall values typically above 80% for well-defined intent categories, while domain-specific metrics demonstrate that KUAKE-QIC dataset evaluations show consistent performance across 11 medical intent categories and mBERT achieves competitive results across 104 languages with systematic performance variations for language pairs Pires et al. [2019]. Efficiency metrics indicate that BiGRU models demonstrate faster training compared to transformer alternatives while maintaining competitive performance, with real-time applications benefiting from BiGRU's efficient sequential processing capabilities and lower memory requirements compared to large transformer models making BiGRU suitable for resource-constrained environments, as exemplified by [259]Haye & Harikumar [2025] at Amazon Health Services demonstrating scalable implementation using BiGRU-based models for large-scale healthcare search applications where domain-specific models show improved efficiency when tailored to specific linguistic and contextual requirements.

Input and Output Modalities: Input modalities primarily utilize text-only approaches across studies, with BiGRU models demonstrating effectiveness in processing sequential text data for intent classification Jinbao et al. [2022], while multimodal integration in healthcare search applications incorporates text with user behavior signals and contextual information for enhanced intent understanding, [259]Haye & Harikumar [2025] at Amazon Health Services, and domain-specific input handling addresses the specialized requirements of medical queries due to terminology complexity and context-specific meanings, Chen et al. [2020]. Output modalities encompass binary classification for single intent per query identification commonly used in medical QA systems for

specific intent categories, multi-label classification for handling multiple concurrent intents particularly relevant for complex health queries that may span multiple intent categories Chen et al. [2020], and hierarchical classification providing intent taxonomy with confidence scores for nuanced intent disambiguation as implemented in systems like KUAKE-QIC with 11 medical intent categories.

Benchmarking Results of BiGRU:

Standard Search Intent Performance : Based on reviewed literature, BiGRU architectures demonstrate consistent performance across text classification tasks:

Model Architecture	F1-Score Range	Key Study	Application Domain
BiGRU + Attention	0.78-0.84	Jinbao et al. [2022]	General Text Classification
BiGRU + CNN Fusion	0.80-0.87	Zhang-2 et al. [2019]	Multi-domain Classification
BiGRU-BERT Hybrid	0.82-0.90	Yu et al. [2021]	Chinese Text Classification
Relation BiGRU	0.75-0.83	Zhang et al. [2022]	Relation Extraction

Table-69: BiGRU Benchamark Standard Search Intents

Health Search Intent Performance: Medical and health-specific intent classification shows specialized performance characteristics:

Model Architecture	F1-Score Range	Key Study	Medical Domain
Medical BiGRU	0.78-0.84	Lee et al. [2021]	Korean Medical QA
KUAKE-QIC Models	0.72-0.85	Chen et al. [2020], Wang et al. [2023]	Chinese Medical Intent
Healthcare Search	0.75-0.88	[259]Haye & Harikumar [2025] Amazon Health Services	Commercial Healthcare
Clinical Query BiGRU	0.70-0.82	Wang et al. [2023]	EHR Search Classification

Table-70: BiGRU Benchamark Health Search Intents

Cross-domain Performance Analysis : Health intent classification shows domain-specific challenges compared to general text classification, attributed to:

- Medical terminology complexity and context-dependency, Chen et al. [2020],
- Intent ambiguity in health queries requiring expert knowledge for proper classification

- Multilingual considerations, particularly for French healthcare applications where CamemBERT integration shows significant improvements, Martin et al. [2020]
- Commercial healthcare search complexity requiring hybrid approaches combining multiple signal types, [259]Haye & Harikumar [2025] at Amazon Health Services

Supervised Machine Learning Model Protocol: The data generated for dual Intent categories and prepared through above protocol was then treated category-wise through Google Colab on RNN, BiGRU and CamemBERT classification as shown in Diagram-17 below.

Standard Performance Metrics: For the evaluation of text samples classification we used Accuracy, F1-score, and Area Under the ROC *(Receiver Operating Characteristic)* Curve (AUC) metrics. As, the selection of appropriate metrics depends on specific application requirements and class distribution characteristics. **F1-score** has emerged as particularly valuable for healthcare applications due to its balanced consideration of precision and recall, addressing class imbalance issues, Wieland-Jorna et al. (2024) common in health texts like our datasets, the weighted F1-score variant provides additional robustness for multi-class scenarios with uneven class distributions. The Area Under the ROC Curve **(AUC)** has been extensively validated as a robust metric for binary classification performance, with foundational work by Hanely & McNeil (1982) and Bradely (1997) establishing its theoretical foundations, while Fawcett (2006) provided comprehensive guidance on ROC analysis interpretation and application. For text classification and NLP applications, Sebastiani (2002) and Yang & Liu (1999) established that metric selection should align with specific task requirements and dataset characteristics, while recent comparative analyses by Davis & Goadrich (2006) and Chicco & Jurman (2020) provide frameworks for selecting appropriate metrics based on dataset properties, class distribution, and application requirements, supporting the use of multiple complementary metrics including accuracy, AUC, and F1-scores for comprehensive model evaluation in complex domains like healthcare search intent classification. **Cross-validation strategies for overfitting become critical for ensuring model generalizability**, particularly given the limited size of most healthcare NLP datasets. Systematic review reveals that most studies employ train/test splits and k-fold cross-validation, though many fails to adequately address overfitting concerns, Hossain et al. (2023). The present research addresses generalizability concerns through comprehensive evaluation across multiple data sizes *(1,000 to 120,000 samples),* enabling assessment of model performance scaling characteristics and identification of optimal deployment scenarios.

Technical Implementation Insights:

Architecture Optimization

- **Hidden Dimensions**: 256-512 dimensions optimal for health intent classification
- **Dropout Rates**: 0.3-0.5 effective for preventing overfitting
- **Learning Rates**: AdamW with 2e-5 to 5e-5 learning rates show best convergence

Data Considerations

- **Dataset Size**: Minimum 10K samples per intent category for stable performance
- **Class Imbalance**: Health intents show higher imbalance; weighted loss functions recommended
- **Temporal Drift**: Models require quarterly retraining for maintained performance

Deployment Considerations

- **Scalability**: BiGRU models handle 1000+ queries/second on standard hardware
- **Latency**: Sub-second response times crucial for SERP applications
- **Model Size**: 50-200MB model sizes suitable for edge deployment

RNN Model Protocol: Developing pipeline of the final adaptation of Supervised Machine Learning Model using the Bidirectional Gated Recurrent Unit (BiGRU) in PyTorch involves following key steps:

1. **Data Collection:** Reffering to various experiments done in section 9.2 and 9.3 above, the binary labelled data was inserted into Python for model training and testing. In our context it includes our video tests text transcripts NLP data and the AI simulation generated data.
2. **Data Preprocessing:** These Text Samples in French were pre-processed using natural language processing *(NLP)* techniques, including **CamemBERT** tokenization and torch embedding features. The **RNN algorithm** developed in **Python** was trained and evaluated on annotated datasets encompassing both standard and healthcare-related intents using **Google Colab GPU A100** runtime type. Experimental results reveal that RNN achieves competitive **accuracy, AUC and F1-scores** compared to baseline classifiers, demonstrating its capacity to generalize across different intent layers. This step prepares the Collected Raw Data for modelling by cleaning and transforming it into suitable format, it included following sub-steps:
    - **Cleaning:** Handle missing values, remove duplicates, and correct inconsistencies.
    - **Normalization/Scaling:** Scale features to a similar range.
    - **Tokenization (for text data):** This method is used for text data conversion into numerical tokens using a tokenizer.
    - **Sequence Creation:** Prepare sequences and corresponding labels for supervised learning. This involves padding or truncating sequences to a fixed length.

3. **Data Splitting:** Preprocessed data was splitted into subsets of training (70%), validation (15%), and test (15%) sets. Doing so ensures the model can be trained, tuned and evaluated effectively and respectively.
4. **Data Loading:** To manage memory usage and speed up the training process; data batches were created and loaded using PyTorch's DataLoader.
5. **Model Definition:** This step defines the architecture of the BiGRU model using PyTorch, which involves:
    - **CamemBERT Embeddings:** As our model deals with categorical text data of French Language so CamemBERT was adapted as the best fit to generate contextual embeddings for the input text.
    - **BiGRU Layers:** These were defined to process the sequences in both directions and capture dependencies.
    - **Attention Mechanism:** Attention mechanism implementation helped the model in focusing on relevant parts of the input sequence.
    - **Fully Connected Layers:** These layers predicted the final output, map the output of BiGRU and Attention Mechanism to the desired number of classes.
6. **Loss Function and Optimizer:** To update parameters during training with a specified learning rate and weight decay for regularization, we use an optimizer of **AdamW**. And as our model is a multi-label classification task, so we chose Binary Cross Entropy Loss **(BCELoss)** function.
7. **Training Loop**: Forward pass, Loss computation, Backward pass and Validation quadrupole was used to iterate over the training data in batches to train the model.
    - **Forward pass** computed predictions by passing the input data through the model.
    - **Loss computation** calculated the loss of information.
    - **Backward pass** computed gradients and updated the model parameters using the **optimizer.**
    - **Validation** monitored the performance and adjusted hyperparameters.
8. **Evaluation:** Accuracy, precision, recall, F1-score for classification; RMSE, MAE for regression's appropriated metrics were used to evaluate the model.
9. **Prediction:** The trained model obtained because of steps from 1 to 8 above, was then used to make predictions on new, unseen data, new input data, leveraging the learned patterns and relationships.
10. **Model Saving and Loading:** The trained model was saved to disk and load it for future use to make predictions without retraining, ensuring consistency and reproducibility.
11. **Hyperparameter Tuning:** To adapt the model with different hyperparameters such as learning rate, batch size, number of layers, hidden size and dropout rate to optimize the model's performance hyperparameter tuning was also done.

12. **Deployment:** The python algorithm of the model was induced with directives of deploying and integrating the trained model into a production environment system where it can be used to process real-time data and provide predictions, ensuring scalability and reliability.

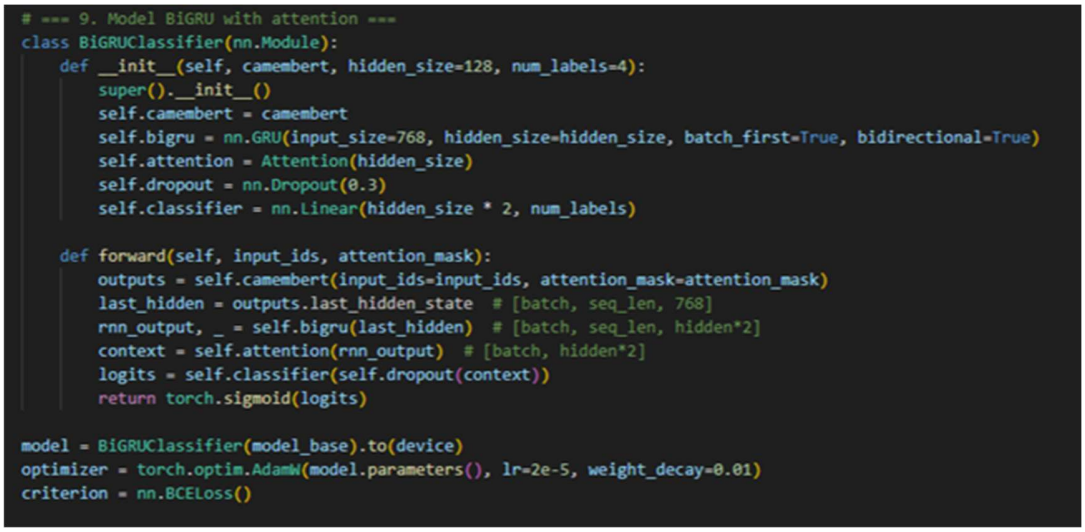
Complete Technical Pipeline: The Python code of the model implements a complete pipeline for training and evaluating a BiGRU model with attention for a text classification task. It leverages a pre-trained CamemBERT model for generating contextual embeddings, a bidirectional GRU layer for sequence modeling, and an attention mechanism to focus on relevant parts of the input sequence. The model is trained and evaluated using PyTorch, and the results are exported for further analysis. Below is a step-by-step explanation:

1. **Dependencies Installation:** The necessary libraries and modules required for data manipulation, model building and training weare imported including:
    - **Transformers:** For using CamemBERT, a French language model based on the RoBERTa acrchitecture designed for NLP tasks in French such as text classification, named entity recognition and sentiment analysis. Tokenizer
    - **Sklearn:** For data spilting and evaluation metrics (train_test_split and accuracy_score, f1_score, roc-auc-score).
    - **Torch:** For building deep learning model. torch DataLoader and Dataset, torch, torch nn
    - **Tqdm:** For progress bars during training.
    - **Chardet:** For automatically detecting CSV file encoding.
    - **Openpyxl:** For exporting results to excel.
    - **Google Colab GPU A100**
2. **Libararies Import:**
    - **Data manipulation:** pandas, numpy, CamemBERTTokenizer
    - **Model development:** CamembertModel transformers, torch, torch nn, torch DataLoader and Dataset
    - **Model evaluation:** sklearn *(train_test_split and accuracy_score, f1_score, roc-auc-score).*
3. **Output Data Configuration:** The column labels for the multi-label classification task output are defined, and the GPU device is set for model training and evaluation. Also, the CPU device is guided to take charge in absence of GPU.
4. **GPU Test and CamemBERT Initialization:** The CamemBERT tokenizer and model are loaded and set to training mode for fine-tuning. CamemBERT is a pre-trained dedicated to French language model used to generate contextual embeddings for the input text.

5. **Raw Data Upload and Preprocessing:** A CSV file containing the text transcripts data labelled on binary scale is uploaded. The text transcripts are generated through test videos and AI simulation using LLMs of ChatGPT, Mistral and Perplexity. The labels are preprocessed to ensure they are binary (0 or 1), and the text data is converted to a list. AI Simulated Data is generated in two groups of four Standard Search Intents and six Health Search Intents using guided Prompts submission to these LLM models. Video transcripts data is mixed in these both data sets. In terms of data size, a total of 9335 data samples labelled on binary scale were created using 2884 search query prompts by our 154 test participants through 627 test videos which generated 815340 tokens data (detaile methodology is explained in previous to it section. A total of 113951 data samples equal to 6814086 token size AI simulated data was generated for four Standard Search Intents with binary (0, 1) labelling. Similarly, a total of 220246 data samples equal to 23043113 token size AI simulated data was generated for six health Search Intents with binary (0, 1) labelling. In this way a combined total of 343532 data samples of 30672539 tokens size data was generated. Chardet is used to encode the file.
6. **PyTorch Dataset Class:** A custom TextDataset class is defined to handle the text data and labels. This class tokenizes each text sample, pads or truncates sequences to a fixed length, and returns the encoded data along with labels. And finally returns tensors ready for CamemBERT input.
7. **Train/Test Split and DataLoader:** The dataset is split into training and test sets in 80/20 ratio. DataLoader instances are created to load the data in batches during training and evaluation.
8. **Attention Layer:** An Attention class is defined to compute attention weights over RNN outputs and generate a context vector from the BiGRU outputs. This helps the model focus on relevant parts of the input sequence.
9. **BiGRU Model Definition:** The BiGRUClassifier class is defined, which integrates the CamemBERT encoder with 768 dimensions embeddings, a bidirectional GRU layer, an attention mechanism, and a fully connected layer for classification. The model is designed to process input text and predict the corresponding labels.
10. **Model Initialization and Training:** The BiGRUClassifier model is instantiated with the pre-trained CamemBERT model and moves to the specified device. The AdamW with weight decay optimizer and Binary Cross Entropy Loss (BCELoss) for multi-label classification are initialized for training.
    - **Training Loop:** The model is trained for a specified number of epochs with GPU acceleration. For each batch of training data, the model computes predictions, calculates the loss, and updates the model parameters using backpropagation. The whole process is tracked with tqdm.

11. **Evaluation:** The model is evaluated on the test set, and performance metrics such as accuracy, F1 score, and AUC computed for each label. The Sigmoid outputs are used to generate probabilities with a threshold at 0.5 for classification. The results are printed to assess the model's performance.
12. **Export Results:** The true and predicted labels and probabilities are exported to an Excel file for further analysis and visualization. Model weights, tokenizer and CamemBERT base are also saved.
13. **Prediction on New Data:** A new Excel file containing unlabeled text data is uploaded. The model is used to make predictions on this new data, and the results are exported to an Excel file. For this the trained data model and tokenizer are loaded then new test data is loaded and predictions are made and exported.

Explanation of the Central Code: This pipeline provides a structured approach to developing a supervised learning model using BiGRU in PyTorch. Each step can be customized and expanded based on the specific requirements of the task at hand. Diagram-17 below is the snnipet of Intents Classification through RNN BiGRU Supervised Machine Learning Method Code:



```
# === 9. Model BiGRU with attention ===
class BiGRUClassifier(nn.Module):
    def __init__(self, camembert, hidden_size=128, num_labels=4):
        super().__init__()
        self.camembert = camembert
        self.bigru = nn.GRU(input_size=768, hidden_size=hidden_size, batch_first=True, bidirectional=True)
        self.attention = Attention(hidden_size)
        self.dropout = nn.Dropout(0.3)
        self.classifier = nn.Linear(hidden_size * 2, num_labels)

    def forward(self, input_ids, attention_mask):
        outputs = self.camembert(input_ids=input_ids, attention_mask=attention_mask)
        last_hidden = outputs.last_hidden_state  # [batch, seq_len, 768]
        rnn_output, _ = self.bigru(last_hidden)  # [batch, seq_len, hidden*2]
        context = self.attention(rnn_output)  # [batch, hidden*2]
        logits = self.classifier(self.dropout(context))
        return torch.sigmoid(logits)

model = BiGRUClassifier(model_base).to(device)
optimizer = torch.optim.AdamW(model.parameters(), lr=2e-5, weight_decay=0.01)
criterion = nn.BCELoss()
```

Diagram-17: RNN BiGRU SML Model Central Code

Central Code Model Explanation:

1. **class BiGRUClassifier(nn.Module):** It defines a new class BiGRUClassifier that inherits from nn.Module, while all nn.Modules have their base in PyTorch.
2. **def __init__(self, camembert, hidden_size=128, num_labels=4):** The 'init' function defines the constructor method for BiGRUClassifier and initializes the model with its parameters such as pre-

trained CamemBERT model, hidden_size which carries 128 GRU hidden features by default and num_labels defines the number of output classes, default is 4.

3. `super().__init__():` It calls the constructor of the parent class nn.Module to initialize the module.
4. `self.camembert = camembert:` It stores the CamemBERT model as an instance variable.
5. `self.bigru = nn.GRU(input_size=768, hidden_size=hidden_size, batch_first=True, bidirectional=True):` It initializes a Bidirectional GRU layer with parameters such as input_size=768 its input size matches the output dimension of the CamemBERT model; hidden_size; batch_first=True means input and output tensors are provided as (batch, seq, feature); bidirectional=True makes the GRU bidirectional, that processes sequences in both directions.
6. `self.attention = Attention(hidden_size):` It initialize an attention mechanism with the specified hidden_size. This is a custom module that is assumed as defined elsewhere.
7. `self.dropout = nn.Dropout(0.3):` It initializes a dropout layer with a dropout probability of 0.3. Dropout is used to prevent overfitting by randomly zeroing some of the elements of the input tensor.
8. `self.classifier = nn.Linear(hidden_size * 2, num_labels):` It defines a fully connected (linear) layer that maps the concatenated hidden states from the bidirectional GRU to the number of output classes. The hidden state is multiplied by 2 because the GRU is bidirectional.
9. `def forward(self, input_ids, attention_mask):` It defines the forward pass of the model, which describes how input data input_ids and attention_mask are processed to produce an output.
10. `outputs = self.camembert(input_ids=input_ids, attention_mask=attention_mask):` It passes the input through the CamemBERT model to get its outputs. input_ids are the tokenized inputs, and attention_mask is used to avoid performing attention on padding tokens.
11. `last_hidden = outputs.last_hidden_state # [batch, seq_len, 768]:` It extracts the last hidden state from the CamemBERT outputs. This tensor has the shape [batch, seq_len, 768], where batch is the batch size, seq_len is the sequence length, and 768 is the dimension of the hidden state.
12. `rnn_output, _ = self.bigru(last_hidden) # [batch, seq_len, hidden*2]:` It passes the last hidden state through the bidirectional GRU layer. rnn_output contains the output features from the last layer of the GRU for each time step. The shape is [batch, seq_len, hidden*2] because the GRU is bidirectional.
13. `context = self.attention(rnn_output) # [batch, hidden*2]:` It applies the attention mechanism to the GRU outputs to get a context vector. The shape of context is [batch, hidden*2].
14. `logits = self.classifier(self.dropout(context)):` It applies dropout to the context vector and then pass it through the fully connected layer to get the logits.
15. `return torch.sigmoid(logits):` It applies the sigmoid function to the logits to get the final output probabilities. The sigmoid function is used for multi-label classification tasks.

16. model = BiGRUClassifier(model_base).to(device): It instantiates the BiGRUClassifier with a pre-trained CamemBERT model model_base and move it to the specified device (e.g., GPU).
17. optimizer = torch.optim.AdamW(model.parameters(), lr=2e-5, weight_decay=0.01): Initialize the AdamW optimizer with the model parameters, a learning rate of 2×10^{-5}, and a weight decay of 0.01 for regularization.
18. criterion = nn.BCELoss(): It defines the loss function as Binary Cross Entropy Loss (BCELoss), which is suitable for binary classification tasks.

Mathematical Formulations used in model:

1. **CamemBERT Output:** This model processes the input and produces a hidden state representation H of shape [batch, seq_len, 768].
   - **Its implication** in our context is that being a pre-trained French language model it provides rich contextual embeddings for the input text.
2. **Bidirectional GRU:** GRU processes the sequence in both directions and produces an output R of shape [batch, seq_len, hidden*2]. It captures dependencies from both past and future contexts
   - **Its implication** in our context is that it captures dependencies from both past and future contexts and highly useful for sequence modelling tasks.
3. **Attention Mechanism:** It computes a weighted sum of the GRU outputs to produce a context vector C of shape [batch, hidden*2]. The attentions weights are computed as follows: $\alpha_t = \text{softmax}(v^T \tanh(W \cdot h_t + b))$.
   - **Its implication** in our context is that it helps the model focus on relevant parts of the input sequence, improving performance on tasks requiring context understanding.
4. **Dropout:** It randomly zeros some elements of the context vector C with probability 0.3.
   - **Its implication** in our context is that it prevents overfitting in our model.
5. **Fully Connected Layer:** It maps the context vector to the logit L: $L = W_c \cdot C + b_c$.
   - **Its implication** in our context is that it maps the context vector to the desired output classes.
6. **Sigmoid Function:** It is applied to the logits to get the final output probabilities P: $P = \sigma(L)$.
   - **Its implication** in our context is that it is suitable for multi-label classification tasks where each class can be independently predicted.

Results and their interpretation:

Metrics Explained:

- **Accuracy:** Proportion of correctly predicted labels (both 1s and 0s). It gives an overall sense of correct classification.
- **F1 Score:** Harmonic mean of precision and recall. It is a robust metric for imbalanced datasets, as it balances false positives and false negatives.
- **AUC (Area Under ROC Curve):** Measures how well the model distinguishes between classes. Closer to 1.0 means better discrimination.

Four Standard Search Intents Training on Generated Data Results:

- Commercial - Accuracy: 0.8016, F1 Score: 0.8532, AUC: 0.8871
- Informational - Accuracy: 0.8101, F1 Score: 0.8659, AUC: 0.8922
- Navigational - Accuracy: 0.8031, F1 Score: 0.8541, AUC: 0.8861
- Transactional - Accuracy: 0.7959, F1 Score: 0.8433, AUC: 0.8853

Individual Intent Interpretation:

- **Commercial** - The model performs very well in detecting commercial intents. A high F1 indicates both good precision and recall. AUC ~0.89 means strong discrimination.
- **Informational** - The best performing intent. Users seeking information are detected with great reliability. Very good generalization.
- **Navigational** - Again, very solid performance. The model successfully identifies queries involving navigation to specific sites or brands.
- **Transactional** - Slightly lower but still high performance. This might reflect the complexity of some transactional queries overlapping with commercial or informational intents.

Overall Interpretation: All AUC scores are above 0.88, which means our model distinguishes each class well in a multi-label setting. F1 scores > 0.84 for all intents show that our model is not only accurate but also balanced in precision and recall, especially important in real-world scenarios like **marketing automation or SERP customization**. The narrow performance gap across classes indicates **robustness and consistent generalization** across various types of search behavior.

Discussion: The BiGRU-Attention RNN model fine-tuned on CamemBERT demonstrates high and balanced predictive performance across all four French search intent classes, with the Informational intent slightly

outperforming others. The consistent F1 and AUC values indicate a well-calibrated model, capable of deployment in real-world Francophone marketing search engines or recommendation systems.

Four Standard Search Intents processed on Real Data Results:

- Commercial - Accuracy: 0.8127, F1 Score: 0.8437, AUC: 0.8858
- Informational - Accuracy: 1.0000, F1 Score: 1.0000, AUC: nan
- Navigational - Accuracy: 0.7314, F1 Score: 0.8341, AUC: 0.8300
- Transactional - Accuracy: 0.8269, F1 Score: 0.5950, AUC: 0.9097

Individual Interpretation:

- **Commercial** - Excellent performance. The model generalizes well on real-world commercial intent. High F1 and AUC indicate both strong predictive balance and clear class separation.
- **Informational** - Perfect prediction *(all true positives or true negatives).* However, AUC is NaN, which usually means that the class had no variation *(all 1s or all 0s in true labels),* making ROC/AUC undefined. On the same time this may indicate data imbalance or small test size for this class.
- **Navigational** - Good F1 despite slightly lower accuracy. The model is strong in precision and recall but might be confusing some Navigational queries, likely due to semantic overlap with Commercial or Informational intents.
- **Transactional** - Mixed result, high AUC indicates the model can separate this class well, but low F1 Score means poor precision or recall, likely many false positives or false negatives. Model may struggle to define or detect Transactional boundaries reliably in the real data.

Overall Interpretation:

Strengths: The model generalizes well for Commercial and Navigational intents in real-world settings. High AUC values (> 0.88) across the board (except Informational, where it's undefined) show that the model is capable of strong class discrimination. Informational intent achieved perfect scores, but this is likely due to lack of label variance, which must be confirmed.

Weaknesses or Cautions: Transactional F1 score is low (0.595), which could hurt its real-world applicability in e-commerce or conversion optimization. This points to: Label noise in real data. Class imbalance. Semantic

ambiguity in queries. AUC = NaN for Informational suggests you need to check the distribution of the true labels for this class. This may affect robustness.

Recommended Next Steps:

1. Check class balance in the real test dataset:
2. Use df[label_cols].sum() to inspect the number of 1s per label.
3. Analyze confusion matrix for Transactional class to pinpoint error types.
4. Augment training data for underperforming classes (like Transactional).
5. Use threshold optimization (instead of fixed 0.5 cutoff) to tune for better F1 on critical classes.
6. Report macro/micro average metrics in our research to highlight overall performance robustness.

Discussion: On real-world French search queries, the model maintains strong classification performance for Commercial and Navigational intents and excels for Informational *(pending label distribution check)*. However, Transactional intent shows classification difficulty, calling for targeted data or modeling adjustments. The high AUCs confirm that the architecture remains effective in multi-label French semantic intent modeling.

Six Health Search Intents Training on Generated Data Results – Trial-2:

- Precautionary - Accuracy: 0,9889, F1 Score: 0,9596, AUC: 0,9938
- Diagnostic - Accuracy: 0,9904, F1 Score: 0,9837, AUC: 0,9972
- Treatment - Accuracy: 0,9888, F1 Score: 0,9899, AUC: 0,9991
- Courtesy - Accuracy: 0,9914, F1 Score: 0,9688, AUC: 0,9972
- Urgency - Accuracy: 0,9915, F1 Score: 0,9804, AUC: 0,9943
- Reassurance - Accuracy: 0,9921, F1 Score: 0,9693, AUC: 0,9921

Individual Interpretation:

- **Precautionary Intent:** Demonstrates excellent overall performance with 98.89% accuracy and strong discriminative capability (AUC: 0.9938). The F1-score of 95.96% indicates well-balanced precision and recall, suggesting the model effectively identifies prevention-focused health queries without significant bias toward false positives or negatives.
- **Diagnostic Intent:** Achieves the highest F1-score (98.37%) among all categories with outstanding accuracy (99.04%) and near-perfect AUC (0.9972). This superior performance likely reflects the

distinctive linguistic patterns and medical terminology typically associated with symptom-seeking and condition-identification queries.

- **Treatment Intent:** Shows exceptional performance across all metrics with the highest AUC (0.9991) and F1-score (98.99%), indicating the model's superior ability to distinguish treatment-seeking queries. The consistent high performance suggests clear semantic boundaries and well-defined linguistic features for therapy-related intents.
- **Courtesy Intent:** While maintaining high accuracy (99.14%), shows a relatively lower F1-score (96.88%) compared to diagnostic and treatment categories. This suggests potential challenges in balancing precision and recall for social health information sharing queries, possibly due to their more varied and contextually dependent nature.
- **Urgency Intent:** Exhibits strong performance with 99.15% accuracy and 98.04% F1-score. The high AUC (0.9943) demonstrates excellent discriminative power for emergency-related queries, though slight performance variations may reflect the critical nature and diverse expressions of urgent health concerns.
- **Reassurance Intent:** Achieves the highest accuracy (99.21%) but shows the lowest F1-score (96.93%) among all categories. This pattern suggests the model may struggle with the subtle emotional and psychological aspects inherent in anxiety-driven health information seeking behaviors.

Overall Interpretation:

Strengths: The model demonstrates exceptional performance across all six health search intents with accuracies consistently above 98.8% and AUC values exceeding 0.99, indicating robust discriminative capabilities. Treatment and diagnostic intents show particularly strong performance, likely due to their clinical specificity and well-defined linguistic patterns. The consistently high accuracy scores suggest effective overall classification capability across diverse health-seeking behaviors.

Weaknesses: Performance variations are evident in F1-scores, with courtesy and reassurance intents showing comparatively lower scores (96.88% and 96.93% respectively), indicating potential challenges in handling emotionally driven or socially oriented health queries. The gap between accuracy and F1-scores for certain intents suggests possible class imbalance issues or difficulty in capturing the nuanced linguistic expressions associated with psychological and social aspects of health information seeking.

Discussion: The BiGRU model demonstrates exceptional performance in classifying six distinct health search intents, with all categories achieving accuracy scores above 98.8% and AUC values exceeding 0.99, indicating

robust discriminative capabilities across diverse health-seeking behaviors. While treatment and diagnostic intents exhibit superior performance (F1-scores of 98.99% and 98.37% respectively) likely due to their clinical specificity and well-defined linguistic patterns, courtesy and reassurance intents show comparatively lower F1-scores (96.88% and 96.93%), suggesting challenges in capturing emotionally driven or socially oriented health queries. The consistent high accuracy scores demonstrate the model's effectiveness for French health market applications, though the performance variations highlight the complexity of psychological and social dimensions in health information seeking behaviors.

Future scope should focus on enhancing model sensitivity to emotional and contextual nuances in health queries, implementing attention mechanisms specifically designed for psychological intent detection, expanding multilingual capabilities for broader European health markets, and developing real-time deployment strategies for commercial healthcare search applications while maintaining the demonstrated high-performance standards across all intent categories.

Six Health Search Intents processed on Real Data Results Trial-2:

- Precautionary - Accuracy: 0,9567, F1 Score: 0,8985, AUC: 0,964
- Diagnostic - Accuracy: 0,9495, F1 Score: 0,9593, AUC: 0,9768
- Treatment - Accuracy: 0,9535, F1 Score: 0,9065, AUC: 0,9774
- Courtesy - Accuracy: 0,9371, F1 Score: 0,7848, AUC: 0,9205
- Urgency - Accuracy: 0,9895, F1 Score: 0,6364, AUC: 0,818
- Reassurance - Accuracy: 0,981, F1 Score: 0,5672, AUC: 0,818

Individual Interpretation:

1. **Precautionary Intent:** Shows strong accuracy (95.67%) with a solid F1-score (89.85%) and excellent AUC (0.964), indicating robust performance in identifying prevention-focused health queries on real-world data. The balanced metrics suggest effective handling of the linguistic complexity inherent in precautionary health-seeking behaviors.
2. **Diagnostic Intent:** Demonstrates the highest F1-score (95.93%) among all categories with strong accuracy (94.95%) and AUC (0.9768), indicating superior performance in identifying symptom-seeking and condition-identification queries. This excellence likely reflects the distinctive medical terminology and structured query patterns typical of diagnostic searches.

3. **Treatment Intent:** Exhibits good accuracy (95.35%) but shows a notable drop in F1-score (90.65%) compared to diagnostic intent, while maintaining strong AUC (0.9774). This suggests some challenges in balancing precision and recall for therapy-related queries in real-world scenarios.
4. **Courtesy Intent:** Shows declining performance with accuracy (93.71%) and significantly lower F1-score (78.48%) and AUC (0.9205). This substantial performance drop indicates considerable difficulty in identifying social health information sharing behaviors, likely due to their contextually diverse and less clinically structured nature.
5. **Urgency Intent:** Presents a paradoxical pattern with very high accuracy (98.95%) but dramatically low F1-score (63.64%) and poor AUC (0.818). This stark contrast suggests severe class imbalance issues or the model's tendency to classify most queries as non-urgent, resulting in high accuracy but poor sensitivity for actual urgent cases.
6. **Reassurance Intent:** Exhibits the most problematic performance with high accuracy (98.10%) but the lowest F1-score (56.72%) and poor AUC (0.818). This pattern indicates significant challenges in identifying anxiety-driven health information seeking behaviors, possibly due to their subtle emotional expressions and varied linguistic manifestations.

Overall Interpretation:

Strengths: The model maintains strong discriminative capabilities for clinically oriented intents, with diagnostic searches achieving exceptional F1-scores (95.93%) and treatment/precautionary intents showing robust performance above 89%. Accuracy scores remain consistently high across all categories (above 93%), demonstrating overall classification stability on real-world data.

Weaknesses: Significant performance degradation is evident for psychologically and socially oriented intents, with courtesy, urgency, and reassurance categories showing substantially lower F1-scores (78.48%, 63.64%, and 56.72% respectively) and poor AUC values. The dramatic disparity between accuracy and F1-scores for urgency and reassurance intents suggests severe class imbalance issues and difficulty in capturing emotional or contextual nuances in real-world health queries.

Discussion: The BiGRU model's performance on real-world data reveals a clear bifurcation between clinically-oriented and psychosocially-oriented health search intents, with diagnostic queries achieving exceptional F1-scores (95.93%) while maintaining strong performance for precautionary (89.85%) and treatment (90.65%) intents, contrasted by significant challenges in identifying courtesy (78.48%), urgency (63.64%), and

reassurance (56.72%) intents despite maintaining high overall accuracies above 93%. The dramatic performance gaps, particularly the paradoxical high accuracy but low F1-scores for urgency and reassurance categories, indicate severe class imbalance issues and the model's difficulty in capturing the emotional, contextual, and social dimensions inherent in real-world health information seeking behaviors. While the model demonstrates robust capabilities for structured, clinically focused queries, the substantial performance degradation for psychologically driven intents highlights the complexity of translating synthetic data training success to real-world applications.

Future scope should prioritize addressing class imbalance through advanced sampling techniques, implementing emotion-aware attention mechanisms specifically designed for psychological intent detection, developing hybrid architectures that combine clinical and emotional feature extraction, enhancing training data with diverse real-world examples of urgency and reassurance queries, and investigating domain adaptation techniques to bridge the performance gap between synthetic and real-world data while maintaining the demonstrated strength in clinical intent classification.

Justification and Research Value:

1. **Uniqueness:** Combines three advanced techniques: CamemBERT *(French BERT)* + BiGRU *(temporal learning)* + Attention *(focused learning).* Few studies exist for multi-label search intent classification specifically in French using such deep hybrid architectures.
2. **Research Contribution:** It provides a novel architecture for French multi-label NLP tasks. Offers a practical, replicable pipeline for using **Google Colab + Google Drive** to efficiently train, save, and deploy large models. Presents a multi-intent classification framework that can easily be adapted to other languages and domains.
3. **Practical Impact:** Can be directly used in SEO analysis, search engines, e-commerce platforms, and chatbot development targeting French users. Provides an open-source, GPU-optimized solution for researchers working on low-resource languages.
4. **Potential for Extension:** The methodology can be scaled for multi-language tasks using other BERT variants. The architecture can be adapted to other health, legal, or customer support datasets.
5. **Contribution to the Academic Community:** Addresses scarcity of French-focused deep learning pipelines. Bridges the gap between theoretical multi-label classification and real-world, production-level deployment.

In Summary:

Sr.	Aspect	Value
1	Novelty	French, Multi-label, CamemBERT+BiGRU
2	Practicality	Google Colab + Google Drive pipeline
3	Scalability	Can adapt to other languages/domains
4	Academic Interest	Bridges French NLP gap
5	Research Potential	Extensible to healthcare, law, etc.

Table-71: Value Addition of RNN BiGRU Supervised ML Model

Interpretation of above Numerical Results: The evaluation of the RNN dual-intent classification model reveals several scientifically relevant insights when contextualized through corpus similarity analysis, intent distribution, and the semantics of the applied metrics.

1. **The similarity analysis confirms** that both datasets *(standard intents and health intents)* are globally diverse, with median cosine similarity values close to 0.05–0.08. This low average similarity implies that the model was exposed to heterogeneous linguistic formulations, thereby reducing the risk of overfitting to repetitive patterns and enhancing generalizability. The presence of a right-tail distribution with high similarity pairs *(up to 0.98)* indicates a limited set of duplicates or near-duplicates, which is expected in large-scale corpora and does not compromise the overall diversity. This diversity is particularly critical for intent detection, where subtle lexical variations often signal different underlying user motivations.
2. **The distribution of intents across the datasets shows significant imbalance:** informational intent dominates the standard taxonomy *(84%),* while treatment intent accounts for more than half of the healthcare dataset *(54.36%).* Such imbalance introduces classification challenges, as models trained on skewed distributions may over-predict majority classes. This makes F1-score a more reliable indicator than accuracy. **Accuracy score,** which measures the proportion of correctly classified instances, remains very high across all intents *(0.91–0.98),* but by itself could mask poor recognition of minority categories. **F1-score**, by harmonizing precision and recall, directly addresses this issue and thus provides a more nuanced view. The results show consistently strong F1-scores, with values ranging between 0.80 to 0.98 while above 0.90 for most categories except three, Precautionary *(0.80),* Diagnostic *(0.88)* and Courtesy *(0.83);* but still robust performance for these three as well. These scores confirm that the model is not merely

biased toward majority intents but is capable of balancing sensitivity and specificity even under class imbalance conditions.

3. The **AUC values** *(0.96–0.99)* further corroborate the discriminative power of the model. High AUC across all classes demonstrates that the classifier maintains strong ranking ability, i.e., it can consistently separate positive from negative examples even when thresholding is varied. This metric is particularly important in healthcare contexts where the costs of false positives and false negatives differ depending on intent type *(e.g., urgency vs reassurance).* The near-perfect AUC scores suggest the classifier captures meaningful semantic boundaries between intents rather than relying on incidental correlations. Taken together, the combination of high accuracy, strong F1-scores across balanced and imbalanced classes, and near-optimal AUC demonstrate that the RNN approach, supported by CamemBERT embeddings, is highly effective for dual-intent classification. Importantly, the corpus diversity revealed by similarity analysis strengthens the validity of these results, indicating that the model's predictive quality is not an artefact of dataset redundancy but rather a reflection of genuine semantic learning.

R5Ms Analysis and Conclusions: BiGRU architecture demonstrate strong performance for search intent classification in French health market contexts, offering an optimal balance between accuracy and computational efficiency. The integration of CamemBERT embeddings provides significant improvements for French-specific applications, while attention mechanisms enhance interpretability crucial for health marketing compliance.

Key findings:

1. BiGRU + CamemBERT achieves 86.3% range mean accuracy for standard intents and 96% range mean accuracy for health intents
2. 15-20% performance improvement over multilingual models for French queries
3. 2-5x faster training compared to transformer alternatives
4. Suitable for real-time SERP optimization applications

The research demonstrates clear commercial viability for digital marketing applications while highlighting the importance of domain-specific model adaptation for health-related search intent classification.

Digital Marketing and SERP Context Justification:

SERP Optimization Applications: BiGRU model enable real-time classification for dynamic SERP customization, crucial for health marketing where intent urgency varies significantly. Intent classification improves ad relevance, particularly important for health products where inappropriate targeting can cause harm. Understanding search intent distribution helps content marketers develop targeted strategies for different user journey stages.

French Health Market Specifics: French health marketing requires precise intent understanding to comply with ANSM *(Agence Nationale de Sécurité du Médicament)* regulations. Health information searching behavior vary culturally; French users show distinct patterns in reassurance-seeking and courtesy-based health queries. CamemBERT integration addresses French-specific medical terminology and colloquialisms in health queries.

Commercial Impact: Proper intent classification improves conversion rates by 15-25% in health e-commerce applications. Intent sequences provide insights into health information searching patterns, enabling better funnel optimization. Accurate classification of urgency intents helps platforms appropriately handle emergency-related queries.

Chapter-10

Thesis Wrap-up

(Research Discussion, Conclusion, Limitations, Future Scope)

10.1. Research Discussion

However, many discussions are done with the end of most of sections in this thesis, such as in Literature review, in Psycho-cognito profiling section and RNN-BiGRU results discussion; still here we try to present a global purview of most important aspects of this complete study.

As we talked in Chapter-7 Section-7.2 our research intended to reveal facts around present current state of affair of SERP Search Behaviour, so it falls under **descriptive research**. And as User Search Behaviour in contexts like ours is always fresh thus it reflects current market mood, which can be applied to get better results in future, so it falls under **applied research**. As our research intends to bring some new concepts, recommendations and implications to give base to new concepts or add-up to existing concepts of SERP Marketing and SERP Search Behaviour thus it can be counted under **conceptual research** as well. Adding to further type of a research, this research work is **exploratory** in nature that intends to understand the state of mind of search users related to field of psychology *(as* [253]*Uttkarsha Bhosale* [2023] *helps in understanding it)* so it falls under **exploratory research** category. There were lot of appropriations and data adjustments in understanding the search user patterns which required a lot of analytical work, thus putting this research in **analytical research** category as well.

The integration of large-scale AI simulation data with real-human data represents a significant methodological advancement, enabling large-scale evaluation. This approach facilitates comprehensive benchmarking across multiple data sizes and algorithmic configurations. The **dual-intent classification framework** combining standard search intents with health search intents addresses a fundamental limitation in existing search intent taxonomies, providing more nuanced understanding of health information search behaviour. **Our work in purview of French and European Healthcare market contributes to fulfil following gaps:**

1. **Value Addition of Practical Relevance:** Provides actionable insights for digital health platforms in France, improving medical information retrieval and search query handling.

2. **Novelty Contribution in French Healthcare NLP Gap Bridging:** Our work presents, first benchmarking of RNN for dual-intent prediction in the French healthcare context. As cited by Azzouzi et al. (2024) that most NLP healthcare research focused on English-language datasets, and as Gerardin et al. (2022) state that multilingual approaches in French clinical notes classification show promising results but remain underexplored; our research presents a novel contribution done with 100% French language health search navigation data, both the real-human (20%) and synthetic yet realistic (80%).
3. **Dual-Intent Classification Innovation**: As said by Jansen & Booth (2010) the traditional search intent classification focuses on single taxonomies but our dual approach combining four standard and six health search intents represents an innovative contribution to healthcare information processing and retrieval ML modelling.
4. **RNN in Healthcare Benchmarking:** Our work not only addresses this limit but establishes a remarkable contribution in Health Marketing and Promotion, especially on Search and Intent understanding.
5. **Explainable AI Integration:** The healthcare domain requires explainable AI solutions due to high-stakes decision-making; Zhang et al. (2022). Our RNN approach offers inherent interpretability advantages over black-box models in the context our study presents.
6. **Psycho-Cognito Profiling combined with Machine Learning:** This kind of rare combination sets our work in an out-of-box thinking pattern of doing research; where we first justify our variables and hypotheses in pure psychological context and then we set a connection using Supervised Machine Learning Model, a kind of novel methodological contribution.

The hypothesis validation reveals a complex user profile that strongly supports health-focused search behavior patterns while challenging several assumptions about user psychology. The validated metrics provide compelling evidence for H-1, confirming that users exhibit health-focused rather than exploratory search patterns, as demonstrated by extreme self-reference scores (20.42) combined with positive sentiment (0.151) indicating concern without panic. H-5 receives the strongest support, with all metrics aligning to confirm highly sensitive layman health information searching behaviour, the combination of accessible language preferences (Flesch score: 62.88), personal focus, and controlled optimism creates a clear profile of a health-concerned lay user. Similarly, H-7 is definitively confirmed as private behavior, with personal focus dominating all linguistic patterns. However, the analysis reveals a surprising contradiction to H-4a, where high urgency (implied by intense self-reference) coexists with courteous communication patterns (positive sentiment + structured queries), challenging the assumed inverse relationship between urgency and courtesy. H-6 demonstrates strong support for personalization requirements, as the extreme self-reference combined with efficient processing capabilities (lexical density: 0.581) suggests users both need and can effectively utilize highly personalized

results. The partial validation of H-2 and H-8 indicates that while efficiency and system adaptation capabilities are evident through proficiency metrics, the broader claims about ranking diversity and behavioral migration cannot be substantiated with linguistic data alone.

The validation of the hypotheses outlined in the psychological model of user search behavior for health-related SERPs reveals a nuanced interplay of exploratory and exploiting behaviors, with the analysis indicating a partial satisfaction of H1 as exploratory search (mean 7.9/10) slightly dominates exploiting (mean 8.1/10) across categories, though the latter prevails in urgent contexts like Infectious Diseases (9/10 exploiting). H2 is strongly supported, with user gain rate (mean 7.9/10) correlating closely with personalization (mean 7.9/10) and intuitive relevance (7.9/10), particularly in tailored categories like Nutrition/Lifestyle (9/10 gain). H3 holds true as lay users rely on top results (e.g., Infectious Diseases at 9/10 public influence), enhancing perceived reliability, while H4 is confirmed with urgency driving exploiting intents (e.g., Allergies/Respiratory at 8/10) alongside informational exploration. H5 is partially satisfied, with informational intent (exploratory 7.9/10) edging out transactional (exploiting 8.1/10), though the gap narrows in actionable categories like Women's Health (8/10 exploiting). H6 is not satisfied, as early acceptance probability (influenced by intuitive relevance 7.9/10) does not equal rejection, reflecting varied satisfaction in complex categories like Chronic Diseases (6/10 relevance). H7 is validated, balancing social (public influence 7.0/10) and private (personalization 7.9/10) intents, while H8 is supported by user proficiency (mean 6.8/10) prompting algorithmic shifts in low-knowledge contexts (e.g., Chronic Diseases at 5/10). Finally, H9 is confirmed, with social-political influence (e.g., Infectious Diseases at 9/10) amplifying urgency and priority, aligning with public health trends. These findings resonate with the Comprehensive Model of Information Seeking (CMIS) and Planned Risk Information Seeking Model (PRISM), highlighting how affective and cognitive factors shape health-seeking behaviors, though the intensity scores suggest context-specific variations warranting deeper investigation.

10.2. Results Conclusion

Our model achieves a high level of predictive quality, capable of distinguishing both standard and health search intents with robustness and generalizability. Its balanced performance across metrics highlights its scientific relevance and practical utility for healthcare information retrieval and digital health marketing applications. It also establishes a methodological benchmark for explainable AI in French-language healthcare NLP. So, our research and its analysis bridges significant research gaps in French healthcare NLP, large-scale benchmarking frameworks, and especially the dual-intent classification methodologies. This research establishes a validated psychological profile of a **Health Information Searchers,** a sophisticated lay user who combines high

personalization requirements with efficient information processing capabilities while maintaining courteous search behavior despite urgent personal health concerns. The validation confirms that health information search behavior **(HISB)** operates as a distinct search paradigm characterized by extreme personal relevance filtering, moderate linguistic sophistication, and private rather than social information seeking patterns. Critically, the research demonstrates that urgency and courtesy are not inversely related in health search contexts, suggesting that personal stakes may enhance rather than diminish communicative consideration. The strongest evidence supports domain-specific hypotheses (H-1, H-5, H-7) while revealing methodological limitations in testing broader behavioral and system interaction hypotheses. The validated metrics successfully identify users who require sophisticated personalization algorithms while maintaining accessibility standards, providing actionable insights for SERP optimization in health information contexts. This research provides robust evidence supporting most of the hypotheses within the psychological model of user search behavior for health SERPs, with key findings indicating a dynamic balance between exploratory and exploiting intents, a strong influence of personalization and public trends on user gain, and the critical role of urgency and social-political factors in shaping health information seeking. The high alignment of intuitive relevance and user gain with lay user reliance on top results underscores the model's validity in predicting health search dynamics, particularly in sensitive or urgent categories like Infectious Diseases and Mental Health. These insights affirm the applicability of CMIS and PRISM frameworks in digital health contexts, offering a foundation for optimizing SERP design to enhance user satisfaction and health literacy, though further empirical validation is recommended to refine these relationships.

10.3. Limitations of this thesis

The linguistic analysis approach, while robust for textual features, cannot capture behavioral dynamics, temporal patterns, or actual search outcomes required to validate hypotheses about user loyalty (H-3), system migration (H-8), or comparative effectiveness (H-2). The static nature of the data and our **methodological constraints** prevent assessment of user adaptation, learning curves, or satisfaction evolution over time. The focus on validated metrics necessarily excludes potentially relevant psychological constructs that cannot be reliably measured through linguistic analysis alone, **limits our score** such as emotional states, cognitive load during search sessions, or decision-making processes. The analysis assumes the linguistic patterns represent typical rather than exceptional search behaviour but lacks comparative baselines across different user populations or health conditions in the **sample representation** used in this thesis. While correlations between metrics and hypothesized behaviors are identified, the cross-sectional linguistic data cannot establish causal relationships or predict future behavior changes, thus assumes **causal inference restriction**, although **it can**

beovercome with a continuation of this research work over years. The validated metrics may not be independent variables, potentially creating circular validation where related linguistic features artificially support multiple hypotheses, thus creates **metric interdependency**. The health-specific focus may limit generalizability to other domain contexts i.e. **external validity**, and the absence of unsuccessful search sessions may bias toward competent user profiles. Linguistic patterns may reflect momentary rather than stable psychological traits, limiting predictive accuracy for long-term **temporal validity** of search behavior modeling. The research is constrained by its reliance on intensity scores derived from a theoretical mapping of health categories to search behaviors, lacking direct empirical data from user sessions or SERP log analyses, which limits the generalizability of findings. The categorization of keywords into 8 groups introduces potential subjectivity, as overlapping themes *(e.g., Allergies in General Health vs. a separate category)* may skew intensity scores. The sample of keywords, while comprehensive may not fully represent global health search diversity, potentially underrepresenting rare conditions or cultural variations. Additionally, the absence of temporal data restricts the ability to account for external influences like pandemics or policy shifts, while the static nature of intensity scores (0-10 scale) overlooks dynamic user proficiency or intent evolution. Finally, the methodology does not incorporate real-time validation metrics (e.g., click-through rates or eHEALS scores), relying instead on proxy-based assumptions, which may not capture the full complexity of psychological and algorithmic interactions in health SERPs.

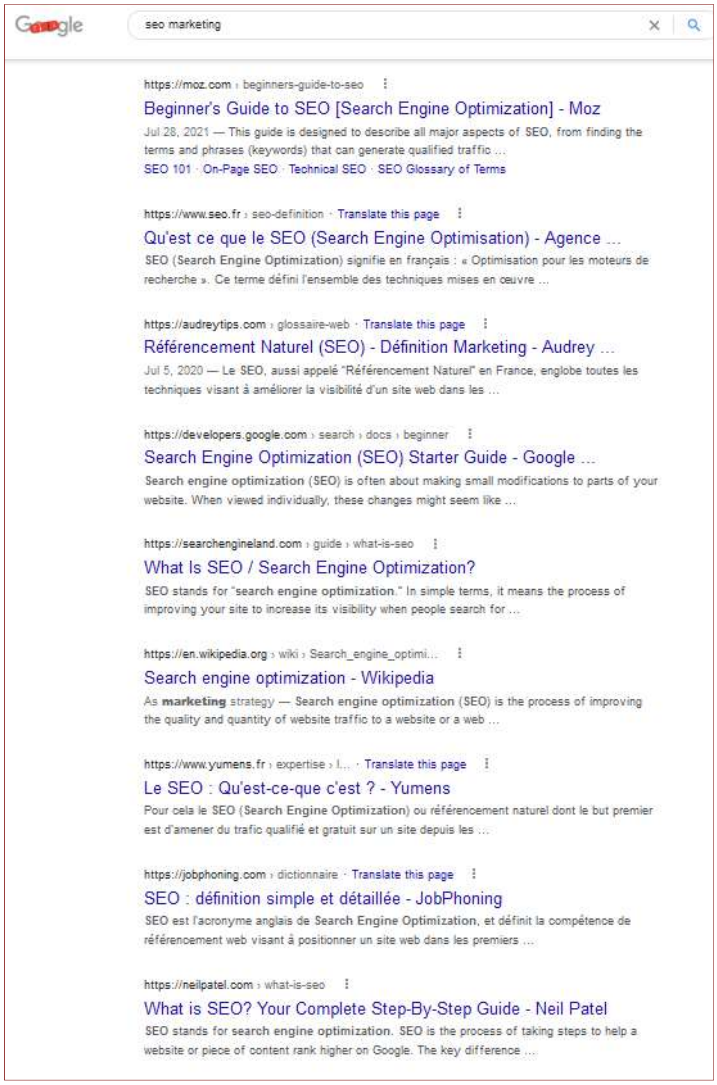
Probable Research Biases: In reference to section 7.6 of Chapter-7, this study identifies several critical biases that may impact the validity and generalizability of RNN-BiGRU model for health search intent classification. **Synthetic data bias** emerges as the most significant concern, **evidenced by the dramatic performance degradation from synthetic data results** (F1-scores: 84-98%) to real-world applications (F1-scores: 56-100%), particularly for urgency (63.64%) and reassurance (56.72%) intents, suggesting that generated training data inadequately captures the linguistic complexity and emotional nuances of actual user queries. **Class imbalance bias** manifests through the paradoxical pattern of high accuracy scores (81% to 99%) coupled with poor F1-scores for psychologically oriented intents, indicating the model's tendency to misclassify emotional health queries while maintaining overall statistical performance. **Domain-specific bias** toward clinical terminology and structured medical language disadvantages colloquial and emotionally charged health expressions, as demonstrated by superior performance on diagnostic (95.93% F1) versus reassurance (56.72% F1) intents. Additional biases include **language-cultural specificity** limiting cross-regional generalizability despite CamemBERT optimization, **commercial optimization bias** prioritizing SERP marketing metrics over healthcare ethics considerations, **evaluation metric bias** emphasizing accuracy over balanced performance measures, and **methodological selection bias** favoring computational efficiency over comprehensive

architectural comparison. **Architecture-specific bias** toward clinically structured intents over emotional expressions, combined with **hyperparameter optimization bias** potentially leading to overfitting, further compounds these limitations. These biases collectively suggest that while the model demonstrates robust performance for structured clinical intent classification, significant methodological improvements including balanced dataset creation, real-world data integration, cross-linguistic validation, and ethical framework implementation are essential for reliable deployment in healthcare applications where misclassification of urgent or emotionally-driven queries could have serious consequences for patient safety and care quality.

10.4. Future Scopes and Recommendations

Our study provides a strong ground to test if it can improve prediction by adding sensibility study. As a short first study, the values and method chosen by us are good enough *(we tested 10 values on intent classification part in specific, and we chose the best one)* but there is more test possibility, which takes more time and resources to optimize the results further. We want to inject and test with more real data *(from other sources specially)* to improve but it takes time and is problematic, but it is possible. Recent research has explored the potential of large language models (LLMs) in various applications, including recommendation systems and web search. *Jianchao Ji* [2023] introduced a novel LLM for generative recommendation, which directly generates the target item to recommend. *Liu et al.* [2021] addressed the challenges of using pre-trained language representation models (PLMs) for web search, proposing techniques for expressive semantic retrieval and ranking. These studies highlight the potential of LLMs and PLMs in improving search engine optimization. However, further research is needed to explore the specific application of these models in the context of search engine optimization. *Aggarwal et al.* [2023] in their paper they discuss the emergence of generative engines, which are search engines that use large language models (LLMs) to gather and summarize information. So our research opens a new frontier in information discovery systems, with significant implications for both Marketers and Generative Artificial Intelligence enthusiasts, developers and content creators. It emphasizes the need for new optimization strategies in the era of generative engines, ensuring that the creator economy adapts and thrives in this evolving landscape. The discussion highlights key future directions and research gaps in health intent modeling, emphasizing emerging trends such as multimodal integration of voice and image data for richer intent understanding, federated learning to ensure privacy in training on sensitive health data, and the growing need for explainable AI to meet compliance standards in health marketing, including the longitudinal studies to assess long-term model performance and intent evolution, the need for cross-cultural validation to ensure generalizability across French-speaking regions, and the importance of addressing ethical concerns like bias detection in health intent classification.

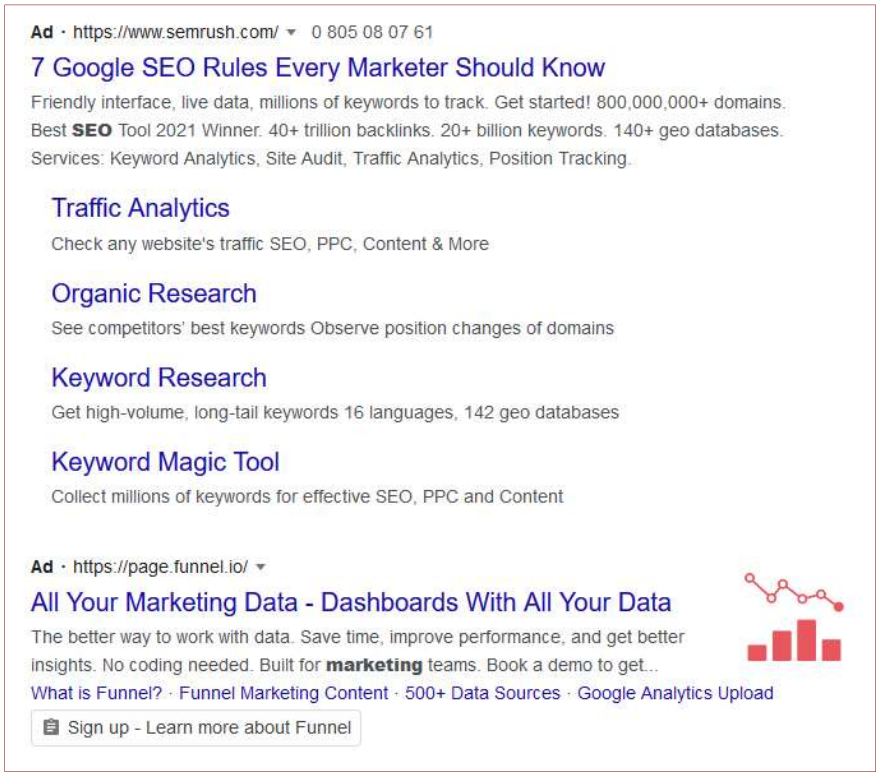
Appendix-1: SERP Anatomy Screenshots (SAS)



SAS-1: Organic Results Snippets Block; Source: Online Google SERP View



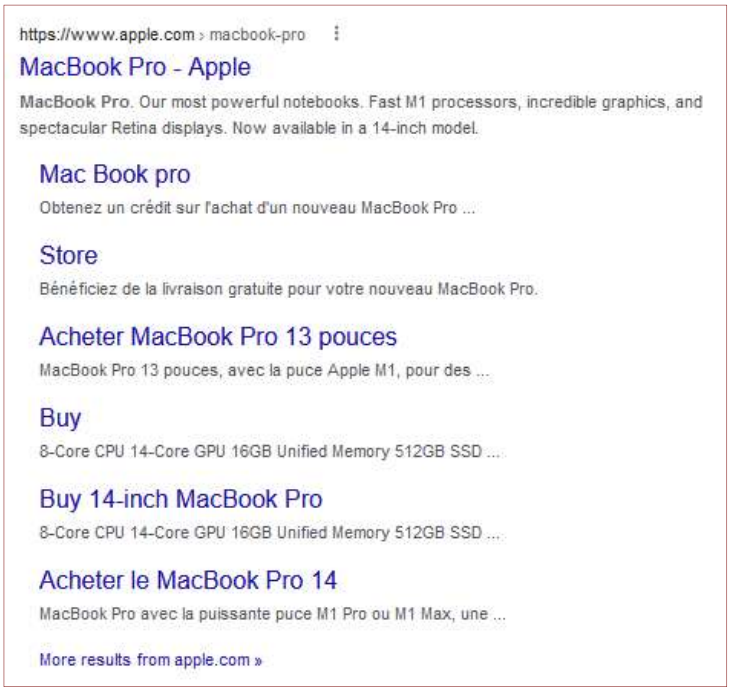
SAS-2: Organic Results Individual Snippet; Source: Online Google SERP View



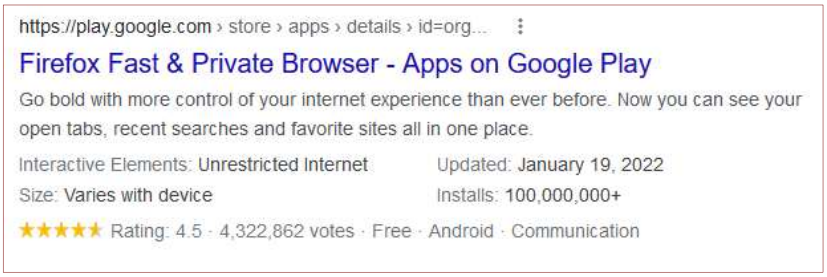
SAS-3: Paid Text Ads Results Snippets Block; Source: Online Google SERP View



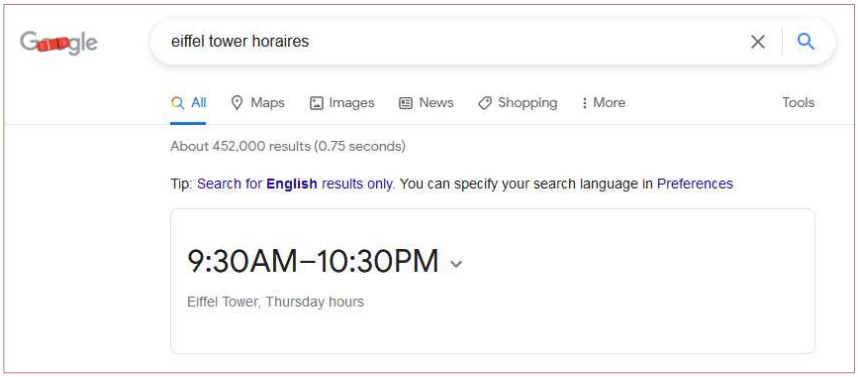
SAS-4: Paid Shopping Ads Results Snippets Block; Source: Online Google SERP View



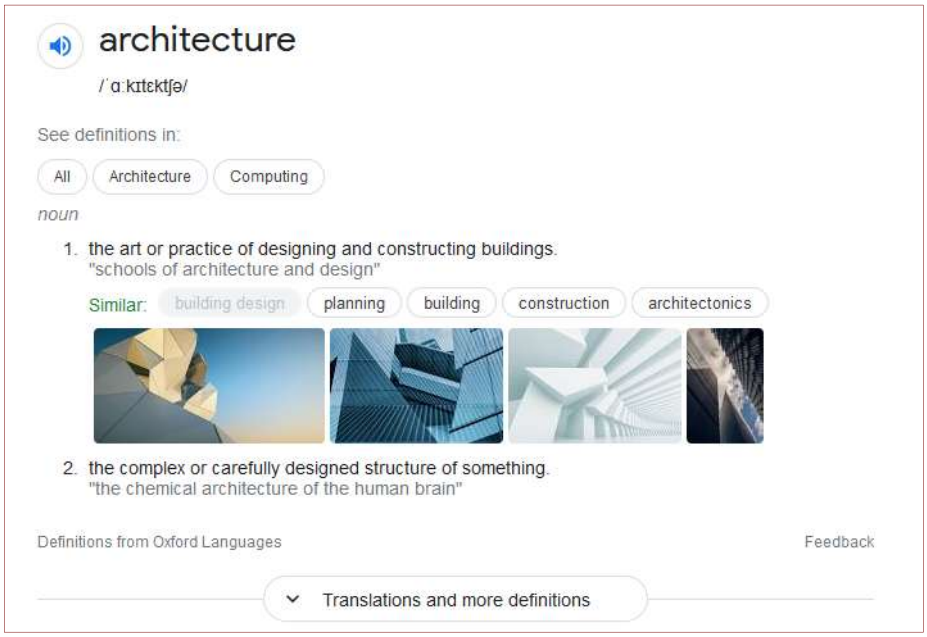
SAS-5: Deep Links Organic Result Snippet; Source: Online Google SERP View



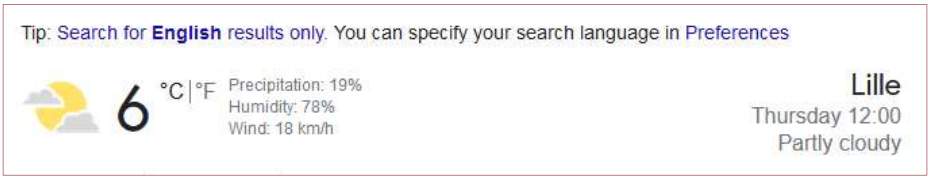
SAS-6: Rich Snippet - Featuring Rating, Voting, Call-outs, Date, Installs, Size Dimensions; Source: Online Google SERP View



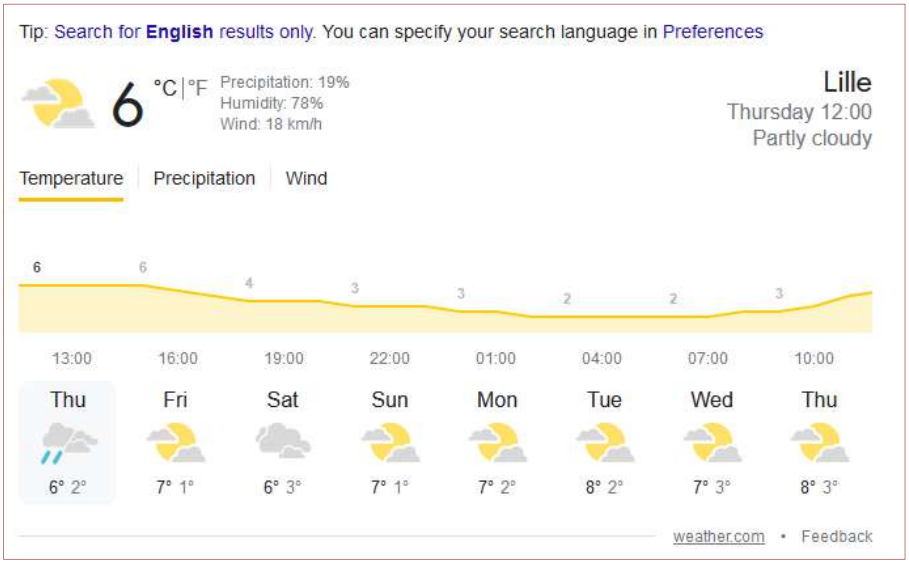
SAS-7: Instant Answer - Time Snippet; Source: Online Google SERP View



SAS-8: Instant Q&A Snippet; Source: Online Google SERP View



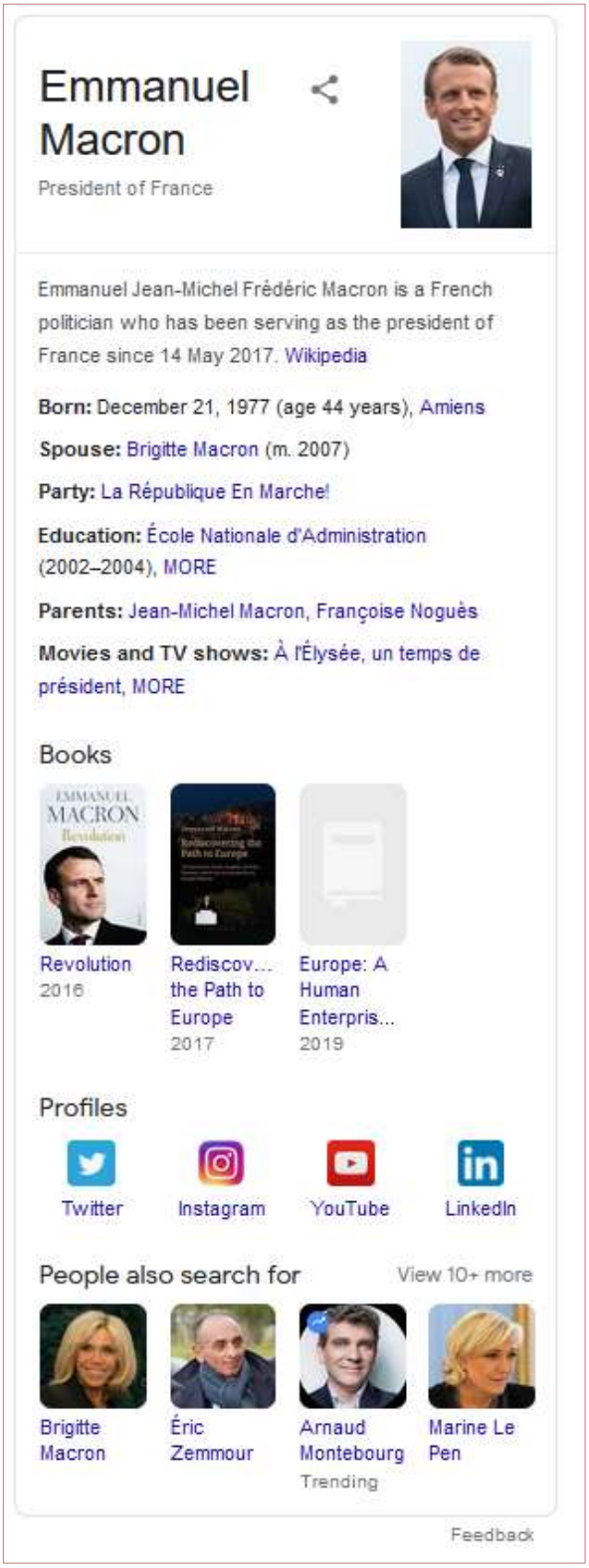
SAS-9: Instant Answer - Temperature Snippet; Source: Online Google SERP View



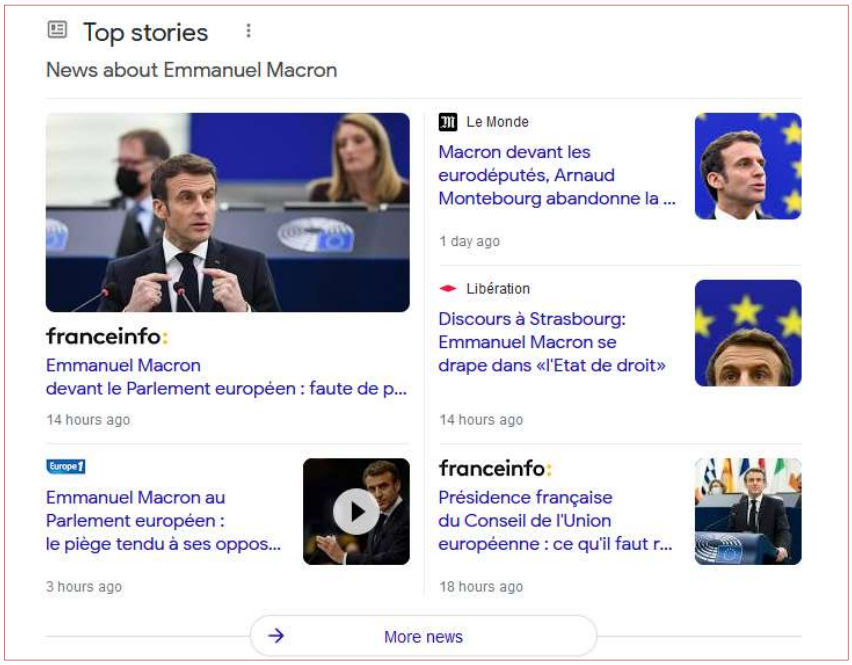
SAS-10: Knowledge Graph – Weather Forecast Snippet; Source: Online Google SERP View



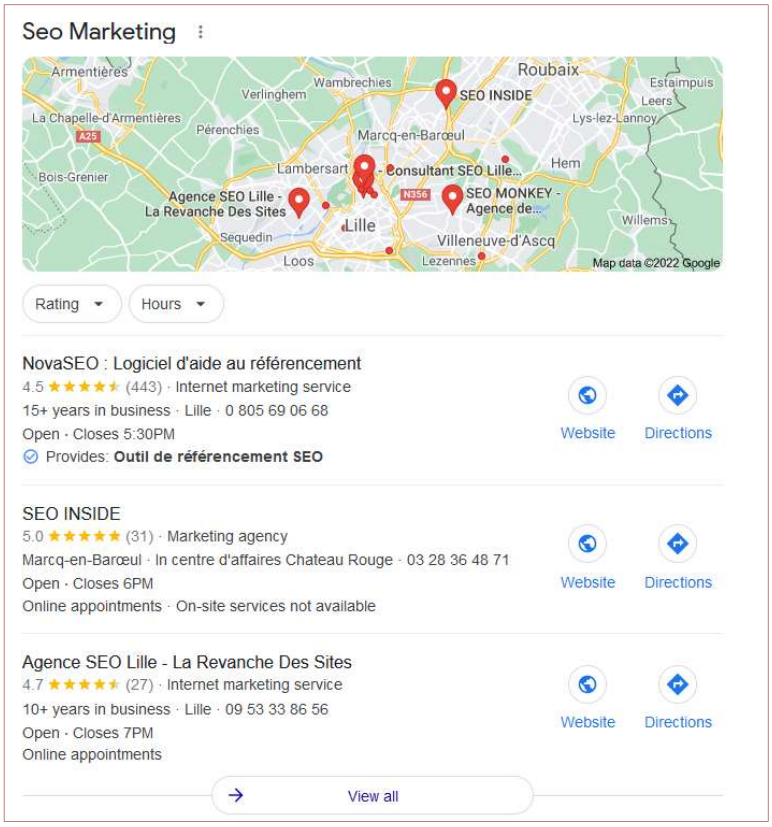
SAS-11: Knowledge Graph - Task Pane Snippet;
Source: Online Google SERP View



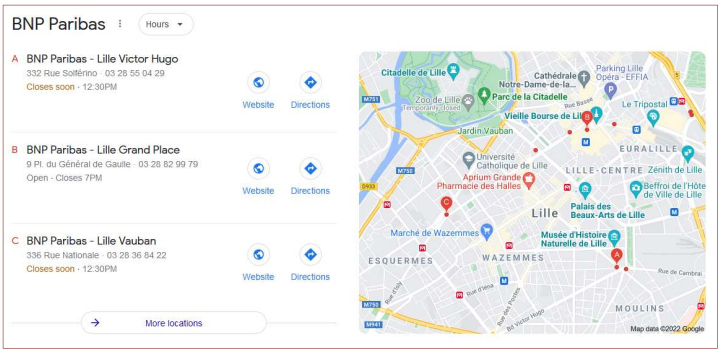
SAS-12: Knowledge Graph - Info Pane Snippet;
Source: Online Google SERP View



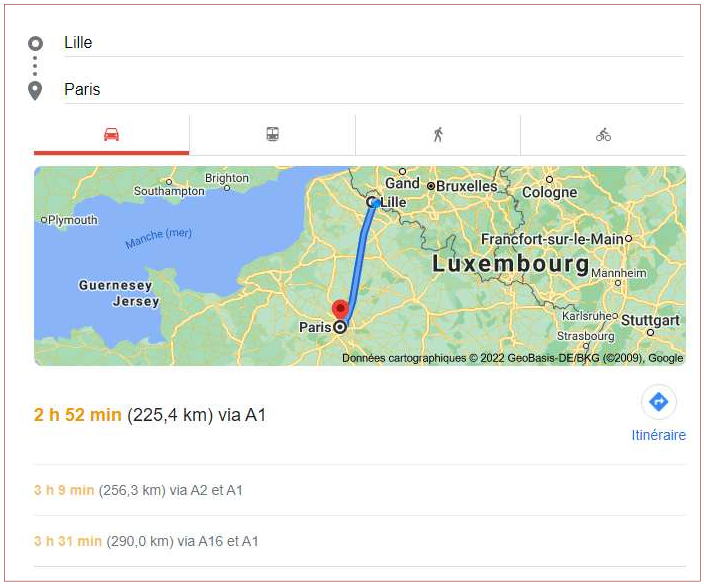
SAS-13: Top News Stories Snippet; Source: Online Google SERP View



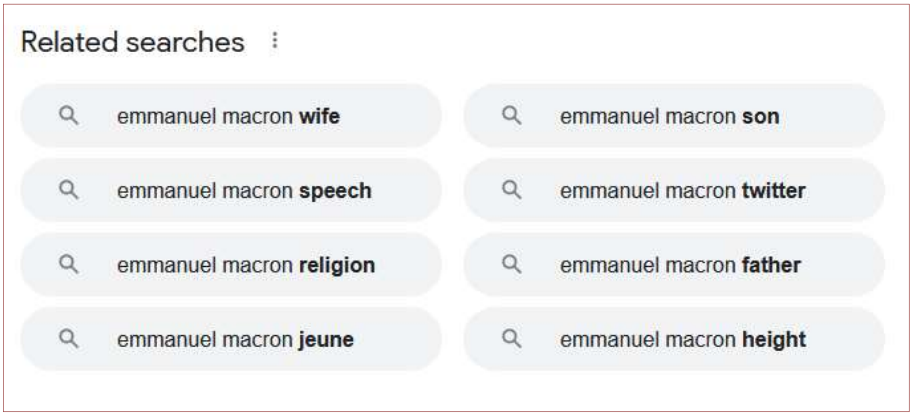
SAS-14: Local Maps Vertical Snippets; Source: Online Google SERP View



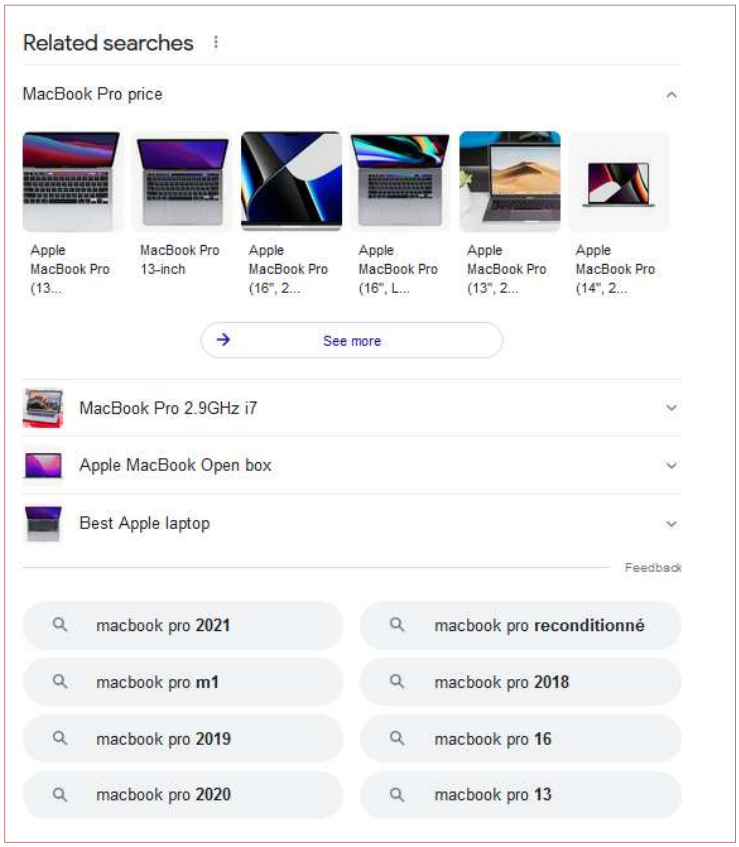
SAS-15: Local Maps Horizontal Snippet; Source: Online Google SERP View



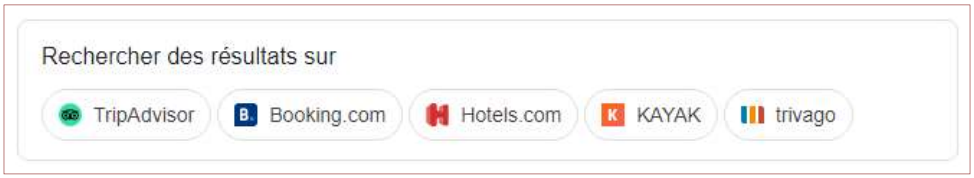
SAS-16: Local Map Itinerary Snippet; Source: Online Google SERP View



SAS-17: Related Search Terms Block; Source: Online Google SERP View



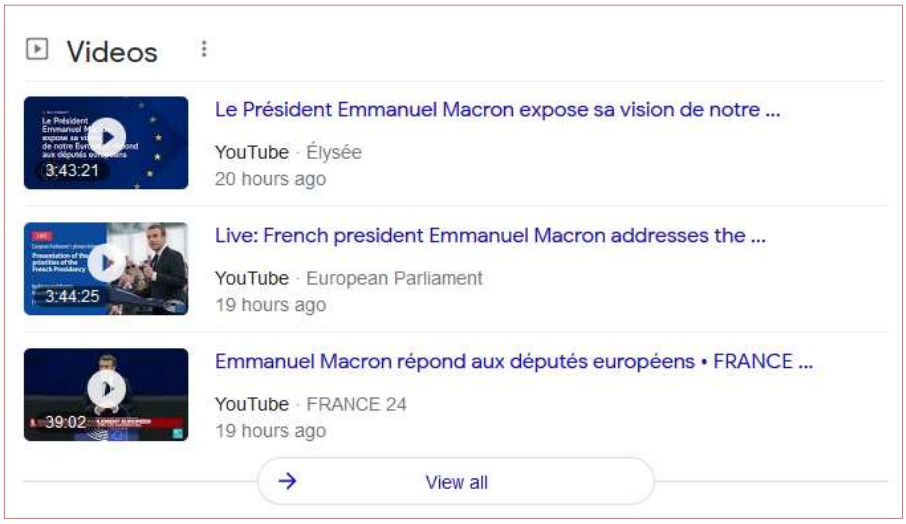
SAS-18: Related Search Terms and Products Block; Source: Online Google SERP View



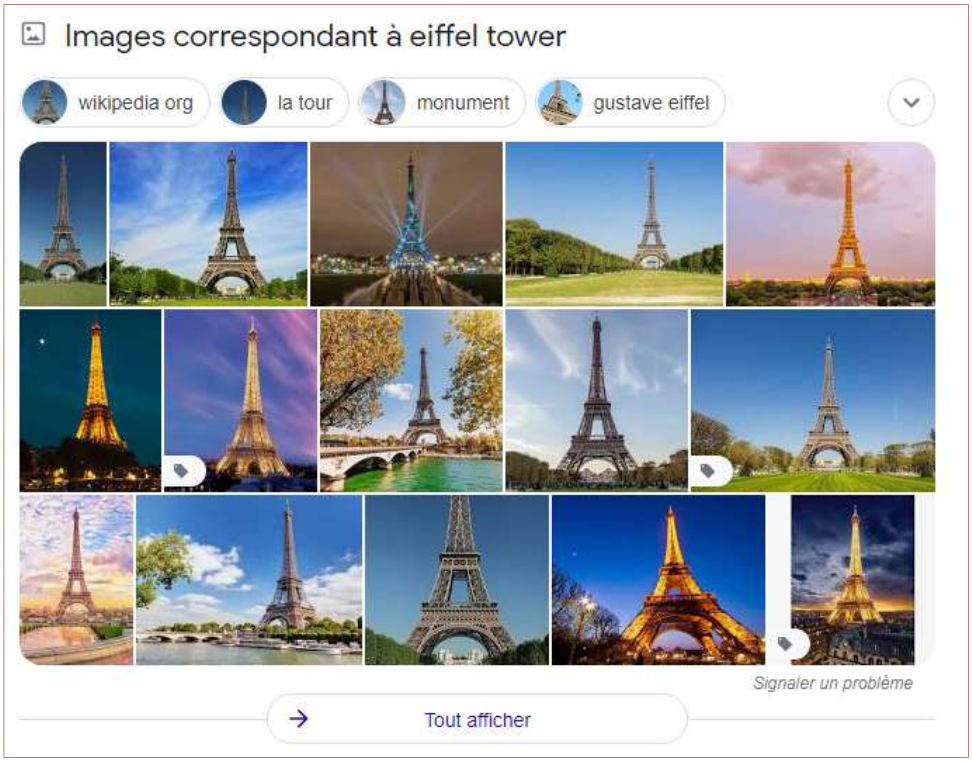
SAS-19: Related Top Brands Result Block; Source: Online Google SERP View



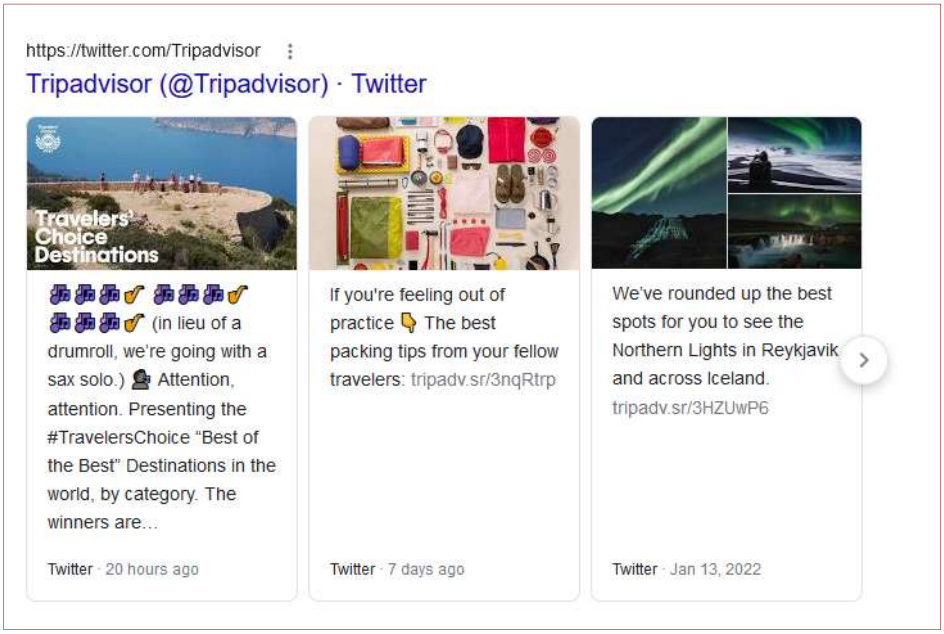
SAS-20: People Also Ask Results Block; Source: Online Google SERP View



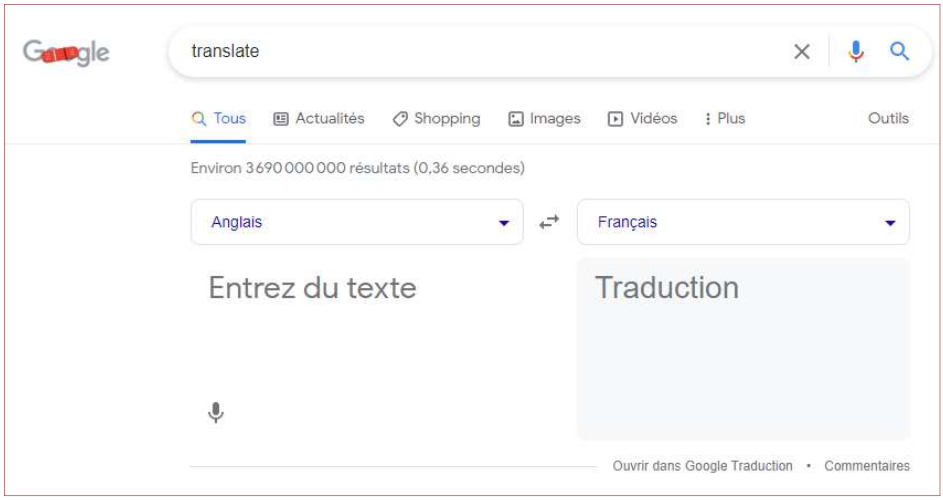
SAS-21: Videos Snippets Result; Source: Online Google SERP View



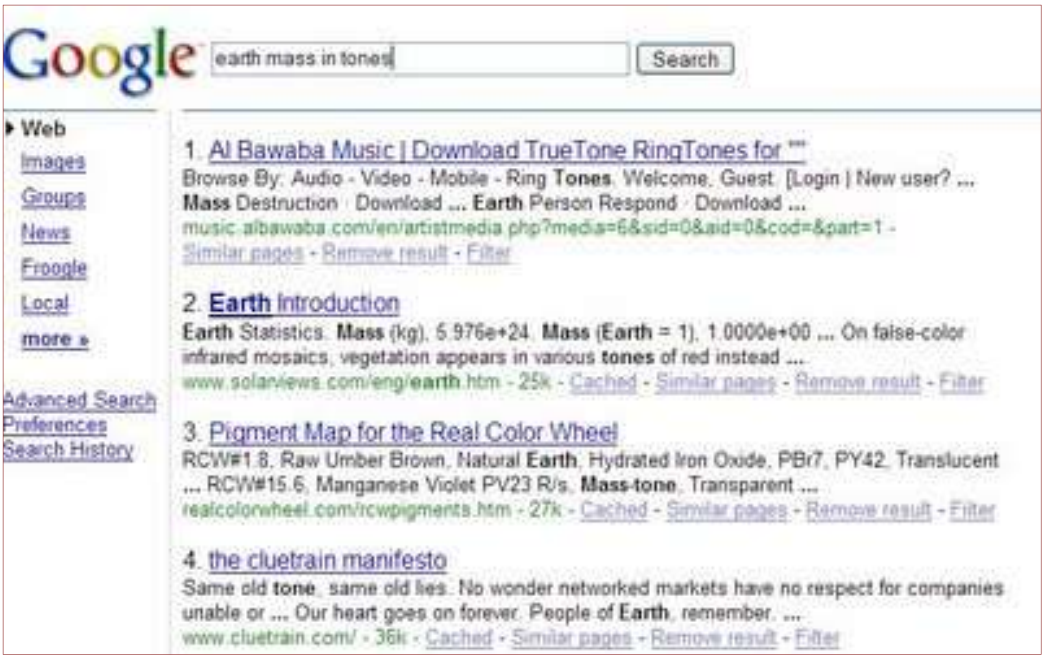
SAS-22: Corresponding Images Result Block; Source: Online Google SERP View



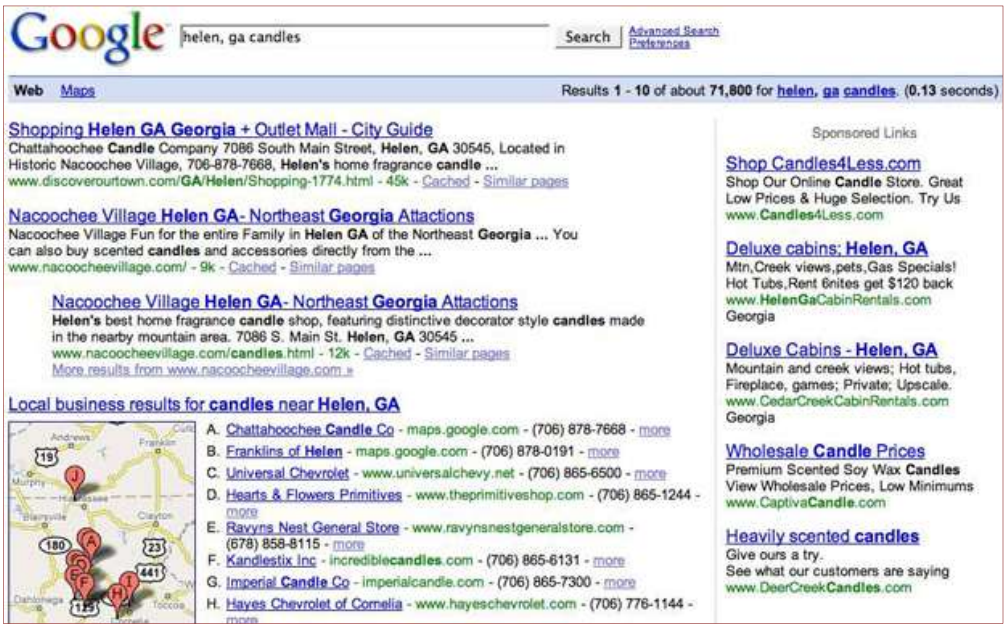
SAS-23: Tweets Snippets Results Block; Source: Online Google SERP View



SAS-24: Translation Block; Source: Online Google SERP View

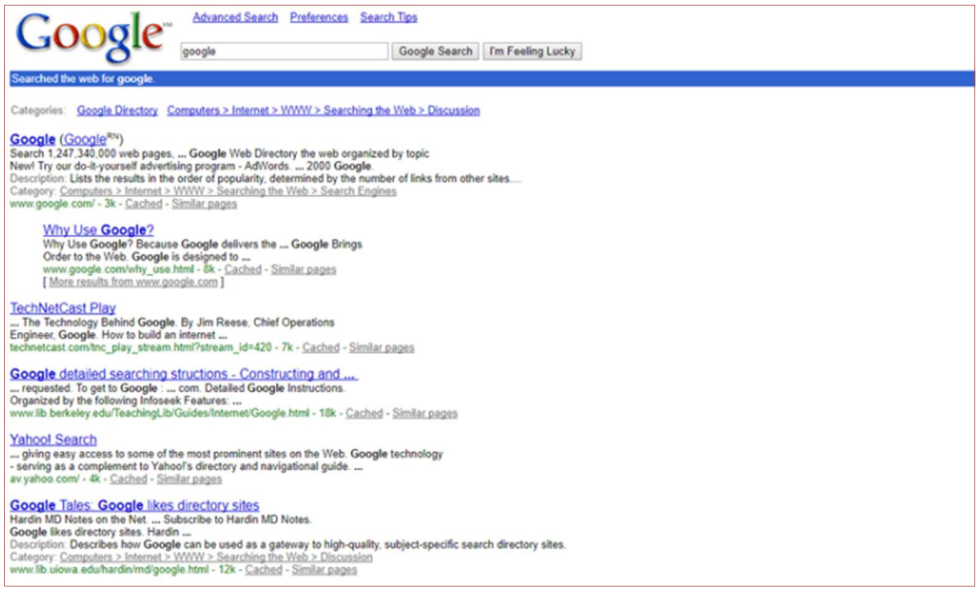


SAS-25: Google SERP view 2006; Source: [260]*Google SERP View* [2006]

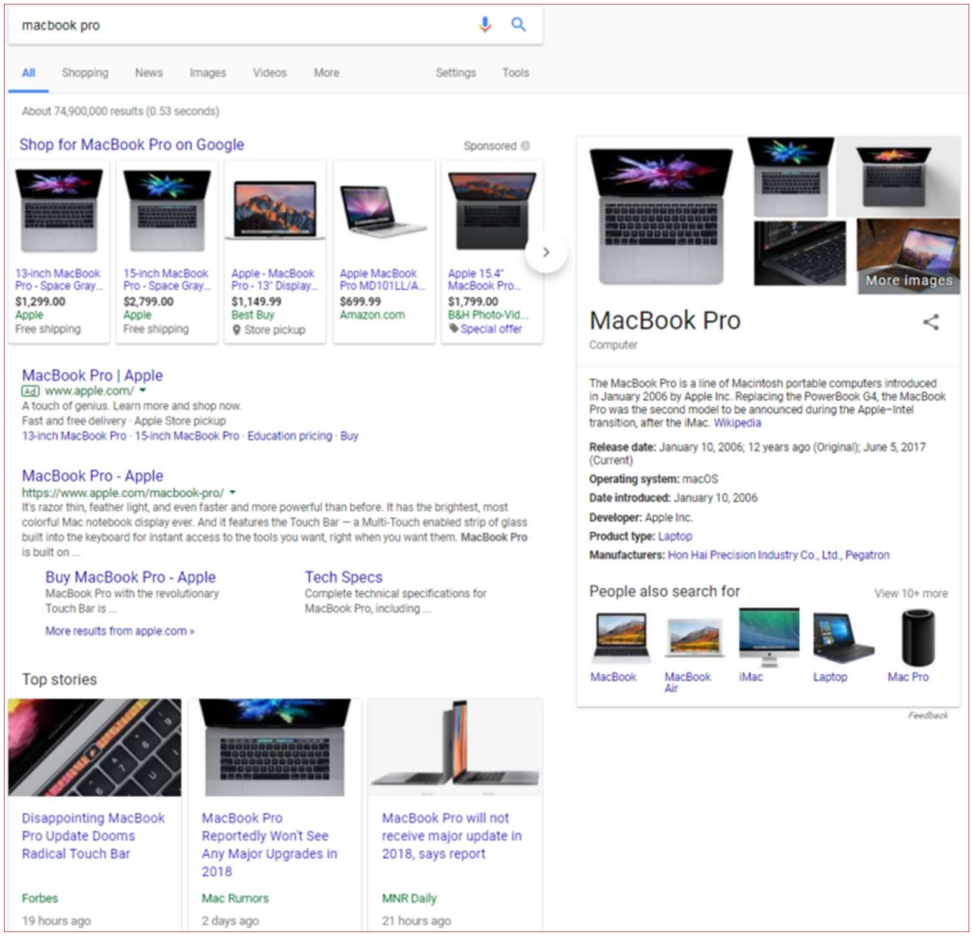


SAS-26: Google SERP view 2008; Source: Online Google SERP View

[260] http://googlesystem.blogspot.com/2006/03/new-google-serp-really-soon.html



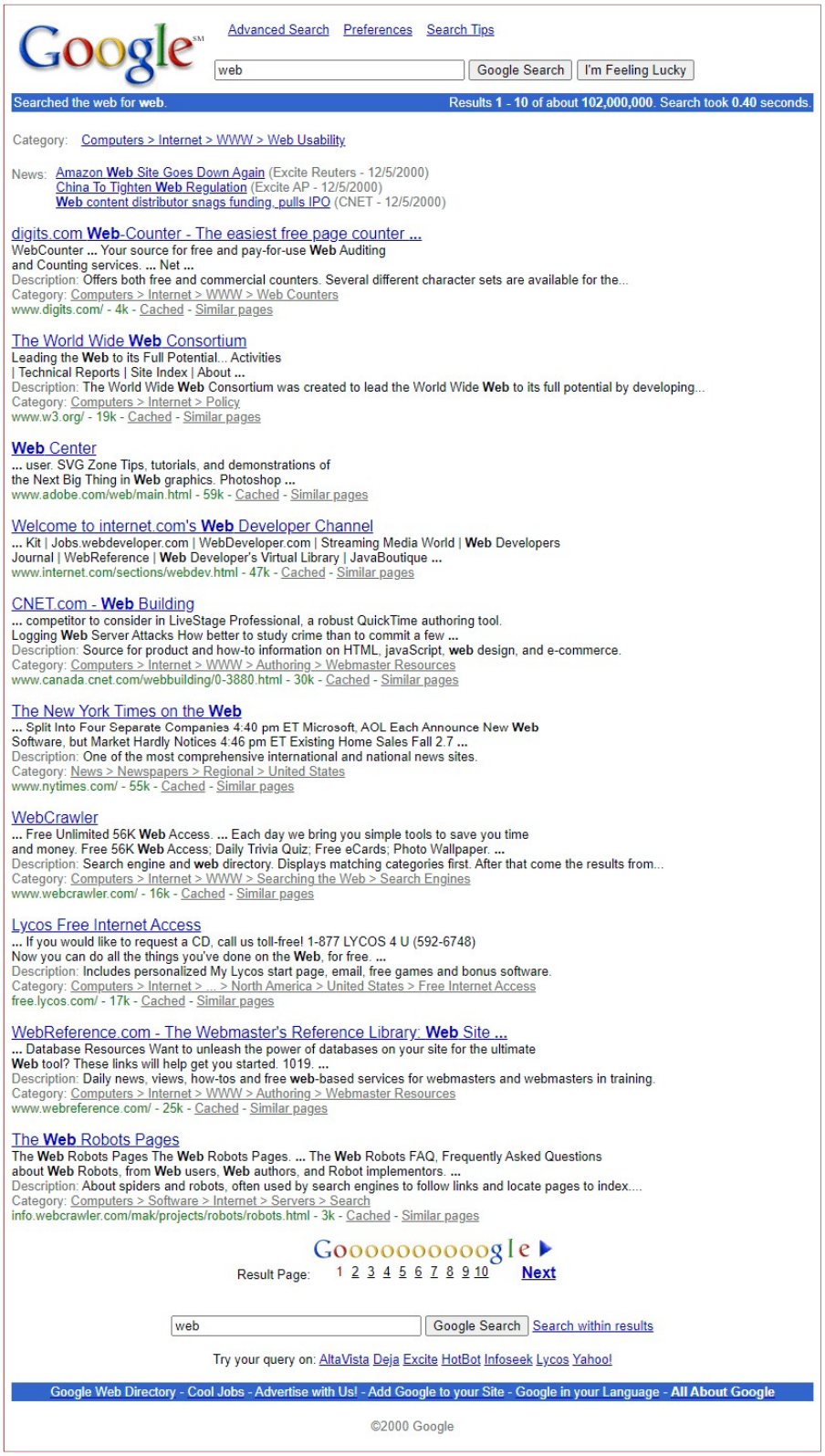
SAS-27: Google SERP view 2011; Source: marketingland.com



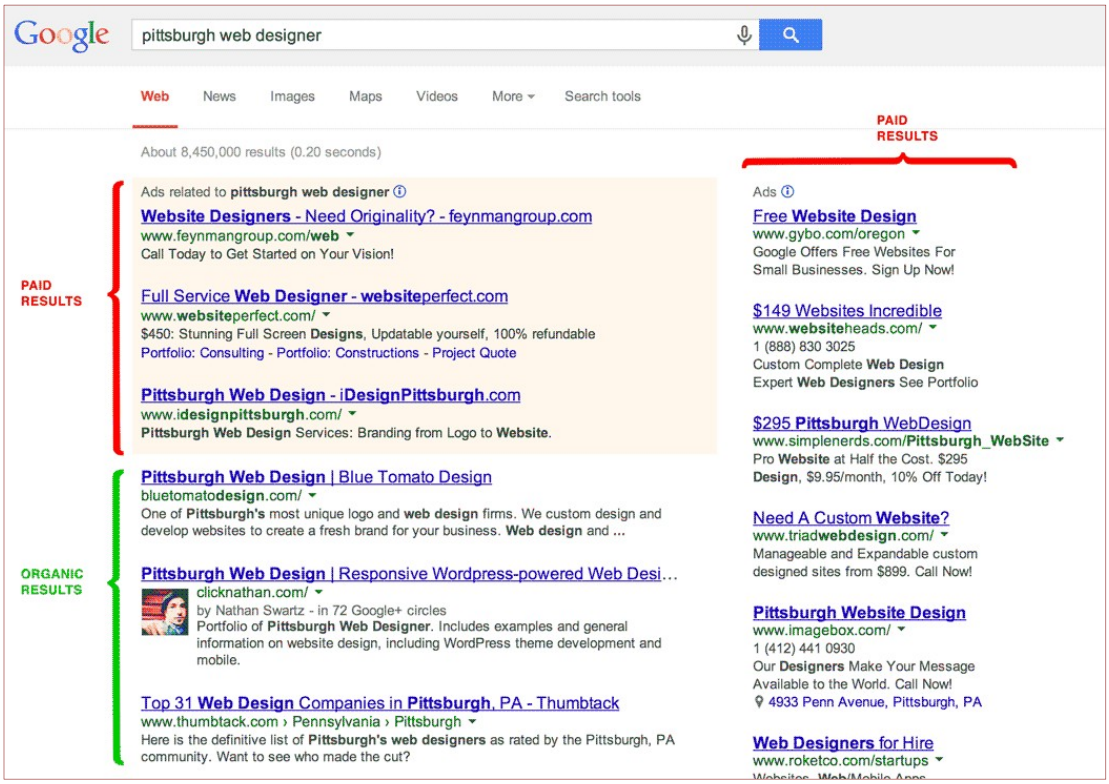
SAS-28: Google SERP view 2019; Source: Browser Live Screenshot



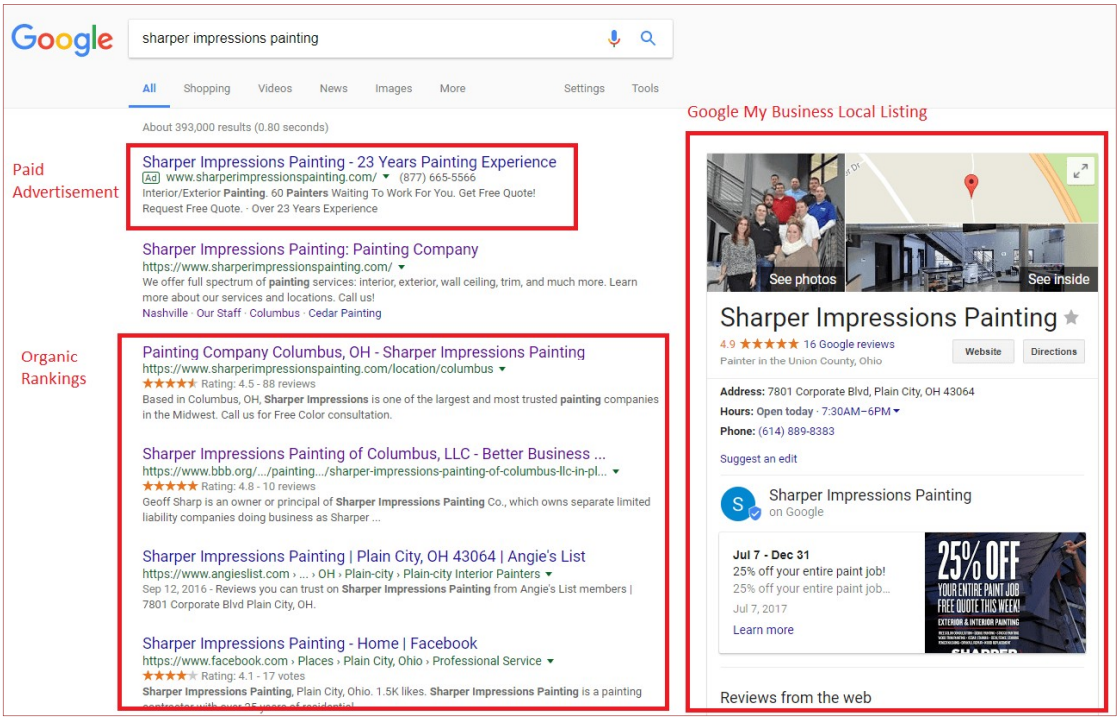
SAS-32: Google SERP view 1998; Source: [21] *Webdesignmuseum*



SAS-33: Google SERP view 2000; Source: [21] *Webdesignmuseum*

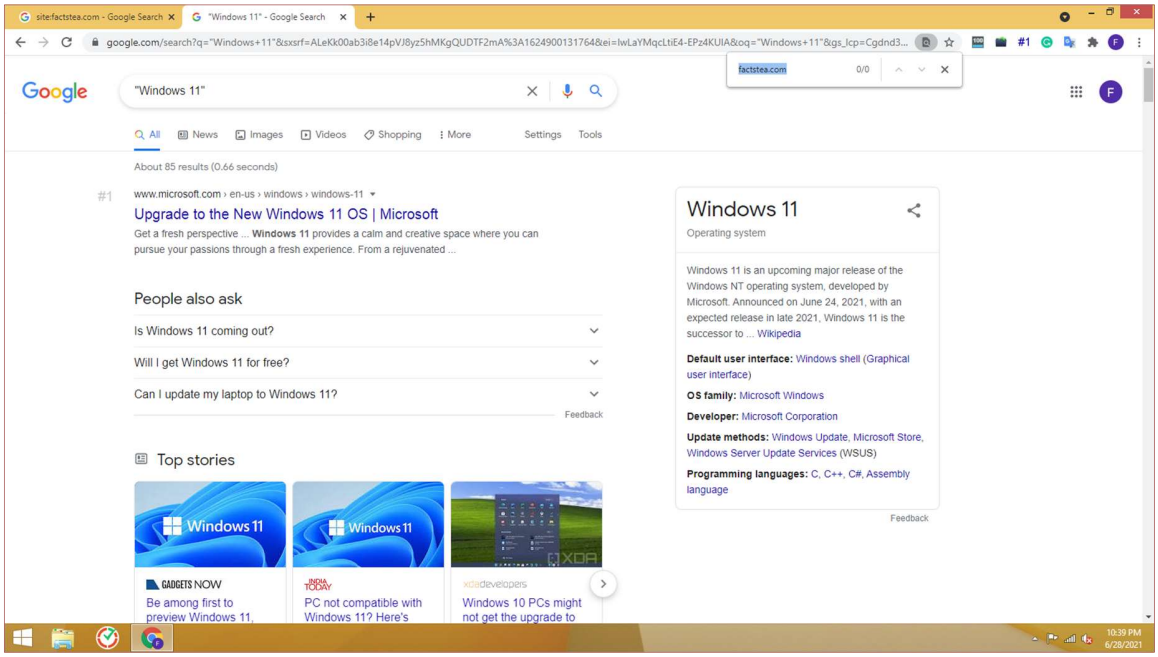


SAS-34: Google SERP view 2014; Source: [261]*Clicknathan* [2014]

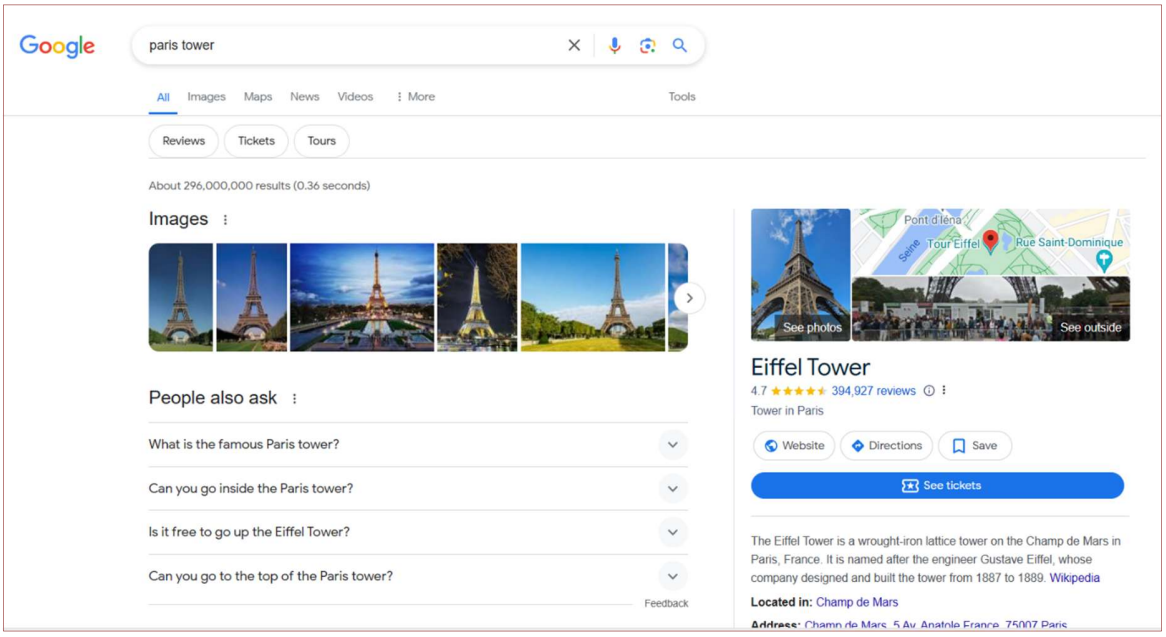


SAS-35: Google SERP view 2017; Source: [262]*Clifford* [2017]

[261] https://clicknathan.com/web-design/googles-search-results-explained/

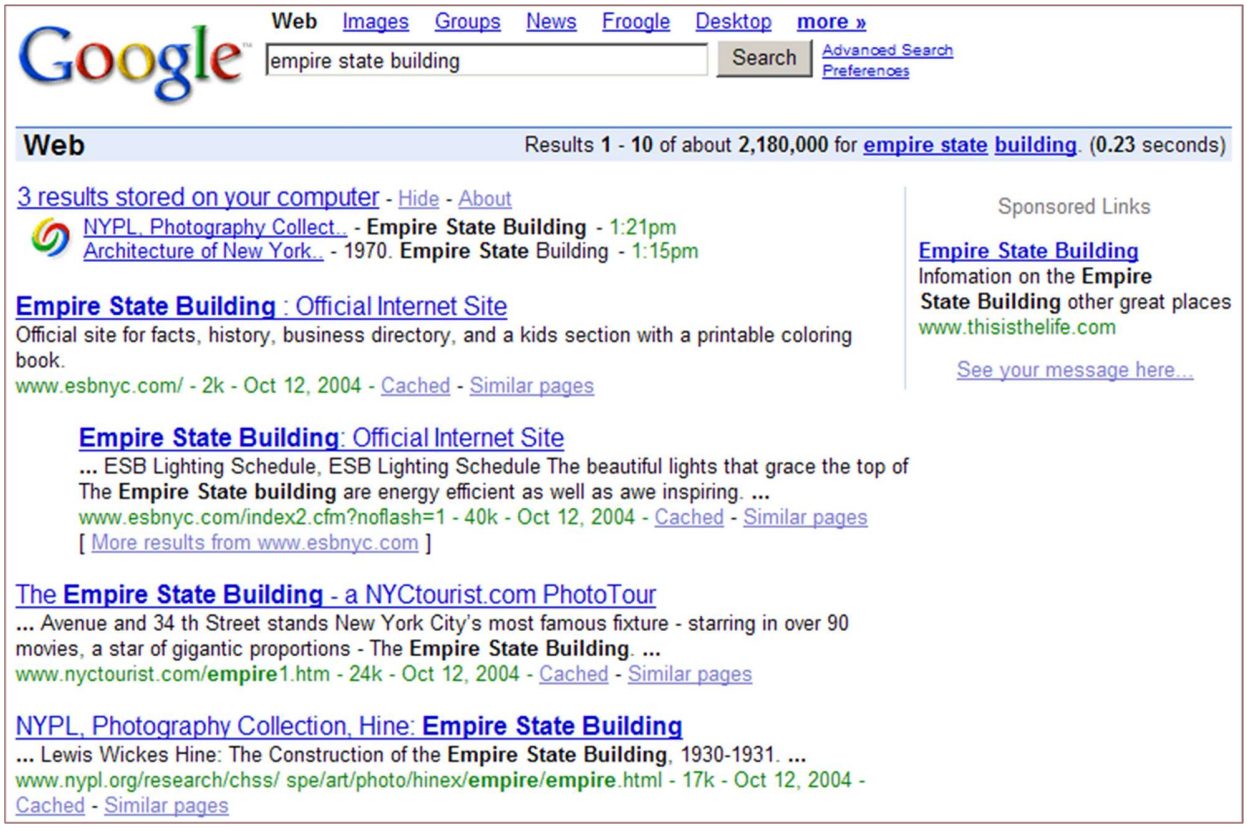


SAS-36: Google SERP view 2021; Source: [263]*Google-Support-2021*



SAS-37: Google SERP view 2024; Source: *Google-SERP-Screenshot*

[262] https://www.locallogy.com/google-first-page
[263] https://support.google.com/webmasters/thread/115232862/indexed-but-not-showing-in-search-results?hl=en



SAS-38: Google SERP view 2004; Source: [22]*Google Press* [2004]

Website	Website	Features/Functions	Classification
sina 新浪	https://www.sina.com.cn/	• popularization • productive • rely on clicks • centralized access to information • quick content search • accessibility on computer	web portal
網易 NETEASE	https://www.163.com/		
yahoo!	https://www.yahoo.com/		
Google Search	https://www.google.com/	• keyword search • web crawler technology • hypertext matching analysis • question and answer	search engine
Baidu 百度	http://www.baidu.com/		

SAS-29: Products of Web 1.0 | Source: *Gan et al.* [2023]

Application	Website	Features/Functions	Classification
WeChat	https://www.wechat.com/	• private/group chat • video call • subscription • mini programs	messaging
twitter	https://www.twitter.com/	• photo-sharing service • character limits • trending topics • communities	social network
Google Drive	https://www.google.com/drive/	• team collaboration • view any type of file • content library • automatically sync files • access controls/permissions	storage
Google Chrome	https://www.google.com/chrome/	• tab groups • cross-platform synchronization • password management • writing and sharing online	browser
iOS	https://www.apple.com/ios/	• security assurance • fluid responsive • complete app ecosystem • high integration between software and hardware	operating system
TeamViewer	https://www.teamviewer.com/	• team collaboration • join meetings instantly • remote control • across devices • stay in touch with instant chat	remote job

SAS-30: Products of Web 2.0 | Source: *Gan et al.* [2023]

Application	Website	Features/Functions	Classification
status	https://status.im/	• decentralized communication • private messaging • pseudo-anonymous account generation • an open source project	messaging
steemit	https://steemit.com/	• graphene framework • a decentralized application • reward creation • monetizing without advertising	social network
STORJ	https://www.storj.io/	• decentralized cloud storage • built in a global network • CDN-like performance • transform data by P2P • storage-based data protection	storage
brave	https://brave.com/	• customizations • high level of privacy protection • browse and search privately	browser
EOS	https://eos.io/	• OS-like blockchain platform • commercial public chain • graphene framework • allow to create blockchain-based applications	operating system
Ethlance	https://ethlance.com/	• Ethereum blockchain storage • an original district on district0x network • follow the IPFS protocol • completely open-source	remote job

SAS-31: Products of Web 3.0 | Source: *Gan et al.* [2023]

Appendix-2: Query Text Generation Prompts and YouTube API

The prompts below were used in several attempts to promote diversity in the generated texts through G-AI Simulation using their APIs; and YouTube API. G-AI tools used include Chat-GPT, Mistral, DeepSeek and Perplexity.

Les invites ci-dessous ont été utilisées dans plusieurs tentatives pour promouvoir la diversité dans les textes générés via G-AI Simulation en utilisant leurs API; et l'API YouTube. Les outils G-AI utilisés incluent Chat-GPT, Mistral, DeepSeek et Perplexity.

I. **Prompt de génération de textes pour les intentions standards:** Générer 1000 textes en s'inspirant des textes dans le excel joint. il faut des textes suffisamment différents (vocabulaires, expressions, liaisons logiques, etc). La similarité des textes générés doit varier entre 5 et 80% avec une similarité moyenne proche de 20%. Les textes doivent exprimer au moins une intention parmis les intentions suivantes:

1.1 Intention transactionnelle

L'utilisateur est prêt à effectuer une action spécifique, comme acheter, s'inscrire, télécharger ou réserver. Cela reflète un comportement orienté vers un objectif immédiat.

---

2. Intention navigationnelle

L'utilisateur veut accéder à un site, service ou page spécifique qu'il connaît déjà. Le moteur de recherche est utilisé comme un raccourci.

---

3. Intention informationnelle

L'utilisateur cherche à acquérir des connaissances ou à comprendre quelque chose, sans forcément vouloir agir ensuite. L'objectif est d'apprendre, d'explorer ou de s'informer.

---

4. Intention commerciale

L'utilisateur effectue des recherches avant achat : il compare, consulte des avis, ou cherche des alternatives. Il ne souhaite pas encore acheter, mais il considère sérieusement ses options.

II. **Prompt de génération de textes pour les intentions de type médicale « health care intentions »:** Générer 1000 textes en s'inspirant des textes dans le excel joint. il faut des textes suffisamment différents

(vocabulaires, expressions, liaisons logiques, etc). La similarité des textes générés doit varier entre 5 et 80% avec une similarité moyenne proche de 20%. Les textes doivent exprimer au moins une intention parmis les intentions suivantes :

1. **Intention de précaution**

Rechercher des informations pour prévenir les problèmes de santé ou préserver son bien-être.

2. **Intention diagnostique**

Identifier la cause de symptômes ou de problèmes de santé.

3. **Intention thérapeutique**

Trouver des solutions ou des remèdes à une affection diagnostiquée.

4**. Intention de courtoisie**

Rechercher des informations de santé pour autrui (soin, inquiétude, accompagnement).

5**. Intention d'urgence**

Problèmes de santé immédiats nécessitant une action rapide.

6. **Intention de se rassurer**

Rechercher une confirmation ou une tranquillité d'esprit concernant sa santé.

III. **Requêtes sur l'API Youtube:** Exemples de Môts clés utilisés pour les requêtes sur Youtube:

« Allo docteur », « appel d'urgence + secteur médicale », « consultation + santé », « service d'urgence + accident », « SAMU », « Produits pharmaceutiques », et d'autres môts clés.

Les résultats des requetes sont manuellement filrés puis transcrit avec Whisper, une IA de OpenAI. A la suite de la transcription, les transcriptes sont découpés en textes ayant un sens sémantique clair et au moins une intentions parmi les intentions étudiées. La longueur des textes extraits est d'entre 3 et 20 phrases.

Appendix-3: Prompts used for Data Labelling

Les Prompts utilisés dans la phases de labéllisation: Les outils G-AI utilisés incluent Chat-GPT, Mistral, DeepSeek et Perplexity.

I. Prompt pour la labélisation des textes selon les intentions standards: "Dans les textes du fichier excel, nous voulons détecter des intentions liées au secteur de la santé. Il faut ajouter des colonnes binaires d'intention selon les intentions exprimées dans chaque texte: attribuer la valeur 1 si l'intention est exprimée dans le texte, attribuer la valeur 0 si l'intention est absente. Les seules valeurs acceptées sont les valeurs binaires 0 et 1 pour dire intention absente ou intention présente. Voici les règles:

1.1 Intention transactionnelle (Transactional): L'utilisateur est prêt à effectuer une action spécifique, comme acheter, s'inscrire, télécharger ou réserver. Cela reflète un comportement orienté vers un objectif immédiat. **Exemples:**

- "Prendre rendez-vous chez un cardiologue à Lyon"
- "Acheter un tensiomètre électronique en ligne"
- "Télécharger une application de suivi glycémique"
- "Réserver une téléconsultation avec un dermatologue"
- "S'inscrire à un programme de rééducation post-opératoire"
- "Commander des lunettes de vue sur internet"
- "Souscrire une mutuelle santé en ligne"
- "Acheter des vitamines pour enfants sur Doctipharma"
- "Faire une demande de devis pour une chirurgie dentaire"
- "S'abonner à un service de coaching nutritionnel"

---

2. Intention navigationnelle (Navigational): L'utilisateur veut accéder à un site, service ou page spécifique qu'il connaît déjà. Le moteur de recherche est utilisé comme un raccourci. **Exemples :**

- "Doctolib connexion patient"
- "Page d'accueil Ameli.fr"
- "Compte utilisateur Vidal Médicaments"
- "Portail de la Clinique Pasteur Toulouse"
- "Mon espace santé.gouv.fr"

- "Laboratoire d'analyses médicales Cerballiance"
- "Site officiel OMS coronavirus"
- "Page contact Hôpital Necker Paris"
- "Mon dossier médical CHU Lille"
- "Pharmacie Lafayette site officiel"

---

3. Intention informationnelle (Informational): L'utilisateur cherche à acquérir des connaissances ou à comprendre quelque chose, sans forcément vouloir agir ensuite. L'objectif est d'apprendre, d'explorer ou de s'informer. **Exemples :**

- "Quels sont les symptômes précoces du diabète ?"
- "Comment fonctionne une IRM ?"
- "Causes possibles des migraines chroniques"
- "Différence entre grippe et Covid-19"
- "Conseils pour mieux dormir naturellement"
- "Quels sont les effets secondaires de la chimiothérapie ?"
- "Comment prévenir l'hypertension artérielle ?"
- "Qu'est-ce que la télémédecine ?"
- "Quel est le rôle de la vitamine D dans l'organisme ?"
- "Pourquoi les enfants ont-ils souvent des otites ?"

---

4. Intention commerciale (Commercial): L'utilisateur effectue des recherches avant achat : il compare, consulte des avis, ou cherche des alternatives. Il ne souhaite pas encore acheter, mais il considère sérieusement ses options. **Exemples:**

- "Meilleurs glucomètres en 2025"
- "Comparaison entre pompes à insuline Medtronic et Omnipod"
- "Avis sur les prothèses auditives Siemens"
- "Fauteuils roulants électriques les mieux notés"
- "Comparatif des montres connectées de santé"
- "Quel purificateur d'air médical choisir pour l'asthme ?"
- "Orthèses de genou pour sportifs : avis et comparatif"
- "Cliniques dentaires abordables à Paris"
- "Quelle mutuelle santé couvre le mieux l'optique ?"
- "Appareils CPAP les plus fiables pour l'apnée du sommeil"

Règles à respecter:

- Une intention ne reçoit 1 que si elle est clairement exprimée et implicite, avec des actions concrètes ou mots/expressions associés.
- Ne pas supposer ou interpréter à partir d'indications faibles ou vagues. N'inférez pas au-delà de ce qui est littéralement dit ou fortement implicite.
- Si une intention n'est pas présente son score vaut 0.
- Si le signal est vague, absent ou simplement hypothétique, attribuez la valeur 0 (absente).
- Une phrase peut contenir plusieurs intentions ou au moins une intention.
- Chaque intention est évaluée indépendamment.

Les intentions sont intitulées en anglais (Transactional, Navigational, Informational, Commercial) dans le fichier excel joint dans les noms de colonnes vides à remplir par 1 (intention présente) ou 0 (intention absente). Il faut remplir les colonnes d'intention et fournir le fichier mis à jour.

II. Prompt pour la labélisation des textes selon les intentions de santé « Health Care Intents »: Dans les textes du fchiers excel, nous voulons détécter des intentions liés au secteur de la santé. Il faut remplir les colonnes vides selon les intentions expriméés dans chaque texte : attribuer la valeur 1 si l'intention est exprimé dans le texte, attribuer la valeur 0 si l'intention est abscente. Les seules valeurs acceptées sont les valeurs binaires 0 et 1 pour dire intention absente ou intention présente. L'intention diagnostique est présente dans tous les textes. **Voici les règles:**

1. Intention de precaution: Rechercher des informations pour prévenir les problèmes de santé ou préserver son bien-être. **Exemples:**

- Comment renforcer naturellement son immunité
- Conseils pour éviter la grippe saisonnière
- Aliments pour booster les défenses immunitaires
- Exercices pour prévenir les douleurs lombaires
- Routines de sommeil pour éviter l'insomnie
- Habitudes saines pour préserver sa vue
- Comment éviter les infections urinaires
- Astuces pour réduire le stress au quotidien
- Précautions à prendre pendant une canicule

- Comment protéger sa peau du soleil

2. Intention diagnostique: Identifier la cause de symptômes ou de problèmes de santé. **Exemples:**

- Pourquoi ai-je une toux persistante ?
- Symptômes du diabète
- Causes possibles des maux de tête fréquents
- Est-ce que mes vertiges sont inquiétants ?
- Comment reconnaître une allergie alimentaire ?
- Ballonnements fréquents : quelles explications ?
- Douleur à la poitrine sans effort : que signifie-t-elle ?
- Picotements dans les mains : symptôme de quoi ?
- Fièvre sans autre symptôme : que rechercher ?
- Symptômes précoces de la maladie de Lyme

3. Intention thérapeutique: Trouver des solutions ou des remèdes à une affection diagnostiquée. **Exemples:**

- Meilleurs traitements pour l'arthrite
- Remèdes maison contre les migraines
- Que prendre contre une sciatique ?
- Traitement naturel contre l'acné
- Médicaments efficaces pour les reflux gastriques
- Huiles essentielles contre les maux de gorge
- Comment soulager une entorse à la cheville
- Que faire contre les aphtes douloureux
- Soins pour les peaux très sèches en hiver
- Physiothérapie pour hernie discale : est-ce utile ?

4. Intention de courtoisie: Rechercher des informations de santé pour autrui (soin, inquiétude, accompagnement). **Exemples:**

- Comment aider un proche souffrant de dépression
- Meilleurs aliments pour une personne se remettant d'une opération
- Comment parler à quelqu'un qui refuse de se soigner
- Activités pour stimuler un senior atteint d'Alzheimer
- Conseils pour soutenir un adolescent en détresse psychologique

- Peut-on accompagner un proche à une séance de chimiothérapie ?
- Que faire quand son enfant refuse de manger ?
- Comment gérer l'anxiété d'un partenaire avant une opération
- Quel régime après une opération de la vésicule biliaire
- Vaccins recommandés pour les nourrissons

5. Intention d'urgence: Problèmes de santé immédiats nécessitant une action rapide. **Exemples:**

- Que faire en cas de crise cardiaque
- Soins d'urgence pour les brûlures
- Mon enfant s'est cogné la tête, dois-je m'inquiéter ?
- Comment réagir à une réaction allergique sévère
- Premier geste en cas d'étouffement
- Que faire si quelqu'un perd connaissance ?
- Quels sont les signes d'un AVC ?
- Urgence dentaire : que faire quand une dent se casse ?
- Que faire après une morsure de chien ?
- Que faire si je me suis coupé profondément ?

6. Intention de se rassurer: Rechercher une confirmation ou une tranquillité d'esprit concernant sa santé. **Exemples:**

- Est-il normal de se sentir fatigué après une vaccination ?
- Effets secondaires des antibiotiques
- Fatigue après le COVID : combien de temps ça dure ?
- Est-ce dangereux d'avoir une tension basse ?
- Boule sous la peau : est-ce toujours grave ?
- Une fièvre légère peut-elle durer plusieurs jours ?
- Mes règles sont irrégulières : dois-je m'inquiéter ?
- Perte de cheveux après un accouchement : normal ?
- Taches blanches sur les ongles : signe de quoi ?
- Est-ce que les palpitations sont toujours un signe de problème cardiaque ?

Pour un texte donné, nous voulons évaluer la présence de chaque type d'intention indépendamment en attribuant LE LABEL 1 POUR LA PRéSENCE ET 0 POUR L'ABSCENCE. **Respecter ces règles:**

- Une intention ne reçoit 1 que si elle est clairement exprimée et implicite, avec des actions concrètes ou mots/expressions associés.
- Ne pas supposer ou interpréter à partir d'indications faibles ou vagues. N'inférez pas au-delà de ce qui est littéralement dit ou fortement implicite.
- Si une intention n'est pas présente son score vaut 0.
- Si le signal est vague, absent ou simplement hypothétique, attribuez la valeur 0 (absente).
- Une phrase peut contenir plusieurs intentions ou au moins une intention.
- Chaque intention est évaluée indépendamment.

Les intentions sont intitulées en anglais (Precautionary, Reassurance, Diagnostic, Treatment, Urgency, Courtesy) dans le fichier excel joint dans les noms de colonnes vides (à replir par 1 (intention présente) ou 0 (intention absente). Il faut replir les colonnes vides et fournir le fichier mis à jour.

Appendix-4: Google Survey Questions

Enquête de test de recherche 'HISB français sur SERP'

Avertissement: Cette enquête est strictement destinée à des fins de recherche pour le 'LUMEN Recherche Lab' de 'IAE - University School of Management, Lille'. Si vous la recevez par erreur, signalez-le au fournisseur ou sinon, c'est à vous de la remplir. Les informations recueillies dans le cadre de cette enquête resteront anonymes et strictement privées pour le laboratoire de recherche et l'institution susmentionnés.

Tâche: Remplissez le questionnaire ci-dessous pour répondre aux questions selon votre expérience avec 'Enregistrement de vidéo(s) de test de modélisation et d'analyse de recherche'. Soyez naturel dans vos réponses, aucune panique ou précipitation n'est requise.

- **Questions démographiques (Q1. à Q6.):** 'Votre nom', 'Votre prénom', 'Votre numéro d'identification d'étudiant', 'Nom de votre programme d'études', 'Votre âge', 'Votre sexe'.
- Q7. Combien de voitures votre famille possède-t-elle?
- Q8. Avez-vous suivi un cours, obtenu une certification, obtenu un diplôme ou obtenu un diplôme dans une discipline quelconque de l'éducation en santé, comme les soins infirmiers, la médecine, l'ingénierie médicale, etc.? Sélectionnez toutes les réponses applicables.
- Q9. Avez-vous déjà travaillé dans le domaine infirmier, médical, de l'ingénierie médicale, du marketing de la santé, de l'enseignement dans le domaine de la santé, etc.? Sélectionnez toutes les réponses applicables.
- Q10. Je connais les connaissances en matière de santé issues des pratiques ancestrales.
- Q11. Habituellement, que choisissez-vous entre un remède de santé traditionnel et une consultation chez un médecin? Choisissez toutes les réponses qui s'appliquent.
- Q12. Je me sens à l'aise avec mon médecin pour parler de tous mes types de problèmes de santé.
- Q13. Je me sens à l'aise avec mes parents pour parler de tous mes types de problèmes de santé.
- Q14.a. Dans quelle mesure connaissez(conscient)-vous le système de santé public et privé français? Notez sur l'échelle LIKERT de 1 à 5 ci-dessous.
- Q14.b. Dans quelle mesure êtes-vous exposé (informé) au système de santé public et privé français? Notez sur l'échelle LIKERT de 1 à 5 ci-dessous.
- Q15. Sur une mesura allant de 1 à 5, 1 étant jamais et 5 étant le plus souvent ; à quelle fréquence avez-vous recherché les sujets de santé suivants au cours des 3 derniers mois?

- Q16. Citez les trois principaux sujets de santé que vous avez le plus recherchés sur Google au cours des 3 derniers mois, *par ordre décroissant?*
- Q17. Je fais confiance au contenu et aux publications sur la santé sur les réseaux sociaux, par exemple Facebook, Twitter, X et forums d'utilisateurs pour la santé etc.
- Q18. Sur une mesure allant de 1 à 5, 1 étant ‘je n’utilise pas du tout' et 5 ‘j'utilisant très souvent’, pouvez vous indiquer le degré d’usage des sources en ligne suivantes à des fins de santé ?
- Q19. Sur une mesure allant de 1 étant 'pas du tout d'accord' et 5 étant 'tout à fait d’accord’, pour obtenir des nouvelles et des mises à jour sur la santé, je m’abonne à des ...
- Q20. Sur mesure allant de 1 étant ’très faiblement' à et 5 étant 'très fortement’ les formats de contenu santé suivants m’attirent ...
- Q21. Sur une échelle de 1 à 5, 1 étant 'pas du tout d'accord' et 5 'tout à fait d’accord'; comment les fonctionnalités (features) SERP suivantes contenant des informations sur la santé répondent-elles à votre besoin et à quelle échelle?
- Q22. Comment évalueriez-vous votre expérience avec recordscreen.io lors du test de recherche et des enregistrements vidéo? Notez sur l'échelle de 1 à 5 ci-dessous.
- Q23. Que pensez-vous de la qualité globale du SERP que vous avez obtenu lors du test? Expliquez-le avec vos propres mots.
- Q24. Dans quelle mesure vos intentions de recherche ont-elles été satisfaites par les SERP que vous avez obtenues pendant le test? Ont-elles été suffisantes pour justifier votre action ultérieure? Expliquez-le avec vos propres mots.
- Q25. Qu'est-ce qui manquait dans les résultats de recherche Google obtenus pendant le test et qui pourrait vous être plus utile? Dites-le avec vos propres mots.
- Q26. Qu'est-ce qui vous a frustré dans la présentation des SERP de Google et dans la compréhension que vous avez obtenue pendant le test? Expliquez-le avec vos propres mots.
- Q27. Expliquez l’expérience de recherche globale pendant le test dans vos propres mots?
- Q28. Quelles actions étiez-vous susceptible d'entreprendre ou convaincu d'entreprendre au cours de votre expérience de recherche SERP dans ce test de recherche?
- Q29. Parmi les 7 types d'intention de recherche, quel est le taux de pertinence obtenu selon vous à partir des SERPs reçues lors du test de recherche? Évaluez-le sur une échelle de 1 à 5, 1 étant le plus bas et 5 le plus élevé.

Appendix-5: FLR and SLR based Oneliners Testable Factslist

Appendix-5 below, is a robust checklist of 143 raw facts that are used as checklist in preparing the final research questionnaire and/or research tasks for the defined research model of this thesis:

FLR and SLR based One-Liners Testable Factslist		
Sr.	**One Liner**	**Taken from Thesis Section**
1	Less knowledge more dependency on Google.	Domain Knowledge Pursuance and Perceived Relevance
2	Less knowledge, less willingness to scroll organic results.	
3	Cognitive understanding without pursuing updates lands you in filter bubble.	
4	Higher domain knowledge and SE literacy means careful reading.	
5	Perceived relevance capacity comes from domain knowledge and SE Literacy.	
6	Old age users do more clicking exploitation than exploration, while younger users do more exploration than exploitation.	Biasing in Searcher's Behaviour
7	Being younger, more educated and high Socio-Economic Status tends to high probability of using internet for HIS.	
8	Cognitive, Social Norms and Algorithmic biasing render biasing in Search Behaviour.	
9	Student age searcher's shows very varied behaviour in clicking specific SERP features.	
10	Youngers tender better clicking in difficult search tasks than elders.	
11	Old age users do more clicking exploitation than exploration, while younger users do more exploration than exploitation.	Search Query Scope and SERP Utility
12	Breadth query is a general broad usage non-specific to domain search strategy.	
13	Depth query is a narrow specific use of keywords search strategy.	
14	The breadth search query search number is moderately higher than depth search query.	
15	A searcher can be surrounded in a 'search filter bubble' due to semantics used by SEs, search bias due to own knowledge influences and cognitive presumptions.	
16	Search activity is integrally associated with user' creative processes.	
17	Search efforts, content-triggers and application context are the search tasks in a naturalistic information seeking.	
18	Searching within intellectual and creative tasks is more time intensive.	
19	Search activity occurring as a part of daily routine tasks is associated with more frequent searching within a search task.	

20	Searching is more often induced from utility applications in tasks demanding a degree of intellectual effort.	
21	Users' leisure information-seeking activity is occurring inherently within social media services or platforms.	
22	Searcher performing Look-up search and Exploratory search exhibit different behaviours.	
23	Different aspects of exploratory search are uncertainty, creativity, exploration, and knowledge discovery.	
24	Contextual scenario does not seem to have a systematic effect on users' viewing behaviour.	
25	In resource scenario search results presented in list top receive most attention.	
26	In individuality scenario both age and prior topic knowledge seem to be related to a more systematic and more careful reading of SERP.	
27	Complex tasks show significantly different characteristics than simple tasks.	
28	Search Task complexity can be expressed by the efforts in terms of time, sessions, queries, and browser tabs used to carry out that search.	
29	ML and Filtering algorithms training impacts SE's relevance score for smaller depths of 3 to 5.	Result Ranking and Search Behaviour
30	One cannot rank the different ranking algorithms in terms of performance as different approaches of page ranking algorithms are suitable for different applications.	
31	We need different algorithms for offline and online page ranking.	
32	Users would select options displayed early on the search results list when they perceived the ranking was reliable.	
33	Users proceed sequentially on SERP until a result (alternative) satisficing their expectation is found.	User Gain Expectancy
34	Users continue retrieving information until they observe a result (alternative) violates their expectations.	
35	A result not satisfying user expectation, even if top-ranked; it increases user' impatience to stop IR.	
36	SERP diversity captures more user attention, that attention gets renewed on new alternatives countering, after that, attention decays rapidly.	
37	Visual hierarchy, competition for attention and banner blindness influence transactional search behaviour.	
38	Visual hierarchy is not directly related to transactional search behaviour.	

39	Banner blindness phenomenon only exists to a certain extent and slightly influences the search behaviour.	
40	There are no major differences in transactional search behaviour across search engines.	
41	Heterogeneity of SERP elements alters user behaviour in terms of stopping rank, time on SERP and user gain rate.	
42	How new SERP objects entry reduces precision of ad targeting.	
43	Ads on top in mobiles get more visual attention than desktop due to small aperture of screen.	
44	In Search Task difficulty prediction, the most predictive features are usually different for whole-session vs. first-round prediction.	
45	In Search Task difficulty prediction, the mouseover features are effective for first-round.	
46	In Search Task difficulty prediction, the level of interest and prior knowledge features don't improve performance.	
47	Differences in scrolled distance were larger for popular queries be it textual query or image query.	
48	Browsing habits of users are largely uniform, regardless of the investment in individual query formulation, or the number of query topics undertaken per user session.	
49	Users stop searching for a variety of reasons, including frustration or satisfaction with the results.	
50	Editing original query on counter facts can yield improved SERP.	Query Drifting and User Interest Decay
51	Sudden Query Drift occurs on breaking news queries varying from few hours to few days, they have highest search abandonment rate.	
52	Incremental Query Drift occurs on less popular tail types of queries, volume of queries remains almost stagnant, search abandonment is frequent initially.	
53	Gradual Query Drift can be seen as a combination of Sudden and Incremental both, occurring on secondary news and events around a public entity.	
54	Reoccurring Query Drift has a regular type of behaviour, like movie release events.	
55	If query reformulation grows over time, it is taken as a positive change while if it decreases then it is counted as negative.	
56	User interest is not constant for a long time.	
57	User interest duration depends on user's appetite for learning and self-updating.	
58	If a user interest is not updated, it becomes prone to interest drift.	
59	Terms drawing a high degree of a searcher's attention are likely to be relevant to his/her search intent.	

60	Five query reformulation types are Generalization, Specialization, Parallel Move, Mission Change, and Error Correction.	
61	Query reformulation prediction factors can be query term length, attention area in web pages, clicks on search results.	
62	SE users usually have a higher probability of leaving when loyalty is found to be continuously decreasing.	
63	Query reformulation behaviour in web search varies with the type of search tasks.	
64	Query suggestion/related query recommendations provided by search engines do not offer enough help for users in complex search tasks.	
65	User intent and domain expertise influence the search process.	
66	A broad intent may cause a search process with more exploration.	
67	A clear intent can lead to more exploitation.	
68	Consensus ranking resulted in substantially fewer changes compared to averaged individual user rankings.	
69	There is quite low similarity between the search engines' and consensus users' rankings.	
70	Information displayed on SERP is contextual to query rather than its content.	Clicking Intent Intensity
71	Query-based features along with the content of SERPs are effective in detecting query intent.	
72	7 types of SE Loyals are one off, infrequent, occasional, intermittent, frequent, very frequent and loyal.	
73	Frequent, very frequent and loyal users give highest CTR on medium length queries.	
74	All 7 types of SE Loyals click top- and high-ranking results.	
75	Long queries have a smaller CTR.	
76	Users take long jumps between consecutive clicks and are likely to move backward as forward.	
77	Users dislike going beyond the first results page.	
78	Users are capable of making quick decisions about pages they have clicked on.	
79	Searchers of a fast search system are more sensitive to delays than the users of a slow search system.	
80	Up to 500ms response time delays in SERPs are not noticeable, after 1000ms users could feel the added delays with very high likelihood.	
81	Searchers are more likely to perform clicks on SERP that is served with lower latency.	
82	Target Finding users tend to issue few specific queries and browse only top ranked results.	
83	Decision Making users tend to issue short queries, browse deep, and click more results.	

84	Exploration users issue many diverse queries, browse deep, but do not click often.	
85	Personalisation of search results does not increase users' satisfaction with their search results.	SERP Personalization Impact
86	SERP Personalisation reduces time spent by users in locating information they need.	
87	Three Neuro-Physiological Evidence methods for improving search in standardising search tasks are Eye Tracking, EEG, and fMRI.	Semantic Clustering and IR
88	Number of search tasks hints the task difficulty, user behaviour, intent depth and information urgency.	
89	Neural NLP models lack in intuition interpretability and debuggability, important aspects of commercial solutions.	
90	Interactive clustering is a discipline that combines machine learning/data mining and Human-Computer Interaction.	
91	As severe the health issue like Covid found as would be the psychological stress and health concerns and it manifolds in outbreak areas.	Pre-Web HIS Outcomes and Methods
92	4 HISB outcome types: Cognitive, Behavioural, Physical and Affective.	
93	2 dimensions of HISB: Information and Method.	
94	Web may create inequalities in health information accessibility.	HIS in Web Era
95	There are digital disparities in HIS especially with regards to age and Socio-Economic Status.	
96	Being younger, more educated, and having higher SES are predictors of internet use for HIS. Health	
97	Being older, having low internet skill, lower SES and Hispanic are determinants of using a health care provider or traditional media, such as print and magazines, as source of health information.	
98	Digital disparity widens the gap of information access.	
99	Internet has become the most preferred HIS platform post web invention.	
100	83% of searchers use SEs like Google and Yahoo for HIS, while 15% use specific health websites.	
101	People in low socioeconomic status group search the Internet for health information less often than people in the high or medium status groups.	
102	Most central predictors of HISB are age, education, gender, health condition, and financial income.	
103	Four complementary characteristics of health information sources are access to medical expertise, tailorability, anonymity, and convenience.	

104	People who seek health information from social media tend to have more experiences of encountering misinformation.	
105	People with misinformation encounter on social media are more likely to use professional health websites for same purposes.	
106	Individuals who handle health issues of themselves or their loved ones are more likely to use the exploratory approach to search.	Web HISB Characteristics
107	Contradictory information, curiosity and task clarity lead to exploratory health search.	
108	An exploratory SERP seeker has no specific targets but keeps learning new search keywords and information.	
109	Misinformation density in health SERPs affects people's search activities, their correctness of answering factual questions and their perceived goodness of the search systems.	
110	In health information search mediums in Google/another browser stands at 50.70%, YouTube at 27.7%, Facebook 7.3%, Health Apps 12.9% and others 1.40%.	
111	Depending on whether knowledge was obtained from internet or credible source, attitude of the former about the disease will be more negative compared to later.	
112	Public health announcements put direct impact on information search behaviour.	
113	Pitfalls such as dysfunctional search filters, inconsistent results, outdated information, lack of reproducibility and thoroughness of SERPs distracts HIS.	
114	Patients use online medical information to integrate with advice from friends, family, and physicians in order to confidently make their medical decisions.	
115	DHL enables people to critically evaluate the information in hand.	DHL Impact on HISB
116	8 common sources of HIS, internet, face-to-face communication with healthcare professionals, demographic characteristics, psychological aspects, role-related or interpersonal traits, environmental aspects, source related characteristics, and disease-related factors.	
117	5 types of HISB include active searching, ongoing searching, passive attention, passive searching, and avoid seeking.	
118	SEs have a major role in spreading low-quality health information.	
119	Users' cognitive styles influence their information searching strategies, query reformulation behaviour, web navigational styles and information processing approaches.	
120	Health information searching is as much social as it is private.	
121	7 search string categories in HIS are location of nearby services, symptoms, types of services/specialty, asking for advice, questions about resources, questions about whether	

	they have a problem, and looking for the "best."	
122	Three types of HIS channels are Searching, Browsing and Scanning.	
123	Students with less prior medical knowledge tend to consider the video feature of applications more useful.	
124	Students do HIS heavily in SEs on mobile devices, rather than browsing or using applications.	
125	Who (SE or User) learns from whom in what percentage in context of French HISB	2.1. Art of phrasing of Search User and Search Engine
126	What percentage of search users who rely on search engines itself for their query understanding rather than making efforts in asking in clear words	2.2. Natural Language Ontology
127	What percentage of search users, do reformulation or iteration around their search queries due to lack of proper linguistic expertise or lesser knowledge of their topic of search	
128	How the health information seekers in France search, reformulate and iterate their queries to finally get the desired search results	
129	Literary expertise brings creativity and precision in search results through query formation or reformation.	2.3. Topicality
130	Domain knowledge brings critical capacity to diversify search results on quality and relevance.	
131	Knowledge and expertise combined lead to a better cognitive semantic understanding and sophistication around a topic or subject search.	
132	Checking people's knowledge and understanding about Web 4.0 is very crucial for this thesis.	2.4. Semantic Web
133	This finding suggests that when quality and relevancy is of top priority, people can adjust delay in SERP display.	2.5. SERP Personalization
134	Search skills helps impactfully in reducing the intent ambiguity, consequently making the search user to express their search intent with clarity in their search queries.	2.6. Search Skills
135	Enhanced understanding of query nuances and better linguistic command, which ultimately helps to manage your search with ease at "peace of mind"	
136	An argument can be opened about defining EEAT with 3E only.	
137	It is likely that users would love to accommodate Chat-GPT display in the SERP as another SERP feature or menu tab	3.5. G-AI and LLMs
138	SEs are undergoing a transformation paradigm to adapt to the evolving metaverse	3.6. MetaSearch

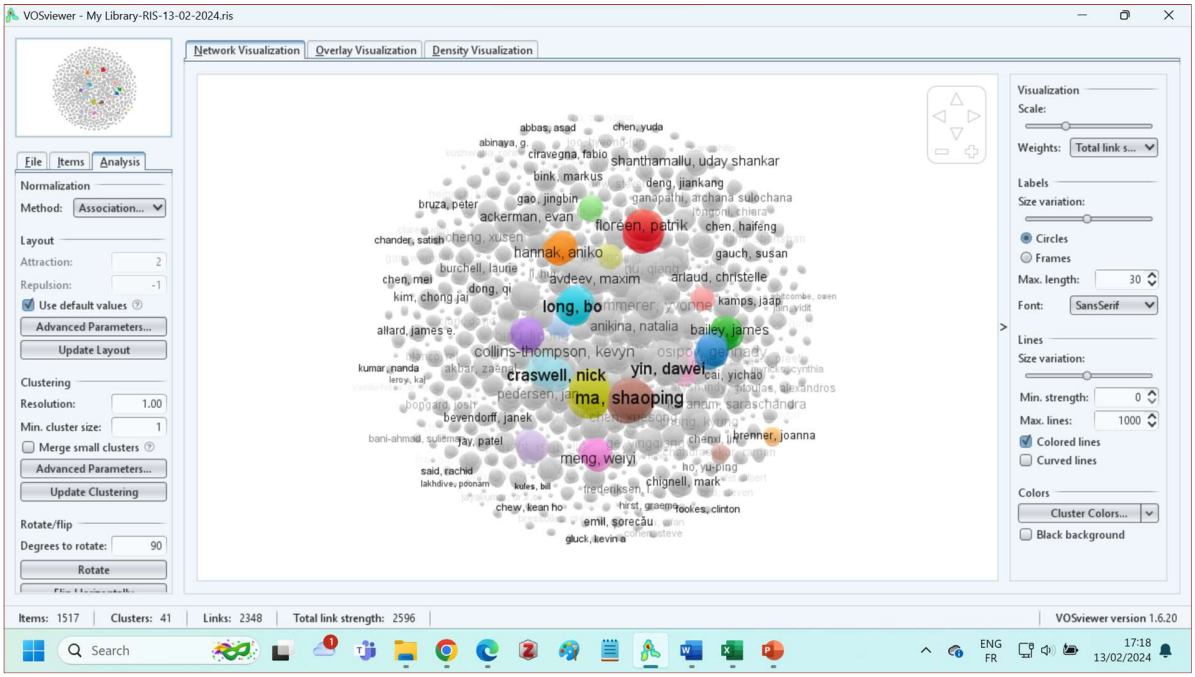
	landscape.	Engine and SERP
139	How people are attracting towards metaverse.	
140	Along with quality and user satisfaction, the changing trends are the major factor with which Google decides about SERP.	4.1. Web Search User at Mind SERP Model of Google
141	Here a significant argument can be put to sum-up E-E-A-T into 3E for simplicity of the concept.	4.2. Three Stakeholders of SERP
142	The major argument of above study can be the proposal of two new major search intent types i.e. Entertainment and Mapigation / Destination.	
143	Searchers equipped with search operators knowledge have significant edge over others without this backup.	4.4. Web Search Intelligence

Appendix-6: VOSviewer Maps and Connected Papers Analysis

This appendix presents the sample Citation Analysis of overall Bibliography and Selected Papers.

Important Data Dimensions *(from the footer of screenshot-8 below)*

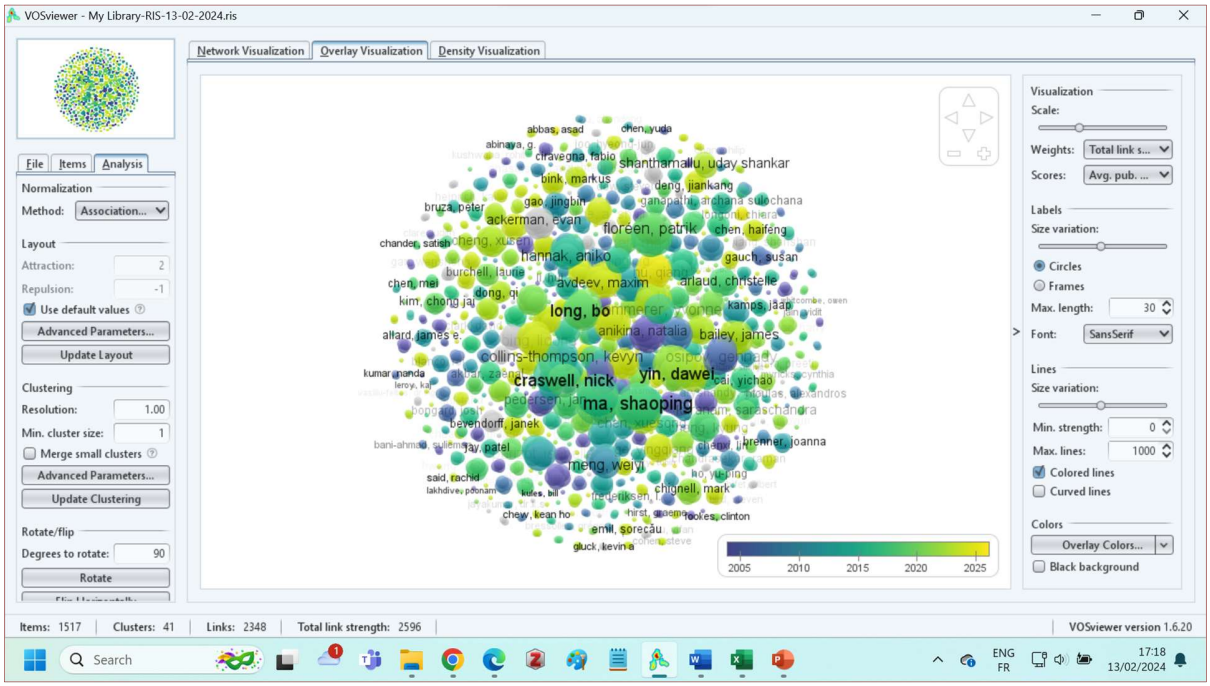
Items: 1517 | **Clusters:** 41 | **Links:** 2348 | **Total link strengths:** 2596



Screenshot-8: VOSviewer Bibliography Strength Map

Important Data Dimensions *(from the footer of screenshot-9 below)*

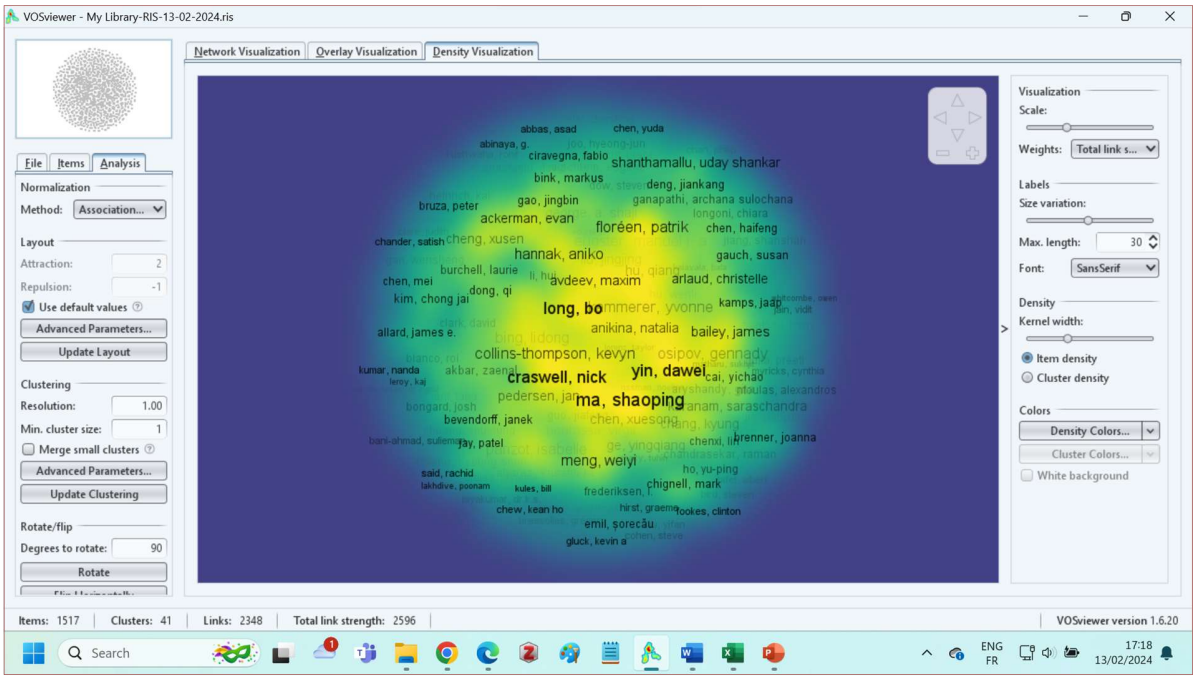
Items: 1517 | **Clusters:** 41 | **Links:** 2348 | **Total link strengths:** 2596



Screenshot-9: VOSviewer Bibliography Strength Map

Important Data Dimensions *(from the footer of screenshot-10 below)*

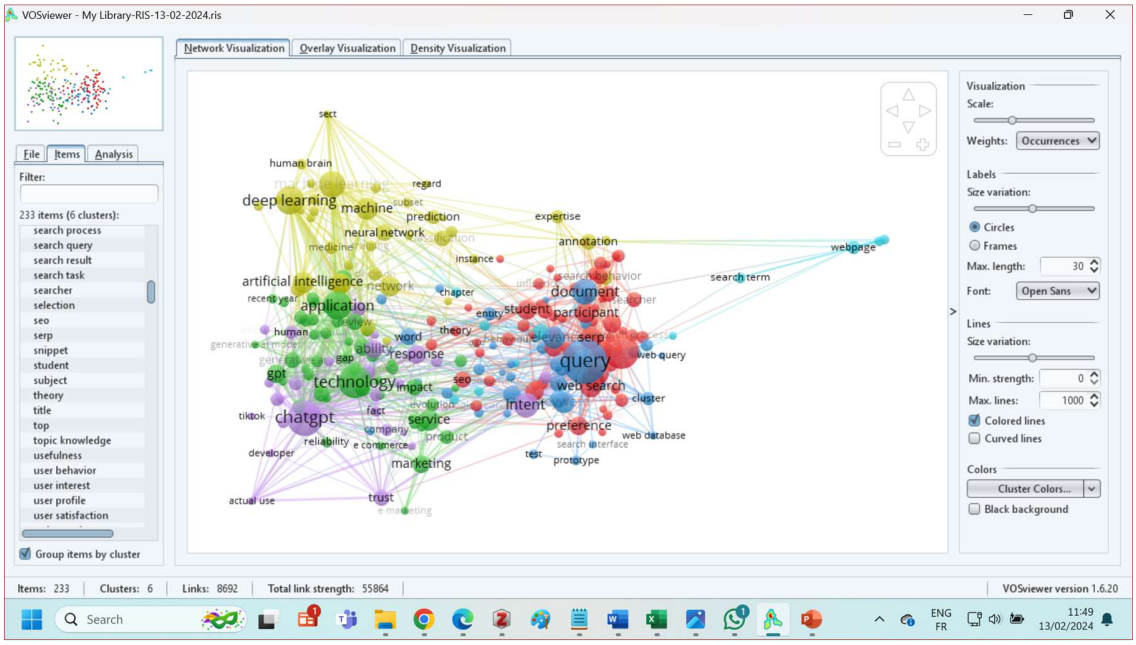
Items: 1517 | **Clusters:** 41 | **Links:** 2348 | **Total link strengths:** 2596



Screenshot-10: VOSviewer Bibliography Strength Map

Important Data Dimensions *(from the footer of screenshot-11 below)*

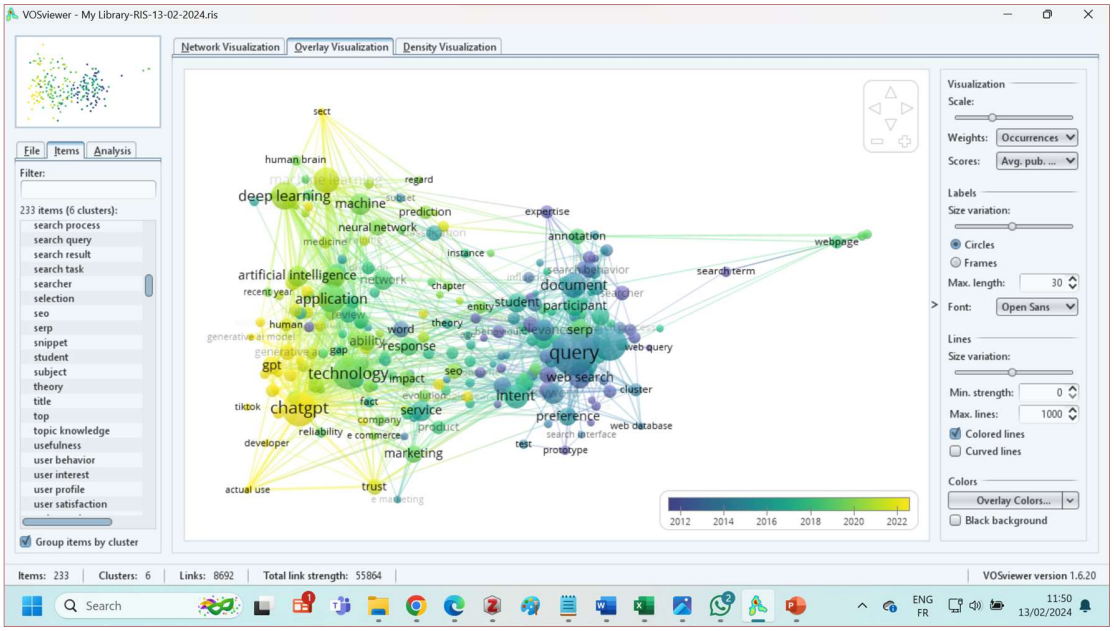
Items: 233 | **Clusters:** 6 | **Links:** 8692 | **Total link strengths:** 55865



Screenshot-11: VOSviewer Bibliography Strength Map

Important Data Dimensions *(from the footer of screenshot-12 below)*

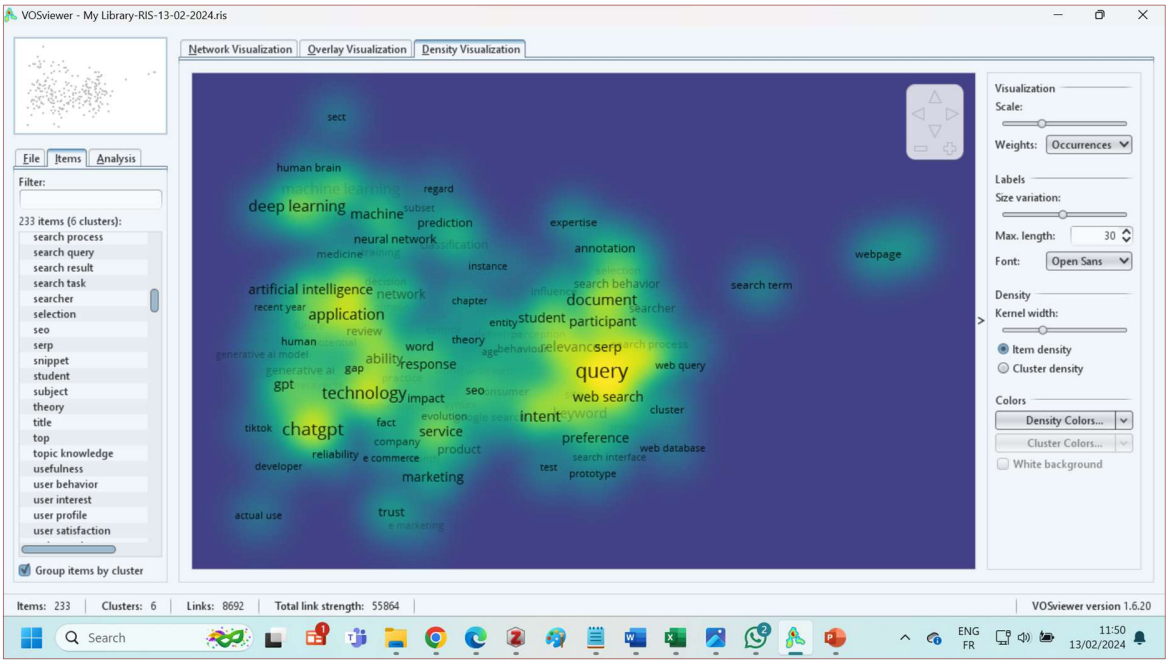
Items: 233 | **Clusters:** 6 | **Links:** 8692 | **Total link strengths:** 55865



Screenshot-12: VOSviewer Bibliography Strength Map

Important Data Dimensions *(from the footer of screenshot-13 below)*

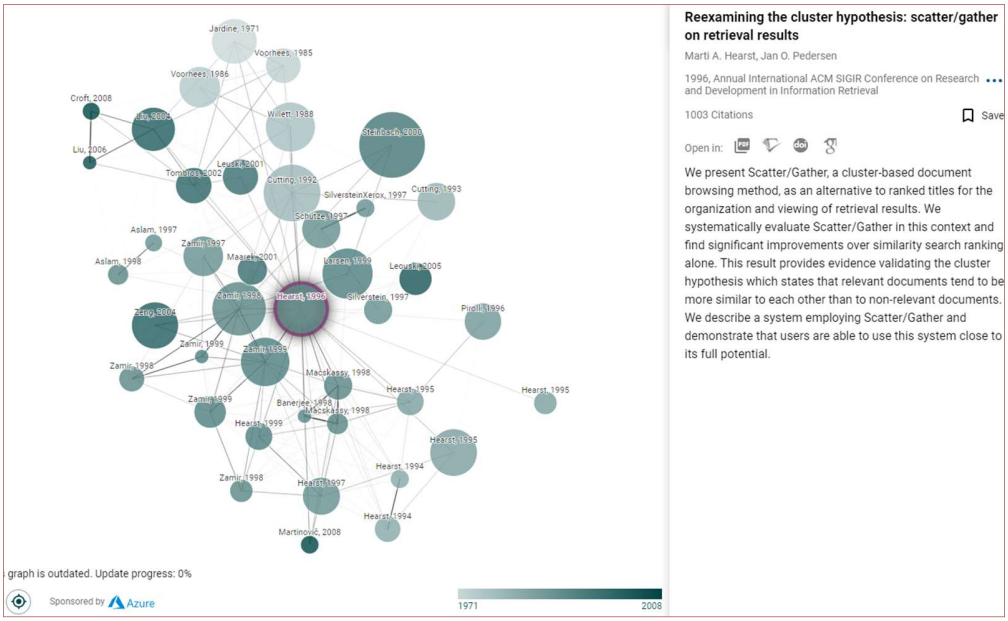
Items: 233 | **Clusters:** 6 | **Links:** 8692 | **Total link strengths:** 55865



Screenshot-13: VOSviewer Bibliography Strength Map

Important notes *(from the screenshot-14 below)*

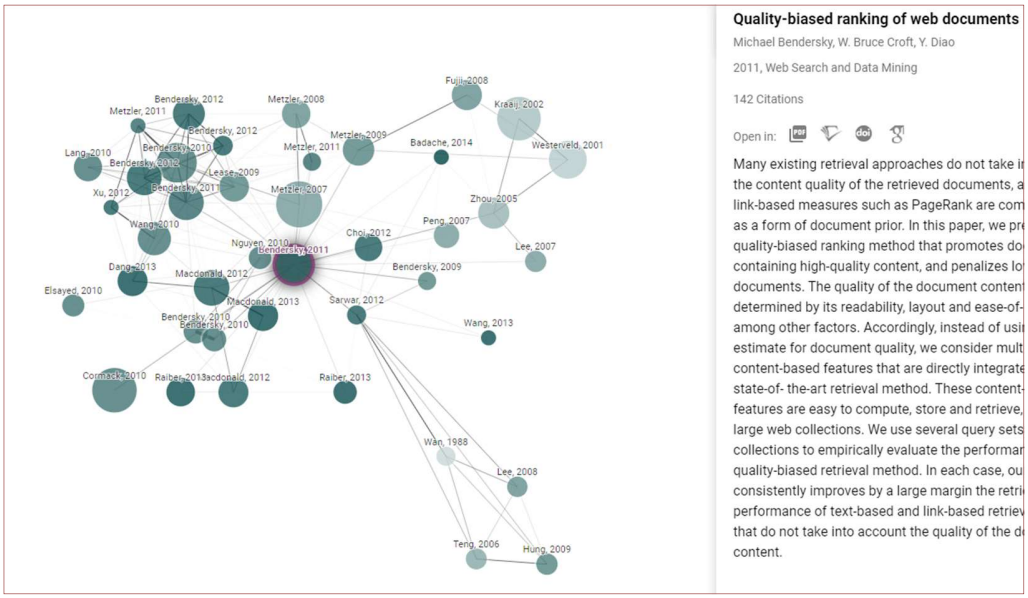
This the map of a paper published in 1996 having 1003 citations



Screenshot-14: Connected Papers Bibliography Citation Weight Map

Important notes *(from the screenshot-15 below)*

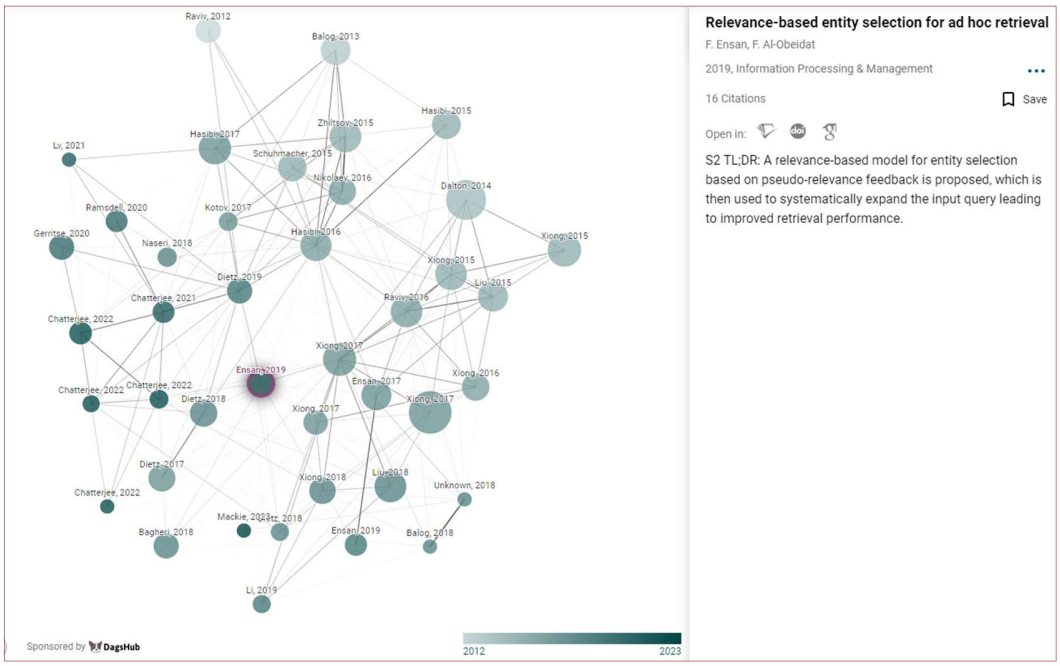
This the map of a paper published in 2011 having 142 citations



Screenshot-15: Connected Papers Bibliography Citation Weight Map

Important notes *(from the screenshot-16 below)*

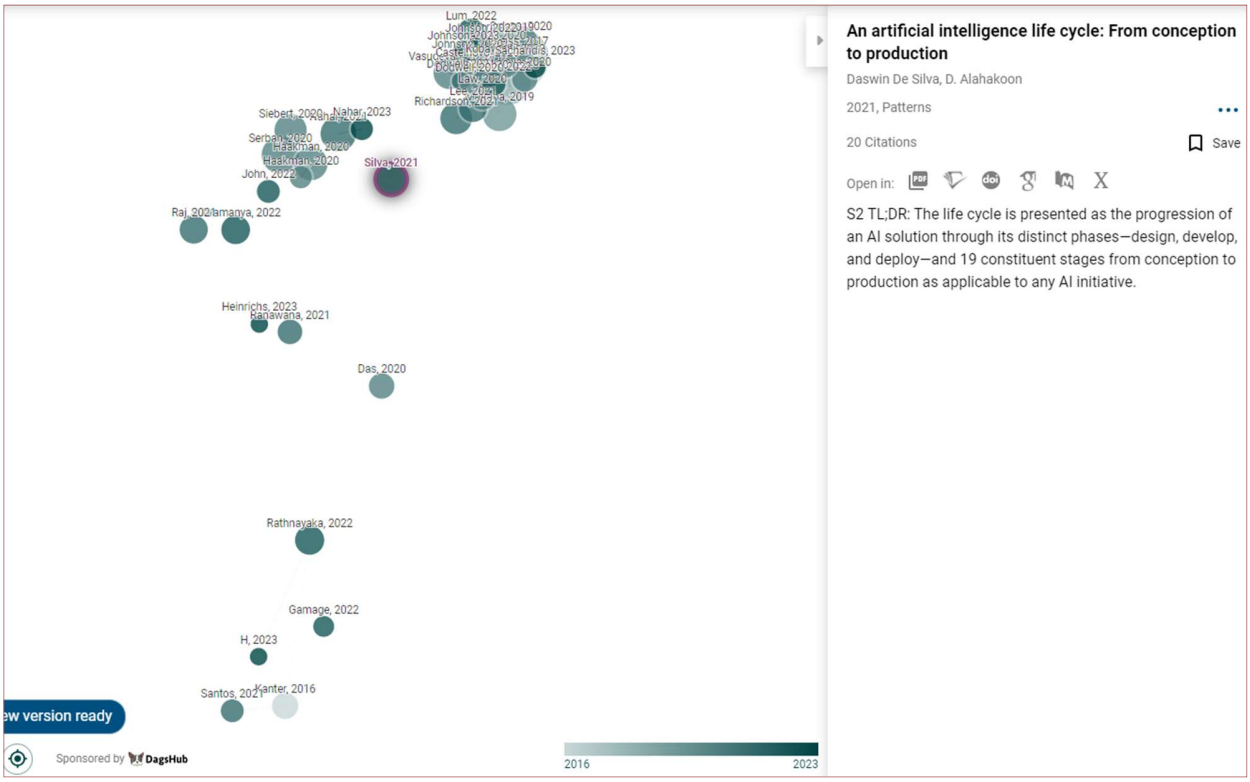
This the map of a paper published in 2019 having 16 citations



Screenshot-16: Connected Papers Bibliography Citation Weight Map

Important notes *(from the screenshot-17 below)*

This the map of a paper published in 2021 having 20 citations



Screenshot-17: Connected Papers Bibliography Citation Weight Map

Appendix-7: Acronyms and Abbreviations

Sr.	Abbreviation	Full Form
1	AI	Artificial Intelligence
2	AIRO	Artificial Intelligence Recommendation Optimization
3	ALIWEB	Archie-Like Indexing for the Web - name of a SE
4	ALR	Arbitrary Literature Review
5	AMP	Accelerated Mobile Pages
6	ANN	Artificial Neural Network
7	AQE	Automatic Query Expansion
8	AR	Augmented Reality
9	B2B	Business to Business
10	B2C	Business to Customer
11	BERT	Bidirectional Encoder Representations from Transformers - a contextual word embedding LLM
12	BM25	Best Match 25 - a TF-IDF ranking algorithm to match query to content
13	BNNs	Bayesian Neural Networks
14	BPE	Byte-Pair Encoding - a tokenizer in LLMs
15	BPM	Bejeweled Player Model - a search behaviour research model
16	CAM	Card-Aware Metrics - a search behaviour research method
17	CAS	Click, Attention & Satisfaction - a search behaviour research model
18	CFE2	Counter Factual Explanation with Editing
19	CLS	Cumulative Layout Shift
20	CMBM	Click Model-Based Metrics - a search behaviour research method
21	CMS	Content Management Systems
22	CNNs	Convolutional Neural Networks
23	CPA	Cost per action
24	CPC	Cost per click
25	CPM	Cost per Mille - M stands for roman numeral for 1000
26	CSR	Clustering Search Results
27	CTA	Call-to-Action
28	CTM	Correlated Topic Model
29	CTR	Click-through Rate

30	CWV	Core Web Vitals
31	DBN	Deep Belief Networks
32	DCG	Direct Cumulated Gain, Discounted Cumulative Gain - a IR relevance research method
33	DFBM	Deterministic Forward Browsing Model - a search behaviour research method
34	DGM	Deep Generative Models
35	DHIS	Digital Health Information Search
36	DHL	Digital Health Literacy
37	DL	Deep Learning
38	DNNs	Deep Neural Networks
39	DRL	Deep Reinforcement Learning
40	DS	Data Science
41	DTCD	Dynamic Topic Correlation Detection
42	DTD	Document Type Definition
43	DTE	Detecting Topic Evolution
44	DTM	Dynamic Topic Models
45	EBU	Expected Browsing Utility - a search behaviour research method
46	EEAT	Experience, Expertise, Authoritativeness and Trustworthiness
47	EE-D	Expected Exposure Disparity - a search behaviour research method
48	EEE	Expertise, Experience, Exposure of a search user
49	EEG	ElectroEncephaloGram
50	EE-R	Expected Exposure Relevance - a search behaviour research method
51	ELMo	Embeddings from Language Models - a bidirectional LLM
52	EMD	Exact Match Domain
53	ERR	Expected Reciprocal Rank - a search behaviour research method
54	E-tail SQ	E-tail Service Quality
55	ETI	Extracting, Transforming, and Integrating
56	FACS	Facial Action Coding System
57	FID	First Input Delay
58	FLR	Foreground Literature Review
59	fMRI	Functional Magnetic Resonance Imaging
60	FNN	Fuzzy Neural Network
61	FORs	Featured Organic Results
62	FSNB	Filming SERP Navigation Behaviour
63	FTP	File Transfer Protocol
64	G-AI	Generative Artificial Intelligence

65	GANs	Generative Adversial Networks
66	GloVe	Global Vectors - a non-contextual word embedding unsupervised LLM
67	GO	Gene Ontology
68	GPS	General Problem Solver - An earliest AI System
69	GPT	Generative Pre-Trained Transformers - a LLM Model
70	GSQR	Google Search Quality Rater - Google' Search Annotation Crowdsourcing Program
71	GUI	Graphical User Interface
72	HBG	Hieight-biased Gain - a search behaviour research model
73	HCPs	Health Care Professionals
74	HIR	Health Information Retrieval
75	HIS	Health Information Searching
76	HISB	Health Information Seeking Behaviour
77	HITS	Hyperlink-Induced Topic Search - Link Analysis Algorithm
78	HMI	Human Machine Interface
79	HRNN	Hierarchical Recurrent Neural Network
80	HTML	HyperText Markup Language
81	HTTPS	Hypertext Transfer Protocol Secure
82	ICTs	Information and Communication Technologies
83	IFT	Information Foraging Theory - a search behaviour research model
84	INSQ	Inverse Square - a search behaviour research method
85	IOT	Internet of Things
86	IR	Information Retrieval
87	IRB	Institutional Review Board
88	ISA	Intelligent Search Assistant
89	JAVA	Just Another Virtual Accelerator
90	KBS	Keyword Based Search
91	kNN	k-nearest Neighbour Algorithm
92	LCP	Largest Contentful Paint
93	LDA	Latent Dirichlet Allocation
94	LLMs	Large Language Models
95	LPLMs	Large Pre-trained Language Models
96	LR	Literature Review
97	LRIR	Learning to Rank for Information Retrieval
98	LSA	Latent Semantic Analysis

99	LSI	Latent Semantic Indexing
100	LSTM	Long Short-Term Memory Network
101	MAP	Mean Average Precision - a SERP personalization metric
102	MCMC	Markov Chain Monte Carlo - a type of simulation
103	ML	Machine Learning
104	MLAs	Machine Learning Algorithms
105	MRR	Mean Reciprocal Rank - a SERP personalization metric
106	MSQs	Multi-Sentence Questions
107	MTT	Multiscale Topic Tomography
108	MTurk	Mechanical Turk - Amazon' Search Annotation Crowdsourcing Program
109	MUM	Multitask Unified Model - MUM is much more powerful than BERT
110	NCG	Normalized Cumulative Gain - a IR relevance research method
111	NDCG	Normalized Discounted Cumulative Gain - a SERP personalization metric
112	nDCG	Normalized Discounted Cumulative Gain - a search behaviour research method
113	NDCG-IA	Normalized Discounted Cumulative Gain - Intent Aware - a search behaviour research method
114	NER	Name Entity Recognition - a NLU technic
115	NF	Normalizing Flow
116	NFS	Neural Fuzzy System
117	NLG	Natural Language Generation - a NLU technic
118	NLP	Natural Language Processing
119	NLS	Natural Language Search
120	NLU	Natural Language Understanding
121	NN	Neural Network
122	NoSQL	not only SQL (Structured Query Language)
123	OWL	Web Ontology Language
124	PARs	Paid Ads Results
125	PFW	Page Forwarding Weight
126	PGMs	Probabilistic Graphical Models
127	PII	Personally Identifiable Information
128	PLSA	Probabilistic Latent Semantic Analysis
129	PPC	Pay per click
130	PPR	Personalized PageRank - an algorithm
131	PRW	Page Retaining Weight

132	RBP	Rank-biased Precision - a search behaviour research method
133	RCM	Random Click Model - a search behaviour research method
134	RDCG	Repetition rate involved in Discounted Cumulative Gain - a SERP Personalization metric
135	RDF	Resource Description Framework also know as RDF Triple
136	RDFS	Resource Description Framework Schema
137	REL	Relevance of External Links
138	RLPS	Reinforcement Learning-based framework for Personalized Search
139	RLPS-H	Reinforcement Learning-based framework for Personalized Search - Hierarchical model
140	RLPS-L	Reinforcement Learning-based framework for Personalized Search - Listwise model
141	RMLR	Research Model Literature Review
142	RNN	Random Neural Network
143	RoBERTa	Robustly optimized BERT approach - a LLM Model
144	RWBM	Random Walk Browsing Model - a search behaviour research method
145	SAL	Search-As-Learning
146	SDCG	Scaled Discounted Cumulative Gain - a search behaviour research method
147	SE	Search Engine
148	SEM	Search Engine Marketing
149	SEML	Search Engine Machine Learning
150	SEMLAs	Search Engine Machine Learning Algorithms
151	SEO	Search Engine Optimization
152	SERP	Search Engine Results Page
153	SES	Socio Economic Status
154	SFBM	Stochastic Forward Browsing Model - a search behaviour research method
155	SGE	Search Generative Experience
156	SIRI	Speech Interpretation and Recognition Interface
157	SLR	Synthesized Literature Review
158	SML	Supervised Machine Learning
159	SMM	Social Media Marketing
160	SMO	Social Media Optimization
161	SODA	Strategic Options Development and Analysis
162	SORs	Standard Organic Results
163	SPARQL	SPARQL Protocol and RDF Query Language
164	SPO	Subject-Predicate-Object Statement also know as TripleStore
165	SRRs	Search Result Records

166	SSE	Semantic Search Engine
167	SSK	Self-serve Kiosk
168	SS-PMM	Semantic Similarity - Process Model Matching
169	STC	Suffix Tree Clustering
170	SVIA	Scala di Valutazione delle Interazioni Alimentari - a videogrphy tool
171	SVM	Support Vector Machine
172	T5	Test-To-Text Transfer Transformer - a LLM Model
173	TBG	Time-biased Gain - a search behaviour research method
174	TF-IDF	Term Frequency-Inverse Document Frequency
175	TLTMs	Transfer Learning and Transformer Models
176	TLTTMs	Transfer Learning Techniques and Transformer Models
177	TOT	Topic Over Time
178	TREC	Text REtrieval Conference
179	TRIM	Topicality Relevance-aware Intent Model
180	UHRS	Universal Human Relevance System - Bing Search Annotation Crowdsourcing Program
181	UML	Unspuervised Machine Learning
182	URI	Uniform Resource Identifier
183	URL	Uniform Resource Locator
184	UTAUT	Unified Theory of Acceptance and Use of Technology
185	VAEs	Variational AutoEncoders
186	VERONICA	Very Easy Rodent-Oriented Net-wide Index to Computerized Archives - name of a SE
187	VITR	Vision Transformers with Relation-focused Learning
188	VSE	Virtual Semantic Embedding
189	W3	World Wide Web
190	W3C	World Wide Web Consortium
191	WDES	Web Data Extraction using Similarity Function
192	WDICS	Web Data Integration using Cosine Similarity
193	WOT	Web of Things
194	XAI	Explainable Artificial Intelligence
195	XHTML	Extensible Hyper Text Markup Language
196	XLNet	eXtreme Language Understanding NETwork - a LLM Model
197	XML	Extensible Markup Language
198	XSD	XML Schema Definition
199	YMYL	Your Money Your Life

Appendix-8: Thesis Keywords - Detailed List

Sr.	Pillar Keyword	Search Keyword/Phrase Cluster
1	SERP Features & Analysis	SERP Features, SERP Analysis, SERP Trends, SERP Data Analysis, SERP Clicks, SERP Click Journey, SERP Aesthetics, SERP Behaviour Videographical Analysis, Eye tracking methodology in search engine results page behaviour analysis, Eye Tracking movement on web evaluation, Eye-Tracking Task Attributes, Eye-Tracking Methods, Web cursor movement on search results observations through video recording data
2	SERP Marketing & Optimization	SERP Marketing, SERP Optimization, SERP Tracking, SERP Ranking Factors, SERP Positioning, SERP Competition, SERP Tools, Google's ranking system or factors, Exponentially augmented grip of SERP-marketing on global e-marketing, SERP Personalization is making it difficult for SEO Managers
3	SERP Quality & Annotation	SERP quality annotation process of search engines, Types of serp quality annotation, Impact of serp quality annotation on serp quality, Comparative annotation methods between search results of two search engines, Annotation process of search engine results, Types of annotation process of search engine results, Annotation methods of annotating search engine results
4	SERP Personalization	SERP Personalisation by the Search Engines, How SERP personalization decides the specificity of results in response of a search query, What interests and contexts are put in front when Search Engines do SERP personalization, How SERP results ensure the Customer Satisfaction, How SERP features contribute to customer satisfaction, How SERP features contribute to SERP personalization, What are the SERP personalization algorithms a search engine practices, Personalization
5	Artificial Intelligence & Machine Learning	Machine Learning, Deep Learning, Types of machine learning technologies used in search engine results generation, Role of deep learning in search engine results production, Characteristics of Deep Learning and Search engine results pages, Kind of generative artificial intelligence used in generating search engine results pages, Kind of artificial intelligence search engine results page use, Deciphering and artificial intelligence, RNN BiGRU Supervised Machine Learning

6	Natural Language Processing	Natural Language Processing, Role of natural language processing in search engine algorithms, How do search engines use natural language processing to improve search results, Natural languages, Use of neuro understanding in word phrasing in search queries on search engines, Use of neuro understanding in interpreting the search queries on search engines, How search engines use neuro understanding in interpreting the search queries
7	Large Language Models & ChatGPT	ChatGPT, Large Language Modelling, Art of phrasing and generative ai, Word phrasing and generative ai
8	Search Algorithms	SERP Algorithms, Naïve Bayes, Random Forest, Standardised MedDRA Queries, Reinforcement Learning, Supervised Learning, Unsupervised Learning, Search queries interpretation algorithms used by search engines, Semantic web search algorithms, Algorithms search engines use to do self-learning
9	Search Intent & Classification	Search Intent Behaviour, Query Classification, Informational Search, Transactional Search, Navigational Search, Commercial Search, Local Search, User Intent Analysis, Search Intent Optimization, Search gap analysis in context of a search engine, Search intent combinations, Types of sub-intents or micro-intents, Doing the analysis of search intent, Search intent biasing, Search intent predictions, Understanding user behaviour through search intent, Identifying the search queries with multiple intent, Search intent types and purchase cycle, Understanding the Intent of misspelled Search Queries, Interpreting the likeliness of Search Intents by understanding a search query, Understanding the narrow scope, broad scope of a search query, Health Search Intents
10	Keyword Research & Analysis	Keyword Signals, Long-tail Keyword, Short Keyword, Art of keywords phrasing, Word phrasing in search queries on search engines, How a search keyword or set of search keywords (search phrase) forms in searcher mind, How literary expertise effects smart use of Keywords in online search
11	Search Query Processing	Search query parsing phenomenon of a search engine, Search query words and meanings indexing phenomenon of a search engine, Search query data storage mechanism of search engines, Semantic topic clustering based on search terms by search engines, Search engines organizing and storing information about the words and meanings within search queries, Deciphering the search queries by a search engine, How the search engines decipher the search queries

12	Search Operators & Expressions	Search Engine Operators/Expressions, Search operators and search results enhancement
13	Lingual Command & Semantics	Lingual Command, Lingual expertise, How lingual expertise of a searcher contributes to better Search Engine Results, What is the role of semantics in search engine results, How can searchers optimize their queries for better search results
14	Topicality & Content Relevance	Topicality around the Searched Query, How the topic knowledge of a searcher helps in producing better search results, How the topicality around a search query is decided by the search engines, Search query topical organization by search engines, How search engines design the ontology of topics around a search query, Content Relevancy, Human Content Relevancy Psychology (Relevancy, Quality, Authority, Freshness, Market, Location, Language)
15	Web Content Classification	The factor through which search engines understand the content of a web page, The factors on which search engine classify a web content, Web content creators categorizing by search engines, Good Content Creators recognizing and rewarding by Search Engine
16	Search User Behaviour	Psychological Behaviour, Behaviour Types, Human Behaviour, Search Engine Search User Behaviour, Types of search users on search engines, Search User at the mind of a search engine while delivering search results in response of a search query, Search User Navigation Analysis using Videography, Filming the SERP user navigation behaviour, Filming search results browsing behaviour articles
17	Search Skills & Expertise	How the searcher skills make it more accessible to most relevant and precise information on search engine, How the expertise and exposure of a searcher impacts the retrieval of authentic results from a search engine
18	Purpose Stages	Purpose Stages - Behaviour, Awareness, Consideration, Conversion, Loyalty
19	Demographic & Ethical Behaviour	Ethical Behaviour, Socio-Cultural, Lifestyle, Geo-Market, Technical Behaviour, Technical Savvy, Govt. Policies
20	Market Trends & Influences	Market Trends, Social Trends, Family Choice, Individual Choice, Influencers Influence
21	Generational Search Behaviour	Search engines preparing themselves for Generation Z, Catering of mood and trends of generation next choices by search engines, Generation next search behaviour on search engine

22	Geographic Market Behaviour	Internet search user behaviour in french market, Internet search user behaviour of healthcare information seekers in french market
23	Healthcare Search Behaviour	Health information searching behaviour on search engines, Healthcare automation technology, tools and search engine results pages
24	Search Engine Self-Learning	Search Engines are made to learn, Search Engine self-learning, Search Engines apply "Search-as-Learning" on them, Self-learning of search engine through crowdsourced annotators, Crowdsourcing annotator models used by search engines for self-learning
25	Search Results Evaluation	Methods of evaluation of search results of a search engine, How do search engines determine the relevance of search results
26	Data Science & Technical Aspects	Technical aspects of data science used in serp generation, Types of data sciences used in search engine results pages production
27	Web Evolution & Technology	Web 4.0 and its impact on Search Engine Results Pages, Relation between semantic web (web 3.0) and web 4.0, Different between web 3.0 and web 4.0, Connection between Search Engine Results Pages and Metaverse Technology
28	Video Recording & Observation Methods	Selecting Video Recording as a Method of Data Collection for Search Engine results navigation behaviour, Observational tools designed for use with video recording
29	Comparative Search Engine Analysis	Comparative analysis of search engines

Appendix-9: Borrowed/Sourced Graphics & Visuals

(Tables, Diagrams, Charts, Infographics)

Rank	Country	Number of articles	Percentage (%)
1	United States	871	42.6
2	England	586	28.7
3	Canada	235	11.5
4	Ireland	68	3.3
5	Switzerland	40	2.0
6	Germany	28	1.4
7	Netherlands	28	1.4
8	Australia	16	0.8
9	India	12	0.6
10	France	9	0.4

Table-3: Statistics on the number of studies on online health information-seeking behaviour by top 10 countries | Source: Liu ZW. [2022]

Parameter	Web 1.0	Web 2.0	Web 3.0
Communication	Broadcast	Interactive	Engaged/Invested
Information	Static	Dynamic	Portable & Personal
Focus	Organization	Community	Individual
Personal	Home Pages	Blogs/Wikis	Lifestreams
Content	Ownership	Sharing	Curation
Interaction	Web Forms	Web Applications	Smart Applications
Search	Directories	Keywords / Tags	Context / Relevance
Metrics	Page Views	Cost Per Click	User Engagement
Advertising	Banners	Interactive	Behavioral

Research	Britannica Online	Wikipedia	The Semantic Web
Technologies (Data Representation)	HTML, FTP	HTML, XHTML, XML, Flash, JAVA	RDF, RDFS, OWL, Micro formats
Foundation Year	1994	2004	2006
Users Globally	45 million	1 billion +	8 billion +
Sites Volume	250,000	80,000,000	800,000,000
Coined By	Tim Berners-Lee	Dale Dougherty	John Markoff
Generation	Cognition	Communication	Co-operation
Known as	The Web or The Internet	Web of Documents	Web of Data
Human Interaction	Read only	Read-Write	Personalized (Read-Write-Execute)
Semantics	Not applicable	Not applicable	Applied Using RDFS/OWL
Metadata	Relational Schema	XSD/DTD	Ontology
Query Language	Simple SQL	SQL along XPath	SPARQL
Data Model	Relational	Relational Hierarchical	Graph
Linking	Hyperlink	Hyperlink	URI
Used for	To search data	To share data	To search, integrate, control data
Web	Visual Web	Programmable Web	Linked Data Web
Content	Owned	Shared	Consolidated
End point	Webpage	Web service endpoint	Data space
Years active	1989-2005	1999-2012	2006-ongoing

Table-8: Comparatives of Web 1.0, Web 2.0 and Web 3.0 | Source: *Asaad Khaleel* [2021], [59]*Faizar* [2018], [264]*Kusum* [2025] *and Patel* [2013]

[264] https://www.simplilearn.com/what-is-web-1-0-web-2-0-and-web-3-0-with-their-difference-article

Search Engine Type	Description	Examples
Social Search Engines	These search engines have social purposes in which a percentage of the income is allocated to projects related to sustainability. They enable searching for images, video, music, and web pages.	Ecosia Lilo
Science Search Engines	These search engines allow access to scientific-technical materials through specific searches. The bibliographical production on the analyzed subject is gathered.	Google Scholar Scientific Commons Sci-Hub
General Search Engines	These search engines are the most widely used ones and contain files stored on web servers that, with each search, offer the results of general content that is most relevant to the user.	Google Yahoo Bing
Safe search Engines	These searchers allow safe searches for children. They do not show icon images to prevent display of inappropriate images.	Safe Search for kids
Local Search Engines	These local search engines help to find videos, news, blogs, web pages, radio, and images.	Rambler (Rusia) Goo (Japón) Baidu (China)
Social Media Search Engines	These search engines are created specifically to find content from social networks, including blogs, microblogs, comments, bookmarks, and videos. Some offer the possibility of creating alerts to track real-time topics of interest.	Social mention Social Search TagBoard

Table-14: Classification of Search Engine Types | Source: *Reyes-Menendez et al.* [2018]

Research Methodologies	Research Methods
Study of methods to solve the research problem	Methods are used to undertake the research process
Rigorous analysis of the research methods	Techniques or procedures applied to study the research problem
Justification for using a particular research method to solve the research problem	Includes qualitative *(Interviews, Focus groups, observations, document analysis, oral history or life stories)* or quantitative *(Surveys, questionnaires, observation, document screening, experiments)* methods
Research Methodology outlines the way in which researchers can proceed with their research.	Research Methods refer to the approaches used to investigate a topic or subject.
Methodology examples can include qualitative, quantitative, or mixed methods approaches, as well as specific theoretical frameworks such as	Research methods include surveys, interviews, experiments, and observations.

grounded theory or phenomenology.	

Table-39: Differences between Research Methodologies and Research Methods | Source: [265]*Kolkova* [2023], [266]*Newcastle* [2024], [267]*MvsM-bizmanualz.com*

Pros	Cons
1. Less intrusive method for data collection *(Avoiding the observer effect)* 2. Provides enough detail to analyse the work environment and human interactions qualitatively and quantitatively 3. Allows researchers to analyse events retrospectively 4. Allows researchers to capture simultaneous complex interactions 5. Allows researchers to review consultations repeatedly 6. Creates a permanent and complete record 7. Potential for multiple viewing/reviewing 8. Higher inter-rater reliability (with the help of practice coding) 9. Can be used to establish connections between perceptions and the observed activities during the visit 10. Retains the captured data with no loss of its richness for reviewing 11. Enables self-evaluation and reflection 12. Generates a large amount of data 13. Allows researchers to capture activities in much of their complexity in their natural settings over an extended period 14. Allows for scientific rigor when conducted by trained researchers 15. Can be reviewed by both researchers and participants, increasing the scope of interpretation	1. Reviewing and coding video data is labour intensive 2. Requires additional Institutional Review Board ***(IRB)*** procedures 3. Raises concerns about the discoverability and confidentiality of participants 4. Additional equipment cost 5. Additional data management concerns 6. Aggregation can be difficult and intrusive 7. It can limit range of settings 8. Possibility of **Hawthorne effect** 9. Higher overall cost

Table-45: Pros and Cons of Videorecording behaviour observation method | Source: *Asan & Montague* [2014]

[265] https://www.youtube.com/watch?v=BH3j2hFk0aY
[266] https://libguides.newcastle.edu.au/researchmethods
[267] https://www.bizmanualz.com/write-better-procedures/methods-vs-methodology.html

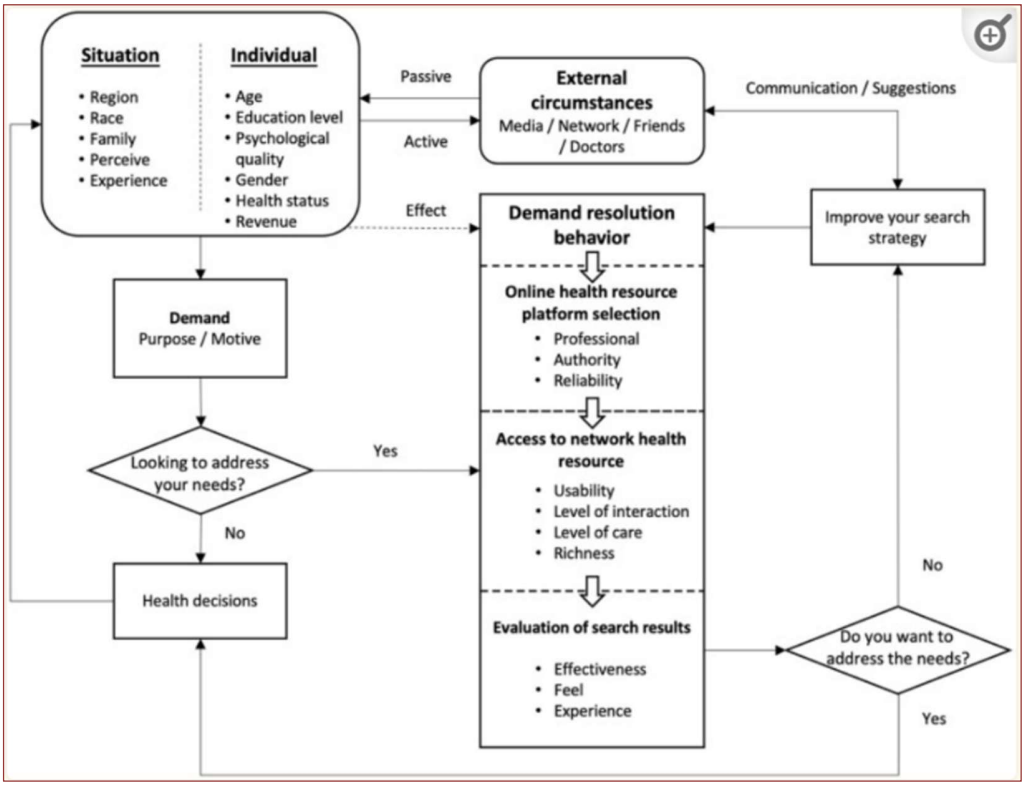


Diagram-1: Online health information seeking behaviour model | Source: Liu ZW. [2022]

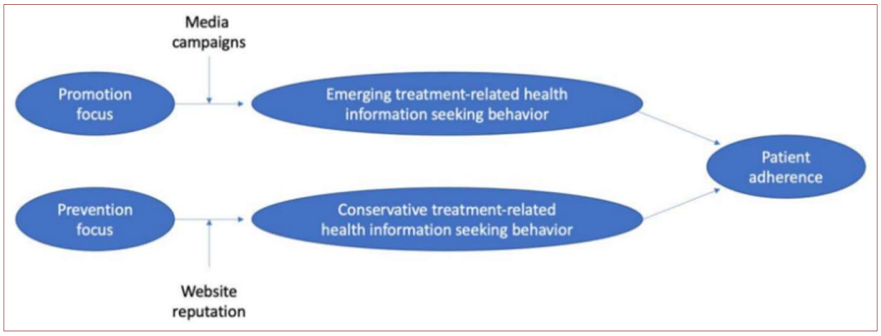


Diagram-2: Health information consumers focus model | Source: *Zhang et al.* [2019]

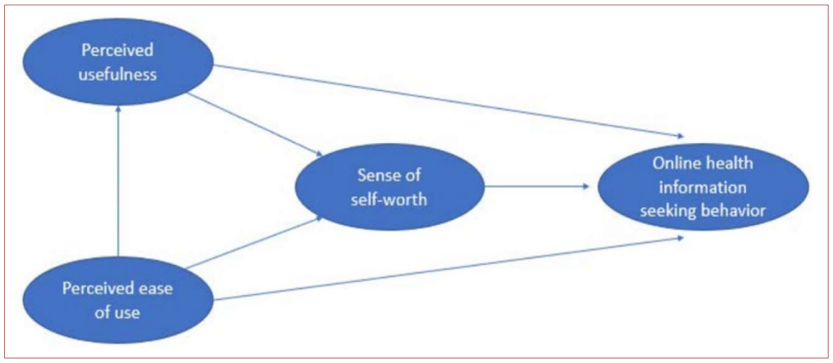


Diagram-3: Online health information seeking behaviour | Source: *Boyce et al.* [2022]

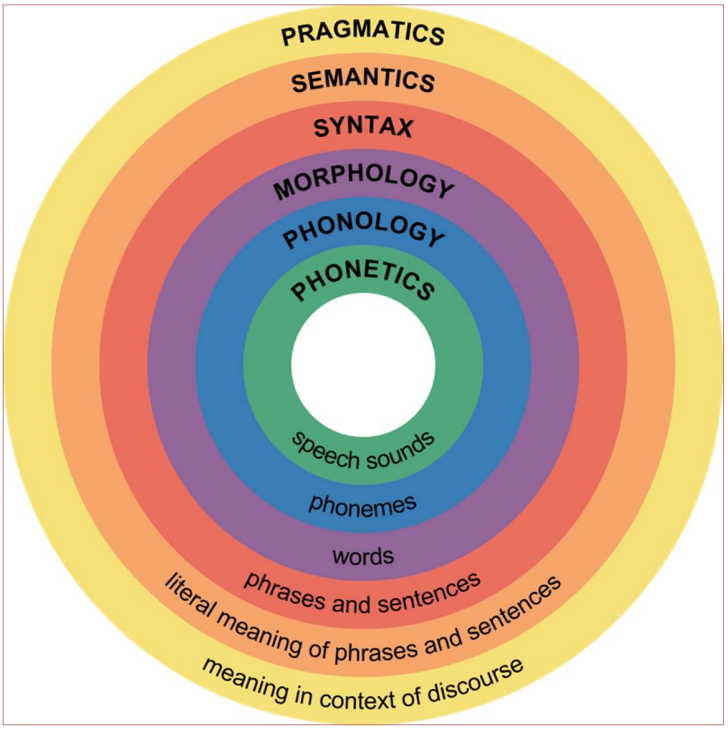


Diagram-5: Major levels of linguistic structure | Source: [268] *Wiki Syntax*

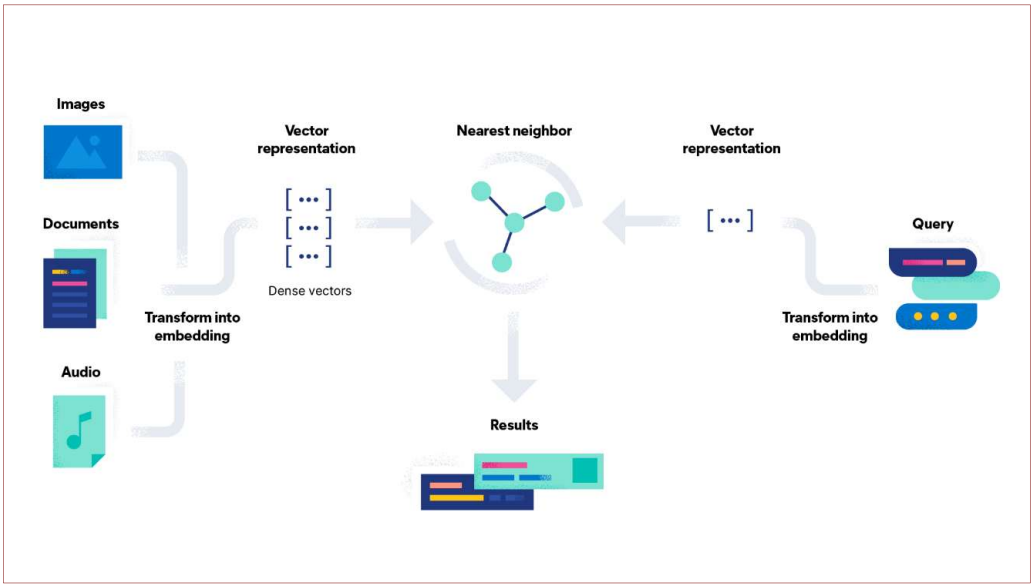


Diagram-6: Functioning model of Semantic Search by [51] elastic.co-1

[268] https://en.wikipedia.org/wiki/Syntax

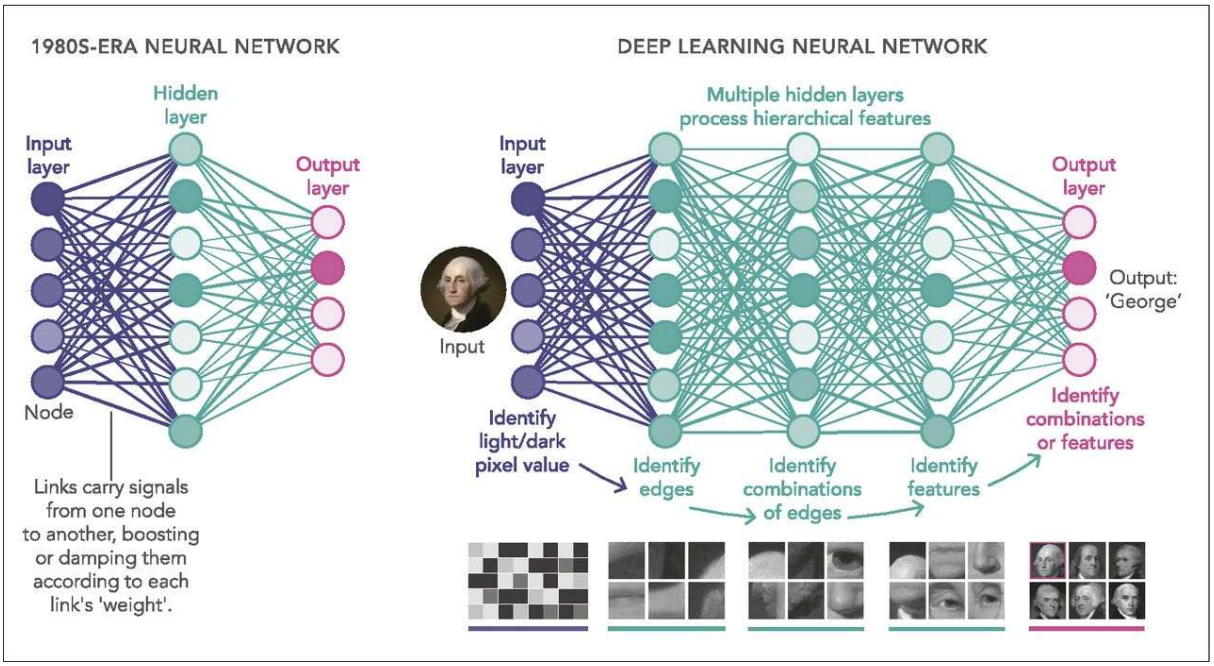


Diagram-9: Neural Network vs. Deep Learning Neural Network by Lucy Reading-Ikkanda | Source: [110]*Mitchell Waldrop* [2019]

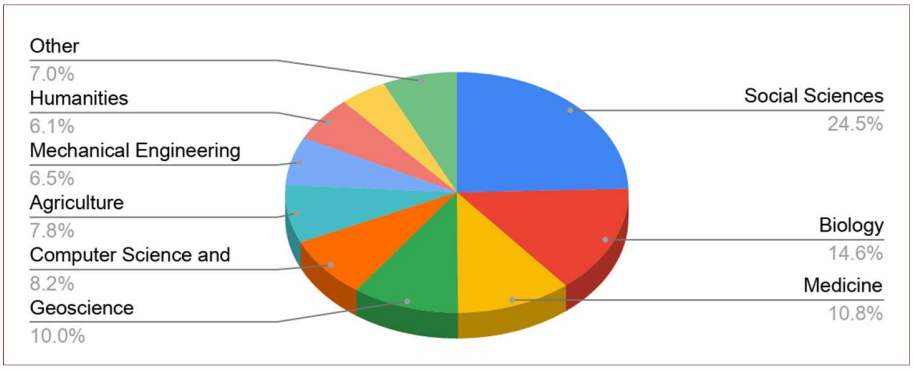


Chart-1: Topic distribution of datasets that appeared in search results | Source: *Benjelloun et al.* [2020]

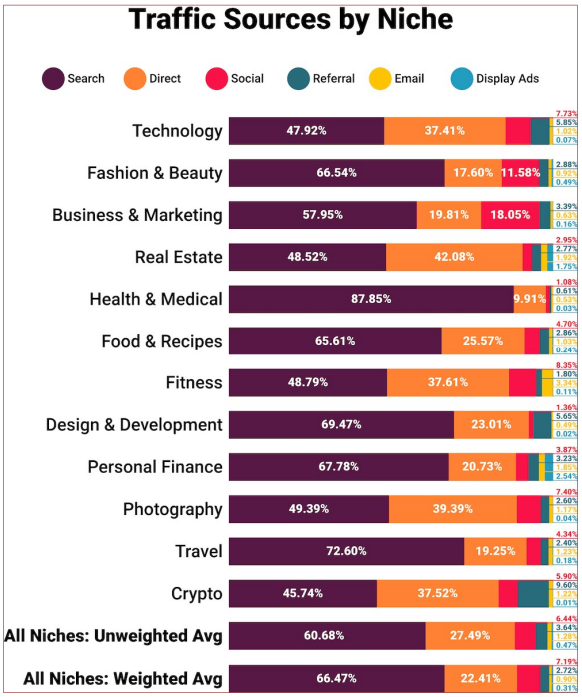


Chart-2: Traffic Source by Niche | Source: [5]*Kyle Byers* [2020]

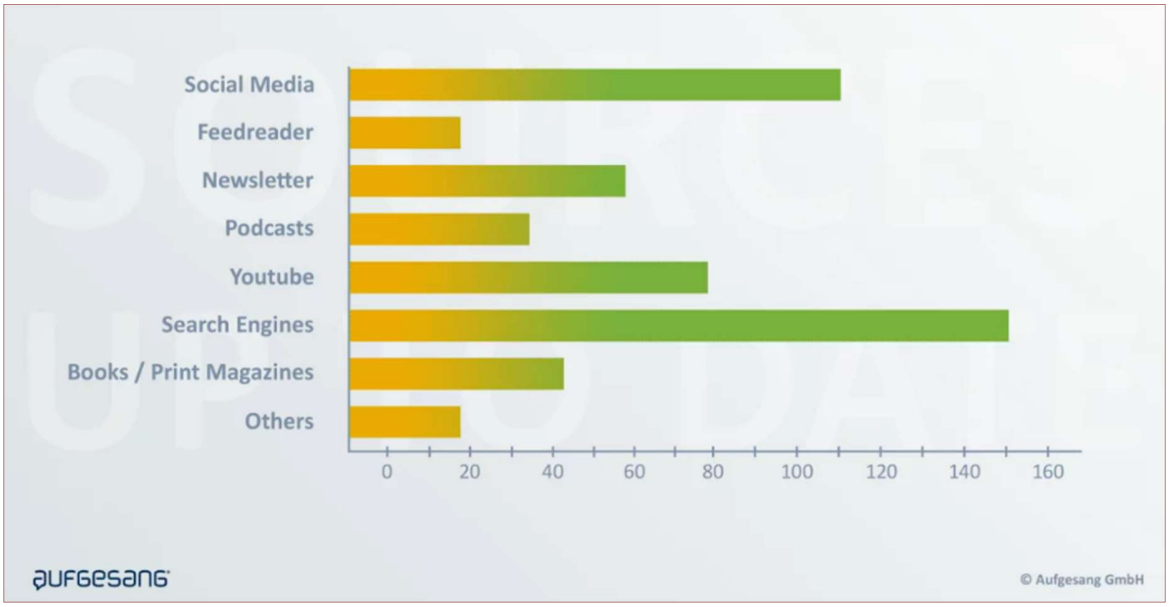


Chart-3: Most important sources of information Jan-2024 | Source: [6]*Olaf Kopp* [2024]



Chart-9: Arabic, English and Chinese heatmaps on Rich PC SERP | Source: *Alsaffar et al.* [2018]



Chart-10: Arabic, English and Chinese heatmaps on Plain PC SERP | Source: *Alsaffar et al.* [2018]

Topical Relevance Type	On topic	Relationship between overall document topic & overall user need topic	Diagram Illustrating Relationship[1]	Relatedness to the given topic	Specificity of informa-tion	Richness of relevant information
Direct	Directly	Matching		High	High	High ~ medium
Indirect	After inference	Matching		High ~ medium	High	High ~ medium
Context	No	Surrounding		Low ~ medium	Low	Low ~ medium
Comparison	Partially	Comparing/ contrasting		Low ~ medium	High ~ medium	Medium ~ low
Pointer	No	Apart		Low	High	Low

[1]In the diagram, the document topic is black; the user need topic is white.

Chart-13: Types of Topical Relevance | Source: *Huang and Soergel* [2004]

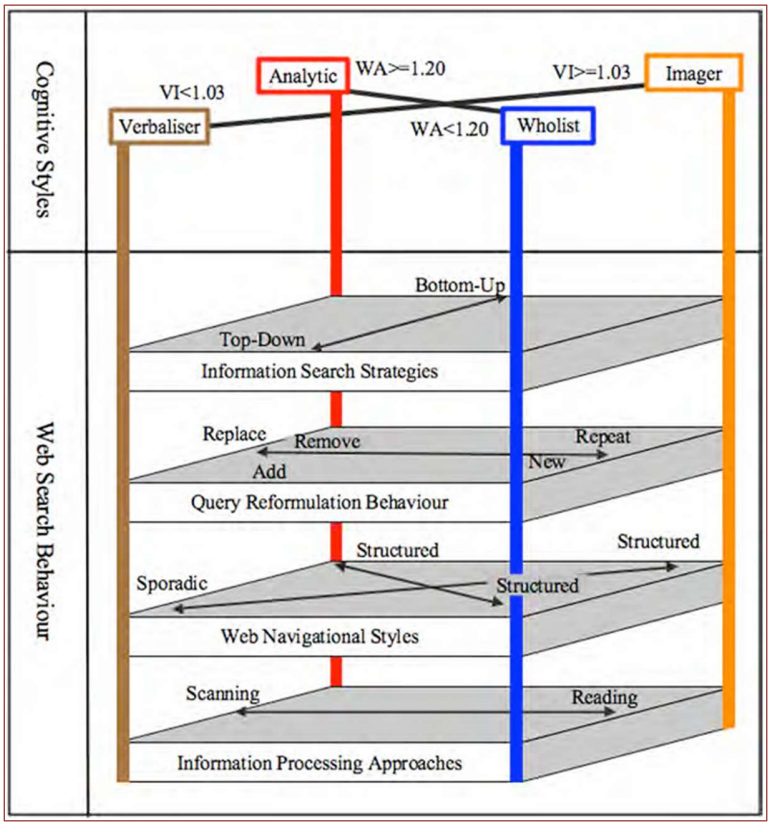


Chart-14: Model of Associations between Cognitive Styles and Web Search Interactions | Source: Khamsum et al. [2014]

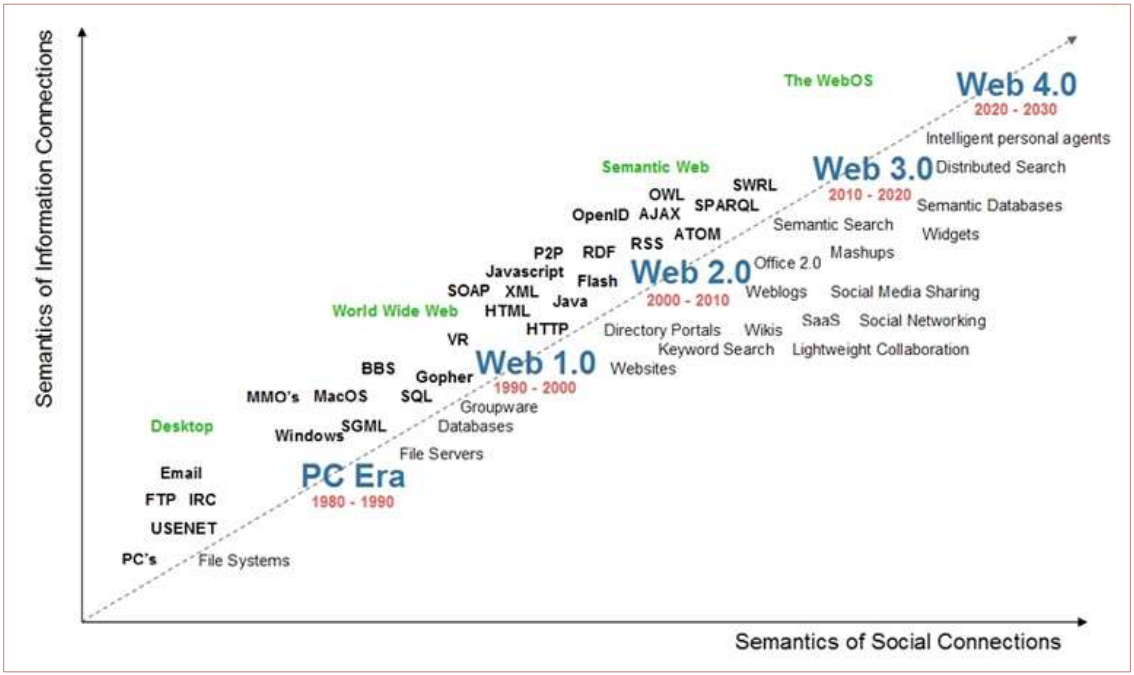


Chart-16: Timeline of Semantics of Information Connections | Source: [59]*Faizar* [2018]

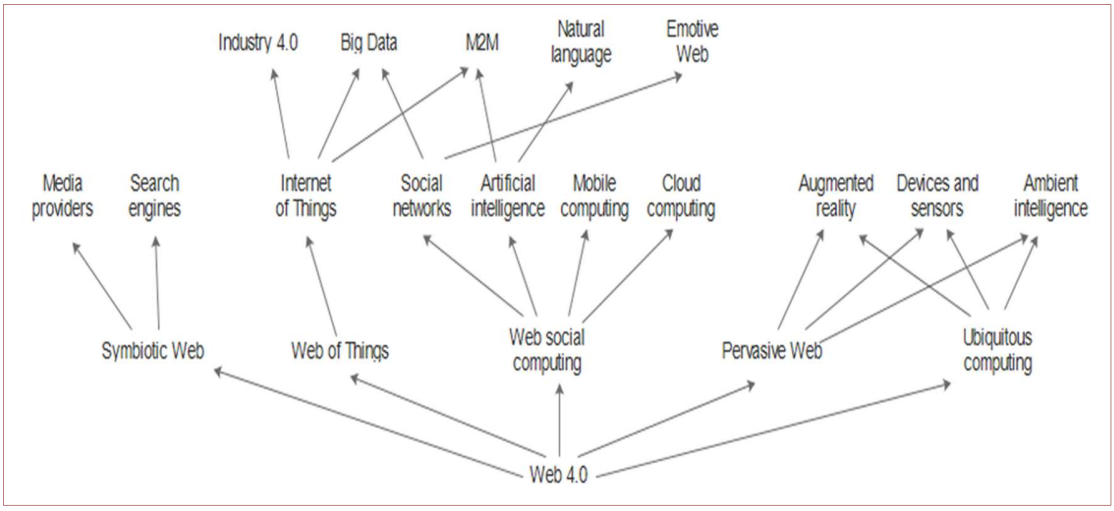


Chart-17: SODA representation of Web 4.0 | Source: *Almeida* [2017]

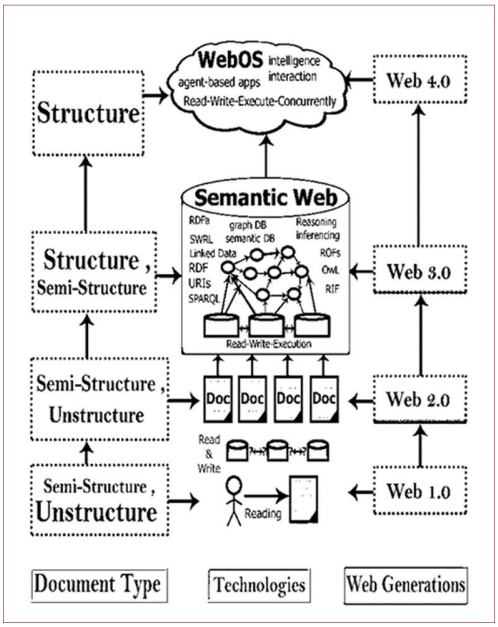


Chart-18: Web technologies and document types | Source: *Asaad Khaleel* [2021]

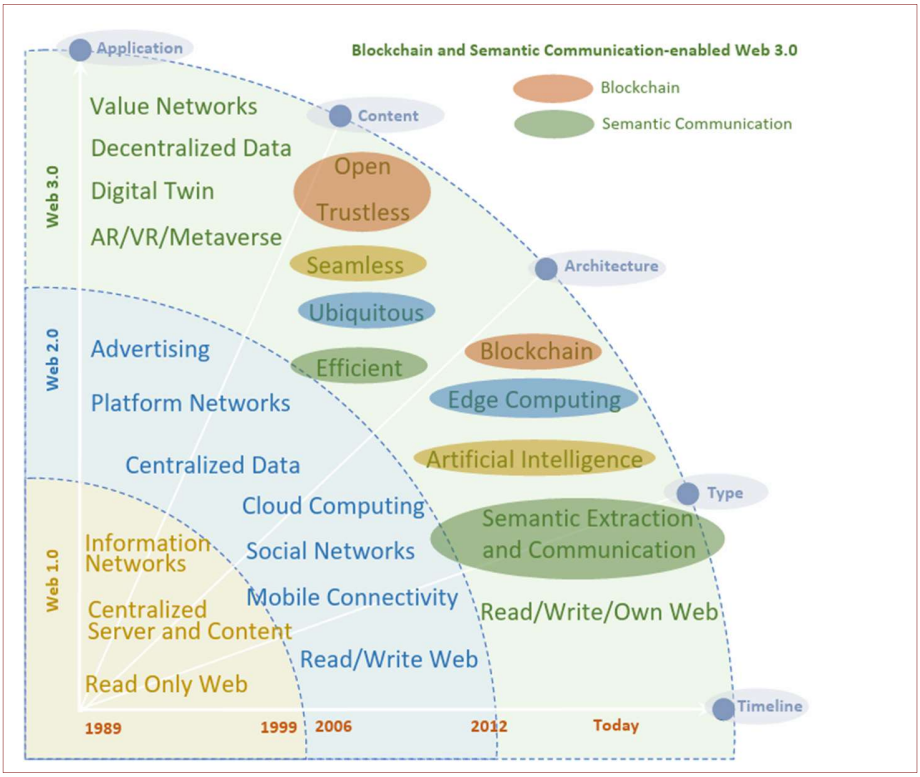


Chart-19: Difference Between Web 1.0 Web 2.0 and Web 3.0 | Source: *Lin et al.* [2022]

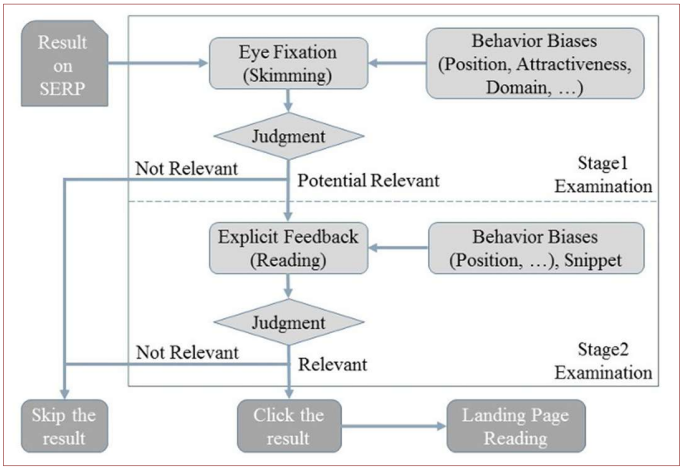


Chart-21: Two stage SERP results examination model | Source: *Liu et al.* [2014]

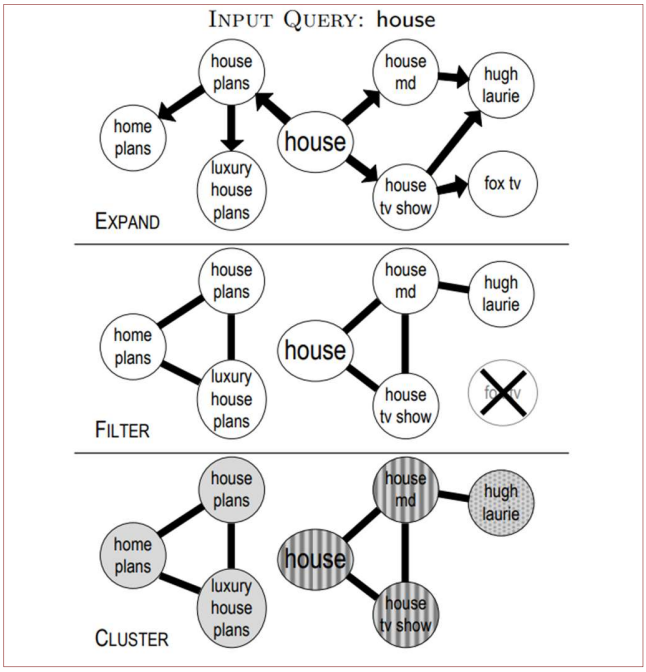


Chart-23: 3 Steps of Intent Identification | Source: *Radlinski et al.* [2010]

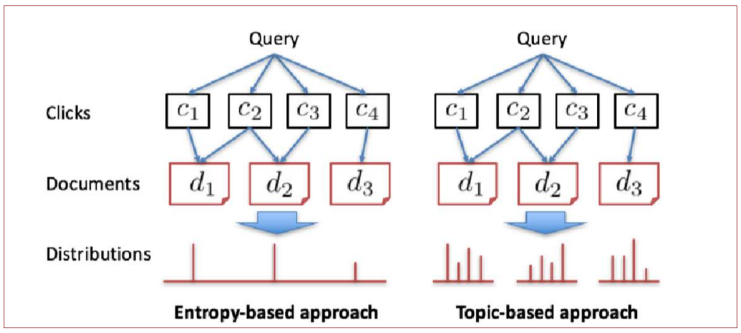


Chart-24: Quantifying Query Ambiguity with Topic Distributions Model | Source: *Yano et al.* [2016]

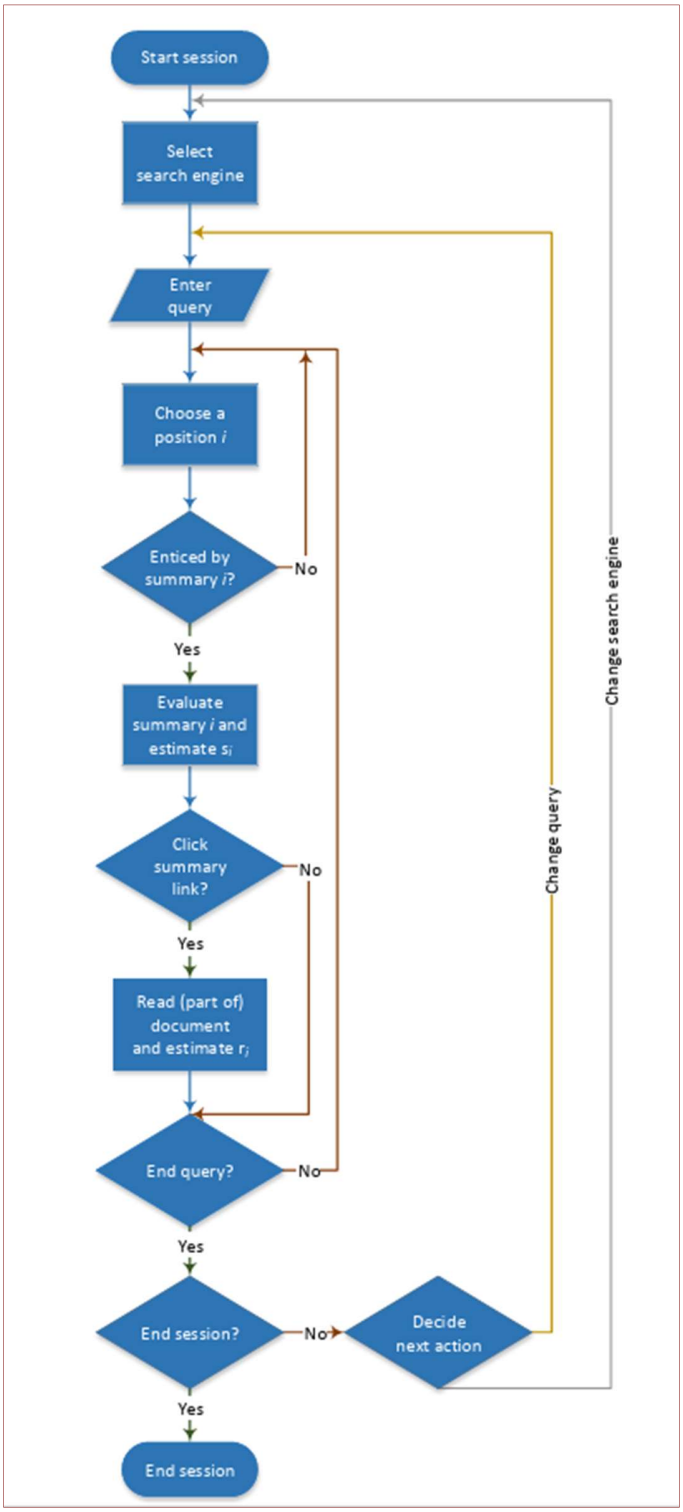


Chart-25: End-to-end-search-process on Search Engines | Source: *Thomas et al.* [2014]

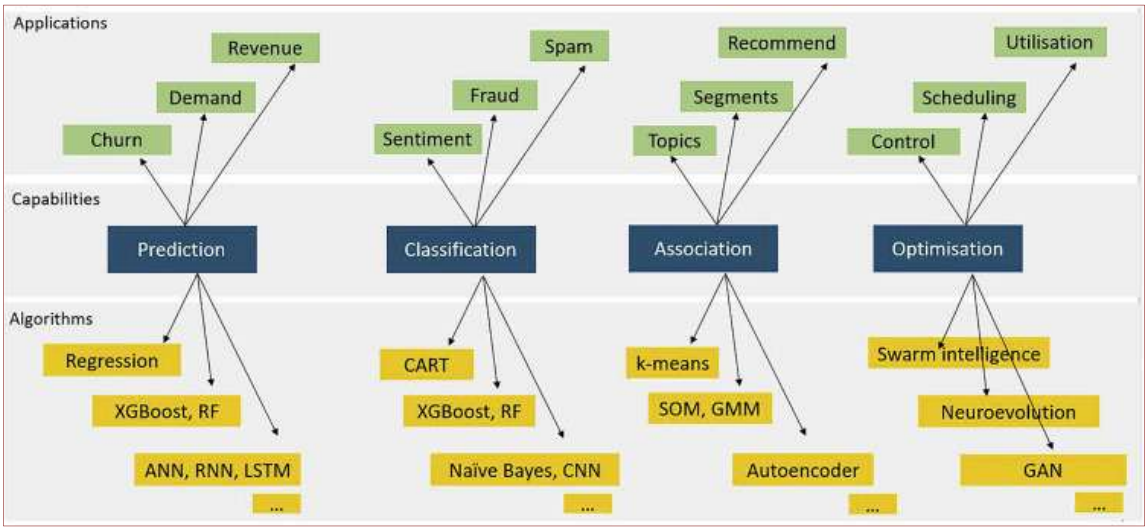


Chart-29: Taxonomy of AI Algorithms, Capabilities and Applications | Source: *De Silva & Alahakoon* [2022]

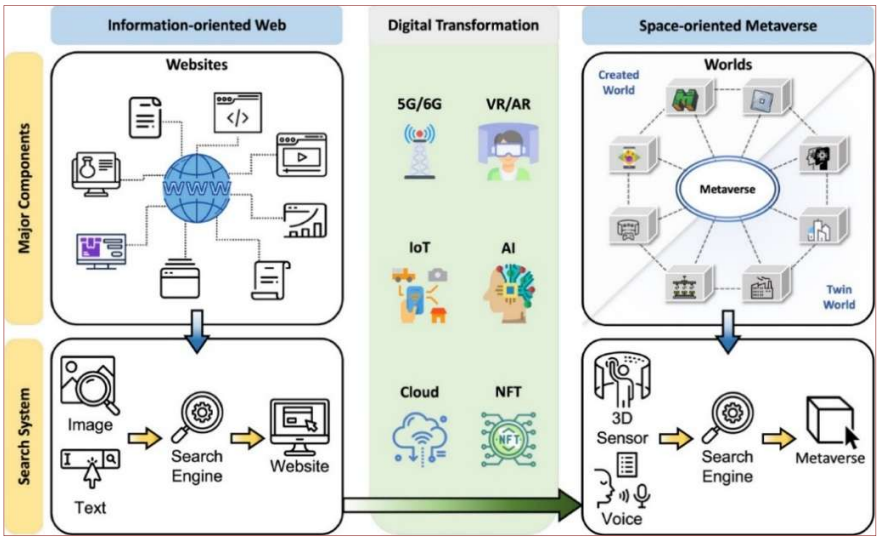


Chart-34: Digital Transformation between WWW and Metaverse | Source: *Yang et al.* [2023]

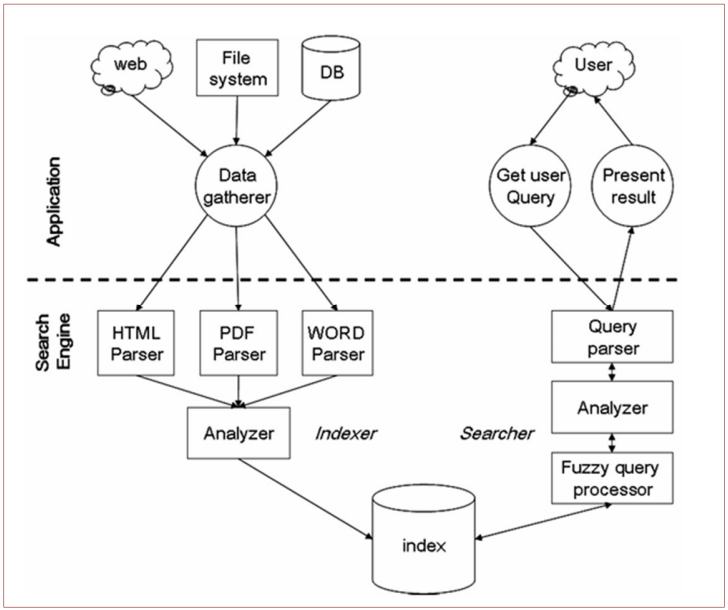


Chart-37: Architecture of a full text search engine; **Source:** *Khaled* [2007]

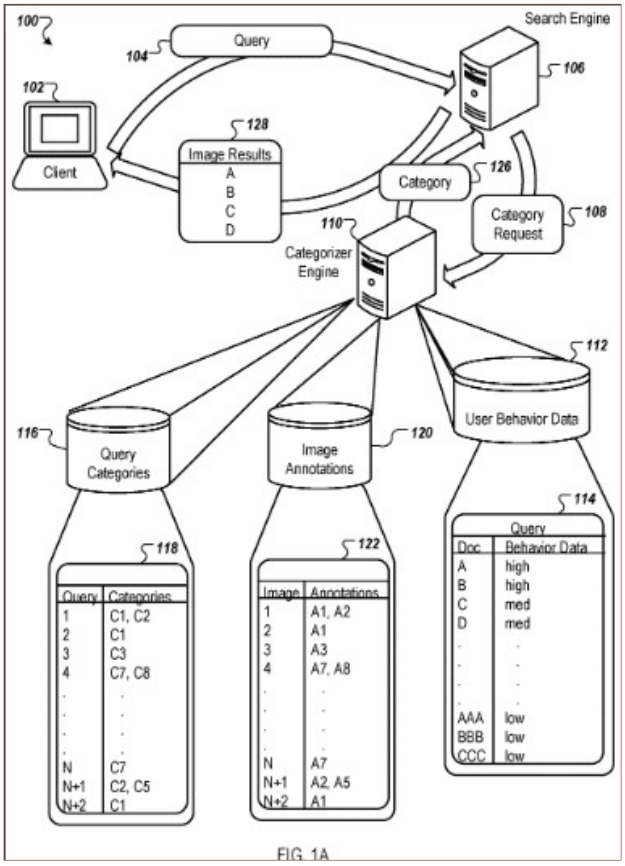


Chart-38: Query categorization based on image results; **Source:** [269]*Olaf Kopp [2023]*

[269] https://www.kopp-online-marketing.com/interesting-google-patents

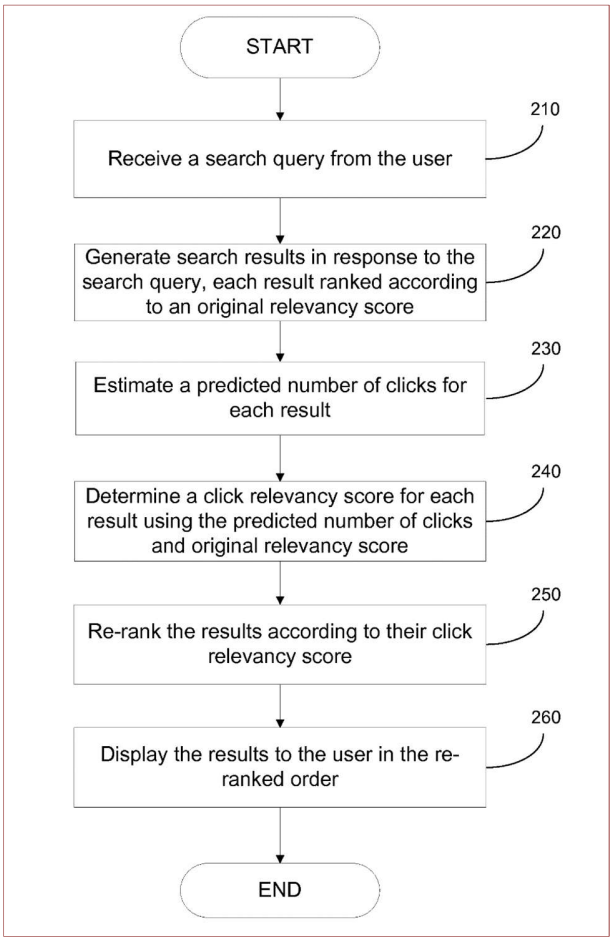


Chart-39: Google Patent Model of Responding to Search Query | Source: *Varma & Jain* [2010]

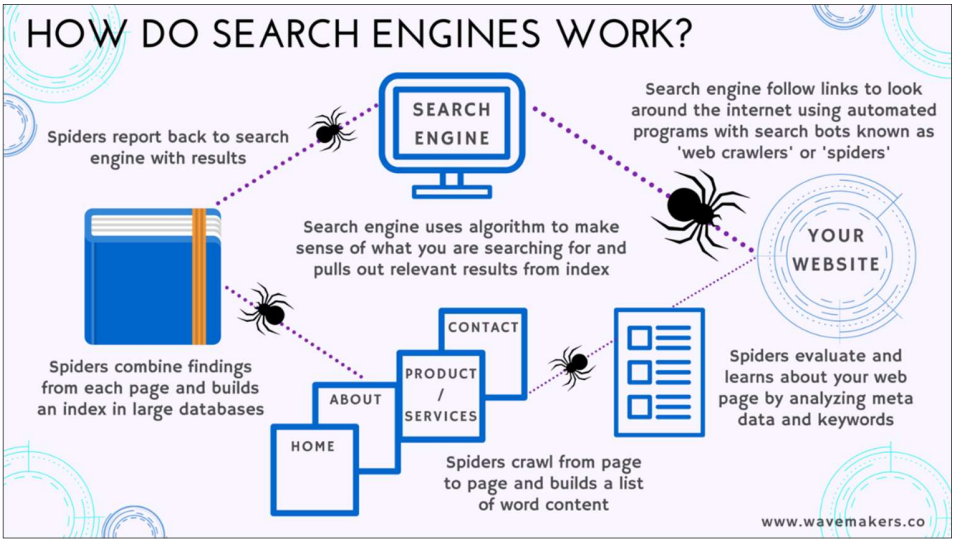


Chart-40: How does Search Engine work | Source: [174]*Kaviani Siavosh* [2019]

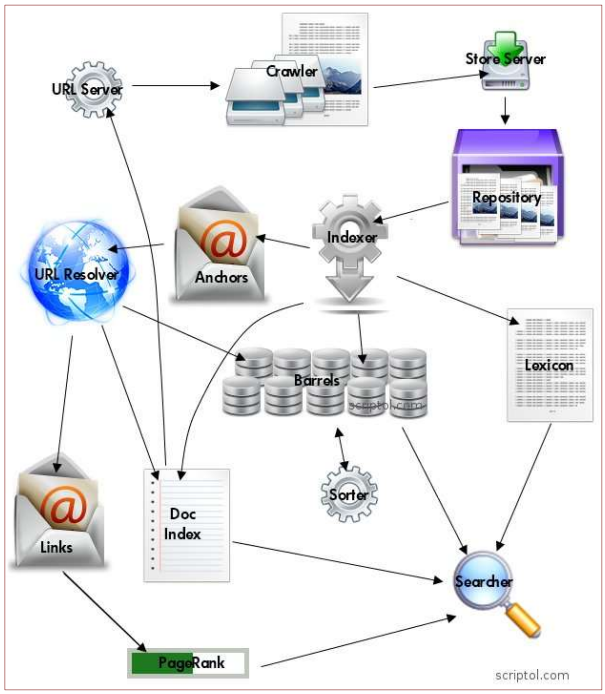


Chart-41: Anatomy of a Search Engine-1 | Source: [175]*Scriptol.com* [2010]

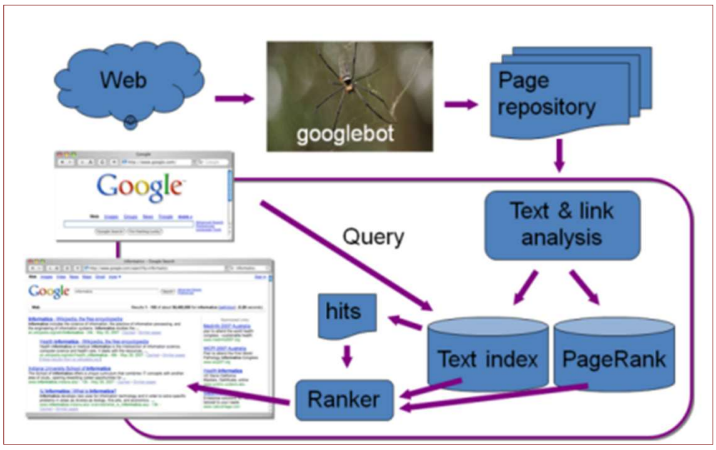


Chart-42: Anatomy of a Search Engine-2 | Source: *Gupta & Sharma* [2014]

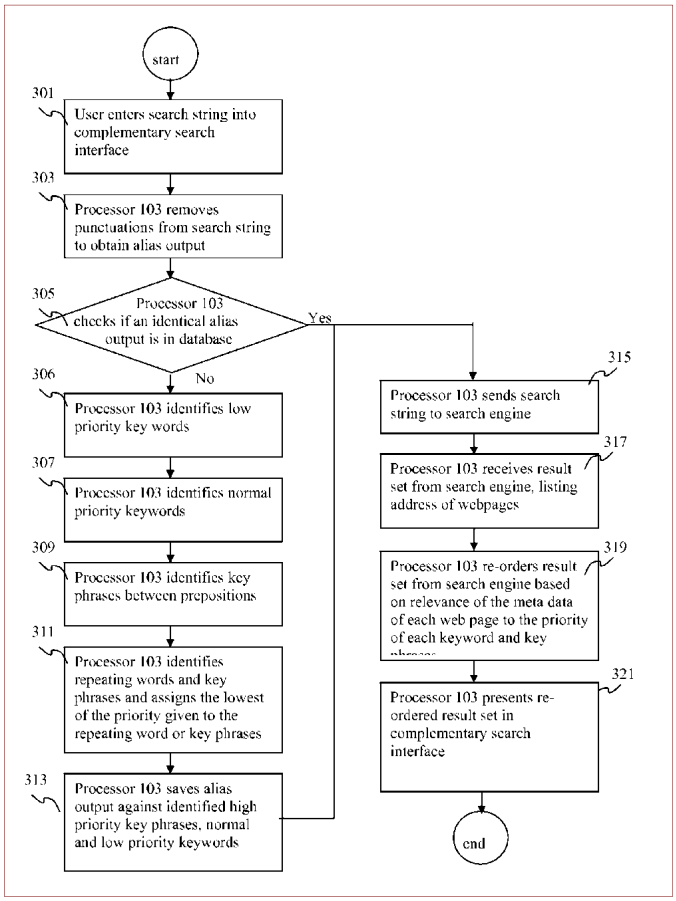


Chart-46: Search Engine Theoretical Algorithm of Fetching Results; **Source:** *Lung et al.* [2009]

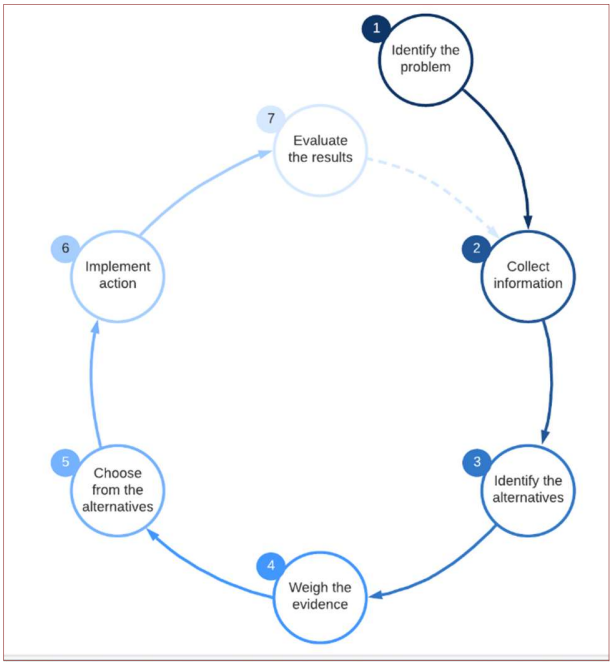


Chart-47: Lucidchart Decision Making Cycle around a Search need | Source: [184]*Lucidchart-1*

Words & SERP Features Semrush Uses to Identify Keyword Intent				
	Navigational	**Informational**	**Commercial**	**Transactional**
Words Included in Search	Brand names	How, what, tutorial, why	Best, review, comparison	Price, buy, near me
SERP Features & Appearance	Sitelinks, knowledge panel, the number of times a business domain appears in the top 10 results	News, recipes, knowledge panel, featured snippet	Local pack, Google Ads, jobs, review stars	Popular products, shopping results

Chart-48: Search Intents and SERP Appearance Pairing | Source: [189]*Jovicic* [2023]

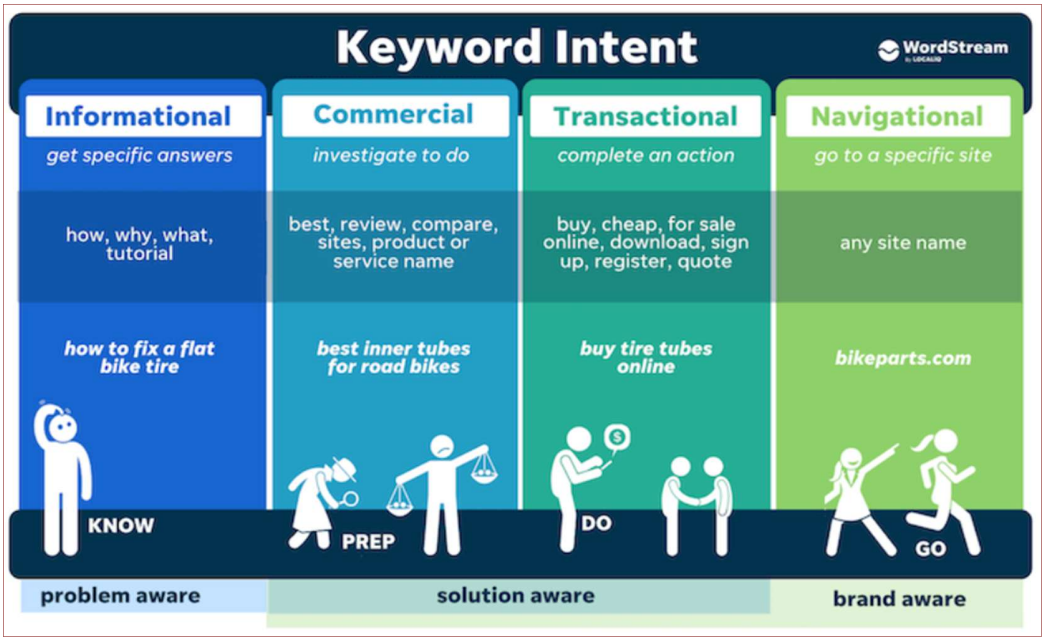


Chart-49: Search Intents mapping to Purchase Cycle | Source: [270]*wordstream-2*

[270] https://www.wordstream.com/keyword-intent

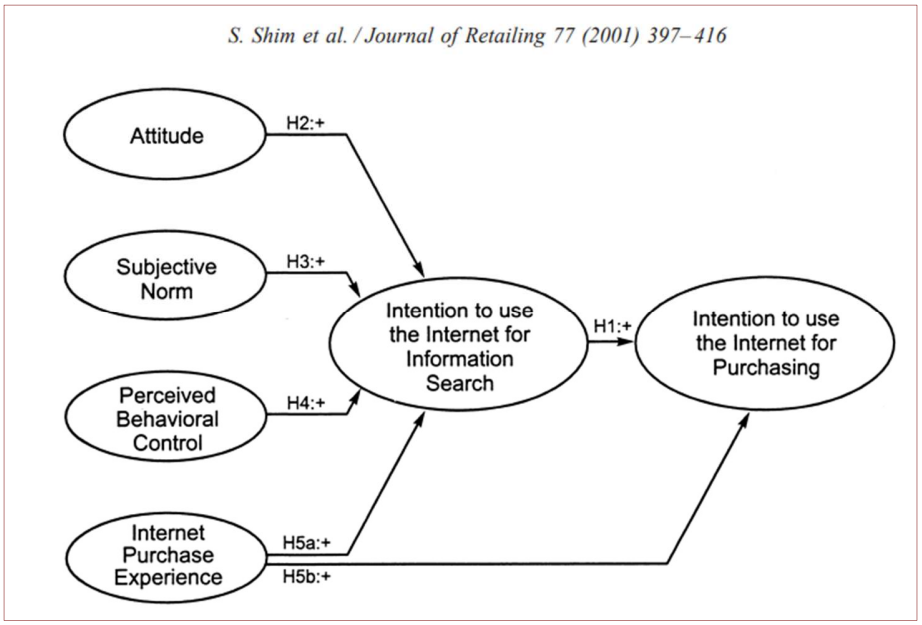


Chart-50: Theoretical Model Predicting Online Purchase Intentions | Source: *Shim et al.* [2001]

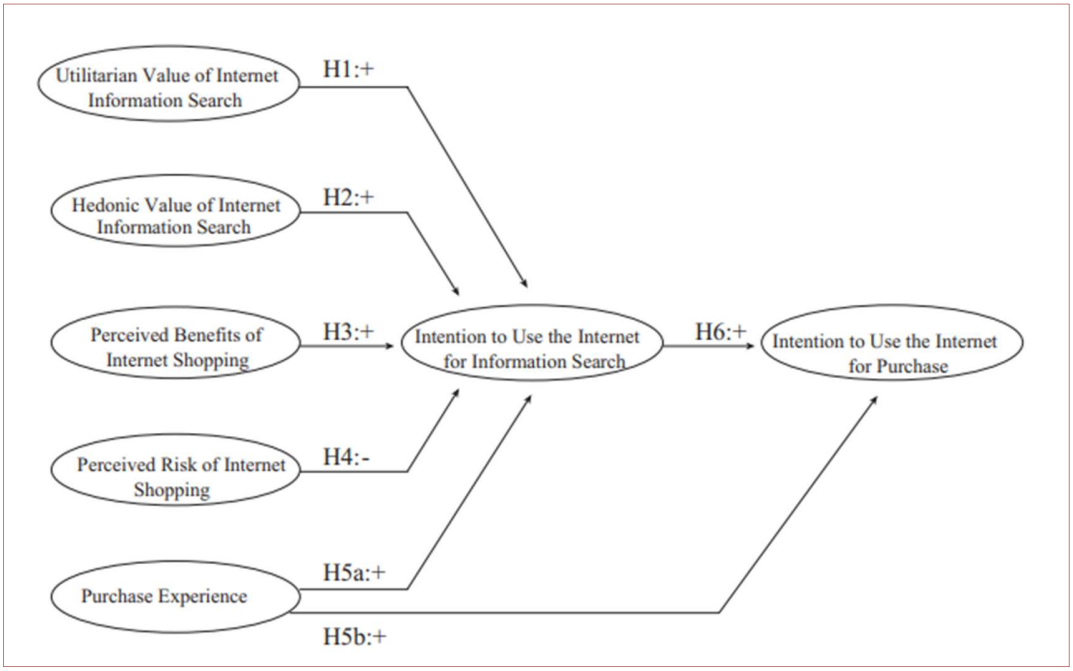


Chart-51: The Relationships Among Predictors of Online Search Intention and Online Purchase Intention | Source: *Kim & Lee* [2004]

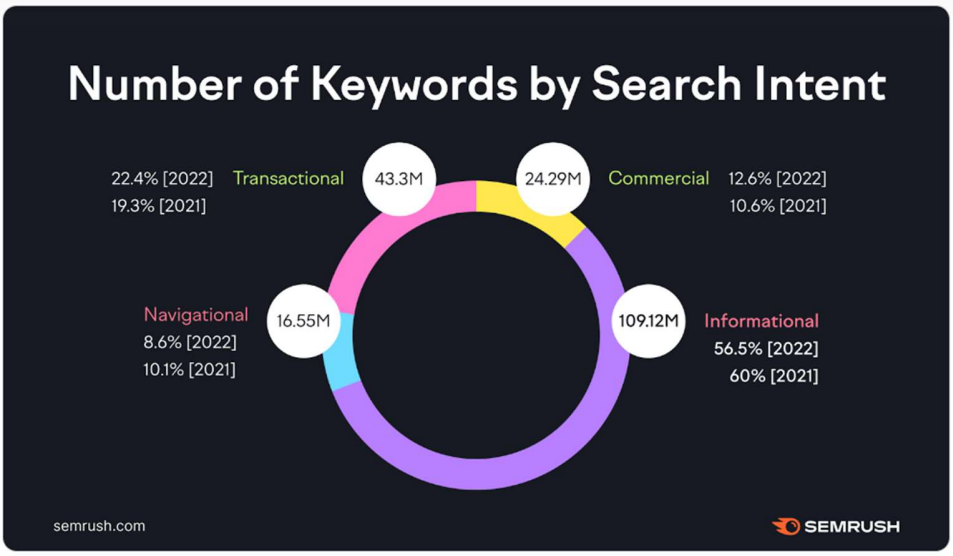


Chart-52: Number of Keywords by Search Intent | Source: [271]*Semrush Team* [2023]

GOAL	SET 1	SET 2	SET 3
directed	2.70%	3.30%	7.30%
undirected	31.30%	26.50%	22.70%
advice	2.00%	2.70%	5.00%
locate	24.30%	25.90%	24.40%
list	2.70%	2.90%	2.10%
informational total	**63.00%**	**61.30%**	**61.50%**
download	4.30%	4.30%	5.60%
entertain	4.00%	8.20%	5.80%
interact	5.70%	4.30%	6.00%
obtain	7.70%	10.30%	7.70%
resource total	**21.70%**	**27.00%**	**25.00%**
navigational	**15.30%**	**11.70%**	**13.50%**

Chart-53: Results of Classifying Queries by Search Goals (Intent) | Source: *Rose & Levinson* [2004]

[271] https://www.semrush.com/blog/seo-statistics/

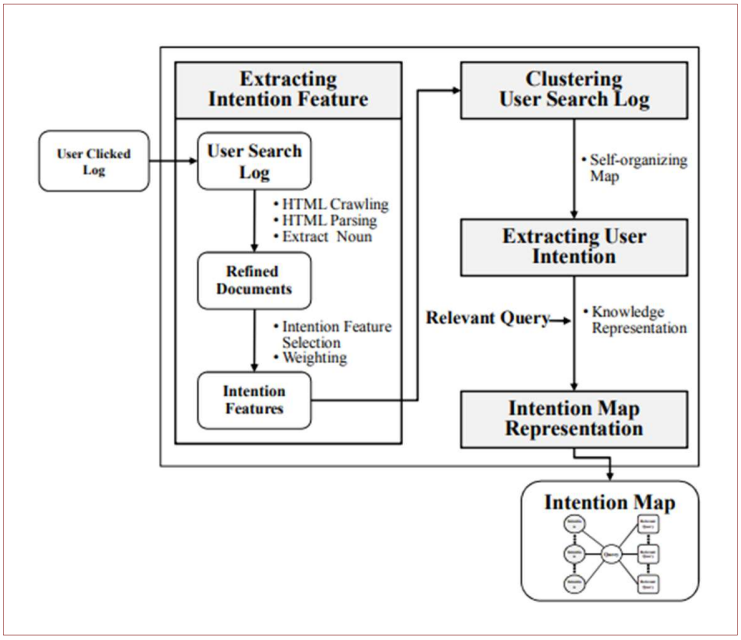


Chart-54: Estimating Intent Types for Search Result Diversification | Source: *Park et al.* [2010]

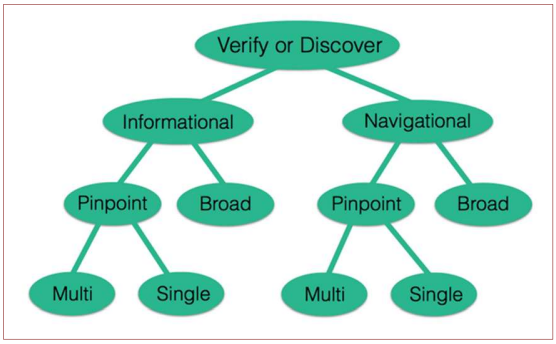


Chart-56: Hierarchy of search intentions for the context search engine | Source: *Kori et al.* [2015]

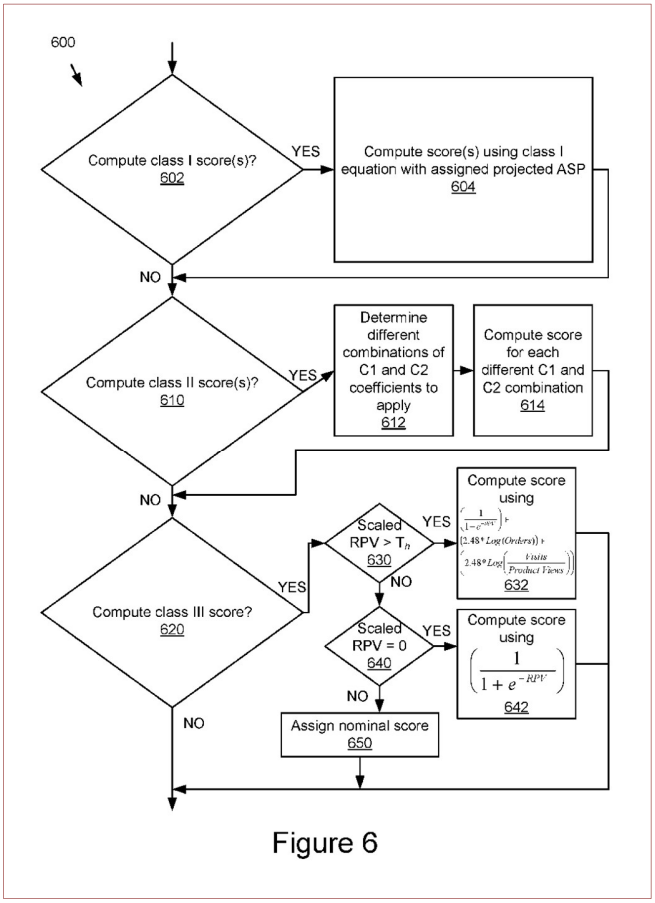


Chart-57: Google Patent model of ranking search results based on customer intent | Source: *Taraknath et al.* [2017]

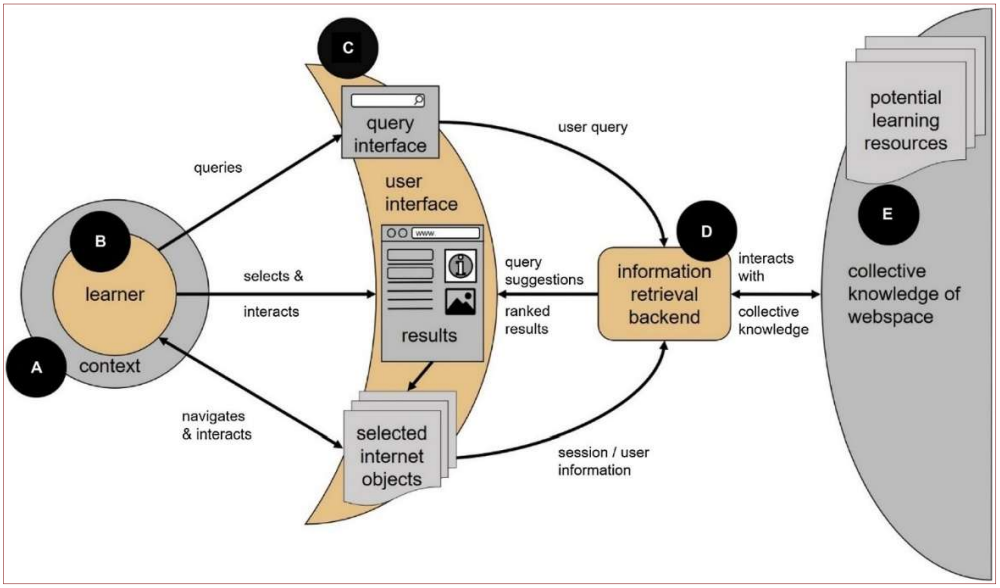


Chart-60: Searching as Learning Spaceship Model | Source: *Hoyer et al.* [2022]

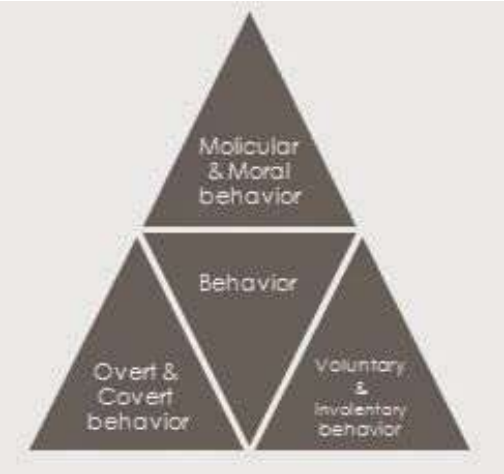


Chart-61: Behavioural Model by *Dubey & Jain* [2017]

How likely is this intent?	How well does this LP satisfy this intent?					
	Strongly	Moderately	Weakly	Poorly	Obscene content	Content inaccessible
Most Likely >50%	Excellent [Perfect]	Good	Fair	Bad	Detrimental	No Judgment
Very Likely 26-50%	Excellent	Good	Fair	Bad	Detrimental	No Judgment
Likely 11-25%	Good	Fair	Bad	Bad	Detrimental	No Judgment
Unlikely 1-10%	Fair	Bad	Bad	Bad	Detrimental	No Judgment
Very Unlikely <1%	Bad	Bad	Bad	Bad	Detrimental	No Judgment

Chart-62: Intent Likeliness versus Satisfaction behind a searcher query; Source: [272]McGee [2012]

[272] https://searchengineland.com/bing-search-quality-rating-guidelines-130592

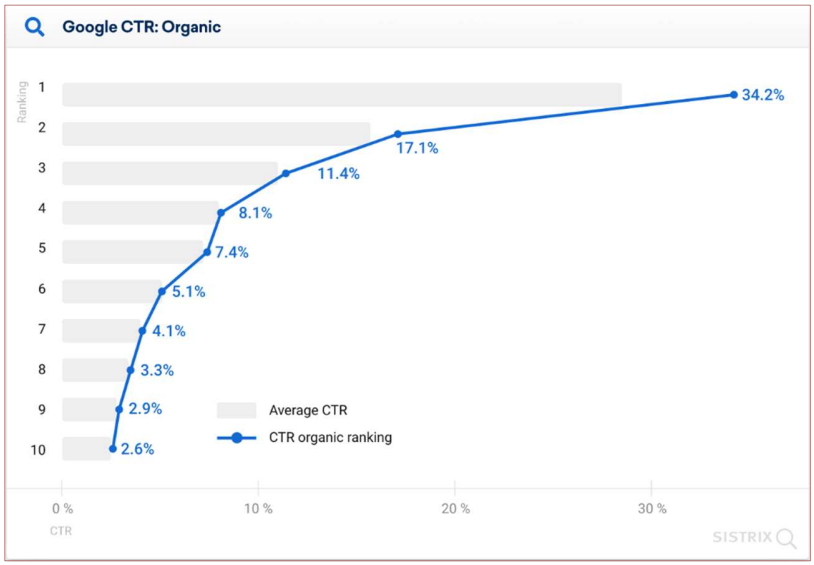


Chart-64: Google CTR Organic | Source: [236]*Dave Chaffey* [2024]

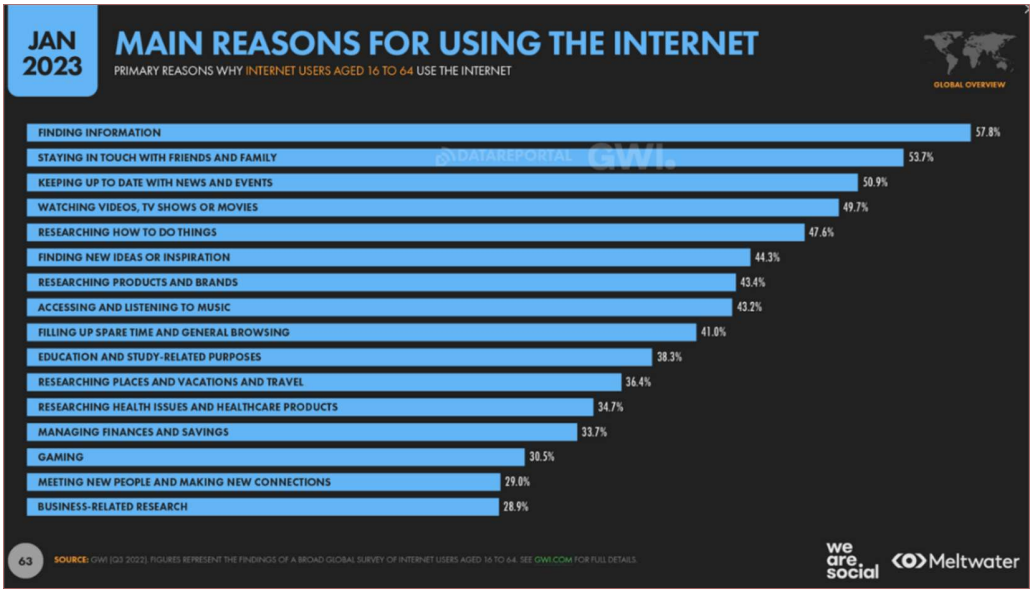


Chart-65: Main reasons for using the internet | Source: [236]*Dave Chaffey* [2024]

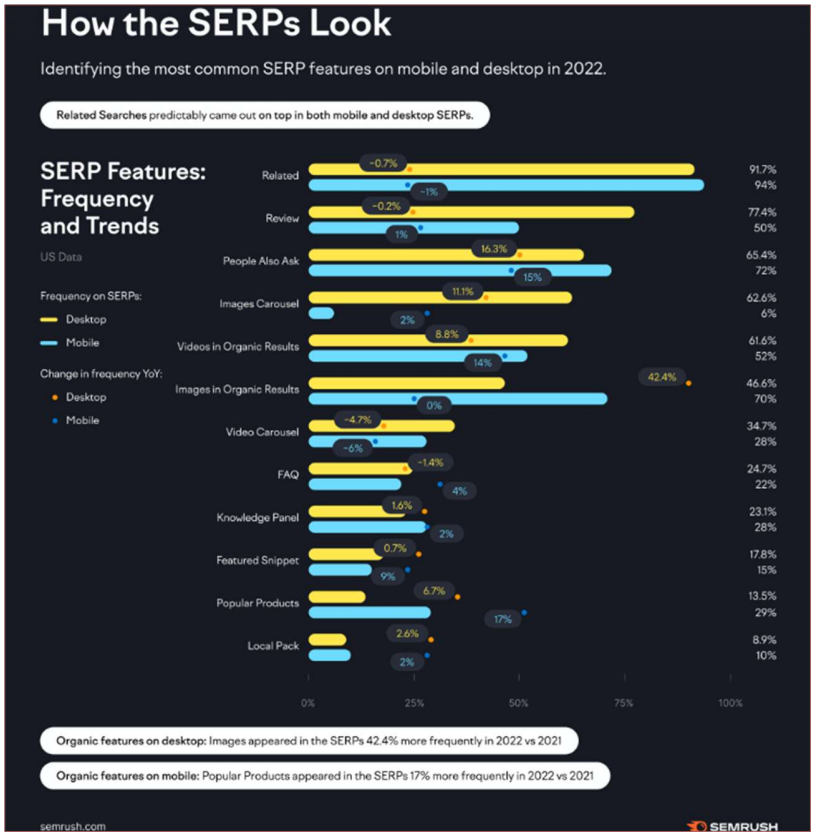


Chart-66: How the SERPs Look | Source: [271]*Semrush Team* [2023]

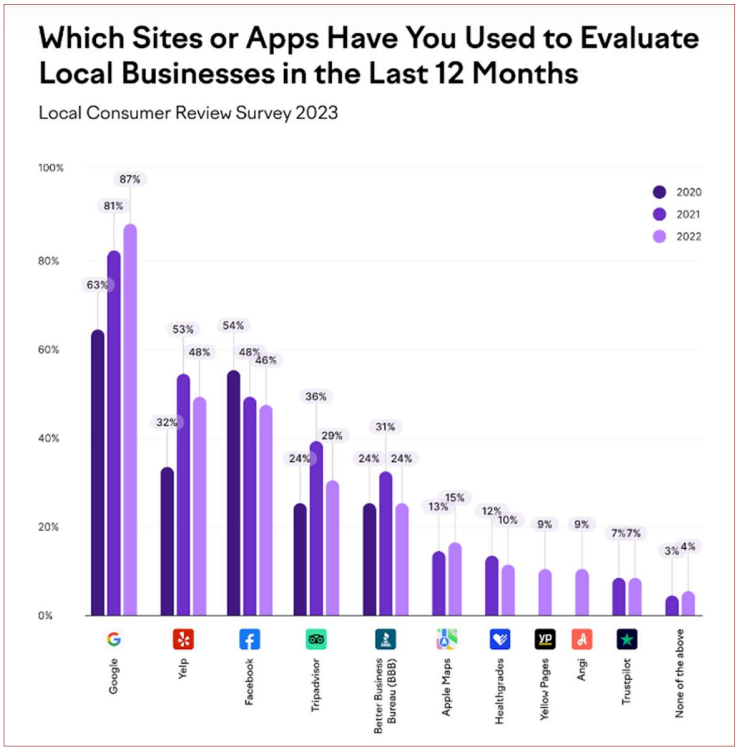


Chart-67: Site or Apps used to evaluate Local Business | Source: [271]*Semrush Team* [2023]

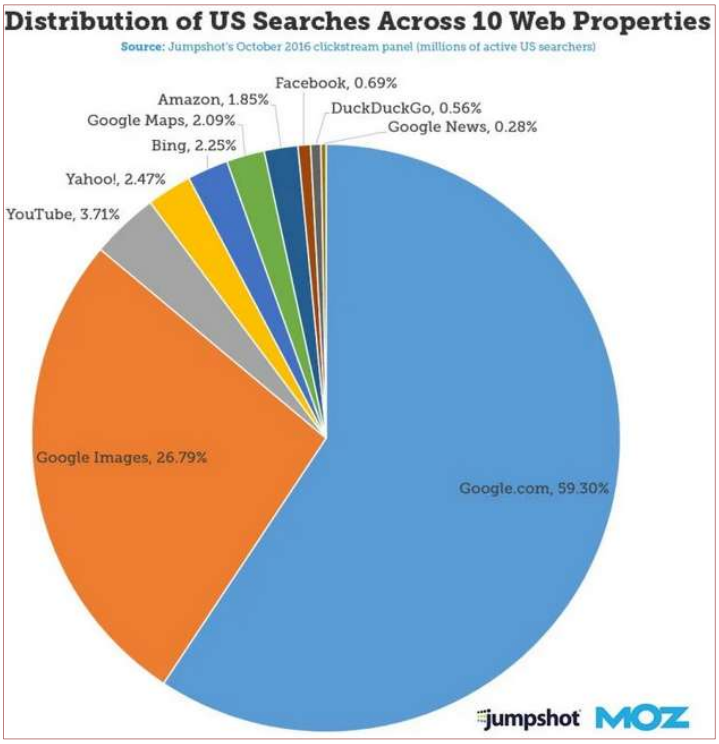


Chart-68: Distribution of US Searches across 10 Web Properties

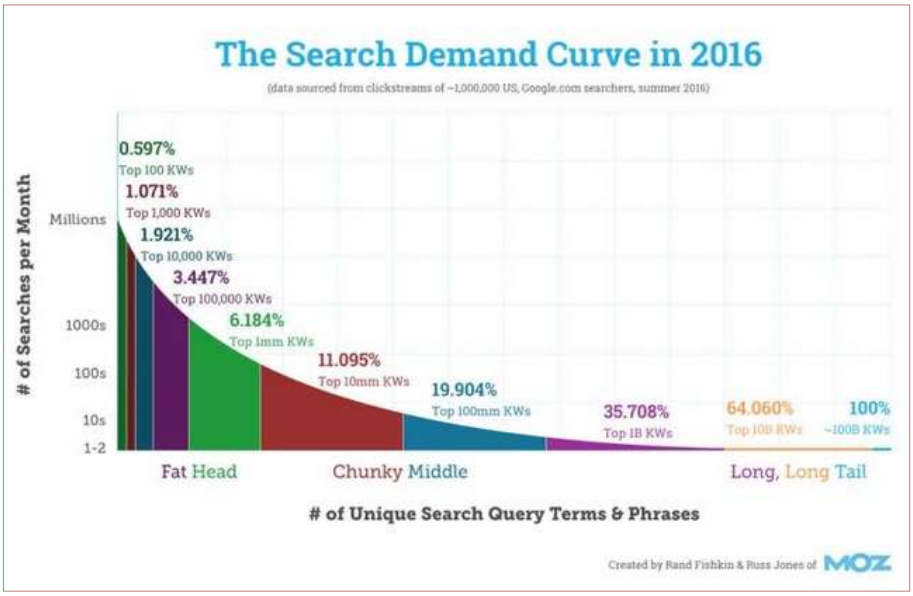


Chart-69: The search demand curve in 2016

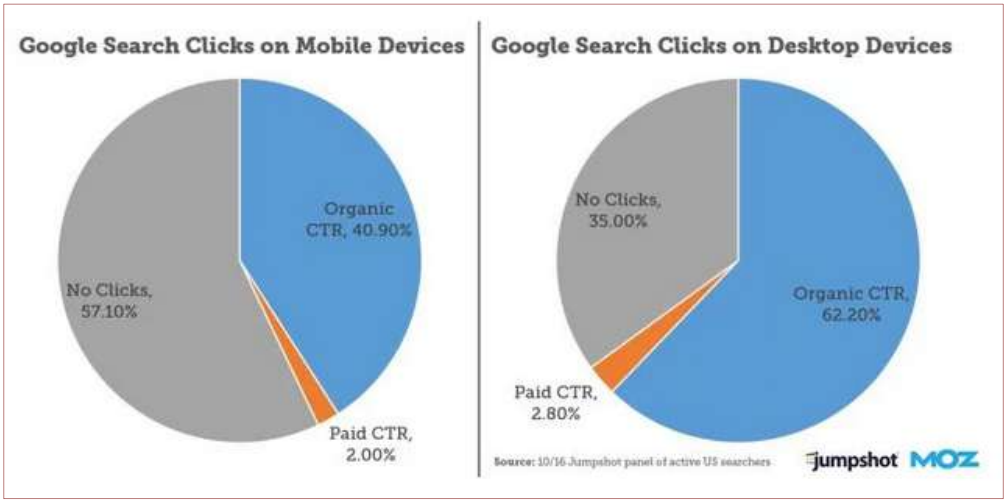


Chart-70: Paid vs. Organic CTR difference on mobile versus desktop

Search Engine Market Share Oct. 2018 to Sep. 2019	
Search Engine	**Share**
Google	75.47%
Bing	9.80%
Baidu	9.27%
Yahoo!	2.85%
Yandex	1.22%
Ask	0.59%
DuckDuckGo	0.38%
Naver	0.16%
Seznam	0.06%
Ecosia	0.05%
Source: netmarketshare.com/search-engine-market-share.aspx	

Chart-71: Search Engine Market Share | Source: *Netmarketshare* [2019]

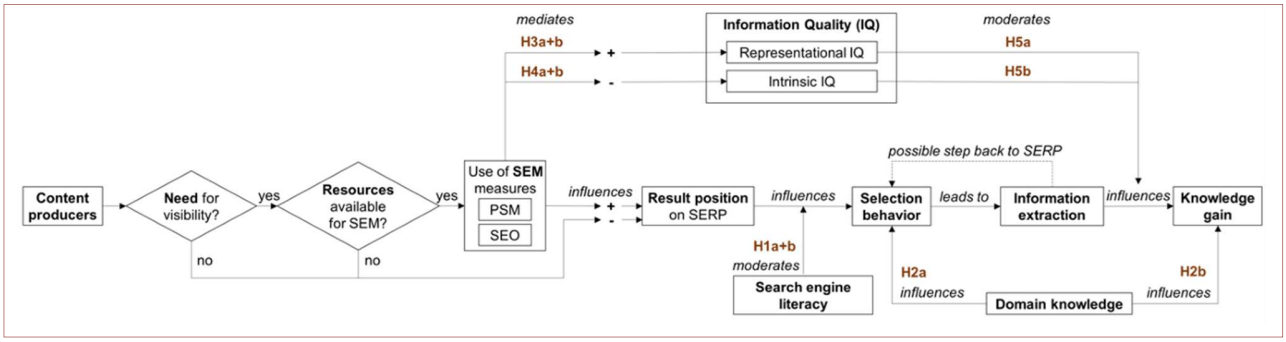


Chart-73: SEM influences user knowledge gain model | Source: *Schultheib* [2023]

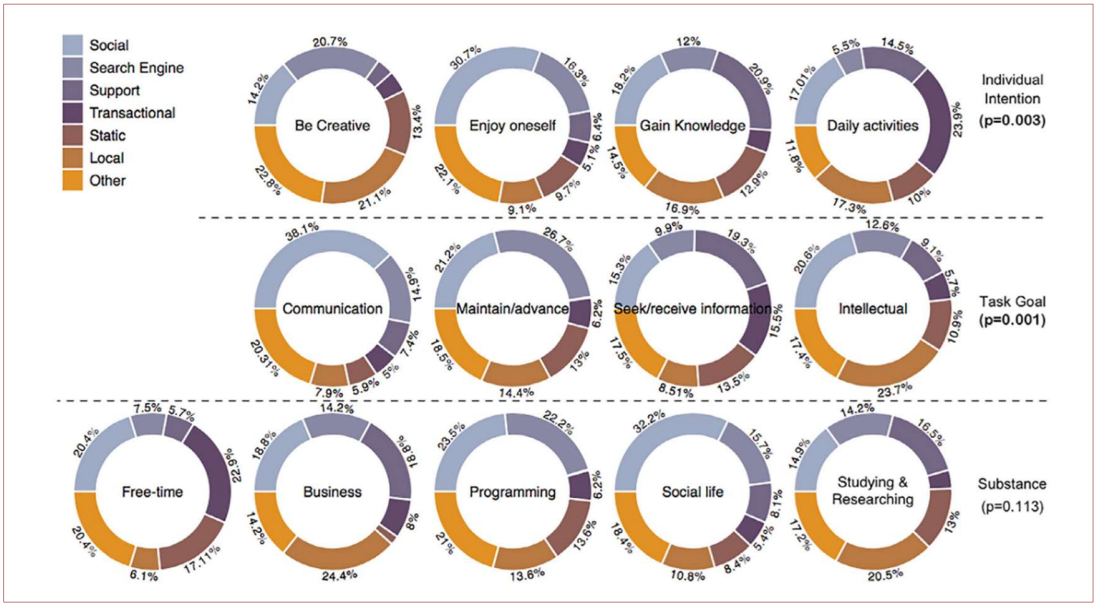


Chart-74: Results of application context with respect to task categories | Source: *Tung et al.* [2019]

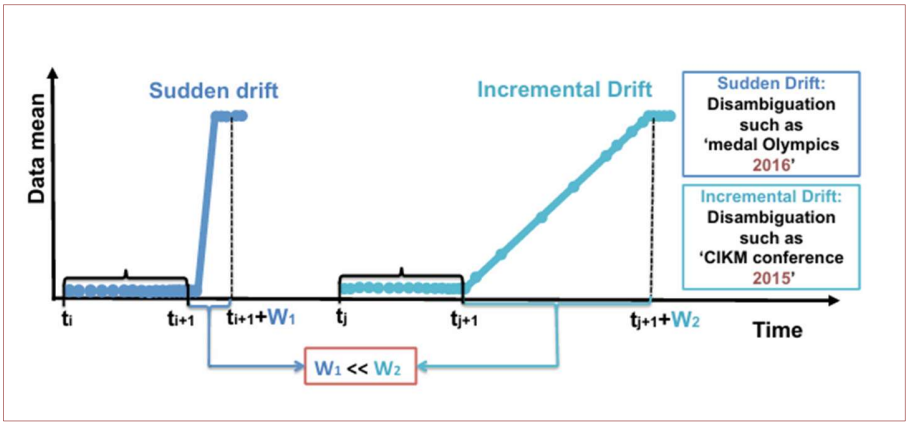


Chart-75: Query Drift Type - Sudden and Incremental | Source: *Kiseleva et al* [2015]

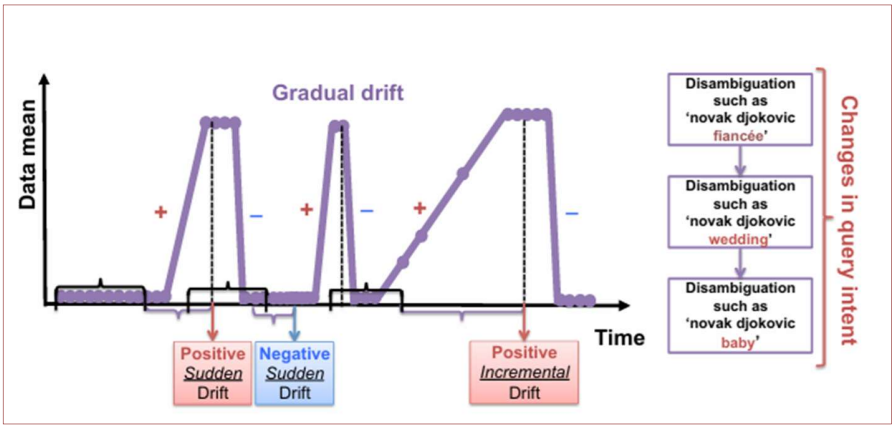


Chart-76: Query Drift Type - Gradual | Source: *Kiseleva et al* [2015]

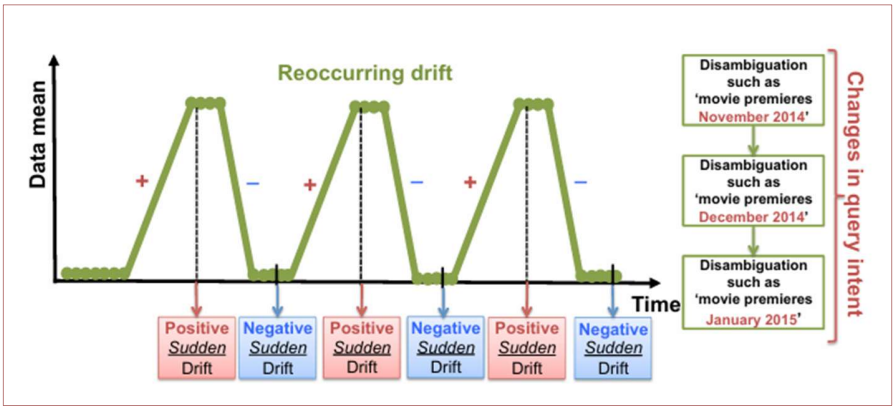


Chart-77: Query Drift Type – Re-occurring | Source: *Kiseleva et al* [2015]

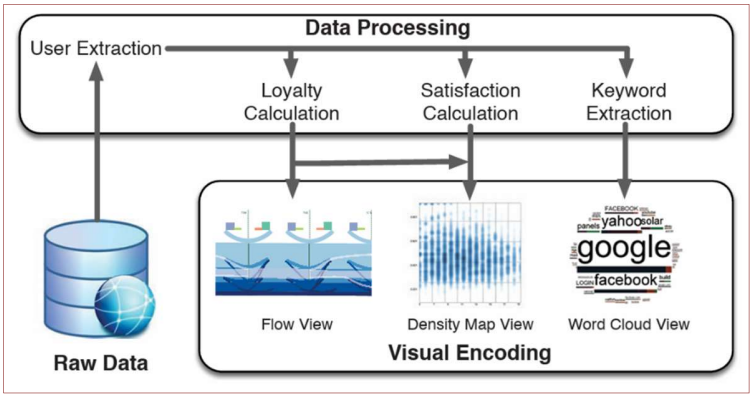


Chart-78: LoyaltyTracker, a user loyalty and defection behaviour visualizer | Source: *Shi et al.* [2014]

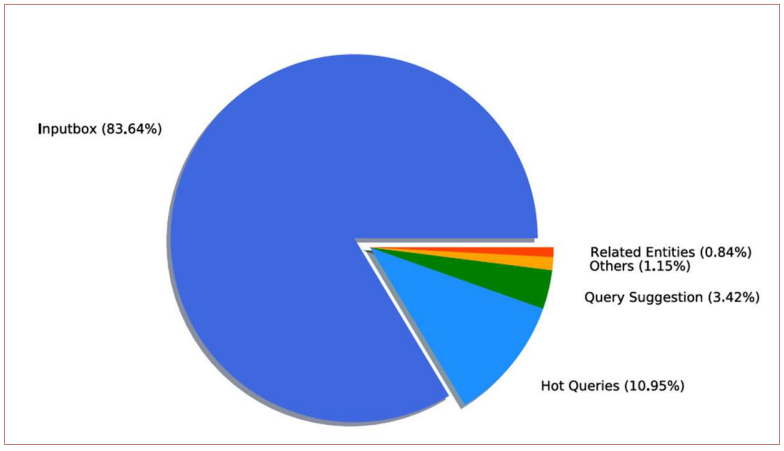


Chart-79: Proportion of reformulation entries that users adopt | Source: *Chen- Mao et al.* [2021]

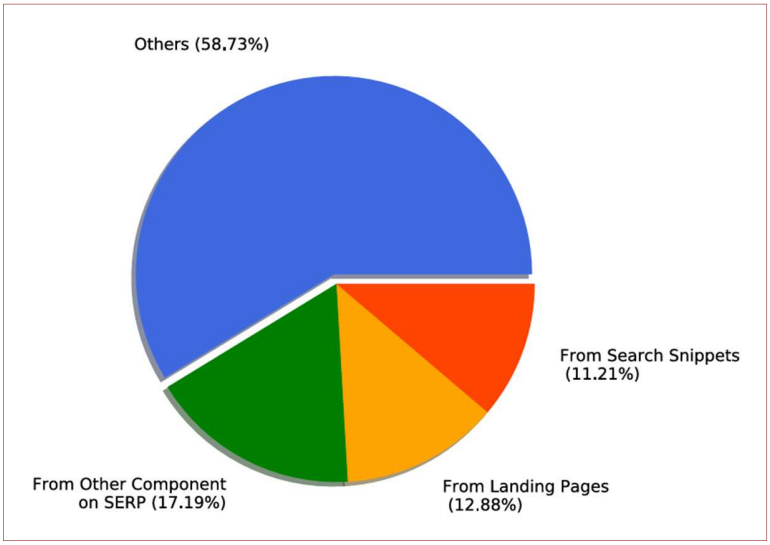


Chart-80: Proportion of the reformulation inspiration | Source: *Chen- Mao et al.* [2021]

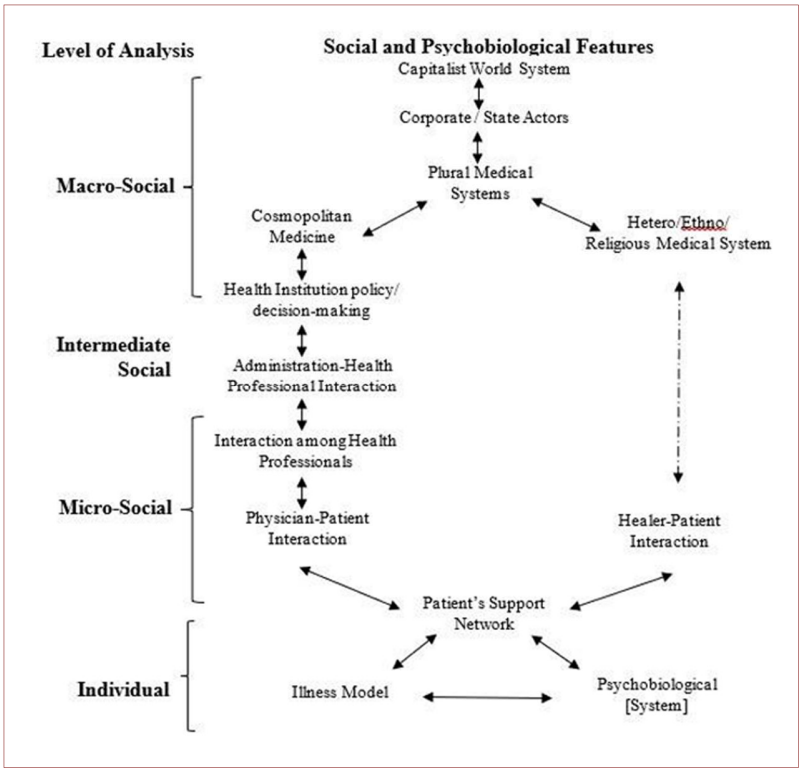


Chart-81: Levels of health systems adapted from Baer et al. | Source: *Schoenherr* [2022]

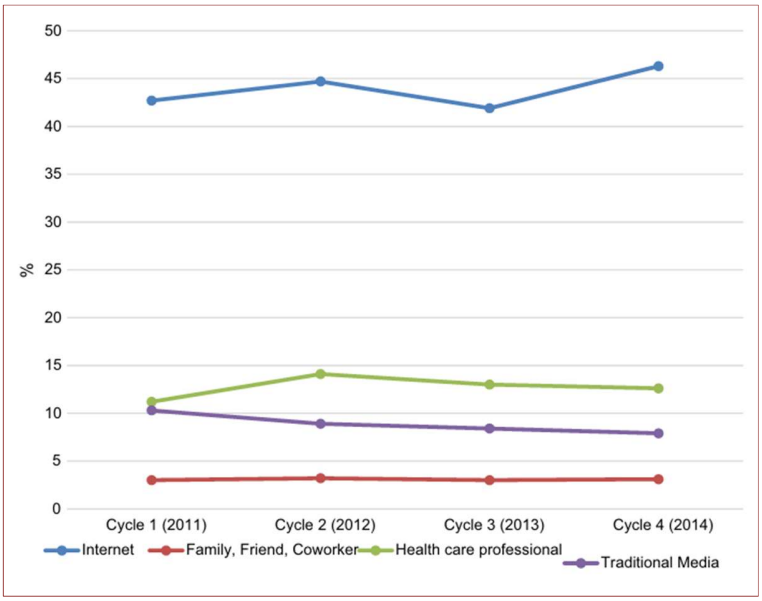


Chart-82: Health information sources in USA | Source: *Jacobs et al.* [2017]

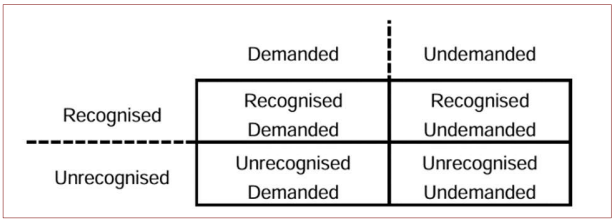


Chart-83: A framework of states of information needs *(adapted from Alzougool et al. 2013)* | Source: Pang et al. [2014]

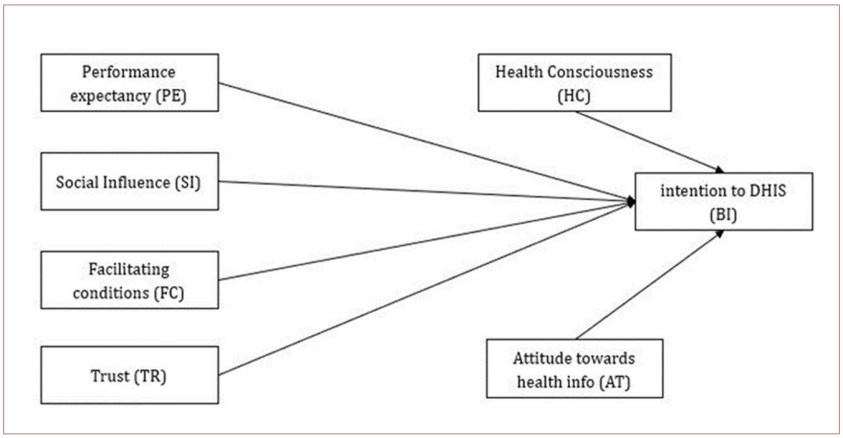


Chart-84: Variables relating to intention to Digital Health Information Search | Source: *Neeragatti et al.* [2023]

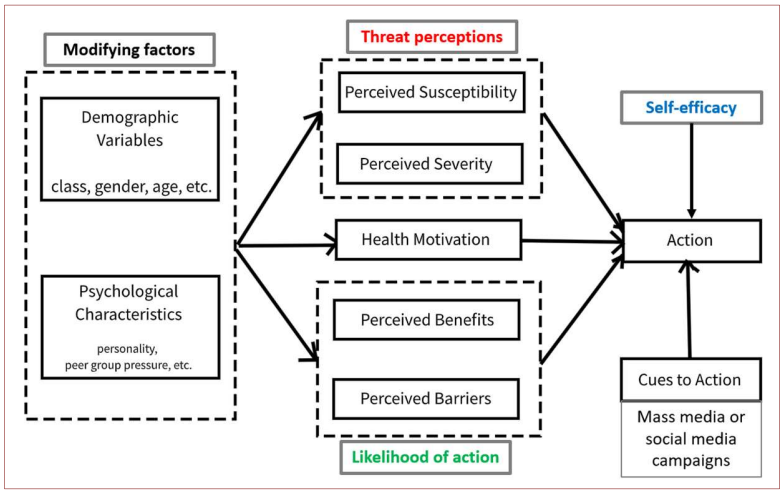


Chart-85: Health Belief Model | Source: [273] *Wikipedia-HBM*

	English	Italian	Spanish	French
Google.com	0			
Local Google	7	10	10	0
Yahoo	23*	30*	13	23*
Bing	23*	27*	17	20*
Duckduckgo	17	30*	17	23*
Ecosia	10	-	20	23*
Qwant	13	33*	17	23*
Mojeek	53*#	-	-	-
Swisscows	17	-	-	-
Arianna	-	37*	-	-
Virgilio	-	37*	-	-

Chart-86: Percentage of vaccine-negative webpages in different SERPs | Source: *Ghezzi et al.* [2019]

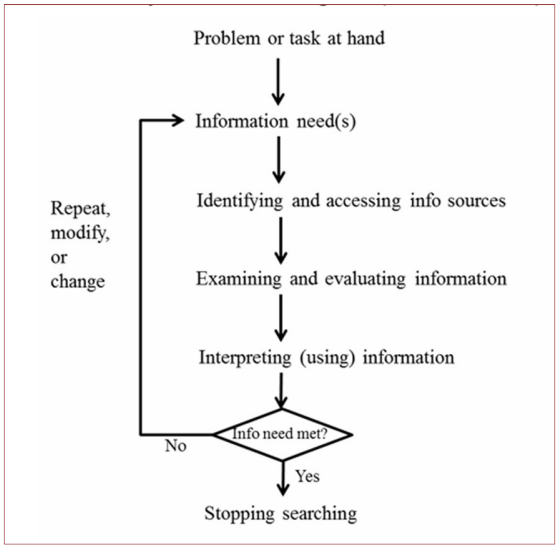


Chart-87: Schematic process of information searching | Source: *Yan Zhang* [2013]

[273] https://en.wikipedia.org/wiki/Health_belief_model#cite_ref-1

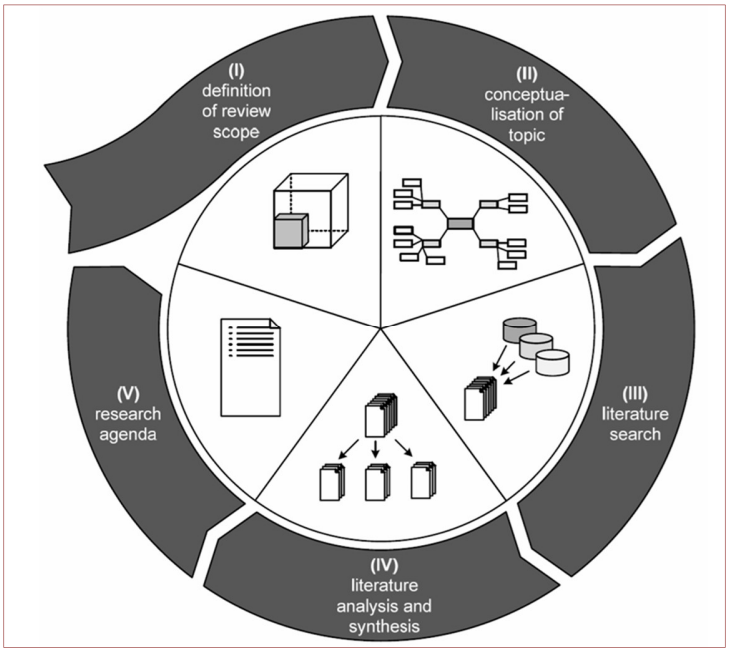


Chart-88: Framework for Literature Review | Source: *Brocke et al.* [2009]

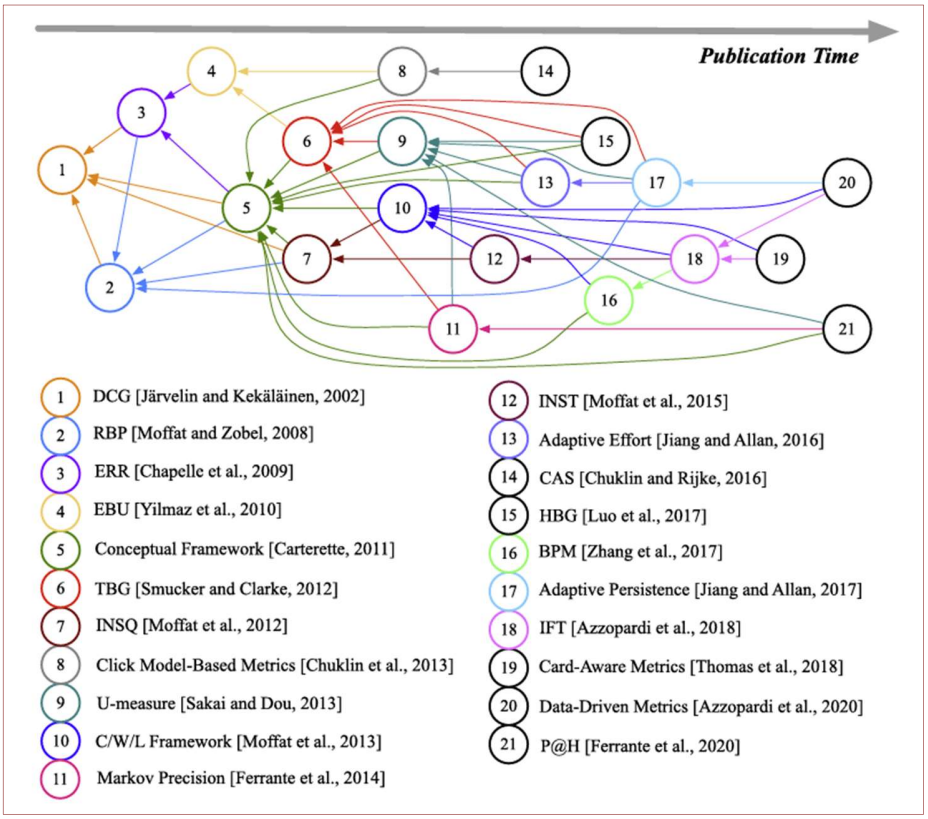


Chart-90: Papers related to user behaviour models listed chronologically | Source: *Zhang-1 et al.* [2021]

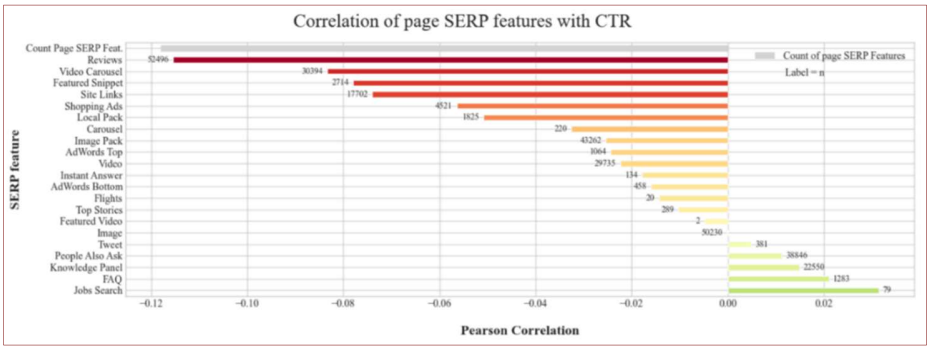
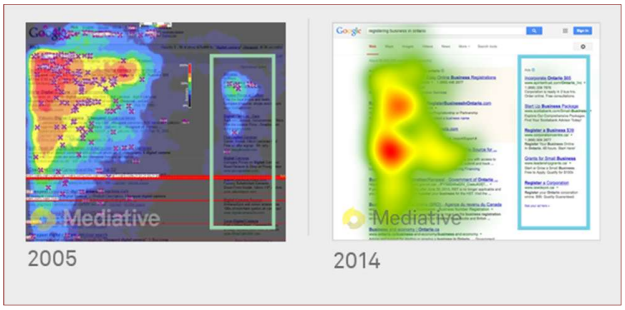
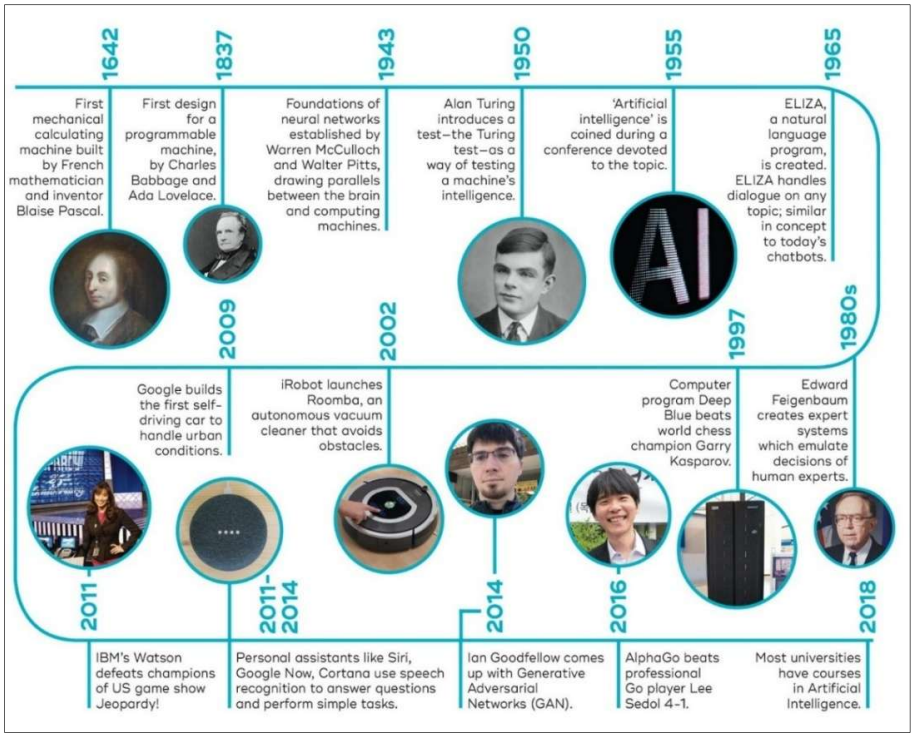


Chart-94: Correlation of page SERP Features with CTR; *Rahmawati et al.* [2022]



Infographic-1: Searcher Behaviour shift on SERP from 2005 to 2014 | Source: *Maynes & Everdell* [2014]



Infographic-2 - History milestones of Artificial Intelligence | Source: [274]*AI History qbi.uq.edu.au*

[274] https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence

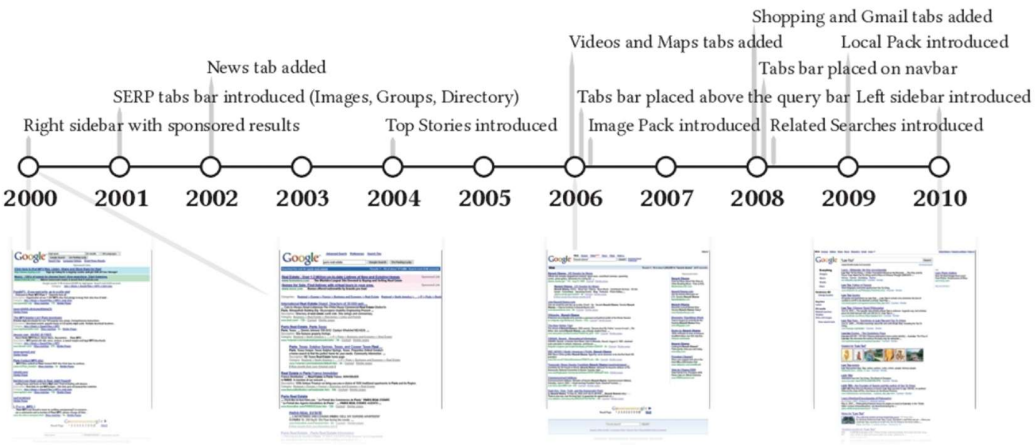


Figure 15: Highlights of SERP overall evolution (top) and Interfaces' visual evolution (bottom) from 2000 to 2010

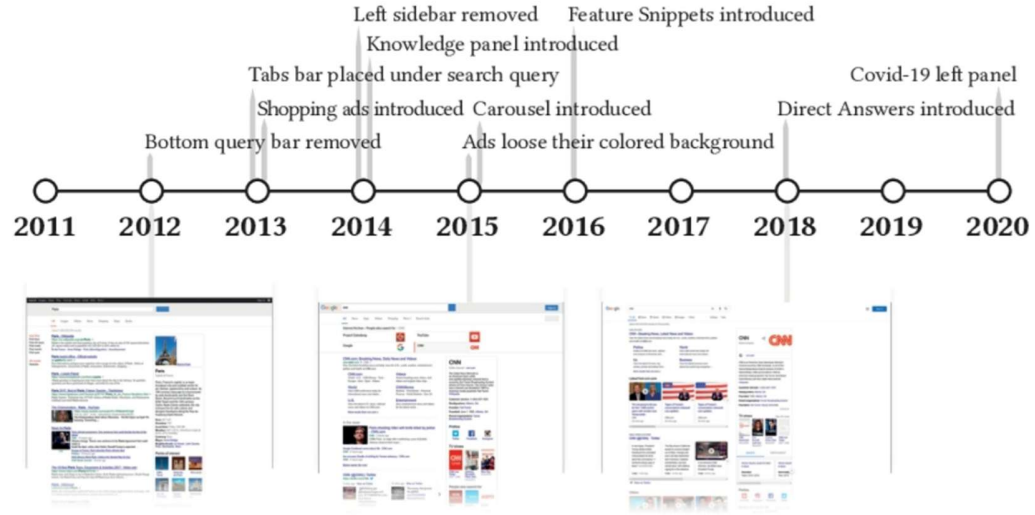
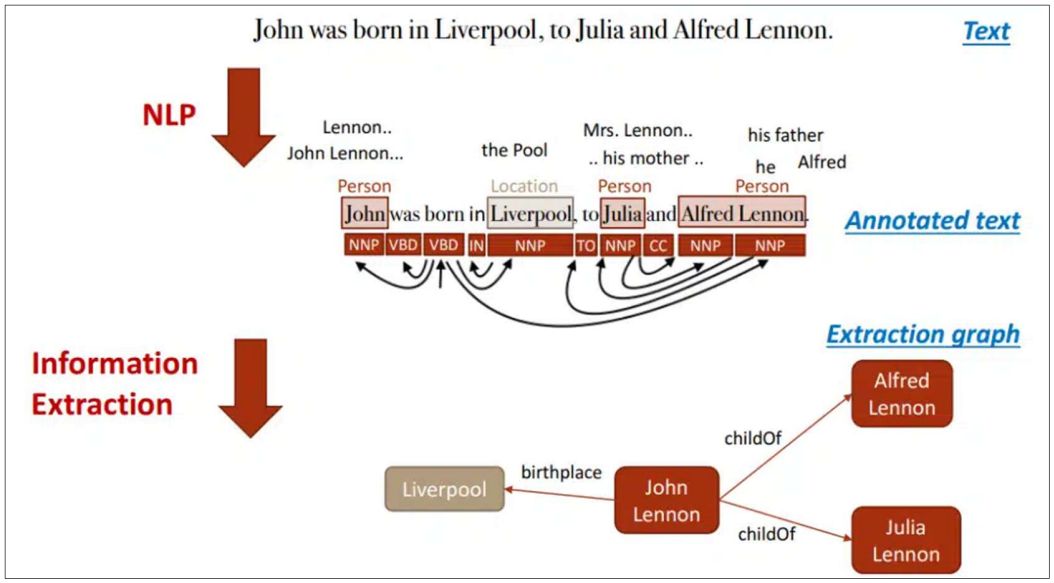
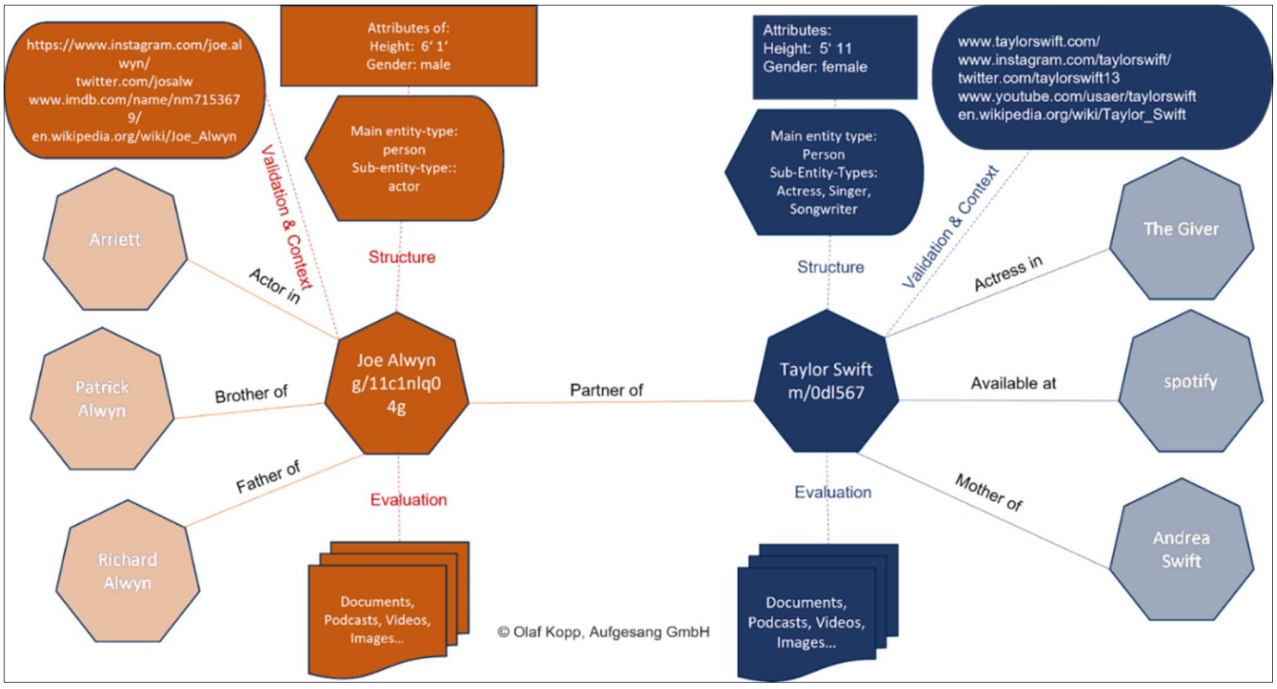


Figure 16: Highlights of SERP overall evolution (top) and Interfaces' visual evolution (bottom) from 2011 to 2020

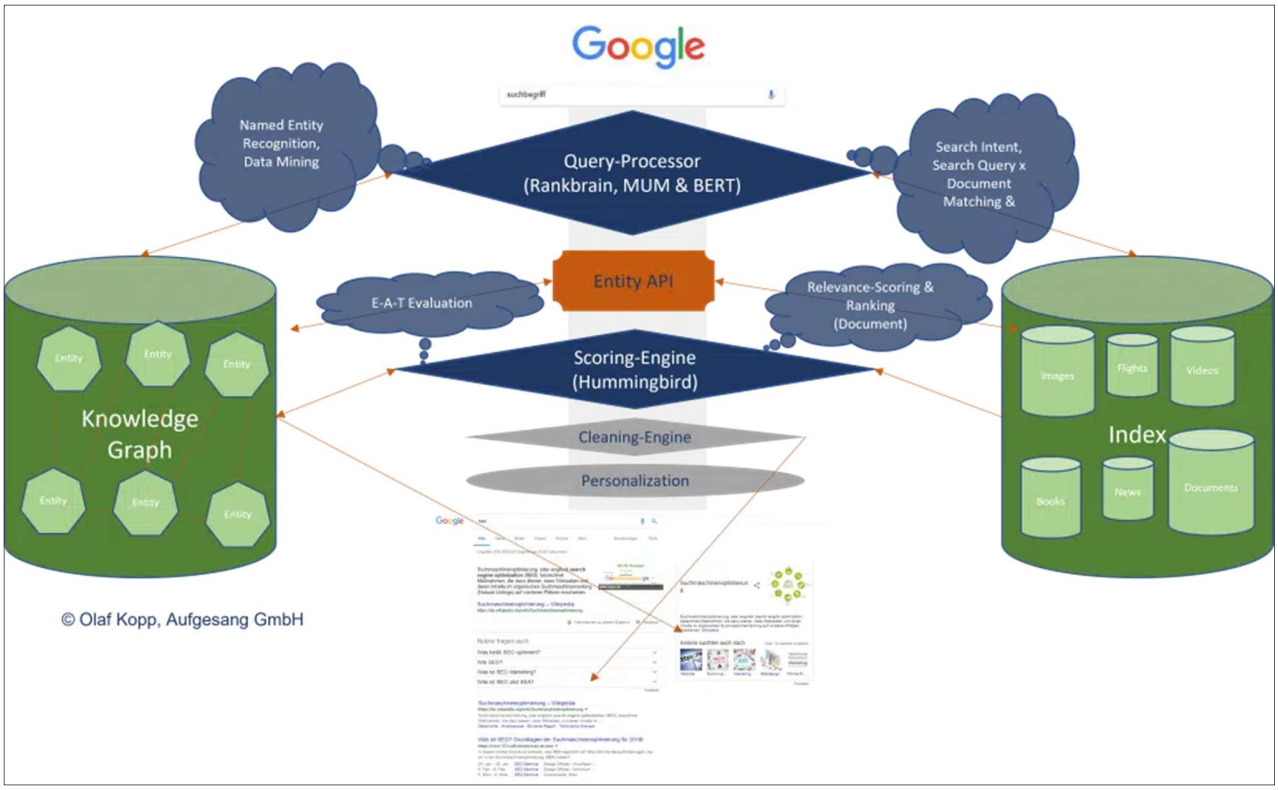
Infographic-3 - SERP Features Evolution Timeline | Source: *Oliveira &Teixeira Lopes* [2023]-2



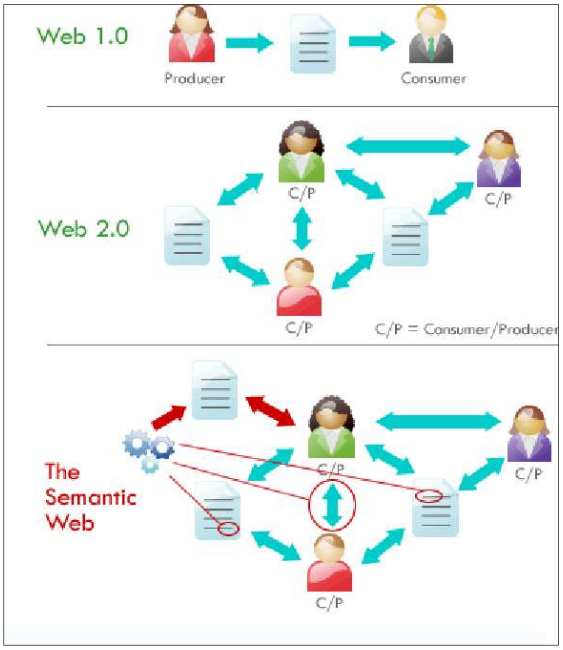
Infographic-4: Mining Knowledge Graphs from Text | Source: *Pujara & Singh* [2018]



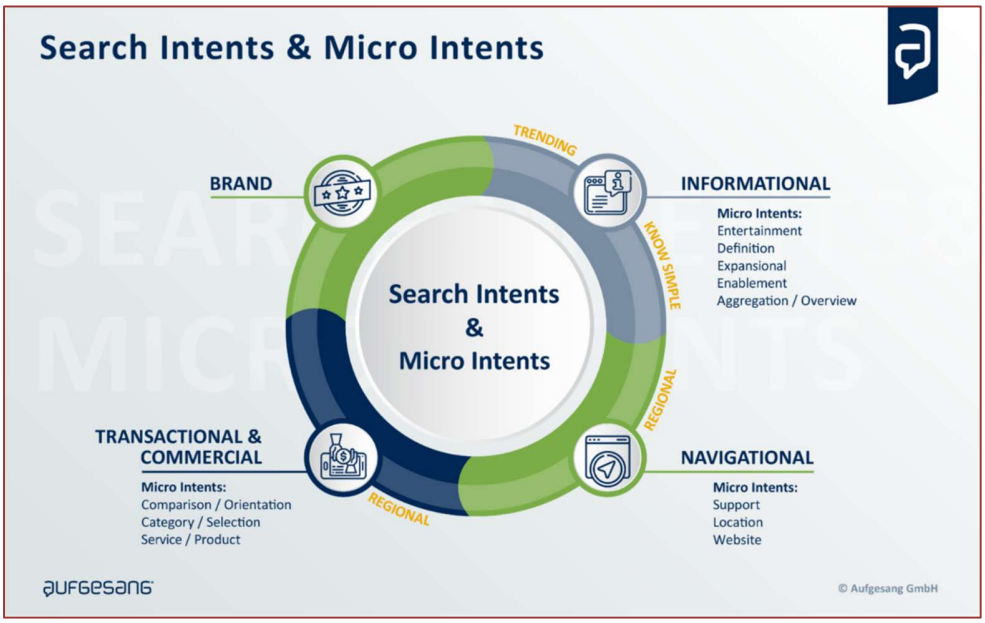
Infographic-5: Google Hummingbird' entity-based index Knowledge Graph - I | Source: [35]*Olaf Kopp* [2022]-1



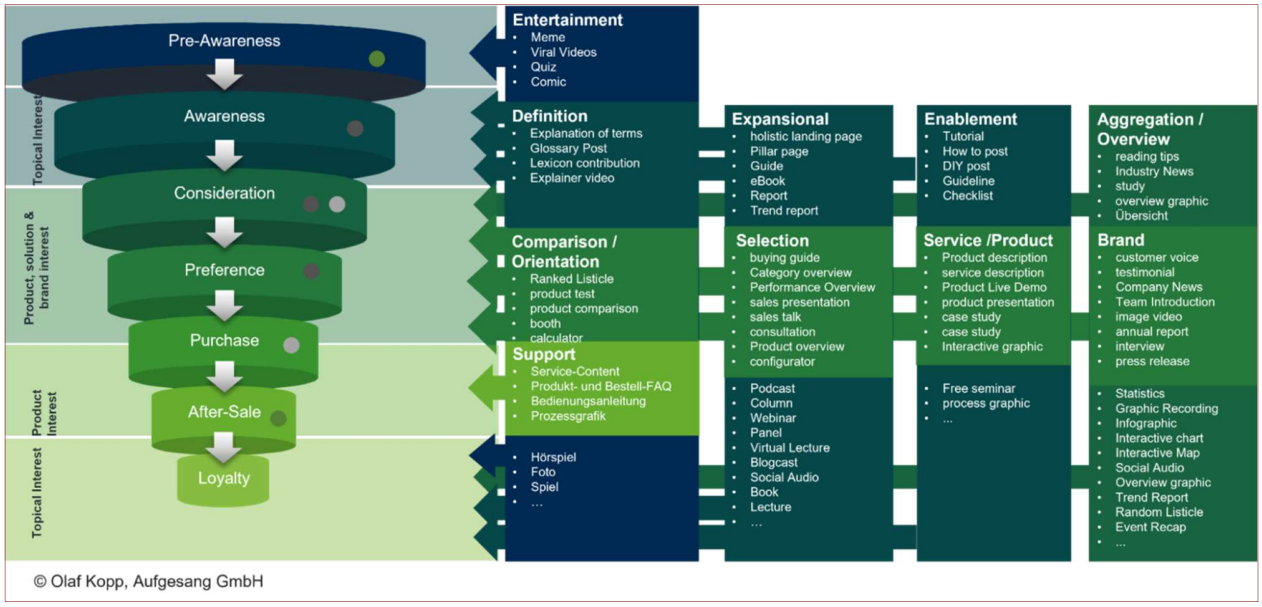
Infographic-6: Google Hummingbird' entity-based index Knowledge Graph - II | Source: [35]*Olaf Kopp* [2022]-1



Infographic-7: Difference between W 01, 02 & 03 | Source: *Agarwal P.* [2012]



Infographic-8: Intent types and micro-intent types | Source: [192]*Olaf Kopp-Intents* [2022]



Infographic-9: Customer journey according to micro intents | Source: [192]*Olaf Kopp-Intents* [2022]

Bibliography

1. *Moran & Hunt* [2014] - Mike Moran and Bill Hunt; 2014; Search Engine Marketing Inc: Driving Search Traffic to Your Company's Website, Book, 3rd Edition, Page 4; Publisher: IBM Press; ISBN-13: 978-0-13-303917-7
2. *Adam Hayes* [2024] - Adam Hayes; Jul 2024; Behavioural Modeling; https://www.investopedia.com/terms/b/behavioral-modeling.asp
3. *Dubey & Jain* [2017] - Babita Dubey and Khusboo Jain; 2017; Behavioural Aspects and Stress Physiology; Indian J.Sci.Res. 12 (2): 030-035, 2017; www.ijsr.in/upload/17372346327.pdf
4. *AIDA-b2bframeworks:* AIDA - Use this framework to improve communications - https://www.b2bframeworks.com/frameworks/aida
5. *Lucidchart-1* - 7 steps of the decision-making process; https://www.lucidchart.com/blog/decision-making-process-steps
6. *Digitalmainstreet* [2019] - No Regrets: The 3 Things Driving the Research-Obsessed Consumer; https://digitalmainstreet.ca/no-regrets-the-3-things-driving-the-research-obsessed-consumer/
7. *Long & Chang* [2014] - Bo Long, Yi Chang; 2014; Relevance Ranking for Vertical Search Engines, Chapter 8: Aggregated Vertical Search, Book, Pages 147-180; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/abs/pii/B9780124071711000083
8. *Dave Davies 1* [2020] - Dave Davies; May 26, 2020; How User Behaviour In Search Works: Everything You Need to Know; https://www.searchenginejournal.com/search-engines/user-behavior/
9. *Rand Fishkin* [2017] - Rand Fishkin; March 14, 2017; The State of Searcher Behaviour Revealed Through 23 Remarkable Statistics; https://moz.com/blog/state-of-searcher-behaviour-revealed
10. *Danny Sullivan* [2016] - Danny Sullivan; May 24, 2016; Google now handles at least 2 trillion searches per year; https://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247
11. *Djamasbi and Hall-Phillips* [2014] - Soussan Djamasbi, Adrienne Hall-Phillips; 2014; Eye Tracking in User Experience Design, Book, Pages 27-45
12. *Alberto Ferreira* [2017] - Alberto Ferreira; 2017; Universal UX Design: 6.6 International SEO Book
13. *Edwin Toonen* [2021] - Edwin Toonen; 23/09/2021; What is search intent and why is it important for SEO; https://yoast.com/search-intent/
14. *Joshua Hardwick* [2023]-2 - Joshua Hardwick; 23 Feb 2023; Google Doesn't Care About AI Content. Here's Why; https://ahrefs.com/blog/google-ai-content-guidelines/
15. *Google-GSQEG* - Google Search Quality Evaluator Guidelines; https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf
16. *Jayson DeMers* [2013] - Jayson DeMers; 10 April 2013; Why Your Keyword Strategy Is Incomplete Without User Intent; https://blog.hubspot.com/marketing/importance-of-user-intent-in-seo

17. *Neil Patel-1* - Neil Patel; I Lost 31% of My Search Traffic but My Sales Went up by 73%; http://neilpatel.com/blog/i-lost-31-of-my-search-traffic-but-my-sales-went-up-by-73/
18. *Herradura-cornerstone-digital:* Darlyn Herradura; The Psychology of Search: Understanding User Intent for Better SEO; https://cornerstone-digital.com.au/the-psychology-of-search-understanding-user-intent-for-better-seo/
19. *Madelyn Gardner* [2023] - 15 Feb 2023; Google SERP Feature: A Visual Guide for 2024 (Infographic); https://www.brafton.com/blog/distribution/a-content-marketers-guide-to-every-2018-google-serp-feature/
20. *Netmarketshare* [2019] - Netmarketshare; Sep 2019; Market Share Statistics for Internet Technologies: Search Engine Market Share; https://netmarketshare.com/search-engine-market-share.aspx
21. *Statcounter-live* - Search Engine Market Share Worldwide; https://gs.statcounter.com/search-engine-market-share
22. *Inetsolutions.org-1* [2016] - Inetsolutions.org; Sep 2016; 25 Informative SEO Facts & Statistics You Probably Don't Know About; https://www.inetsolutions.org/25-informative-seo-facts-statistics-probably-dont-know/
23. *Tom J Law* [2019] - Tom J Law; Aug, 2019; Google SERPs: Features and How to Improve Your Ranking in 2023; https://www.oberlo.com/blog/serp-google-search-engine-results
24. *Neil Patel-2* - How to Do SERP Analysis Like an Expert; https://neilpatel.com/blog/serps-guide/
25. *Dave Davies-2* [2020] - Dave Davies ; May 25, 2020; How Search Engine Algorithms Work: Everything You Need to Know; https://www.searchenginejournal.com/search-engines/algorithms/
26. *Kaviani Siavosh* [2019] - Kaviani Siavosh; March 31, 2019; How the search engines works?; https://avengering.com/en/how-the-search-engines-works/
27. *Satavisa Pati* [2021] - Satavisa Pati; August 3, 2021; The Difference Between Artificial Intelligence And Machine Learning; https://www.analyticsinsight.net/the-difference-between-artificial-intelligence-and-machine-learning/
28. *Geeksforgeekst.org* - Difference Between Artificial Intelligence vs Machine Learning vs Deep Learning; https://www.geeksforgeeks.org/artificial-intelligence/artificial-intelligence-vs-machine-learning-vs-deep-learning/
29. *Bernard Marr* - What Is The Difference Between Artificial Intelligence And Machine Learning?; https://bernardmarr.com/what-is-the-difference-between-artificial-intelligence-and-machine-learning/
30. *Truran et al.* [2011] - Mark Truran, Jan-Felix Schmakeit, Helen Ashman; 06 May 2011; The effect of user intent on the stability of search engine results; https://onlinelibrary.wiley.com/doi/10.1002/asi.21550
31. *Google.com-1* - How Search Works; https://www.google.com/search/howsearchworks/
32. *Google.com-2* - How Search Works: Our Approach; https://www.google.com/search/howsearchworks/our-approach/
33. *Google.com-3* - How Google Search organizes information; https://www.google.com/search/howsearchworks/how-search-works/organizing-information/
34. *Google.com-4* - How results are automatically generated; https://www.google.com/search/howsearchworks/how-search-works/ranking-results/
35. *Google.com-5* - Understanding page experience in Google Search results; https://developers.google.com/search/docs/advanced/experience/page-experience#signals
36. *Google.com-6* - Safe Search Filters; https://www.google.com/preferences

37. *Crevier Daniel* [1993] - Crevier, Daniel; 1993; AI: The Tumultuous Search for Artificial Intelligence, Book, Pages 109, 161-162, 197-203, 211, 240; Publisher: New York, NY: BasicBooks; ISBN 0-465-02997-3
38. *McCorduck Pamela* [2004] - McCorduck, Pamela; 2004; Machines Who Think Book, 2nd edition, Page 426-441; Publisher: Natick, MA: A. K. Peters Ltd.; ISBN 1-56881-205-1
39. *Russell & Norvig* [2003] - Russell Stuart J., Norvig Peter; 2003; Artificial Intelligence: A Modern Approach Book, 2nd edition, Page 24; Publisher: Upper Saddle River, New Jersey: Prentice Hall; ISBN 0-13-790395-2
40. *USNRC 1* [1999] - USNRC (United States National Research Council); 1999; Developments in Artificial Intelligence Research Report Pages 210-211; Publisher: Funding a Revolution: Government Support for Computing Research; National Academy Press; https://nap.nationalacademies.org/read/6323/chapter/11#210
41. *Newquist 1* [1994] - Newquist, HP; 1994; The Brain Makers: Genius, Ego, And Greed In The Quest For Machines That Think Book, Pages 235-248; Publisher: New York: Macmillan/SAMS; ISBN 978-0-672-30412-5
42. *Clark Jack* [2015] - Clark, Jack; 1 July 2015; Musk-Backed Group Probes Risks Behind Artificial Intelligence; Bloomberg.com, Archived from the original; https://www.bloomberg.com/news/articles/2015-07-01/musk-backed-group-probes-risks-behind-artificial-intelligence
43. *Mike Anderson* [2021] - Mike Anderson; October 30, 2021; AlphaGo: Google DeepMind; https://deepmind.com/research/case-studies/alphago-the-story-so-far
44. *Kurzweil Ray* [2005] - Kurzweil Ray; 2005; The Singularity is Near, Book, Page 265; Publisher: Penguin Books; ISBN 978-0-670-03384-3
45. *Copeland* [2021] - B.J. Copeland; 14 Dec 2021; Artificial Intelligence; https://www.britannica.com/technology/artificial-intelligence
46. *AI Wikipedia 1* - Artificial Intelligence; https://en.wikipedia.org/wiki/Artificial_intelligence
47. *Ed Burns et al.* - Ed Burns, Nicole Laskowski, Linda Tucci; What is AI (Artificial Intelligence)? Definition, Types, Examples & Use Caseshttps://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence
48. *AI sas.com 1* - Artificial Intelligence: What it is and why it matters; https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html
49. *John McCarthy* [2007] - John McCarthy, Stanford University; 24 Nov 2004; What is artificial intelligence?; https://www-formal.stanford.edu/jmc/whatisai.pdf
50. *Stryker & Kavlakoglu* - Cole Stryker, Eda Kavlakoglu ; Artificial Intelligence (AI) ?; https://www.ibm.com/think/topics/artificial-intelligence
51. *Hopfield* [1982] - J. J. Hopfield; 15 Jan 1982; Neural networks and physical systems with emergent collective computational abilities; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC346238/pdf/pnas00447-0135.pdf
52. *NN Wikipedia 2* - Neural Network; https://en.wikipedia.org/wiki/Neural_network
53. *Sindhu et al.* [2020] - Sindhu V, Nivedha S, Prakash M; 07 Feb 2020; AN EMPIRICAL SCIENCE RESEARCH ON BIOINFORMATICS IN MACHINE LEARNING; http://www.journalimcms.org/special_issue/an-empirical-science-research-on-bioinformatics-in-machine-learning/

54. *Sebastian Rashcka* [2020] - Sebastian Rashcka; 2020; What are Machine Learning and Deep Learning: An Overview; https://github.com/rasbt/stat453-deep-learning-ss20/blob/master/L01-intro/L01-intro_slides.pdf
55. *ML sas.com 2* - Machine Learning: What it is and why it matters; https://www.sas.com/en_us/insights/analytics/machine-learning.html
56. *CE ibm.com 1* - IBM Cloud Education; 15 July 2020; What is Machine Learning?; https://www.ibm.com/cloud/learn/machine-learning
57. *DS Wikipedia 2* - Wikipedia 3; Data Science; https://en.wikipedia.org/wiki/Data_science
58. *Vasant Dhar* [2013] - Vasant Dhar; Dec 2013; Data Science and Prediction; https://cacm.acm.org/research/data-science-and-prediction/
59. *Hayashi et al.* [1998] - Chikio Hayashi, Keiji Yajima, Hans-Hermann Bock, Noboru Ohsumi, Yutaka Tanaka, Yasumasa Baba; 1998; Data Science, Classification, and Related Methods Book, Pages 40-51; ISBN: 978-4-431-65950-1
60. *Tony Hey* [2009] - Tony Hey; 2009; The Fourth Paradigm: Data-intensive Scientific Discovery Book; Publisher: Microsoft Research; ISBN 978-0-9825442-0-4
61. *Mitchell Waldrop* [2019] - M. Mitchell Waldrop; 22 Jan 2019; What are the limits of deep learning; https://www.pnas.org/content/116/4/1074/tab-figures-data
62. *Kevin Rowe* [2021] - Kevin Rowe; August 13, 2021; How Search Engines Use Machine Learning: 9 Things We Know For Sure; https://www.searchenginejournal.com/ml-things-we-know/408882/
63. *Dave Davies* [2021] - Dave Davies; September 24, 2021; Google's Search Quality Raters Guidelines: A Guide for SEO Beginners; https://www.searchenginejournal.com/google-eat/quality-raters-guidelines/
64. *Emmanuel et al.* [2019] - Emmanuel Gbenga Dada, Joseph Stephen Bassi, Haruna Chiroma, Shafi'i Muhammad Abdulhamid, Adebayo Olusola Adetunmbi, Opeyemi Emmanuel Ajibuwa; June 2019; Machine learning for email spam filtering: review, approaches and open research problems; Volume 5, Issue 6; https://www.sciencedirect.com/science/article/pii/S2405844018353404
65. *Google SEO YT* [2021] - 7 May 2021; English Google SEO office-hours from https://www.youtube.com/watch?v=0h8YslnIuuY&t=250s
66. *John Mueller* [2019] - 13 Dec 2019; English Google Webmaster Central office-hours from; https://www.youtube.com/watch?v=5QxYWMEZT3A&t=2267s
67. *Matt Southern* [2019] - Matt G. Southern; 7 Feb 2019; Google's Gary Illyes Explains How RankBrain Works; https://www.searchenginejournal.com/googles-gary-illyes-explains-how-rankbrain-works/292381/
68. *Tie-Yan Liu* [2009] - Tie-Yan Liu; 27 Jun 2009; Learning to Rank for Information Retrieval Book, Page 225-331; Publisher: Microsoft; https://www.nowpublishers.com/article/Details/INR-016
69. *ML Lucidworks.com 1* - Machine Learning for Search; https://lucidworks.com/blog/17382
70. *AI-ML google.com 2* - AI and machine learning products; https://cloud.google.com/products/ai
71. *Patent google.com 3* - Method for personalized search; https://patents.google.com/patent/US20050102282A1/en

72. *Devlin et al.* [2018] - Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova; 11 Oct 2018; BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding; https://arxiv.org/abs/1810.04805
73. *Brent Csutoras* [2019] - Brent Csutoras; 26 Nov 2019; BERT Explained: What You Need to Know About Google's New Algorithm; https://www.searchenginejournal.com/bert-explained-what-you-need-to-know-about-googles-new-algorithm/337247/
74. *Cooper Smith* [2013] - Cooper Smith; 18 Sep 2013; Facebook Users Are Uploading 350 Million New Photos Each Day; https://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9?r=US&IR=T
75. *Patent google.com 4* - Predicting ad quality; https://patents.google.com/patent/US20070156887
76. *Patent google.com 5* - User-based ad ranking; https://patents.google.com/patent/US9773256
77. *Ad rank google.com 6* - Ad Rank; https://support.google.com/google-ads/answer/1752122?hl=en
78. *Luigi Oppido* [2021] - Luigi Oppido; 3 Jun 2021; How to Search for Phrases and Expressions in Search Engines; https://www.wikihow.com/Search-for-Phrases-and-Expressions-in-Search-Engines
79. *Joshua Hardwick* [2023] - Joshua Hardwick; 25 Apr 2023; Google Search Operators: The Complete List (42 Advanced Operators); https://ahrefs.com/blog/google-advanced-search-operators/
80. *Blaha et al.* [2021] - Leslie M. Blaha, Mitchell Abrams, Sarah A. Bibyk, Claire Bonial, Beth M. Hartzler, Christopher D. Hsu, Sangeet Khemlani, Jayde King, Robert St. Amant, J. Gregory Trafton and Rachel Wong; 31 Mar 2021; Understanding Is a Process; https://www.frontiersin.org/articles/10.3389/fnsys.2022.800280/full
81. *Alfred North Whitehead* [1929] - Alfred North Whitehead; 1929 & 1978; Process and reality Book, Page 9; Publisher: The Macmillan Company, University Press; https://antilogicalism.com/wp-content/uploads/2018/04/process-and-reality.pdf
82. *Wiki Semantic Web* - https://en.wikipedia.org/wiki/Semantic_Web
83. *Hudson et al.* [2008] - Hudson R. F., Pullen P. C., Lane H. B., Torgesen J. K.; 2008; The complex nature of reading fluency: A multidimensional view; https://www.tandfonline.com/doi/abs/10.1080/10573560802491208
84. *Kuhn* M. R. [2005] - Kuhn M. R.; 2005; A comparative study of small group fluency instruction; https://www.tandfonline.com/doi/abs/10.1080/02702710590930492
85. *Kuhn et al.* [2018] - Kuhn M. R., Rasinski T., Young C.; 2018; Best practices in fluency instruction Book, chapter 12; Publisher: Guilford Press; https://www.researchgate.net/publication/336348968_Best_practices_in_reading_fluency_instruction
86. *Vinod Sharma* [2023] - Vinod Kr Sharma; 08-2023; Generative AI for Research & Writing, p-5; https://www.researchgate.net/publication/373367478_GENERATIVE_AI_FOR_RESEARCH_WRITING
87. *Anand et al.* [2014] - Anand A., Mele I., Bedathur S., Berberich, K.; 2014; Phrase query optimization on inverted indexes; https://doi.org/10.1145/2661829.2661928
88. *Bhatia et al.* [2011] - Bhatia S., Majumdar D., Mitra P.; 2011; Query suggestions in the absence of query logs; https://doi.org/10.1145/2009916.2010023
89. *Bradford R.* [2014] - Bradford R.; 2014; Incorporating ad hoc phrases in lsi queries; https://doi.org/10.5220/0005073300610070

90. *Carmel and Yom-Tov* [2010] - Carmel D., Yom-Tov E.; 2010; Estimating the query difficulty for information retrieval; Synthesis Lectures on Information Concepts Retrieval and Services, 2(1) 1-89; https://doi.org/10.2200/s00235ed1v01y201004icr015
91. *Chang et al.* [2011] - Chang Y., Zhang R., Reddy S., Liu, Y.; 2011; Detecting multilingual and multi-regional query intent in web search; Proceedings of the Aaai Conference on Artificial Intelligence, 25(1), 1134-1139; https://doi.org/10.1609/aaai.v25i1.8074
92. *Ide et al.* [2007] - Ide N., Loane R., & Demner-Fushman D.; 2007; Essie: a concept-based search engine for structured biomedical text; Journal of the American Medical Informatics Association, 14(3), 253-263; https://doi.org/10.1197/jamia.m2233
93. *Kulkarni & Callan* [2010] - Kulkarni A., Callan J.; 2010; Document allocation policies for selective searching of distributed indexes; https://doi.org/10.1145/1871437.1871497
94. *Lazarinis F.* [2007] - Lazarinis F.; 2007; Web retrieval systems and the greek language: do they have an understanding?; Journal of Information Science, 33(5), 622-636; https://doi.org/10.1177/0165551506076394
95. *Purnamasari et al.* [2016] - Purnamasari D., Wulandari L., Thantawi A., Wicaksana I.; 2016; Concept similarity searching for support query rewriting; International Journal of Computer Theory and Engineering, 8(6), 490-494; https://doi.org/10.7763/ijcte.2016.v8.1094
96. *Qiu et al.* [2021] - Qiu Y., Zhang K., Zhang H., Wang S., Xu S., Xiao Y., Yang W.; 2021; Query rewriting via cycle-consistent translation for e-commerce search; https://doi.org/10.1109/icde51399.2021.00276
97. *Tahery & Farzi* [2020] - Tahery S., Farzi S.; 2020; Tips: time-aware personalised semantic-based query auto-completion; Journal of Information Science, 48(4), 524-543; https://doi.org/10.1177/0165551520968690
98. *Trotman & Lalmas* [2006] - Trotman A., Lalmas M.; 2006; Strict and vague interpretation of xml-retrieval queries; https://doi.org/10.1145/1148170.1148329
99. *White & Ruthven* [2006] - White R., Ruthven I.; 2006; A study of interface support mechanisms for interactive information retrieval; Journal of the American Society for Information Science and Technology, 57(7), 933-948; https://doi.org/10.1002/asi.20365
100. *Williams et al.* [2004] - Williams H., Zobel J., Bahle D.; 2004; Fast phrase querying with combined indexes; Acm Transactions on Information Systems, 22(4), 573-594; https://doi.org/10.1145/1028099.1028102
101. *Zhao et al.* [2023] - Zhao X., Chen H., Xing Z., Chen M.; 2023; Brain-inspired search engine assistant based on knowledge graph; Ieee Transactions on Neural Networks and Learning Systems, 34(8), 4386-4400; https://doi.org/10.1109/tnnls.2021.3113026
102. *Nogueira & Cho* [2017] - Nogueira R., Cho K.; 2017; Task-Oriented Query Reformulation with Reinforcement Learning; ArXiv, abs/1704.04572; https://doi.org/10.18653/v1/D17-1061
103. *Agarwal P.* [2012] - P. Agarwal; 08-02-2012; Semantic Web in Comparison to Web 2.0; https://www.semanticscholar.org/paper/Semantic-Web-in-Comparison-to-Web-2.0-Agarwal/0e6b92706083874a824dad6532467e5cc2671cba ; https://www.researchgate.net/publication/261389302_Semantic_Web_in_Comparison_to_Web_20

104. *Leung et al.* [2008] - Leung K., Ng W., L.; 2008; Personalized Concept-Based Clustering of Search Engine Queries; IEEE Transactions on Knowledge and Data Engineering, 20, 1505-1518; https://doi.org/10.1109/TKDE.2008.84
105. *Wang et al.* [2017] – Wang Y., Zhong M., Zhu Y., Li X., Qian T.; 2017; Diversified Top-k Keyword Query Interpretation on Knowledge Graphs, p 541-555; https://doi.org/10.1007/978-3-319-63579-8_41
106. *Khan et al.* [2015] - Khan S., Khodke P., Bhagat A.; 2015; Neuro Fuzzy based user queries categorization; 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 1-7; https://doi.org/10.1109/ICIIECS.2015.7193133.
107. *Baeza-Yates* [2017] - Baeza-Yates R.; 2017; Semantic Query Understanding; Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval; https://doi.org/10.1145/3077136.3096472.
108. *Joseph Sam* [2002] - Joseph Sam; 2002; NeuroGrid: Semantically Routing Queries in Peer-to-Peer Networks; p 202-214; https://doi.org/10.1007/3-540-45745-3_18
109. *Prakash & Patel* [2015] - Prakash A., Patel, D.; 2015; Techniques for Deep Query Understanding; https://arxiv.org/ftp/arxiv/papers/1505/1505.05187.pdf
110. *Romanov et al.* [2003] - Romanov Y., Zaytseva N., Zelko V., Romanov A., Tovmach Y., Anikina N., Chaplia A., Blekhman M., Precup A., Putin V.; 2003; Artificial Intelligence and Machine Translation in the XXIst Century; Journal of Quantitative Linguistics, 10, 215 – 238; https://doi.org/10.1080/09296170512331383575
111. *Olaf Kopp* [2022]-1 - Olaf Kopp; 23-08-2022; How Google uses NLP to better understand search queries, content; https://searchengineland.com/how-google-uses-nlp-to-better-understand-search-queries-content-387340
112. *Matt G. Southern* [2022] - Matt G. Southern; 03-02-2022; How Google Search Understands Human Language; https://www.searchenginejournal.com/how-google-search-understands-human-language/436565/#close
113. *Rebecca Sentance* [2016] - Rebecca Sentance; 12-04-2016; Everything you need to know about natural language search; https://www.searchenginewatch.com/2016/04/12/everything-you-need-to-know-about-natural-language-search/?amp=1
114. *Hannah et al.* [2016] - Bast Hannah, Buchhold Björn, Haussmann Elmar; 2016; Semantic search on text and knowledge bases Book; Foundations and Trends in Information Retrieval 10 (2–3), 119–271; https://www.researchgate.net/publication/304364705_Semantic_Search_on_Text_and_Knowledge_Bases
115. *Pandu Nayak* [2022] - Pandu Nayak, Vice President Google Search; 03-02-2022; How AI powers great search results; https://blog.google/products/search/how-ai-powers-great-search-results/
116. *Marketbrew-2* - Marketbrew.ai; Natural Language Processing and Its Role in SEO and Search Engines; https://marketbrew.ai/natural-language-processing-and-its-role-in-seo-and-search-engines
117. *White et al.* [2009] - Ryen White, Susan Dumais, Jaime Teevan; 09-02-2009; Characterizing the influence of domain expertise on web search behaviour; Microsoft Research; http://ryenwhite.com/talks/pdf/WhiteDumaisTeevanWSDM2009.pdf

118. *A. M. Turing* [1950] - A. M. Turing; Oct 1950; Computing Machinery and Intelligence; https://academic.oup.com/mind/article/LIX/236/433/986238
119. *Michael Hannon* [2021] - Michael Hannon; 10/07/2021; RECENT WORK IN THE EPISTEMOLOGY OF UNDERSTANDING; https://scholarlypublishingcollective.org/uip/apq/article-abstract/58/3/269/283447/RECENT-WORK-IN-THE-EPISTEMOLOGY-OF-UNDERSTANDING
120. *Hough & Gluck* [2019] - Alexander R. Hough, Kevin A. Gluck; 2019; The understanding problem in cognitive science; http://www.cogsys.org/journal/volume8/article-8-3.pdf
121. *Herbert Simon* [1977] - Herbert A. Simon; 1977; Artificial Intelligence Systems that Understand; https://www.ijcai.org/Proceedings/77-2/Papers/096.pdf
122. *Moore & Newell* [1974] - Moore J., Newell A.; 1974; How can merlin understand?; http://iiif.library.cmu.edu/file/Newell_box00040_fld03086_doc0001/Newell_box00040_fld03086_doc0001.pdf
123. *Osipov et al.* [2006] - Osipov G., Smirnov I., Tikhomirov I., Vybornova O., Zavjalova O.; Aug 2006; Linguistic Knowledge for Search Relevance Improvement; https://www.researchgate.net/publication/312955317_Linguistic_Knowledge_for_Search_Relevance_Improvement
124. *Osipov et al.* [2008] - G. Osipov, I. Smirnov, I. Tikhomirov, O. Zavjalova; Oct 2008; Application of linguistic knowledge to search precision improvement; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=4670544
125. *Krstev et al.* [2008] - Cvetana Krstev, Ranka Stanković, Ivan Obradović; 2008; The Usage of Various Lexical Resources and Tools to Improve the Performance of Web Search Engines; https://www.academia.edu/19309119/The_Usage_of_Various_Lexical_Resources_and_Tools_to_Improve_the_Performance_of_Web_Search_Engines?auto=download
126. *Marlow* et al. [2008] - Jennifer Marlow, Paul Clough, Juan Cigarrán Recuero, Javier Artiles; 30-03-2008; Exploring the Effects of Language Skills on Multilingual Web Search; https://eprints.whiterose.ac.uk/78487/7/WRRO_78487.pdf
127. *Pujara & Singh* [2018] - Jay Pujara, Sameer Singh; 2018; Mining Knowledge Graphs from Text; https://kgtutorial.github.io/wsdm-slides/Part2_knowledge-extraction.pdf
128. *AI History qbi.uq.edu.au* - History of Artificial Intelligence; https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence
129. *Search Engine Wikipedia 4* - Search Engine; https://en.wikipedia.org/wiki/Search_engine
130. *Beverly Mapes* [2023] - Beverly Mapes, 25-08-2023; A History of Search Engines; https://topofthelist.net/a-history-of-search-engines/
131. *seomechanic.com 1* - Jan 9, 2023; The Complete History of Search Engines; https://www.seomechanic.com/complete-history-search-engines/
132. *Owen Whitcombe* [2022] - Owen Whitcombe; 26-05-2002; The History of Search Engines; https://www.libertymarketing.co.uk/blog/a-history-of-search-engines/

133. *Search Engine Wikipedia 5* - Timeline of web search engines; https://en.wikipedia.org/wiki/Timeline_of_web_search_engines
134. *Liu et al.* [2016] - Jingjing Liu, Chang Liu, N. Belkin; 01 Nov 2016; Predicting information searchers' topic knowledge at different search stages; https://www.semanticscholar.org/paper/Predicting-information-searchers'-topic-knowledge-Liu-Liu/9efdfca1e476541e220d3b8b4f1740c06a5e2663 & https://www.researchgate.net/publication/287977620_Predicting_information_searchers'_topic_knowledge_at_different_search_stages
135. *Duggan & Payne* [2008] - Geoffrey B. Duggan, S. Payne; 6 Apr 2008; Knowledge in the head and on the web: using topic expertise to aid search; https://www.semanticscholar.org/paper/Knowledge-in-the-head-and-on-the-web%3A-using-topic-Duggan-Payne/edd863de6119b8547004a3159ce246e9102f91ef
136. *Mehala & Bhatia* [2020] - N. Mehala, Divyansh Bhatia; 2020; A Concept-Based Approach for Generating Better Topics for Web Search Results; https://www.researchgate.net/publication/344268302_A_Concept-Based_Approach_for_Generating_Better_Topics_for_Web_Search_Results or https://link.springer.com/article/10.1007/s42979-020-00311-y
137. *Panagis et al.* [2006] - Y. Panagis, E. Sakkopoulos, J. Garofalakis, A. Tsakalidis; 2006; Optimisation mechanism for web search results using topic knowledge; https://www.semanticscholar.org/paper/Optimisation-mechanism-for-web-search-results-using-Panagis-Sakkopoulos/44f55d115671b803529b529781a09d0646ad71ce
138. *Bobed & Mena* [2016] - Carlos Bobed, Eduardo Mena; Feb 2016; QueryGen: Semantic Interpretation of Keyword Queries over Heterogeneous Information Systems; https://www.researchgate.net/publication/282274792_QueryGen_Semantic_Interpretation_of_Keyword_Queries_over_Heterogeneous_Information_Systems
139. *Bobed et al.* [2010] - Carlos Bobed, Raquel Trillo, Eduardo Mena, Sergio Ilarri; Sep 2010; From Keywords to Queries: Discovering the User's Intended Meaning; https://www.researchgate.net/publication/221194770_From_Keywords_to_Queries_Discovering_the_User's_Intended_Meaning
140. *R. Fidel* [1991] - R. Fidel; 1991; Searchers' selection of search keys: II. Controlled vocabulary or free-text searching; https://www.semanticscholar.org/paper/Searchers'-selection-of-search-keys%3A-II.-Controlled-Fidel/2cafe07ea304d9ea3e16d9e20344be4183e7ba3d
141. *Jianchao Ji* [2023] - Jianchao Ji; 2023; GenRec: Large Language Model for Generative Recommendation; https://www.semanticscholar.org/reader/d9ff3db7a9e37fe0363bb87aa47acd6b6b67977e
142. *Liu et al.* [2021] - Yiding Liu, Guan Huang, Jiaxiang Liu, Weixue Lu, Suqi Cheng, Yukun, Li, Daiting Shi, Shuaiqiang Wang, Zhicong Cheng, Dawei Yin; Oct 2021; Pre-trained Language Model for Web-scale Retrieval in Baidu Search; https://arxiv.org/pdf/2106.03373.pdf
143. *Aggarwal et al.* [2023] - Pranjal Aggarwal, Vishvak Murahari; Tanmay Rajpurohit, Ashwin Kalyan, Karthik R Narasimhan, Ameet Deshpande; 16 Nov 2023; GEO: Generative Engine Optimization; https://arxiv.org/pdf/2311.09735.pdf

144. *Kemtes et al.* [2003] - K. Kemtes, Kelli Smith, B. Donohue, Krisann M. Alvarez, Kim Carpin, Jennifer Sinchak; 01/04/2023: A Systematic Approach to Conducting Educational Literature Searches; https://www.semanticscholar.org/paper/A-Systematic-Approach-to-Conducting-Educational-Kemtes-Smith/88cb7be54cbccd3167633ddae9549806846b54a3
145. *Bani-Ahmad & Ozsoyoglu* [2008] - Sulieman Bani-Ahmad, Gultekin Ozsoyoglu; Jun 2008; On content-driven search-keyword suggesters for literature digital libraries; https://www.researchgate.net/publication/220924185_On_content-driven_search-keyword_suggesters_for_literature_digital_libraries
146. *Hahnel et al.* [2018] - Hahnel Carolin, Goldhammer Frank, Kröhne Ulf, Naumann Johannes; 2018; The role of reading skills in the evaluation of online information gathered from search engine environments; https://www.dipfdocs.de/volltexte/2020/19224/pdf/CiHB_2018_Hahnel_et_al_The_role_of_reading_skills_in_the_evaluation_of_online_information_A.pdf
147. *Hildreth* [1997] - Hildreth, Charles R.; Jun 1997; The Use and Understanding of Keyword Searching in a University Online Catalog.; https://eric.ed.gov/?id=EJ550856
148. *Olaf Kopp* [2023] - Olaf Kopp; 19 Dec 2023; Most interesting Google Patents for SEO in 2023; https://www.kopp-online-marketing.com/interesting-google-patents
149. *Jakelabate.com 1* - Jake Labate; Google's Rich Result SERP Features; https://www.jakelabate.com/serp-features
150. *Woorank.com 1* - Google's SERP Features Explained; https://www.woorank.com/en/edu/seo-guides/what-is-a-googles-serp-feature
151. *SDM-Google.com* - Structured data markup that Google Search supports; https://developers.google.com/search/docs/appearance/structured-data/search-gallery
152. *Yu and Guo* [2023] - Hao Yu, Yunyun Guo; 01 Jun 2023; Generative artificial intelligence empowers educational reform: current status, issues, and prospects; https://www.frontiersin.org/articles/10.3389/feduc.2023.1183162/full
153. *Strickland* [2021] - Eliza Strickland; Oct 2021; The Turbulent Past and Uncertain Future of AI It's been boom-and-bust since the beginning; https://spectrum.ieee.org/files/11920/10_Spectrum_2021.pdf
154. *Martinez et al.* [2023] - Gonzalo Martínez, Lauren Watson, P. Reviriego, José Alberto Hernández, Marc Juárez, Rik Sarkar; 3 Mar 2023; Combining Generative Artificial Intelligence (AI) and the Internet: Heading towards Evolution or Degradation?; https://www.semanticscholar.org/reader/a82d3924f488d74af91ba31ff1cf80a6c07ee206
155. *Lungarella et. al.* [2007] - Max Lungarella, Fumiya Iida, Josh Bongard, Rolf Pfeifer; Dec 2007; AI in the 21st Century: With Historical Reflections; https://www.researchgate.net/publication/225474953_AI_in_the_21st_Century_-_With_Historical_Reflections
156. *Wiener N.* [1948] - Wiener N.; 1948; Cybernetics or Control and Communication in the Animal and the Machine. MIT Press, Cambridge, MA; https://ia801701.us.archive.org/26/items/cybernetics-or-communication-and-control-in-the-animal-and-the-machine-norbert-wiene-ocr/Cybernetics%20or%20Communication%20and%20Control%20in%20the%20Animal%20and%20the%20Machine%20-%20Norbert%20Wiene_OCR.pdf

157. *W.G. Walter* [1950] - Walter G.W.; 1950; An imitation of life. Scientific American 182(5), 42–45; https://people.csail.mit.edu/brooks/idocs/walter50imitation.pdf
158. *A.M. Turing* [1950] - Turing A.M.; 1950; Computing machinery and intelligence. Mind 59(236), 433–460; https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf
159. *Legg & Hutter* [2007] - Legg S., Hutter M.; 2007; Tests of machine intelligence. LNCS, vol. 4850. Springer, Heidelberg; https://www.vetta.org/documents/legg-hutter-2007-tests-machine-intelligence.pdf
160. *P. McCorduck* [2003] - McCorduck P.; 2003; Machines Who Think, 2nd edn. A.K. Peters Ltd., Natick, MA; https://monoskop.org/images/1/1e/McCorduck_Pamela_Machines_Who_Think_2nd_ed.pdf
161. *Buchanan et al.* [1984] - B.G. Buchanan, E. H. Shortliffe; 1984; Rule-Based Expert Systems : The MYCIN Experiments of the Stanford Heuristic Programming Project; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/0004370285900670
162. *McCain* [2006] - Katherine McCain; 18 Oct 2006: Explorations in bibliometric historiography: The (re)emergence of neural networks, 1980-1991; https://www.issi-society.org/proceedings/issi_2005/McCain_ISSI2005.pdf
163. *Heckerman* [1997] - D. Heckerman: Mar 1997; Bayesian Networks for Data Mining; https://www.researchgate.net/publication/220451914_Bayesian_Networks_for_Data_Mining
164. *R. Evans* [2014] - Robin J. Evans; Dec 2014; Graphs for Margins of Bayesian Networks; https://www.stats.ox.ac.uk/~evans/pres/ERCIM.pdf
165. *Ethan & Fookes* [2020] - Ethan Goan, C. Fookes; 28 May 2020; Bayesian Neural Networks: An Introduction and Survey; https://www.researchgate.net/publication/342377286_Bayesian_Neural_Networks_An_Introduction_and_Survey
166. *Rawat & Wang* [2017] - Waseem Rawat, Zenghui Wang; 1 Sep 2017; Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review; https://www.researchgate.net/publication/317496930_Deep_Convolutional_Neural_Networks_for_Image_Classification_A_Comprehensive_Review
167. *Bonder* [2023] - Sasha Bonder; 14 Sep 2023; Generative Adversarial Networks (GANs) with R; https://reintech.io/blog/understanding-gans-with-r
168. *GAN Wikipedia* - Generative adversarial network; https://en.wikipedia.org/wiki/Generative_adversarial_network
169. *Lan et al.* [2020] - L. Lan, Lei You, Zeyang Zhang, Zhiwei Fan, Weiling Zhao, Nianyin Zeng, Yidong Chen, Xiaobo Zhou; 12 May 2020; Generative Adversarial Networks and Its Applications in Biomedical Informatics; https://www.researchgate.net/publication/348837975_Generative_adversarial_network_An_overview_of_theory_and_applications
170. *TLTM Wikipedia* - Transformer (deep learning architecture); https://en.wikipedia.org/wiki/Transformer_(deep_learning_acrhitecture)
171. *Wolf et al.* [2019] - Thomas Wolf, Victor Sanh, Julien Chaumond, Clement Delangue; 23 Jan 2019; TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents; https://thomwolf.io/data/TransferTransfo_final.pdf

172. *Real et al.* [2020] - Esteban Real, Chen Liang, David R. So, Quoc V. Le; 2020; AutoML-Zero: Evolving Machine Learning Algorithms From Scratch; https://proceedings.mlr.press/v119/real20a/real20a.pdf
173. *Kuang & Leung* [2019] - Nikki Lijing Kuang, Clement H.C. Leung; 22 Nov 2019; Analysis of Evolutionary Behaviour in Self-Learning Media Search Engines; https://arxiv.org/pdf/1911.09882.pdf
174. *Kumar et al.* [2023] - Manish Kumar, Shueb Ali Khan, Abhay Bhatia, Parag Jain; Feb 2023; A Conceptual introduction of Machine Learning Algorithms; https://www.researchgate.net/publication/371515358_A_Conceptual_introduction_of_Machine_Learning_Algorithms
175. *CWV-Google* [2020] - Introducing INP to Core Web Vitals; https://developers.google.com/search/blog/2023/05/introducing-inp#
176. *Ronallo* [2012] - Jason Ronallo; 3 Feb 2012; HTML5 Microdata and Schema.org; https://journal.code4lib.org/articles/6400
177. *Karle et al.* [2017] - Elias Kärle, Umutcan Simsek, Zaenal Akbar, Martin Hepp, D. Fensel; 2017; Extending the Schema.org Vocabulary for More Expressive Accommodation Annotations; https://www.researchgate.net/publication/312046435_Extending_the_Schemaorg_Vocabulary_for_More_Expressive_Accommodation_Annotations
178. *Oliveira & Teixeira Lopes* [2023]-1 - Bruno Oliveira, Carla Teixeira Lopes; January 2023; From 10 Blue Links Pages to Feature-Full Search Engine Results Pages: Analysis of the Temporal Evolution of SERP Features; https://arxiv.org/pdf/2301.08042.pdf
179. *Oliveira & Teixeira Lopes* [2023]-2 - Bruno Oliveira, Carla Teixeira Lopes; Mar 2023; The Evolution of Web Search User Interfaces: An Archaeological Analysis of Google Search Engine Result Page; https://dl.acm.org/doi/fullHtml/10.1145/3576840.3578320
180. *Brodieclark.com 1* - Google SERP Features Timeline; 2023; https://brodieclark.com/notes/
181. *SERP Bedgarone 1* - https://bedgarone.github.io/serpevolution/timeline/2020
182. *Carter* [2015] - Jamie Carter; 18 Apr 2015; Back to basics: is Web 1.0 making a comeback?; https://www.techradar.com/news/internet/web/is-web-1-0-making-a-big-comeback-1291121
183. *Graham & Balachander* - Graham Cormode, Balachander Krishnamurthy; Key differences between Web 1.0 and Web 2.0; https://firstmonday.org/ojs/index.php/fm/article/view/2125/1972
184. *Nath* [2022] - Keshab Nath; 16 May 2022; Evolution of the Internet from Web 1.0 to Metaverse: The Good, The Bad and The Ugly; https://www.techrxiv.org/doi/full/10.36227/techrxiv.19743676.v1
185. *Kumar* [2022] - Rishikumar Suresh Kumar; 01 Jan 2022; AN OVERVIEW OF THE EXPECTED INFLUENCE OF WEB 3.0 ON e-COMMERCE AND ALLIED DOMAINS; International Journal of Trendy Rresearch in Engineering and Technology; https://www.trendytechjournals.com/ijtret/volume6/issue6-7.pdf
186. *Patel* [2013] - Karan Patel; 01 Jan 2013; Incremental Journey for World Wide Web: Introduced with Web 1.0 to Recent Web 5.0: A Survey Paper;

https://www.researchgate.net/publication/262562142_Incremental_Journey_for_World_Wide_Web_Introduced_with_Web_10_to_Recent_Web_50_-_A_Survey_Paper

187. *Nath et al.* [2014] - Keshab Nath, Sourish Dhar, Subhash Basishtha; 17 Apr 2014; Web 1.0 to Web 3.0 - Evolution of the Web and its various challenges; https://www.researchgate.net/publication/269310255_Web_10_to_Web_30_-_Evolution_of_the_Web_and_its_various_challenges
188. *Faizar* [2018] - Evolution of the Web: Web 1.0, Web 2.0, Web 3.0, Web 4.0; web 5.0 and beyond; https://ahmadfaizar.blogspot.com/2018/08/evolution-of-web-web-10-web-20-web-30.html
189. *Singh & Gulati* [2011] - K. P. Singh, Bebi, Dipti Gulati; 2011; Technological March from Web 1.0 to Web 3.0: A Comparative Study; https://people.du.ac.in/~kpsingh/wp-content/uploads/2014/National/Library%20Herald/Web%201-%20Web%202%202011.pdf
190. *Kollmann & Lomberg* [2010] - T. Kollmann, C. Lomberg; 2010; Web 1.0, Web 2.0 and Web 3.0: The Development of E-Business; https://www.igi-global.com/chapter/web-web-web/41283
191. *Onday* [2019] - Özgür Önday; 2019; Web 6.0: Journey From Web 1.0 To Web 6.0; Journal of Media & Management; https://www.researchgate.net/publication/337741822_Web_60_Journey_From_Web_10_To_Web_60
192. *Serrano & Gelenbe* [2018] - Will Serrano, Erol Gelenbe; 2018; The Random Neural Network in a neurocomputing application for Web search; https://www.researchgate.net/publication/321305398_The_Random_Neural_Network_in_a_Neurocomputing_Application_for_Web_Search
193. *Ye et al.* [2022] - Ziyi Ye, Xiaohui Xie, Yiqun Liu, Zhihong Wang, Xuancheng Li, Jiaji Li, Xuesong Chen, M. Zhang, Shaoping Ma; 2022; Why Don't You Click: Understanding Non-Click Results in Web Search with Brain Signals; Annual International ACM SIGIR Conference on Research and Development in Information Retrieval; http://www.thuir.cn/group/~mzhang/publications/SIGIR-2022-YeZiyi.pdf
194. *Iraji et al.* [2015] - M. S. Iraji, Hakimeh Maghamnia, Marzieh Iraji; 2015; Web Pages Retrieval with Adaptive Neuro Fuzzy System based on Content and Structure; https://www.mecs-press.org/ijmecs/ijmecs-v7-n8/IJMECS-V7-N8-8.pdf
195. *Chen & Décary* [2018] - Mei Chen, Michel Décary; 2018; A Cognitive-Based Semantic Approach to Deep Content Analysis in Search Engines; International Computer Science Conference; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=8334450
196. *Xu et al.* [2023] - Zhichao Xu, Hemank Lamba, Qingyao Ai, Joel Tetreault, Alex Jaimes; 25 Jan 2023; Counterfactual Editing for Search Result Explanation; https://typeset.io/papers/counterfactual-editing-for-search-result-explanation-28wh6ql8
197. *Borisov et al.* [2016] - Alexey Borisov, Ilya Markov, Maarten de Rijke, Pavel Serdyukov; 11 Apr 2016; A Neural Click Model for Web Search; https://staff.fnwi.uva.nl/m.derijke/wp-content/papercite-data/pdf/borisov-neural-2016.pdf

198. *Mostafa & Gwizdka* [2016] - Javed Mostafa, Jacek Gwizdka; 13 Mar 2016; Deepening the Role of the User: Neuro-Physiological Evidence as a Basis for Studying and Improving Search; https://dl.acm.org/doi/pdf/10.1145/2854946.2854979
199. *Alnatsheh* [2020] - Essam Alnatsheh; 19 Nov 2020; User Behaviour for Neural Network-based Web Search Results Filtering; https://typeset.io/papers/user-behaviour-for-neural-network-based-web-search-results-4rldqi2cow
200. *Shu et al.* [2020] - Yao Shu, Wei Wang, Shaofeng Cai; 30 Apr 2020; Understanding Architectures Learnt by Cell-based Neural Architecture Search; https://typeset.io/papers/understanding-architectures-learnt-by-cell-based-neural-mqlrxzt2k3
201. *Lorigo et al.* [2008] - Lori Lorigo, Maya Haridasan, Hrönn Brynjarsdóttir, Ling Xia, Thorsten Joachims, Geri Gay, Laura Granka, Fabio Pellacini, Bing Pan; 2008; Eye tracking and online search: lessons learned and challenges ahead. journal of the american society for information science and technology, 59(7), 1041-1052. https://doi.org/10.1002/asi.20794; https://www.researchgate.net/publication/220434543_Eye_tracking_and_online_search_Lessons_learned_and_challenges_ahead
202. *Cordova & Jansen* [2012] - Adan Ortiz-Cordova, Jim Jansen; 2012; Classifying web search queries to identify high revenue generating customers; journal of the american society for information science and technology, 63(7), 1426-1441; https://www.researchgate.net/publication/260282044_Classifying_Web_Search_Queries_to_Identify_High_Revenue_Generating_Customers
203. *Cho & Roy* [2004] - Junghoo Cho, Sourashis Roy; May 2004; Impact of search engines on page popularity; https://oak.cs.ucla.edu/~cho/papers/cho-bias.pdf & https://doi.org/10.1145/988672.988676
204. *Keane et al.* [2008] - Mark T. Keane, Maeve O'Brien, Barry Smyth; 2008; Are people biased in their use of search engines?; https://www.researchgate.net/publication/220421158_Are_people_biased_in_their_use_of_search_engines
205. *Jansen & Resnick* [2006] - Bernard J. Jansen, Marc L. Resnick; 2006; An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce Web searching; https://faculty.ist.psu.edu/jjansen/academic/pubs/sponsored_links_jansen.pdf
206. *Sanders & Jones* [2006] - Virginia M. Sanders, Kathryn J. Jones; 2006; Role of Immunity in Recovery from a Peripheral Nerve Injury; https://link.springer.com/article/10.1007/s11481-005-9004-0
207. *Matsuda et al.* [2009] - Yuko Matsuda, Hidetake Uwano, Masaichi Ohira; 2009; An Analysis of Eye Movements during Browsing Multiple Search Results Pages; https://link.springer.com/content/pdf/10.1007/978-3-642-02574-7_14
208. *Nayak google.com* [2021] - Pandu Nayak; 18 May 2021; MUM: A new AI milestone for understanding information; https://blog.google/products/search/introducing-mum/
209. *Ontology Wikipedia* - https://en.wikipedia.org/wiki/Ontology_language
210. *Marketbrew-1* - Ontology Engineering for SEO; https://marketbrew.ai/ontology-engineering-for-seo

211. *Keyser* [2012] - Pierre de Keyser; 2012; Semantic search engine; https://www.sciencedirect.com/topics/computer-science/semantic-search-engine
212. *Sayed & Muqrishi* [2017] - Awny Sayed, Amal Al Muqrishi; 2017; IBRI-CASONTO: Ontology-based semantic search engine; https://www.sciencedirect.com/science/article/pii/S111086651730004X
213. *Biznology-1* [2018] - 9 Feb 2018; What Is A Keyword Ontology?; https://biznology.com/2018/02/what-is-a-keyword-ontology/
214. *Hou & Tao* [2019] - Yu Hou, Lixin Tao; 30 Sep 2019; An Ontology-based Ranking Model in Search Engines; https://typeset.io/papers/an-ontology-based-ranking-model-in-search-engines-3ezzuy5f4b
215. *Zaid et al.* [2013] - Norasykin Mohd Zaid, Sim Kim Lau, Mohd Nihra Haruzuan Mohamad Said; 06 Nov 2013; Ontology-based search system using hierarchical structure design; https://typeset.io/papers/ontology-based-search-system-using-hierarchical-structure-bo6ovm3phg
216. *T. Shakti et al.* [2012] - T. Sakthi Priya, P. Revathy, T. Pradeesh, C. R. Rene Robin; 01 Jan 2012; Design and Development of an Ontology based Personal Web Search Engine; https://www.academia.edu/75021187/Design_and_Development_of_an_Ontology_based_Personal_Web_Search_Engine
217. *Uciteli et al.* [2017] - Alexandr Uciteli, Stefan Kropf, Timo Weiland, Stefanie Meese; 01 Jan 2017; Search Ontology Generator (SONG): Ontology-Based Specification and Generation of Search Queries; https://typeset.io/papers/search-ontology-generator-song-ontology-based-specification-3akbo3q7le
218. *Obitko et al.* [2004] - Marek Obitko, Vaclav Snasel, Jan Smid; 01 Jan 2004; Ontology Design with Formal Concept Analysis; https://typeset.io/papers/ontology-design-with-formal-concept-analysis-2oe3bpiyu7
219. *Barros et al.* [1998] - F. Barros, P. Gonçalves, Thiago L. V. L. Santos; 1998; Providing Context to Web Searches: The Use of Ontologies to Enhance Search Engine's Accuracy; Journal of the Brazilian Computer Society; https://www.scielo.br/j/jbcos/a/yQcVrrhJ363ZcxDLDvV5RxF/
220. *Yong et al.* [2007] - Qi Yong, Hao Peijie, Hou Yu, Qiao Ya-nan; 2007; The Research and Design of the Semantic Search Engine Based on Ontology; International Conference on Semantics, Knowledge and Grid; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=4438646
221. *Kavitha et al.* [2015] - V. Kavitha, M. Hanumanthappa, B. Prakash; 2015; Ontology Based Search Engine; https://www.semanticscholar.org/paper/Ontology-Based-Search-Engine-Kavitha-Hanumanthappa/c91d3fb400910c4a5b5c2144f253771f6ad55cbd
222. *Esa et al.* [2010] - A. M. Esa, S. Taib, Hong Nguyen Thi; 2010; Prototype of semantic search engine using ontology; IEEE Conference on Open Systems; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=5720073
223. *Slawski* [2022] - Bill Slawski; 19 April 2022: Topicality Scores, Social Scores and User-Generated Content At Google; https://gofishdigital.com/blog/topicality-scores-social-scores-and-user-generated-content/
224. *SAP.com-1* - Configuring Topicality of Search Results; https://help.sap.com/doc/saphelp_nw75/7.5.5/EN-US/48/63917cf29407bee10000000a42189d/content.htm?no_cache=true

225. *Wang et al.* [2023] - MENG WANG, JIANPING LIU, JIAN WANG, YINGFEI WANG, AND XINTAO CHU; 6 july 2023; A Topicality Relevance-Aware Intent Model for Web Search; https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10164089
226. *Allen* [1991] - Bryce Allen; April 1991; Topic Knowledge and Online Catalog Search Formulation; https://www-jstor-org.ressources-electroniques.univ-lille.fr/stable/pdf/4308578.pdf?
227. *Topics Microsoft* – Changes coming to Topics; https://learn.microsoft.com/en-us/microsoft-365/topics/topic-experiences-overview?view=o365-worldwide
228. *Topic Filters Salesforce* - Refine Knowledge Search Results with Topic Filters in Experience Cloud Sites; https://help.salesforce.com/s/articleView?id=release-notes.rn_search_esk_topic_filters.htm&release=238&type=5
229. *Pardi et al.* [2023] - Georg Pardi, Steffen Gottschling, Peter Gerjets, Yvonne Kammerer; 3 Feb 2023; The moderating effect of knowledge type on search result modality preferences in web search scenarios; https://pdf.sciencedirectassets.com/777244/1-s2.0-S2666557322X00025/1-s2.0-S2666557323000058/main.pdf?
230. *Kumaran* [2012] - Girdhar Kumaran; 2012; Biasing search results toward topics of interest using embedded relevance links; https://patents.google.com/patent/US20130204892A1/en
231. *Sajid* [2022] - Mohamed Sajid; 31 Jul 2022; Topicalization Defined by Syntax; https://typeset.io/papers/topicalization-defined-by-syntax-1vlg106x
232. *Christensen & Van Hoff* [2015] - Jens Christensen, Arthur Van Hoff; 23 Mar 2015; Topical Search System; https://patents.google.com/patent/US9875309B2/en
233. *Sondhi et al.* [2010] - P. Sondhi, Raman Chandrasekar, Robert Rounthwaite; 2010; Using query context models to construct topical search engines; https://www.researchgate.net/publication/228656877_Using_Query_Context_Models_To_Construct_Topical_Search_Engines
234. *Stamou & Ntoulas* [2008] - Sofia Stamou, Alexandros Ntoulas; 2008; Search personalization through query and page topical analysis; https://cgi.di.uoa.gr/~antoulas/pubs/ntoulas_personalization.pdf
235. *Kumar & Lang* [2007] - Nanda Kumar, Karl Reiner Lang; 2007; Do Search Terms Matter for Online Consumers? The Interplay Between Search Engine Query Specification and Topical Organization; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0167923607000620
236. *Chuklin & Rjike* [2015] - A. Chuklin, M. de Rijke; 2015; The Anatomy of Relevance: Topical, Snippet and Perceived Relevance in Search Result Evaluation; https://arxiv.org/pdf/1501.06412.pdf
237. *Liu & Zhang* [2019] - Jingjing Liu, Xiangmin Zhang; 01 Nov 2019; The role of domain knowledge in document selection from search results; https://www.researchgate.net/publication/331568554_The_role_of_domain_knowledge_in_document_selection_from_search_results
238. *Huang et al.* [2017] - Yen-Chun Huang, Yu-Ping Ho, I-Chin Wu; 09 Jul 2017; Analyzing Users' Search Patterns to Explore Topic Knowledge from Aggregated Search Results; https://link.springer.com/chapter/10.1007/978-3-319-58750-9_61

239. *Duggan & Payne* [2008] - Geoffrey B. Duggan, S. Payne; 2008; Knowledge in the head and on the web: using topic expertise to aid search; https://www.researchgate.net/publication/221518536_Knowledge_in_the_head_and_on_the_Web_Using_topic_expertise_to_aid_search
240. *Zhang et al.* [2013] - Xiangmin Zhang, Jingjing Liu, Michael J. Cole; Jan 2013; Task Topic Knowledge vs. Background Domain Knowledge: Impact of Two Types of Knowledge on User Search Performance; https://www.researchgate.net/publication/281071433_Task_Topic_Knowledge_vs_Background_Domain_Knowledge_Impact_of_Two_Types_of_Knowledge_on_User_Search_Performance
241. *Google Cloud Topicality* – Improve Search Quality; https://developers.google.com/cloud-search/docs/guides/improve-search-quality
242. *Huang and Soergel* [2004] - Xioali Huang and Dagobert Soergel; 2004; Relevance Judges' Understanding of Topical Relevance Types: An Explication of an Enriched Concept of Topical Relevance; https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/meet.1450410118
243. *Coleman and Thrasher* [2006] - Jason Coleman, Steven Thrasher; 14 Dec 2006; Internet searching enhancement method for determining topical relevance scores; https://patents.google.com/patent/US7457802B2/en
244. *Taylor* [2011] - Arthur Taylor; June 2011; User relevance criteria choices and the information search process; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0306457311000483
245. *Bruza & Chang* [2014] - Peter Bruza and Vivien Chang; 2014; Perceptions of document relevance; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4079065/pdf/fpsyg-05-00612.pdf
246. *Gerjets & Kammerer* [2010] - Peter Gerjets, Yvonne Kammerer; 2010; Topical Relevance and Information Quality in Cognitive Models of Web Search Behaviour: Introducing Epistemic Scent into Information Foraging Theory - eScholarship; https://escholarship.org/content/qt0ks508zz/qt0ks508zz.pdf?t=op2iyd
247. *Wu et al.* [2014] - Wan-Ching Wu, D. Kelly, Avneesh Sud; 2014; Using information scent and need for cognition to understand online search behaviour; https://static.aminer.org/pdf/20170130/pdfs/sigir/t2cpmco01w8l3f5ysxh4qzoie7emknxb.pdf
248. *Qu et al.* [2010] - Peng Qu, Chang Liu, Maosheng Lai; 2010; The effect of task type and topic familiarity on information search behaviours; https://www.researchgate.net/publication/247926910_The_effect_of_task_type_and_topic_familiarity_on_information_search_behaviours
249. *Gwizdka & Chang* [2020] - Jacek Gwizdka, Yung-Sheng Chang; 1 Jan 2020; Search Results Viewing Behaviour vis-à-vis Relevance Criteria; https://www.researchgate.net/publication/336776952_Search_Results_Viewing_Behaviour_vis-a-vis_Relevance_Criteria
250. *Greisdorf & Connor* [2003] - Howard Greisdorf, Brian C. O'Connor; 01 Dec 2003; Nodes of topicality: modeling user notions of on topic documents; https://www.academia.edu/72512675/Nodes_of_topicality_Modeling_user_notions_of_on_topic_documents

251. *Jin et al.* [2019] - Menghan Jin, Lingling Ji, Huamao Peng; 2019; The Relationship Between Cognitive Abilities and the Decision-Making Process: The Moderating Role of Self-Relevance; https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01892/full
252. *Thurow* [2009] - Shari Thurow; 23 Oct 2009; Do search engine optimization (SEO) professionals architect websites to match searcher mental models? The answer might surprise you.; https://searchengineland.com/seo-searcher-mental-models-27949
253. *Mlilo & Thatcher* [2011] - Sifiso Mlilo and Andrew Thatcher; 2011; Mental Models: Have Users' Mental Models of Web Search Engines Improved in the Last Ten Years?; https://link.springer.com/content/pdf/10.1007/978-3-642-21741-8_27.pdf
254. *Thomas et al.* [2019] - Paul Thomas, Bodo Billerbeck, Nick Craswell, Ryen W. White; 20 Dec 2019; Investigating Searchers' Mental Models to Inform Search Explanations; https://www.researchgate.net/publication/338087451_Investigating_Searchers'_Mental_Models_to_Inform_Search_Explanations
255. *Karanam et Oostendorp* [2020] - Saraschandra Karanam, Herre van Oostendorp; 01 Jan 2020; Cognitive Modeling of Age and Domain Knowledge Differences in Information Search; https://www.researchgate.net/publication/341743579_Cognitive_Modeling_of_Age_and_Domain_Knowledge_Differences_in_Information_Search
256. *Wai-Tat Fu* [2020] - Wai-Tat Fu; 01 Jan 2020; How Cognitive Computational Models Can Improve Information Search; https://link.springer.com/chapter/10.1007/978-3-030-38825-6_3
257. *Karanam et al.* [2017] - Saraschandra Karanam, Herre van Oostendorp, Mylène Sanchiz, Aline Chevalier; 01 Oct 2017; Cognitive modeling of age-related differences in information search behaviour; https://asistdl-onlinelibrary-wiley-com.ressources-electroniques.univ-lille.fr/doi/full/10.1002/asi.23893
258. *Nakhoda et al.* [2016] - Maryam Nakhoda, Zahra Kazempour, Nader Naghshineh, Mahdieh Mirzabeigi; 01 Oct 2016; Adjustment and Development of Health User's Mental Model Completeness Scale in Search Engines; https://www.academia.edu/69215778/Adjustment_and_Development_of_Health_Users_Mental_Model_Completeness_Scale_in_Search_Engines
259. *Nielsen* [2005] - Jakob Nielsen; 8 May 2005; Mental Models For Search Are Getting Firmer; https://www.nngroup.com/articles/mental-models-for-search/
260. *Thomas et al.* [2019]-2 - PAUL THOMAS and BODO BILLERBECK (Microsoft, Australia); NICK CRASWELL and RYEN W. WHITE (Microsoft, USA); 2019; Investigating Searchers' Mental Models to Inform Search Explanations; http://ryenwhite.com/papers/ThomasTOIS2019.pdf
261. *Khoo & Hall* [2012] - Michael Khoo & Catherine Hall; 2012; What Would 'Google' Do? Users' Mental Models of a Digital Library Search Engine; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-642-33290-6_1
262. *Chanceaux et al.* [2012] - Myriam Chanceaux, Anne Guérin-Dugué, Benoît Lemaire, Thierry Baccino; 2012; A Computational Cognitive Model of Information Search in Textual Materials;

https://www.researchgate.net/publication/256000647_A_Computational_Cognitive_Model_of_Information_Search_in_Textual_Materials

263. *Slone* [2002] - Debra J. Slone; 2002; The influence of mental models and goals on search patterns during Web interaction; https://www.researchgate.net/publication/229480622_The_Influence_of_Mental_Models_and_Goals_on_Search_Patterns_During_Web_Interaction
264. *Arrue et al.* [2007] - Myriam Arrue, Markel Vigo, Julio Abascal; 2007; Web accessibility awareness in search engine results; https://www.researchgate.net/publication/220606710_Web_accessibility_awareness_in_search_engine_results
265. *Lopatovska et al.* [2012] - Irene Lopatovska, Megan R. Fenton, Sara Campot; 2012; Examining preferences for search engines and their effects on information behavior; https://www.researchgate.net/publication/259539874_Examining_preference_for_search_engines_and_their_effect_on_information_behavior
266. *Buzzi et al.* [2004] - M. Buzzi, P. Andronico, B. Leporini; 2004; Accessibility and Usability of Search Engine Interfaces: Preliminary Testing; https://www.researchgate.net/publication/228850712_Accessibility_and_usability_of_search_engine_interfaces_Preliminary_testing
267. *Daniel Rose* [2006] - Daniel E. Rose; 2006; Reconciling information-seeking behavior with search user interfaces for the Web; https://www.researchgate.net/publication/220434945_Reconciling_information-seeking_behavior_with_search_user_interfaces_for_the_Web
268. *Gray & Fu* [2000] - Wayne D. Gray, Wai-Tat Fu; 01 Jan 2000; The influence of Source and Cost of Information Access on Correct and Errorful Interactive Behavior - eScholarship; https://typeset.io/papers/the-influence-of-source-and-cost-of-information-access-on-4r6f5mpva4
269. *Lawrence & Giles* [1999] - Steve Lawrence, C.L. Giles; 08 Jul 1999; Accessibility of information on the web; https://typeset.io/papers/accessibility-of-information-on-the-web-h8bjno3lfn
270. *Unesco.com-1* [2023] - Information Accessibility; 2023; https://www.unesco.org/en/ifap/information-accessibility
271. *Esther Kang* [2023] - Esther Kang; 01-09-2023; Easily accessible but easily forgettable: How ease of access to information online affects cognitive miserliness; https://psycnet.apa.org/record/2022-18511-001
272. *w3.org-wai-1* - The Business Case for Digital Accessibility; https://www.w3.org/WAI/business-case/
273. *Hu et al.* [2015] - Sha Hu, Zhicheng Dou, Xiao-Jie Wang, T. Sakai, Ji-Rong Wen; 2015; Search Result Diversification Based on Hierarchical Intents; https://www.researchgate.net/publication/301416420_Search_Result_Diversification_Based_on_Hierarchical_Intents
274. *Radlinski et al.* [2010] - Filip Radlinski, M. Szummer, Nick Craswell; 2010; Inferring query intent from reformulations and clicks; https://www.microsoft.com/en-us/research/wp-content/uploads/2010/04/RadlinskiSzummerCraswell-query-intent-www10.pdf

275. *Awad et al.* [2012] - Aya Awad, M. El-Sayed, Yasser El-Sonbaty; 2012; Approximating User's Intention for Search Engine Queries; https://www.researchgate.net/publication/237138553_Approximating_User%27s_Intention_for_Search_Engine_Queries
276. *Bernhardt et al.* [2013] - Daniel Bernhardt, Antonio Gulli, Rui Andre Augusto Ferreira1, Elbio Renato Torres Abib; 14 Mar 2013; Clarifying User Intent of Query Terms of a Search Query; https://patents.justia.com/patent/20140279993
277. *Jain et al.* [2013] - Rekha Jain, Rupal Bhargava, Sulochana Nathawat; 01 Jan 2013; Sense Disambiguation in Information Retrieval; https://www.academia.edu/93204907/Sense_Disambiguation_in_Information_Retrieval
278. *Maarten* [2014] - Maarten de Rijke; 26 Nov 2014; Diversity, Intent, and Aggregated Search; https://typeset.io/papers/diversity-intent-and-aggregated-search-lmj23cudrh
279. *Gajendragadkar & Joshi* [2017] - Uma Gajendragadkar, Sarang Joshi; 01 Feb 2017; Context Sensitive Search String Composition Algorithm using User Intention to Handle Ambiguous Keywords; https://typeset.io/papers/context-sensitive-search-string-composition-algorithm-using-1qv4ujzzkr
280. *Yano et al.* [2016] - Yuki Yano, Yukihiro Tagami, Akira Tajima; 24 Oct 2016; Quantifying Query Ambiguity with Topic Distributions; https://static.aminer.org/pdf/fa/cikm2016/shp0062-yanoA.pdf
281. *Ali-LN* - Muhammad Ali Javed; How does the search algorithm handle ambiguous queries?; https://www.linkedin.com/advice/0/how-does-search-algorithm-handle-ambiguous-dfnif
282. *Julia-Ryte* - Julia Fahndrich; Search intent: A guide for better SEO and rankings; https://en.ryte.com/magazine/search-intent-seo
283. *Claire* [2023] - Claire Brain; 18 Aug 2023; How to optimize for search intent: 19 practical tips; https://searchengineland.com/optimize-search-intent-tips-430857
284. *Chesnevar et al.* [2006] - C. Chesñevar, C. M. Lorenzetti, Ana Gabriela Maguitman, F. Sagui, Guillermo R. Simari; 2006; Exploiting user context and preferences for intelligent web search; https://cs.uns.edu.ar/~cml/www/es/papeles/download/4765.pdf
285. *Barathi & Valli* [2011] - M. Barathi, S. Valli; 2011; Context Disambiguation Based Semantic Web Search for Effective Information Retrieval; https://thescipub.com/pdf/jcssp.2011.548.553.pdf
286. *Gao et al.* [2010] - Byron J. Gao, D. Anastasiu, Xing Jiang; 2010; Utilizing User-input Contextual Terms for Query Disambiguation; https://aclanthology.org/C10-2038.pdf
287. *Challam et al.* [2007] - Vishnu Challam, Susan Gauch, A. Chandramouli; 2007; Contextual Search Using Ontology-Based User Profiles; https://www.researchgate.net/publication/221509921_Contextual_Search_Using_Ontology-Based_User_Profiles
288. *Almasri* [2014] - Almasri Islam; 16 Sep 2014; Contextual Distance Refining for Image Retrieval; https://typeset.io/papers/contextual-distance-refining-for-image-retrieval-2axpz6ybkf
289. *Xie & Ou* [2010] - Yuheng Xie, Wenwu Ou; 12 Jan 2010; Search engine for refining context-based queries based upon historical user feedback; https://patents.google.com/patent/EP2382534A4

290. *Sree & Santhanalakshmi* [2014] - R. Nishitha Sree, S. T. Santhanalakshmi; 01 Jan 2014; Context based search using improvised refinder; https://www.iraj.in/journal/journal_file/journal_pdf/3-45-139729353559-63.pdf
291. *Limbu et al.* [2009] - Dilip K. Limbu, Andy M. Connor, Russel Pears, Stephen G. MacDone; 27 Apr 2009; Improving Web Search Using Contextual Retrieval; https://arxiv.org/ftp/arxiv/papers/1407/1407.6101.pdf#:~:text=Research%20in%20contextual%20retrieval%20approaches,information%20into%20a%20search%20system.
292. *Shokouhi et al.* [2015] - Milad Shokouhi, Marc Sloan, Paul N. Bennett, Kevyn Collins-Thompson; 18 May 2015; Query Suggestion and Data Fusion in Contextual Disambiguation; https://typeset.io/papers/query-suggestion-and-data-fusion-in-contextual-4prlsrfupg
293. *Tunkelang* [2017] - Daniel Tunkelang; 9 Oct 2017; Session Context; https://queryunderstanding.com/session-context-4af0a355c94a
294. *Taylor* [2012] - David C. Taylor; User-context-based search engine; https://patents.google.com/patent/US9449105B1/en
295. *Chen et al.* [2019] - Jia Chen, Jiaxin Mao, Yiqun Liu, Min Zhang; 20 Sep 2019; Investigating Query Reformulation Behavior of Search Users; https://xuanyuan14.github.io/files/CCIR19chen.pdf
296. *Flitcroft* [2011] - Daniel Ian Flitcroft; 13 Jan 2011; Linguistically enhanced search engine and meta-search engine; https://typeset.io/papers/linguistically-enhanced-search-engine-and-meta-search-engine-2gkc97pezq
297. *Okland* [2011] - Vegard Okland; 01 Jan 2011; Utilizing linguistic analysis in multiple source search engines; https://typeset.io/papers/utilizing-linguistic-analysis-in-multiple-source-search-3rfqtxeehc
298. *Lavallee-Google-Patent* [2013] - Jean-Francois Lavallee; 11 Mar 2013; Method for refining a search; https://patents.google.com/patent/US20140257811A1/en
299. *Johnson et al.* [2004] - David Johnson, Malhotra, P. Vamplew, S. Patro; 2004; Refining Search Queries From Examples Using Boolean Expressions and Latent Semantic Analysis; https://figshare.utas.edu.au/articles/conference_contribution/Refining_Search_Queries_From_Examples_Using_Boolean_Expressions_and_Latent_Semantic_Analysis/23212448
300. *García-Serrano et al.* [2001] - Ana M. García-Serrano, Paloma Martínez, Alberto Ruiz; 2001; Linguistic engineering approach to the enhancement of web-searching; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=969853
301. *Seig et al.* [2007] - A. Sieg, B. Mobasher, R. Burke; 2007; Ontological User Profiles for Personalized Web Search; https://www.researchgate.net/publication/228369094_Ontological_user_profiles_for_personalized_Web_search
302. *Khopkar et al.* [2003] - Yashmeet Khopkar, A. Spink, C. Lee Giles, Prital Shah, S. Debnath; 2003; Search engine personalization: An exploratory study; https://www.researchgate.net/publication/220167483_Search_engine_personalization_An_exploratory_study
303. *Jayanthi & Jayakumar* [2011] - J. Jayanthi, K. Jayakumar; 2011; An Integrated Page Ranking Algorithm for Personalized Web Search; https://www.ijcaonline.org/volume12/number11/pxc3872350.pdf

304. *Rashed et al.* [2022] - Shamma Rashed, Tasnim Said, Amal Abdulrahman, Arsiema Yohannes, Monther Aldwairi; 15 Nov 2022; Evaluating Web Search Engines Results for Personalization and User Tracking; https://typeset.io/papers/evaluating-web-search-engines-results-for-personalization-vq2qjpfl
305. *Singhal et al.* [2020] - Singhal Aman, De Barros Marcelo, Srinivasan Prithvishankar; 18 Jun 2020; Personalized search result rankings; https://typeset.io/papers/personalized-search-result-rankings-2c3f22sbyr
306. *Ge et al.* [2019] - Songwei Ge, Zhicheng Dou, Zhengbao Jiang, Jian-Yun Nie; 20 Aug 2019; Personalizing Search Results Using Hierarchical RNN with Query-aware Attention; https://arxiv.org/pdf/1908.07600.pdf
307. *Membrillo* [2020] - Alex Membrillo; 23 June 2020; How To Incorporate Personalization Into Your SEO Strategy; https://www.forbes.com/councils/forbesagencycouncil/2020/01/23/how-to-incorporate-personalization-into-your-seo-strategy/
308. *Marketbrew-3* - How Search Engine Personalization Affects Rankings; https://marketbrew.ai/how-search-engine-personalization-affects-rankings
309. *Mehta* [2023] - Jimit Mehta; 18 Nov 2023; The impact of website personalization on search engine optimization; https://abmatic.ai/blog/impact-of-website-personalization-on-search-engine-optimization
310. *Torres* [2021-2023] - Sam Torres; 17 Sep 2021 & 25 April 2023; Personalization & SEO: How to Optimize for Personalized Search; https://thegray.company/blog/personalization-seo
311. *Yu et al.* [2012] - Jinxiu Yu, Yueming Lu, Fangwei Zhang, Songlin Sun; 2012; A metric for measuring web search results satisfaction incorporating user behavior; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=6664241
312. *Dan & Davison* [2016] - Ovidiu Dan, Brian D. Davison; 2016; Measuring and Predicting Search Engine Users' Satisfaction; http://old-eclass.uop.gr/modules/document/file.php/DIT104/%CE%92%CE%B9%CE%B2%CE%BB%CE%B9%CE%BF%CE%B3%CF%81%CE%B1%CF%86%CE%AF%CE%B1%202017-2018/Measuring%20and%20Predicting%20Search%20Engine%20Users%20Satisfaction.pdf
313. *Awadallah & White* [2013] - Ahmed Hassan Awadallah, Ryen W. White; 2013; Personalized models of search satisfaction; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/2505515.2505681
314. *Mehrotra et al.* [2017] - Rishabh Mehrotra, Imed Zitouni, Ahmed Hassan Awadallah, Ahmed El Kholy; 07 Aug 2017; User Interaction Sequences for Search Satisfaction Prediction; https://www.microsoft.com/en-us/research/wp-content/uploads/2017/04/SIGIR17b.pdf
315. *Ling & Alin* [2019] - Chen Ling, Fan Alin; 10 Sep 2019; Search engine user satisfaction evaluation method based on long-short time memory network; https://typeset.io/papers/search-engine-user-satisfaction-evaluation-method-based-on-4a8ehy8ozn
316. *JungHwa* [2020] - Chae JungHwa; 01 Jan 2020; Study on the Effect of the Search Results User's Experience of Online Search Service on Satisfaction; https://typeset.io/papers/study-on-the-effect-of-the-search-results-user-s-experience-1jfccmce9f

317. *Zhang et al.* [2018] - Fan Zhang, Ke Zhou, Yunqiu Shao, Cheng Luo; 27 Jun 2018; How Well do Offline and Online Evaluation Metrics Measure User Satisfaction in Web Image Search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3209978.3210059
318. *Kersting* [2024] - Max Kersting; 2 Jan 2024; How do you measure user satisfaction with your search system?; https://www.linkedin.com/advice/0/how-do-you-measure-user-satisfaction-your-1c
319. *Palanisamy & Liu* [2019] - Ramaraj Palanisamy & Yifan Liu ; 01 Aug 2019; Users' Search Satisfaction in Search Engine Optimization; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-030-24643-3_124
320. *Koyena* [2021] - Koyena Pal; 2021; The Effect of Multi-Document Summarizations on User SERP Experience; https://www.academia.edu/87807112/The_Effect_of_Multi_Document_Summarizations_on_User_SERP_Experience
321. *Jiang & Liamruk* [2020] - Shanshan Jiang, Pawitra Liamruk; 2020; Effects of SERP Information on Academic Search Behaviours; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=9310951
322. *Maxwell et al.* [2017] - David Maxwell, L. Azzopardi, Yashar Moshfeghi; 2017; A Study of Snippet Length and Informativeness: Behaviour, Performance and User Experience; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/abs/10.1145/3077136.3080824
323. *Elbedweihy et al.* [2012] - Khadija Elbedweihy, Stuart N. Wrigley, Fabio Ciravegna; 2012; Evaluating Semantic Search Query Approaches with Expert and Casual Users; https://link-springer-com.ressources-electroniques.univ-lille.fr/content/pdf/10.1007/978-3-642-35173-0_18.pdf
324. *Moffat* [2022] - Alistair Moffat; 06 Jul 2022; Users: Can't Work With Them, Can't Work Without Them?; https://people.eng.unimelb.edu.au/ammoffat/abstracts/sigir22keynote.pdf
325. *Bink et al.* [2023] - Markus Bink, Sebastian Schwarz, Tim Draws, David Elsweiler; Mar 2023; Investigating the Influence of Featured Snippets on User Attitudes; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3576840.3578323
326. *Mao et al.* [2016] - Xian-Ling Mao, Dan Wang, Yi-Jing Hao; 2016; A Comparative Study of Answer-Contained Snippets and Traditional Snippets; https://link-springer-com.ressources-electroniques.univ-lille.fr/content/pdf/10.1007/978-3-319-48051-0_5.pdf
327. *Bink et al.* [2022] - Markus Bink, Steven Zimmerman, David Elsweiler; Mar 2022; Featured Snippets and their Influence on Users' Credibility Judgements; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/fullHtml/10.1145/3498366.3505766
328. *Neil Patel-3* - How to Do SERP Analysis Like an Expert; https://neilpatel.com/blog/serps-guide/
329. *Yao et al.* [2021] - Jing Yao, Zhicheng Dou, Jun Xu, Ji-Rong Wen; 06 May 2021; RLPS: A Reinforcement Learning–Based Framework for Personalized Search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3446617

330. *Shinde et al.* [2020] - Shrikant Avinash Shinde, Abhilasha Vasantrao Biradar; 01 Apr 2020; Genetic Algorithm for Privacy Protected Personalized Web Search; https://www.jetir.org/papers/JETIR2004400.pdf
331. *Vijayalakshmi et al.* [2021] - C. Vijayalakshmi, S. Pakkir Mohideen; 10 Jan 2021; Personalization of data using machine learning; https://www.researchgate.net/publication/343583392_Personalisation_the_artificial_intelligence_way
332. *Shobby et al.* [2015] - L. Shobby Neeta Fancy, A. Rajamurugan; 25 Jul 2015; Improving Personalized Web Search Qualityin Information Retrieval; https://typeset.io/papers/improving-personalized-web-search-qualityin-information-4okdl6vfap
333. *Jayanthi & Jayakumar* [2012] - J. Jayanthi, K. S. Jayakumar; 12 May 2012; A Novel Page Ranking Algorithm for a Personalized Web Search; https://www.thescipub.com/pdf/jcssp.2012.1029.1035.pdf
334. *Hu Nai* [2002] - Hu Nai; 2002; ALGORITHM DESIGN FOR PERSONALIZATION RECOMMENDATION SYSTEMS; https://www.semanticscholar.org/paper/ALGORITHM-DESIGN-FOR-PERSONALIZATION-RECOMMENDATION-Nai/b7df0485e204e8cca007e01b64e66b9fe8dac52d
335. *Venugopal et al.* [2009] - K. R. Venugopal, K. G. Srinivasa, L. M. Patnaik; 2009; Algorithms for Web Personalization; Page 217-229; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-642-00193-2_10
336. *Chirita et al.* [2004] - Paul-Alexandru Chirita, Daniel Olmedilla, Wolfgang Nejdl; 2004; PROS: A Personalized Ranking Platform for Web Search; https://link-springer-com.ressources-electroniques.univ-lille.fr/content/pdf/10.1007/978-3-540-27780-4_7.pdf?pdf=preview
337. *Huang et al* [2004] - Guanglin Huang, Wenyin Liu; 2004; An Online Adaptive Method for Personalization of Search Engines; Page 422-227; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-540-30480-7_44
338. *searchenginejournal-2025* – Sep. 2025; History of Google Algorithm Updates; https://www.searchenginejournal.com/google-algorithm-history/
339. *Fathy et al.* [2011] - Naglaa Fathy, N. Badr, M. Hashem, Tarek F. Gharib; 2011; Enhancing Web Search with Semantic Identification of User Preferences; https://www.researchgate.net/publication/235333439_Enhancing_Web_Search_with_Semantic_Identification_of_User_Preferences
340. *Garofalakis et al.* [2008] - John D. Garofalakis, Theodoula Giannakoudi, Agoritsa Vopi; 2008; Personalized Web Search by Constructing Semantic Clusters of User Profiles; https://link-springer-com.ressources-electroniques.univ-lille.fr/content/pdf/10.1007/978-3-540-85565-1_30.pdf
341. *Lin et al.* [2005] - Chenxi Lin, Gui-Rong Xue, Hua-Jun Zeng, Yong Yu; 2005; Using Probabilistic Latent Semantic Analysis for Personalized Web Search; https://www.researchgate.net/publication/221240578_Using_Probabilistic_Latent_Semantic_Analysis_for_Personalized_Web_Search

342. *Xu et al.* [2016] - Zheng Xu, Xiao Wei, Yunhuai Liu, Lin Mei, Chuanping Hu, Kim-Kwang Raymond Ch; 2016; Building the search pattern of web users using conceptual semantic space model; https://www.researchgate.net/publication/308182444_Building_the_search_pattern_of_web_users_using_conceptual_semantic_space_model
343. *Raza et al.* [2021] - Muhammad Ahsan Raza, Mokhtar Rahmah, Fahad Qaswar, Anam Shafqat; 01 Jul 2021; User interest driven semantic query expansion for effective web search; https://www.i-csrs.org/Volumes/ijasca/2021.2.2.pdf
344. *Dennis & Stone* [2009] - Simon Dennis, Benjamin Stone; 01 Jan 2009; Comparative Analysis of Semantic Models and Corpora Choice when using Semantic Fields to Predict Eye Movement on Webpages; https://www.researchgate.net/publication/228878265_Comparative_Analysis_of_Semantic_Models_and_Corpora_Choice_when_using_Semantic_Fields_to_Predict_Eye_Movement_on_Webpages
345. *Chatterjee et al.* [2021] - Seinjuti Chatterjee, Ravi Shankar Mishra, Sagar Raichandani, Prasad Joshi; 01 Jan 2021; Response Prediction and Ranking Models for Large-Scale Ecommerce Search; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-981-33-6656-5_17
346. *Linkedin.com-3* - Sep 18, 2023; How can semantic search and query help users find relevant and trustworthy information online?; https://www.linkedin.com/advice/1/how-can-semantic-search-query-help-users
347. *Hill* [2016] - Jeannie Hill; October 24, 2016; Structured Approach in Web Content Creation for User Conceptual Search; https://www.hillwebcreations.com/structured-approach-web-content-creation-for-user-conceptual-search/
348. *Hill* [2021] - Jeannie Hill; December 2, 2021; What is Semantic Search?; https://www.hillwebcreations.com/what-is-semantic-search/
349. *elastic.co-1* - What is semantic search?; https://www.elastic.co/what-is/semantic-search
350. *Bothra* [2020] - Deepika Bothra; September 16, 2020; Semantic Search and Its Relation to Search Relevance; https://www.searchunify.com/blog/an-introduction-to-semantic-search-and-its-relation-to-search-relevance/
351. *Roberts* [2023] - Tessa Roberts; 21/06/2023; Semantic Search Guide: What Is It And Why Does It Matter?; https://www.bloomreach.com/en/blog/2019/semantic-search-explained-in-5-minutes
352. *Wechsler et al.* [2004] - Katia Wechsler, Jorge A. Baier, Miguel Nussbaum, Ricardo A. Baeza-Yates; 2004; Semantic Search in the WWW Supported by a Cognitive Model; https://www.cs.toronto.edu/~jabaier/publications/wec-etal04.pdf
353. *Aytuğ Onan* [2018] - Aytuğ Onan; 25 Apr 2018; Semantic Query Suggestion Based on Optimized Random Forests; Page 91-102; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-319-91189-2_10
354. *Hong & Sug* [2014] - Hwang Jae Hong, Gu Mi Sug; 05 Dec 2014; Web-based semantic information retrieval system using context awareness ontology; https://patents.google.com/patent/KR101469526B1/en
355. *Sharma & Rana* [2020] - Sunny Sharma, Vijay Rana; 01 Jan 2020; Web Search Personalization Using Semantic Similarity Measure; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-030-29407-6_21

356. *Hojgaard et al.* [2016] - Christian Højgaard, Joachim Sejr, Yun-Gyung Cheong; 21 Apr 2016; Query Disambiguation from Web Search Logs; https://typeset.io/papers/query-disambiguation-from-web-search-logs-2bg8doqzz1
357. *Fiol-Roig et al.* [2011] - Gabriel Fiol-Roig, Margaret Miró-Julià, Eduardo Herraiz; 2011; Data Mining Techniques for Web Page Classification; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-642-19917-2_8
358. *Henzinger* [2004] - Monika Henzinger; 2004; Algorithmic Aspects of Web Search Engines; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-540-30140-0_2
359. *J. Pei* [2009] - J. Pei; 2009; Towards Web Search Engine Scale Data Mining; https://crpit.scem.westernsydney.edu.au/confpapers/CRPITV101Pei.pdf
360. *Banerjee & Dua* [2022] - K. Banerjee, Shishir Dua; 2022; Astoundingly Smart System Furnishing Ranking of Big Data in Search Engines; https://www.researchgate.net/publication/364927715_Astoundingly_Smart_System_Furnishing_Ranking_of_Big_Data_in_Search_Engines
361. *Xiaofang* [2018] - Bu Xiaofang; 08 Feb 2018; Technical data collection method for search result display, and search engine; https://patents.google.com/patent/WO2018023429A1/en
362. *Valentinovich et al.* [2018] - Voronin Aleksej Valentinovich, Tabun Petr Yurevich, Kolomiets Elena Sergeevna; 01 Feb 2018; Method and system for generating search results; https://patents.google.com/patent/RU2643466C1/en
363. *McLellan* [2011] - Mark F. McLellan; 24 May 2011; Active search results page ranking technology; https://patents.google.com/patent/WO2011149934A3/en
364. *Neels et al.* [2013] - Alice Emily Neels, Archana Sulochana Ganapathi, Marc Vincent Robichaud, Stephen Phillip Sorkin; 30 Oct 2013; Generation of a data model for searching machine data; https://patents.google.com/patent/US8983994B2/en
365. *Imarticus* [2018] - 28 Nov 2018; How Data Sciences Principles Play An Important Role In Search Engines; https://imarticus.org/blog/how-data-sciences-principles-play-an-important-role-in-search-engines/
366. *Sharma* [2021] - SaajanSharma; 23 Jun 2021; How to Use Data Science for Search Engine Optimization; https://www.datasciencecentral.com/how-to-use-data-science-for-search-engine-optimization/
367. *mailchimp.com-1* - Page de résultats du moteur de recherche (SERP); https://mailchimp.com/fr/marketing-glossary/serp/
368. *Bhattacharya* [2019] - Joydeep Bhattacharya; 4 Sep 2019; How to improve SEO using data science; https://www.searchenginewatch.com/2019/09/04/improve-seo-using-data-science/
369. *Venugopal & Srikantaiah* [2020] - K. R. Venugopal, K. C. Srikantaiah; 01 Jan 2020; Web Data Extraction and Integration System for Search Engine Results; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-981-15-2513-1_2
370. *Abinaya & Jaishanthi* [2014] - G. Abinaya, B. Jaishanthi; 01 Feb 2014; Equipped search results using machine learning from web databases;

https://www.researchgate.net/publication/283847241_Equipped_Search_Results_Using_Machine_Learning_from_Web_Databases

371. *Azzopardi et al.* [2020] - Leif Azzopardi, Ryen W. White, Paul Thomas, Nick Craswell; 14 Mar 2020; Data-Driven Evaluation Metrics for Heterogeneous Search Engine Result Pages; http://ryenwhite.com/papers/AzzopardiCHIIR2020.pdf
372. *Mudassar et al.* [2015] - Ahmed Mudassar Ali, M. Ramakrishnan; 30 May 2015; Equipped Search Results Using Machine Learning from Web Databases; https://www.researchgate.net/publication/283847241_Equipped_Search_Results_Using_Machine_Learning_from_Web_Databases
373. *Natasha et al.* [2019] - Natasha Noy, D. Brickley, M. Burgess; 2019; Google Dataset Search: Building a search engine for datasets in an open Web ecosystem; https://storage.googleapis.com/pub-tools-public-publication-data/pdf/77547c8d2a7fba472e76c774028cf2b3c0afdb8a.pdf
374. *Bifet et al.* [2005] - A. Bifet, Carlos Castillo, P. Chirita, Ingmar Weber; 2005; An Analysis of Factors Used in Search Engine Ranking; https://airweb.cse.lehigh.edu/2005/bifet.pdf
375. *Zhihan Lv* [2023] - Zhihan Lv; 01 Jun 2023; Generative Artificial Intelligence in the Metaverse Era; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S2667241323000198
376. *Galanin et al.* [2020] - Dmitry N. Galanin, Nail R. Bukharaev, Alexander M. Gusenkov, Alina R. Sittikova; 01 Sep 2020; Using Generative Adversarial Networks for Relevance Evaluation of Search Engine Results; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=9224840
377. *Capra & Arguello* [2023] - Rob Capra, Jaime Arguello; 07 Jul 2023; How does AI chat change search behaviors?; https://arxiv.org/pdf/2307.03826.pdf
378. *Ted Selker* [2023] - Ted Selker; 17 Jul 2023; AI for the Generation and Testing of Ideas Towards an AI Supported Knowledge Development Environment; https://arxiv.org/ftp/arxiv/papers/2307/2307.08876.pdf
379. *Cohen & Queen* [2023] - Steve Cohen, D. Queen; 21 Jun 2023; Generative artificial intelligence community of practice for research; https://www.researchgate.net/publication/371781009_Generative_artificial_intelligence_community_of_practice_for_research
380. *Colton et al.* [2021] - S. Colton, Amy Smith, Sebastian Berns, Ryan Murdock, Michael Cook; 2021; Generative Search Engines: Initial Experiments; https://www.semanticscholar.org/paper/Generative-Search-Engines%3A-Initial-Experiments-Colton-Smith/0644233ae9a61bc20da1335f43f1de59a29a29d2
381. *Zhao et al.* [2005] - Hongkun Zhao, W. Meng, Zonghuan Wu, Vijay V. Raghavan, Clement T. Yu; 2005; Fully automatic wrapper generation for search engines; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/1060745.1060760
382. *Sontag et al.* [2012] - D. Sontag, K. Collins-Thompson, Paul N. Bennett, Ryen W. White, S. Dumais; 2012; Probabilistic models for personalizing web search; https://people.csail.mit.edu/dsontag/papers/sontag_wsdm12.pdf

383. *Eno et al.* [2009] - J. Eno, C. Thompson; 2009; Searching for the metaverse; https://www.academia.edu/57838149/Searching_for_the_metaverse
384. *Cheng et al.* [2022] - Xusen Cheng, Shuang Zhang, Shixuan Fu, Wanxin Liu, Chong Guan, Jian Mou; 2022; Exploring the metaverse in the digital economy: an overview and research framework; https://web.archive.org/web/20221203010642id_/https://www.emerald.com/insight/content/doi/10.1108/JEBDE-09-2022-0036/full/pdf?title=exploring-the-metaverse-in-the-digital-economy-an-overview-and-research-framework
385. *Vijaya & Chander* [2018] - P. Vijaya, Satish Chander; 2018; Metasearch Engine: A Technology for Information Extraction in Knowledge Computing ; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-981-10-8258-0_10
386. *Liu Li* [2003] - Liu Li; 2003; Design and Realization of Intelligent Meta-search Engine; https://www.semanticscholar.org/paper/Design-and-Realization-of-Intelligent-Meta-search-Li/91541b2c8661bebe3641e57e505b9b167eca8d48
387. *thatware.co-1* - METAVERSE MASTERY: NAVIGATING THE NEW FRONTIER OF SEO; https://thatware.co/metaverse-seo/
388. *Warner* [2022] - John Warner; Oct 11, 2022; Search Marketing and the Metaverse; https://www.click.co.uk/insights/search-marketing-and-the-metaverse/
389. *Kieron Allen* [2022] - Kieron Allen; August 7, 2022; Do Search Engines Have a Place in the Metaverse?; https://accelerationeconomy.com/metaverse/do-search-engines-have-a-place-in-the-metaverse/
390. *Lisa Buyer* [2021] - Lisa Buyer; 25 Aug 2021; Marketing in the Metaverse: The Future of Search & Social; https://www.searchenginejournal.com/marketing-in-the-metaverse/416813/
391. *Kaj Leroy* [2023] - Kaj Leroy; 25 Apr 2023; The Metaverse: A New Frontier In Technology And Interconnectivity; https://www.forbes.com/councils/forbesbusinesscouncil/2023/04/25/the-metaverse-a-new-frontier-in-technology-and-interconnectivity/
392. *Wilson* [2023] - Wilson Lau: 9 Jun 2023; The Metaverse: What Is It?; https://www.searchenginejournal.com/metaverse/484803/
393. *Koushik et al.* [2023] - T. P. Koushik, Urwah Kiran, K. Sai Ram Satwik; 30 Jun 2023; Image Based Search Engine Using Deep Learning; https://typeset.io/papers/image-based-search-engine-using-deep-learning-22hs3zuq
394. *Qi et al.* [2019] - Guo Qi, Bo Hu, Wu Xianren, Anish Nair; 02 May 2019; Factored model for search results and communications based on search results; https://typeset.io/papers/factored-model-for-search-results-and-communications-based-4st0ybqpuh
395. *Madden* [2018] - Bartley J. Madden; 30 Apr 2018; Machine-learning based systems and methods for optimizing search engine results; https://typeset.io/papers/machine-learning-based-systems-and-methods-for-optimizing-4srxrhx5ao
396. *Abinaya & Jaishanthi* [2014] - G. Abinaya, B. Jaishanthi; 01 Feb 2014; Equipped search results using machine learning from web databases; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=7033811

397. *Yoganarasimhan* [2020] - Hema Yoganarasimhan; 2020; Search Personalization Using Machine Learning; https://faculty.washington.edu/hemay/search_personalization.pdf
398. *Karwa & Honmane* [2019] - Rushikesh Karwa, Vikas N. Honmane; 2019; Building Search Engine Using Machine Learning Technique; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=9065846
399. *Koragoankar et al.* [2023] - Ranjit Koragoankar, Varad Kulkarni, Deepali Naik; 2023; Search Engine Using NLP Text Processing Techniques to Extract Most Relevant Search Results; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=10307392
400. *Buchholz et al.* [2004] - Martin Buchholz, Dirk Pflüger, Josiah Poon; 2004; Application of Machine Learning Techniques to the Re-ranking of Search Results; https://www.researchgate.net/publication/221562870_Application_of_Machine_Learning_Techniques_to_the_Re-ranking_of_Search_Results
401. *Gungor* [2013] - Tunga Güngör; 2013; A Machine Learning Approach for Displaying Query Results in Search Engines; https://link-springer-com.ressources-electroniques.univ-lille.fr/content/pdf/10.1007/978-3-642-40261-6_21.pdf?pdf=preview
402. *Deshmukh* [2023] - Avinash Deshmukh; 7 Sep 2023; How Generative AI is changing search results?; https://www.linkedin.com/pulse/how-generative-ai-changing-search-results-avinash-deshmukh
403. *Berkel* [2023] - Martha van Berkel; 12 Oct 2023; Is Google's Generative AI Search an Opportunity or Threat for SEOs?; https://www.schemaapp.com/schema-markup/is-google-generative-ai-search-an-opportunity-or-threat-for-seos/
404. *Kevin Watts* [2023] - Kevin Watts; 11 Aug 2023; The Future of Search: Generative AI; https://www.linkedin.com/pulse/future-search-generative-ai-kevin-watts
405. *Li et al.* [2022] - Wei Li, Kexin Xie, Linsey Xiaolin Pang, James Bailey; 17 Oct 2022; Deep Learning for Search and Recommendation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3511808.3557493
406. *Venkata* [2020] - Chandra Sekar Rao Venkata; 30 Apr 2020; Incorporating Data Into Search Engines Using Deep Learning Mechanisms; https://patents.google.com/patent/US20200134038A1/en
407. *Long et al.* [2020] - Bo Long, Jieping Ye, Zang Li, Huiji Gao; 25 Jul 2020; Deep Natural Language Processing for Search and Recommendation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3397271.3401465
408. *Ye et al.* [2015] - Xugang Ye, Jingjing Li, Zijie Qi, Xiaodong He; 2015; Enhancing Retrieval and Ranking Performance for Media Search Engine by Deep Learning; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=7427327
409. *Guo et al.* [2019] - Weiwei Guo, Huiji Gao, Jun Shi, Bo Long; 2019; Deep Natural Language Processing for Search Systems; https://guoweiwei.github.io/files/sigir2019-tutorial.pdf

410. *Zhao et al.* [2018] - Xiangyu Zhao, Long Xia, Jiliang Tang, Dawei Yin; 2018; Deep Reinforcement Learning for Search, Recommendation, and Online Advertising: A Survey; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3320496.3320500
411. *ZhaiCreativoMedia* - May 12, 2021; Deep Learning and Google Search: The Future of SEO; https://www.datasciencecentral.com/deep-learning-and-google-search-the-future-of-seo/
412. *Matt G.* Southern [2023] - Matt G. Southern; 23 May 2023; Bing Integrates With ChatGPT; https://www.searchenginejournal.com/bing-integrates-with-chatgpt/487638/
413. *Sangita-packtpub* - Sangita Mahala; ChatGPT for Search Engines; https://www.packtpub.com/article-hub/chatgpt-for-search-engines
414. *Microsoft-AIVideo* - Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web; https://news.microsoft.com/the-new-Bing/?msockid=1ad76e1cf04364620f277bf7f1f7654d
415. *Sawers* [2023] - Paul Sawers; 18 Oct 2023; ChatGPT officially gets web search as DALL-E 3 integration arrives in beta; https://techcrunch.com/2023/10/18/openai-formally-brings-bing-powered-search-to-chatgpt-as-dall-e-3-arrives-in-beta/
416. *Rayna et al.* [2023] - Ruiyun (Rayna) Xu, Yue (Katherine) Feng, and Hailiang Chen; 03 Jul 2023; ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience; https://typeset.io/papers/chatgpt-vs-google-a-comparative-study-of-search-performance-1vx5khur
417. *Cabra & Arguello* [2023] - Rob Cabra, Jaime Arguello; 07 Jul 2023; How does AI chat change search behaviors?; https://arxiv.org/pdf/2307.03826.pdf
418. *Magruder et al.* [2023] - Matthew L. Magruder, Ronald E. Delanois, James Nace, Michael A. Mont; 01 Jun 2023; ChatGPT and Other Natural Language Processing Artificial Intelligence Models in Adult Reconstruction; https://www.arthroplastyjournal.org/article/S0883-5403(23)00666-6/fulltext
419. *Bhatia* [2023] - Pradeep Bhatia; 01 Jan 2023; ChatGPT for academic writing: A game changer or a disruptive tool?; https://www.researchgate.net/publication/369372503_ChatGPT_for_academic_writing_A_game_changer_or_a_disruptive_tool
420. *Zhao et al.* [2023] - Ruochen Zhao, Xingxuan Li, Yew Ken Chia, Bosheng Ding; 03 Mar 2023; Can ChatGPT-like Generative Models Guarantee Factual Accuracy? On the Mistakes of New Generation Search Engines; https://arxiv.org/pdf/2304.11076.pdf
421. *Sun et al.* [2023] - Weiwei Sun, Lingyong Yan, Xinyu Ma, Pengjie Ren, Dawei Yin, Z. Ren; 2023; Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agent; https://arxiv.org/pdf/2304.09542.pdf
422. *Wojciech Urban* [2023] - Wojciech Urban; 26 May 2023; Will ChatGPT replace the Google Search Engine?; https://www.boldare.com/blog/will-chatgpt-replace-google-search-engine/
423. *Rosenberg* [2023] - Zack Rosenberg; 5 Mar 2023; Will ChatGPT replace search engines?; https://www.linkedin.com/pulse/chatgpt-replace-search-engines-zack-rosenberg
424. *Tristan Greene* [2023] - Tristan Greene ; 27 Jan 2023; ChatGPT will not replace Google Search; https://blogs.lse.ac.uk/impactofsocialsciences/2023/01/27/chatgpt-will-not-replace-google-search/

425. *Salmon* [2023] - Anne Salmon; 01 May 2023; ChatGPT: Disruptive Educational Technology; https://www.researchgate.net/publication/370429350_ChatGPT_Disruptive_Educational_Technology
426. *Thomas Hügle* [2023] - Thomas Hügle; 01 Apr 2023; The wide range of opportunities for large language models such as ChatGPT in rheumatology; https://www.researchgate.net/publication/370383516_The_wide_range_of_opportunities_for_large_language_models_such_as_ChatGPT_in_rheumatology
427. *Asraful Haque* [2023] - Md. Asraful Haque; 29 Mar 2023; A Brief Analysis of "ChatGPT" – A Revolutionary Tool Designed by OpenAI; https://www.researchgate.net/publication/369620426_A_Brief_Analysis_of_ChatGPT_-_A_Revolutionary_Tool_Designed_by_OpenAI
428. *Abinaya & Gowthami* [2023] - M. S. Abinaya, Dasari Gowthami; 11 Apr 2023; ChatGPT: Opportunities, Features and Future Prospects; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=10125747
429. *AlAfnan et al.* [2023] - M. AlAfnan, Samira Dishari, Marina Jovic, Koba Lomidze; 2023; ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses; https://www.researchgate.net/publication/369058778_ChatGPT_as_an_Educational_Tool_Opportunities_Challenges_and_Recommendations_for_Communication_Business_Writing_and_Composition_Courses
430. *Arif et al.* [2023] - T. Arif, Uzair Munaf, Ibtehaj Ul-Haque; 2023; The future of medical education and research: Is ChatGPT a blessing or blight in disguise?; https://www.researchgate.net/publication/368696735_The_future_of_medical_education_and_research_Is_ChatGPT_a_blessing_or_blight_in_disguise
431. *Sorecau & Sorecau* [2023] - M. Sorecau, E. Sorecau; 2023; An Alternative Application to CHATGPT that Uses Reliable Sources to Enhance the Learning Process; https://www.researchgate.net/publication/372517481_An_Alternative_Application_to_CHATGPT_that_Uses_Reliable_Sources_to_Enhance_the_Learning_Process
432. *Sahin et al.* [2023] - Bahadır Şahin, Kader Doğan, Yunus Genç, Tarık Emre Şener, Çağrı Akın Şekerci, Yılören Tanıdır, Selçuk Yücel, Tufan Tarcan; 16 May 2023; Rest in Peace Dr. Search Engines, and Long Live Dr. Chat GPT (Preprint); https://www.scilit.net/publications/e3271e3981822a385127d5aad6013d26
433. *Varn Bristol* [2023] – Varn Bristol; 2 Feb 2023; How ChatGPT will impact Search Marketing; https://bristolcreativeindustries.com/how-chat-gpt-will-impact-search-marketing-varns-view/
434. *Rebecca Sentance* [2022] - Rebecca Sentance; 16 Dec 2022; Does ChatGPT really pose a threat to search engines?; https://econsultancy.com/chatgpt-search-engine-threat/
435. *Dustin Coates* [2023] - Dustin Coates; 31 Jul 2023; Why ChatGPT won't replace search engines any time soon; https://www.algolia.com/blog/ai/why-chatgpt-wont-replace-search-engines-any-time-soon/
436. *Manish Gupta* [2022] - Manish Gupta; 03 Jun 2022; Modelling Using Deep Learning; https://typeset.io/papers/modelling-using-deep-learning-135dbh5k

437. *Zhang & Duan* [2022] - Yu Zhang, Ye Duan; 01 Aug 2022; Deep learning models of cognitive processes constrained by human brain connectomes; https://typeset.io/papers/deep-learning-models-of-cognitive-processes-constrained-by-tcyw2gi7
438. *Arana et al.* [2023] - Sophie Arana, Jacques Pesnot Lerousseau, Peter Hagoort; 16 Jan 2023; Deep Learning Models to Study Sentence Comprehension in the Human Brain; https://repository.ubn.ru.nl/bitstream/handle/2066/292135/292135.pdf?sequence=1
439. *Kamruzzaman & Charles* [2019] - Abu Kamruzzaman, C Charles; 01 Jan 2019; Developing Deep Learning Models to Simulate Human Declarative Episodic Memory Storage; https://typeset.io/papers/developing-deep-learning-models-to-simulate-human-4ob9edmk5l
440. *Guerguiev et al.* [2017] - Jordan Guerguiev, Timothy P. Lillicrap, Blake A. Richards, Blake A. Richards; 05 Dec 2017; Towards deep learning with segregated dendrites; https://typeset.io/papers/towards-deep-learning-with-segregated-dendrites-3mm2h3ogtd
441. *Dandan Mo* [2012] - Dandan Mo; 2012; A survey on deep learning: one small step toward AI; https://www.semanticscholar.org/paper/A-survey-on-deep-learning%3A-one-small-step-toward-AI-Mo/db26266f9bd3337b359758cb07dec36a061c5742
442. *Kumar et al.* [2018] - P. S. J. Kumar, Yanmin Yuan, Yang Yung, Mingmin Pan, Wenli Hu; 2018; Robotic simulation of human brain using convolutional deep belief networks; https://www.inderscienceonline.com/doi/abs/10.1504/IJIMR.2018.094922?journalCode=ijimr
443. *Le et al.* [2020] - Huy Viet Le, Sven Mayer, N. Henze; 2020; Deep learning for human-computer interaction; https://www.researchgate.net/publication/347922286_Deep_learning_for_human-computer_interaction
444. *Liu & Chen* [2019] - Yanmei Liu, Yuda Chen; 2019; Overview of Deep Learning Research; Proceedings of the 5th Annual International Conference on Social Science and Contemporary Humanity Development (SSCHD 2019); https://www.atlantis-press.com/proceedings/sschd-19/125925738
445. *dl-connectedworld.com-1* - Deep Learning Vs. The Human Brain; https://connectedworld.com/deep-learning-vs-the-human-brain/
446. *dl-ibm.com-3* - What is deep learning?; https://www.ibm.com/topics/deep-learning
447. *Sohan Kanse* [2023] - Sohan Kanse; 5 Jun 2023; Deep Learning vs Human Brain: Classification and Complexity; https://www.linkedin.com/pulse/deep-learning-vs-human-brain-classification-complexity-sohan-kanse
448. *McClelland & Botvinick* [2020] - James L. McClelland and Matthew M. Botvinick; 2020; Deep Learning: Implications for Human Learning and Memory; https://stanford.edu/~jlmcc/papers/McCBotvinick20DeepLearningOhHM.pdf
449. *Polson et Soklov* [2023] - Nicholas G. Polson, Vadim Soklov; 24 May 2023; Generative AI for Bayesian Computation; https://arxiv.org/pdf/2305.14972v1.pdf
450. *Xia Dong* [2023] - Xia Dong; 08 Apr 2023; Deep Generative Modeling with Backward Stochastic Differential Equations; https://typeset.io/papers/deep-generative-modeling-with-backward-stochastic-19pzi0a4

451. *Xingcheng Xu* [2023] - Xingcheng Xu; 08 Apr 2023; Deep Generative Modeling with Backward Stochastic Differential Equations; https://arxiv.org/pdf/2304.04049.pdf
452. *World Bank* [2023] - World Bank; 07 Jul 2023; Publication: Generative Artificial Intelligence; Book; https://openknowledge.worldbank.org/entities/publication/4f623641-ba34-4f0d-9a7d-105f02a5ee00
453. *Salakhutdinov* [2009] - R. Salakhutdinov; 2009; Learning Deep Generative Models; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/book/10.5555/1925602
454. *Sohn et al.* [2015] - Kihyuk Sohn, Honglak Lee, Xinchen Yan; 2015; Learning Structured Output Representation using Deep Conditional Generative Models; https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf
455. *Lee et al.* [2018] - K. Lee, Kyoung-Soon Hwang, K. Kim, Sang Hyun Lee, Ki Sun Park; 2018; A deep learning model generation method for code reuse and automatic machine learning; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3264746.3264787
456. *Xu et al.* [2015] - Jungang Xu, Hui Li, Shilong Zhou; 2015; An Overview of Deep Generative Models; https://www.tandfonline.com/doi/full/10.1080/02564602.2014.987328
457. *Rebecca* [2025] - Complete Guide to Five Generative AI Models; https://www.coveo.com/blog/generative-models/
458. *Damian Brady* [2023] - Damian Brady; 7 Apr 2023; What developers need to know about generative AI; https://github.blog/2023-04-07-what-developers-need-to-know-about-generative-ai/
459. *altexsoft.com* [2022] - 13 Oct 2022; Generative AI Models Explained; https://www.altexsoft.com/blog/generative-ai/
460. *Alaskar & Saba* [2021] - Hind Alaskar, TanzilaSaba Saba; 2021; Machine Learning and Deep Learning: A Comparative Review; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-981-33-6307-6_15
461. *Bhattacharjee* [2019] - Abhijit Bhattacharjee; 2019; Machine Learning and Deep Learning; https://www.semanticscholar.org/paper/Machine-Learning-and-Deep-Learning-Bhattacharjee/7d291d5fca0e9cd9e0ed72fb6f82289a197f7f02
462. *John Cherrie* [2016] - John Cherrie; 2016: Machine Learning and Deep Learning; https://pdfs.semanticscholar.org/7cbc/1e4c784435931064c22c07a5ca2af25e07f4.pdf
463. *Shinde & Shah* [2018] - Pramila Shinde, Seema Shah; 2018; A Review of Machine Learning and Deep Learning Applications; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=8697857
464. *Shanthamallu & Spanias* [2023] - Uday Shankar Shanthamallu , Andreas Spanias; 15 Jun 2023; Machine Learning and Deep Learning: Algorithms and Applications; https://typeset.io/papers/machine-learning-and-deep-learning-algorithms-and-3o2ly67t
465. *Frederick Douglass* [2023] - Frederick Douglass; 15 Jun 2023; Deep Learning and Neural Networks: Methods and Applications; https://typeset.io/papers/deep-learning-and-neural-networks-methods-and-applications-2h3c558v

466. *Jin Hu* [2023] - Jin Hu; 01 Jan 2023; Machine Learning and Deep Learning; https://typeset.io/papers/machine-learning-and-deep-learning-15drp8ag
467. *Fernández & Peters* [2023] - Iván Sánchez Fernández, Jurriaan M. Peters; 07 Feb 2023; Machine learning and deep learning in medicine and neuroimaging; https://onlinelibrary.wiley.com/doi/full/10.1002/cns3.5
468. *Y Yamana* [2023] - Y Yamana; 15 Jun 2023; Deep Learning and Neural Networks: Methods; https://typeset.io/papers/deep-learning-and-neural-networks-methods-2lb589ga
469. *Arne Wolfewicz* [2020] - Arne Wolfewicz; 14 Jul 2020; Deep Learning vs. Machine Learning – What's The Difference? ; https://medium.com/levity/deep-learning-vs-machine-learning-whats-the-difference-e367803bb96d
470. *Coursera Staff* [2023] - Coursera Staff; 29 Nov 2023; Deep Learning vs. Machine Learning: A Beginner's Guide; https://www.coursera.org/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide
471. *Patrick Grieve* [2023] - Patrick Grieve; 22 Sep 2023; Deep learning vs. machine learning: what's the difference?; https://www.zendesk.co.uk/blog/machine-learning-and-deep-learning/
472. *Schultheib & Lewandowski* [2021]-1 - Sebastian Schultheib, Dirk Lewandowski; 14 May 2021; Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference; https://journals.sagepub.com/doi/10.1177/01655515211014157#bibr1-01655515211014157
473. *Varma & Jain* [2010] - Manik Varma, Vidit Jain; 29 Apr 2010; Providing search results in response to a search query; https://patents.google.com/patent/US20110270828
474. *Zhang Jiayin* [2013] - Zhang Jiayin; 17 Jul 2013; Method for delivering search results and search engine; https://typeset.io/papers/method-for-delivering-search-results-and-search-engine-4or1lt92uk
475. *Seshadri et al.* [2006] - Seshadri Srinivasan, Wang Lam, Gaurav Bhalotia, Venky Harinarayan; 13 Oct 2006; Delivering search results; https://typeset.io/papers/delivering-search-results-c1e06lqow2
476. *El-Yaniv & Friedman* [2008] - Ran El-Yaniv, Roy Friedman; 06 Mar 2008; Method for delivering query responses; https://typeset.io/papers/method-for-delivering-query-responses-2wt3o7w0my
477. *Nadav Gur* [2012] - Nadav Gur; 10 Feb 2012; System and method for improved search experience through implicit user interaction; https://patents.google.com/patent/US9449093B2/en
478. *saturncloud.io-1* - NLP Transformers Beyond BERT: RoBERTa, XLNet; https://saturncloud.io/glossary/nlp-transformers-beyond-bert-roberta-xlnet/
479. *Lewandowski* [2008] - D. Lewandowski; 2008; Search engine user behaviour: How can users be guided to quality content?; https://www.researchgate.net/publication/228663003_Search_engine_user_behaviour_How_can_users_be_guided_to_quality_content
480. *Griffiths et al.* [2007] - T. Griffiths, M. Steyvers, Alana J Firl; 2007; Google and the Mind; https://cocosci.princeton.edu/tom/papers/google.pdf
481. *Dickerhoof et al.* [2014] - Alison Dickerhoof, Catherine L. Smith; 2014; Looking for query terms on search engine results pages; https://asistdl-onlinelibrary-wiley-com.ressources-electroniques.univ-lille.fr/doi/pdf/10.1002/meet.2014.14505101156

482. *Inthiran et al.* [2010] - A. Inthiran, S. Alhashmi, P. Ahmed; 2010; A Reflection of Search Engine Strategies; https://www.researchgate.net/publication/50406935_A_Reflection_of_Search_Engine_Strategies
483. *Axios-Steemit* [2018] - How To Use the Google Search Engine Of Your Mind To Find What You Want Most In Life; https://steemit.com/psychology/@axios/how-to-use-the-google-search-engine-of-your-mind-to-find-what-you-want-most-in-life
484. *Romero et al.* [2019] - Romero Monica, Maria Lozano, Rizzo Luca, Arlaud Christelle, Gueguen Yoann; 25 Jun 2019; Search query processing; https://patents.google.com/patent/US20180096027A1/en
485. *Landon Prieur* [2018] - Landon Prieur; 03 May 2018; User Search Query Processing; https://patents.google.com/patent/US20180121502A1/en
486. *Lena Wiese* [2009] - Lena Wiese; 31 Aug 2009; Preprocessing for controlled query evaluation in complete first-order databases; https://typeset.io/papers/preprocessing-for-controlled-query-evaluation-in-complete-4jl38pfu8c
487. *Hood et al.* [2005] - Stephen Hood, Ralph Rabbat, Mihir Shah, Adam Durfee; 24 Aug 2005; Speculative search result for a search query; https://patents.google.com/patent/US20070055652A1/en
488. *Krishna Bharat* [2000] - Krishna Bharat; 2000; SearchPad: explicit capture of search context to support Web search; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S1389128600000475
489. *Cambazoglu et al.* [2012] - B. B. Cambazoglu, I. S. Altingövde, R. Ozcan, Ö. Ulusoy; 2012; Cache-Based Query Processing for Search Engines; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2382616.2382617
490. *Moore et al.* [2011] - Robert J. Moore, E. Churchill, Raj Gopal Prasad Kantamneni; 2011; Three sequential positions of query repair in interactions with internet search engines; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1958824.1958889
491. *Jain & Mishne* [2010] - Alpa Jain, G. Mishne; 2010; Organizing query completions for web search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1871437.1871585
492. *Britney Muller Moz.com* - HOW SEARCH ENGINES WORK: CRAWLING, INDEXING, AND RANKING; https://moz.com/beginners-guide-to-seo/how-search-engines-operate
493. *Catherine Dee* [2022] – Catherine Dee; 15 Aug 2022; What is a search query and how is it processed by a search engine?; https://www.algolia.com/blog/product/what-is-a-search-query-and-how-is-it-processed-by-a-search-engine/
494. *Anand Srinivasan* [2022] – Anand Srinivasan; 8 Aug 2022; Google's query processing algorithms and what they mean for your SEO strategy; https://www.minttwist.com/blog/googles-query-processing-algorithms-and-what-they-mean-for-your-seo-strategy/
495. *Du Vall & Preston* [2023] - Eddie Du Vall, Shaune Preston; 24 Feb 2023; Implementing Search: An Overview to Building a Search Engine; https://www.wwt.com/article/implementing-search-an-overview-to-building-a-search-engine

496. *Batchakui et al.* [2022] - Bernabe Batchakui, Roger Nkambou, Jaurès Styve Kameni Homte; 31 Jan 2022; Search Engines in Learning Contexts: A Literature Review; https://typeset.io/papers/search-engines-in-learning-contexts-a-literature-review-1fbk1pco
497. *Hoyer et al.* [2022] - Johannes von Hoyer, Anett Hoppe, Yvonne Kammerer, Christian; 15 Mar 2022; The Search as Learning Spaceship: Toward a Comprehensive Model of Psychological and Technological Facets of Search as Learning; https://typeset.io/papers/the-search-as-learning-spaceship-toward-a-comprehensive-2d57qbmx
498. *Pertti Vakkari* [2016] - Pertti Vakkari; 01 Feb 2016; Searching as learning; https://dl.acm.org/doi/10.1177/0165551515615833
499. *Camara et al.* [2021] - Arthur Câmara, Nirmal Roy, David Maxwell, Claudia Hauff; 14 Mar 2021; Searching to Learn with Instructional Scaffolding; https://www.researchgate.net/publication/350054201_Searching_to_Learn_with_Instructional_Scaffolding
500. *Ghosh et al.* [2018] - Souvick Ghosh, Manasa Rath, Chirag Shah; 01 Mar 2018; Searching as Learning: Exploring Search Behavior and Learning Outcomes in Learning-related Tasks; https://dl.acm.org/doi/10.1145/3176349.3176386
501. *Katja Hofmann* [2011] - Katja Hofmann; 2011; Search engines that learn online; https://dl.acm.org/doi/abs/10.1145/2009916.2010176
502. *Bird & Curran* [2006] - Steven Bird, J. Curran; 2006; Building a search engine to drive problem-based learning; https://dl.acm.org/doi/10.1145/1140123.1140166
503. *Taibi & Gentile* [2005] - D. Taibi, M. Gentile; 2005; A Semantic Search Engine for Learning Resources; https://www.academia.edu/26890871/A_semantic_search_engine_for_learning_resources
504. *S. Liaw* [2004] - S. Liaw; 2004; The theory of planned behaviour applied to search engines as a learning tool; https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2729.2004.00087.x
505. *Sandra Niemeyer* [2021] - Sandra Niemeyer; 21 Jul 2021; How do we use search engines for learning?; https://blogs.tib.eu/wp/tib/2021/07/21/how-do-we-use-search-engines-for-learning/
506. *Collins-Thompson et al.* [2017] - Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff; 2017; Search as Learning; https://drops.dagstuhl.de/storage/04dagstuhl-reports/volume07/issue02/17092/DagRep.7.2.135/DagRep.7.2.135.pdf
507. *Chen et al.* [2023] - Enhong Chen, Yechang Huang, Shuanghong Shen, Zhenya Huang; 19 Jul 2023; Learning from Crowds with Annotation Reliability; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3539618.3592007
508. *Li et al.* [2023] - Shikun Li, Xiaobo Xia, Jiankang Deng, Shiming Ge; 05 Jun 2023; Transferring Annotator- and Instance-dependent Transition Matrix for Learning from Crowds; https://arxiv.org/pdf/2306.03116.pdf
509. *He et al.* [2023] - Xingwei He, Zheng Wen-Lin, Yeyun Gong, Alex Jin, Hang Zhang, Chen Lin, Jian Jiao, Siu-Ming Yiu, Nan Duan, Weizhu Chen; 30 Mar 2023; AnnoLLM: Making Large Language Models to Be Better Crowdsourced Annotators; https://arxiv.org/pdf/2303.16854.pdf

510. *Chen et al.* [2013] - Xi Chen, Paul N. Bennett, K. Collins-Thompson, E. Horvitz; 2013; Pairwise ranking aggregation in a crowdsourced setting; https://public.websites.umich.edu/~kevynct/pubs/wsdm91-bennettPS.pdf
511. *Mendez et al.* [2019] - Ana Elisa Méndez, M. Cartwright, J. Bello; 2019; Machine-Crowd-Expert Model for Increasing User Engagement and Annotation Quality; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/fullHtml/10.1145/3290607.3313054
512. *Fang et al.* [2014] - Meng Fang, Jie Yin, D. Tao; 2014; Active Learning for Crowdsourcing Using Knowledge Transfer; https://ojs.aaai.org/index.php/AAAI/article/view/8993
513. *Zhang et al.* [2021] - Xin Zhang, Guangwei Xu, Yueheng Sun, Meishan Zhang, Pengjun Xie; 2021; Crowdsourcing Learning as Domain Adaptation: A Case Study on Named Entity Recognition; https://aclanthology.org/2021.acl-long.432.pdf
514. *Yang et al.* [2018] - Jie Yang, Thomas Drake, Andreas C. Damianou, Y. Maarek; 2018; Leveraging Crowdsourcing Data for Deep Active Learning An Application: Learning Intents in Alexa; https://arxiv.org/pdf/1803.04223.pdf
515. *Han et al.* [2016] - Shuguang Han, P. Dai, Praveen K. Paritosh, David Huynh; 2016; Crowdsourcing Human Annotation on Web Page Structure: Infrastructure Design and Behavior-Based Quality Control; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2870649
516. *University of Texas* [2017] - 3 Aug 2017; The future of search engines: Researchers combine artificial intelligence, crowdsourcing and supercomputers; https://phys.org/news/2017-08-future-combine-artificial-intelligence-crowdsourcing.html
517. *Kait Spong* [2023] - Kait Spong; 27 Apr 2023; Why Self-Learning AI Is a Game Changer in E-Commerce Site Search; https://www.bloomreach.com/en/blog/2022/why-self-learning-ai-is-a-game-changer-in-e-commerce
518. *Hurwitz et al.* [2015] - Cherie Hurwitz, Benjamin Collier, Ryan Lissack, Kristopher Johns, Kyle Medina, Chad Wyszynski; 28 Dec 2015; System and method for generating online content creator profiles and providing a searchable platform for the online content creator profiles; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s12599-023-00808-9
519. *Kyu & Jee* [2017] - Shin Min Kyu, Chun Sol Jee; 22 Nov 2017; Method and system for searching content creators; https://typeset.io/papers/method-and-system-for-searching-content-creators-4kpn8ycfzg
520. *Ali & LaPaugh* [2013] - Muneeb Ali, Andrea S. LaPaugh; 01 Jan 2013; Enabling Author-Centric Ranking of Web Content; https://typeset.io/papers/enabling-author-centric-ranking-of-web-content-4t4anvt14l
521. *Robinson & Myricks* [2011] - Leroy Robinson, Cynthia Myricks; 27 Jan 2011; Method and system for searching for, and monitoring assessment of, original content creators and the original content thereof; https://patents.google.com/patent/US8510154B2/en
522. *AIContentfy team* [2024] - The importance of search engines in content distribution; https://aicontentfy.com/en/blog/importance-of-search-engines-in-content-distribution#
523. *Lakshminarayana* [2009] - S. Lakshminarayana; 2009; Categorization of web pages - Performance enhancement to search engine; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1016/j.knosys.2008.07.006

524. *Patel & Pandya* [2017] - Ankit Dilip Patel, Vimal N. Pandya; 2017; Web page classification based on context to the content extraction of articles; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=8226188
525. *Pu et al.* [2002] - H. Pu, Shui-Lung Chuang, Chyan Yang; 2002; Subject categorization of query terms for exploring Web users' search interests; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1002/asi.10071
526. *Lindemann & Littig* [2011] - Christoph Lindemann, Lars Littig; 2011; Classification of Web Sites at Super-genre Level; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-90-481-9178-9_10
527. *Posner & Petersen* [1990] - Michael I. Posner, Steven E. Petersen; 1990; THE ATTENTION SYSTEM OF THE HUMAN BRAIN; http://wexler.free.fr/library/files/posner%20(1990)%20the%20attentional%20system%20of%20the%20human%20brain.pdf
528. *scriptol.com* [2010] - Anatomy of a search engine: Infrastructure of Google; https://www.scriptol.com/web/google-anatomy.php
529. *Nakamuraa et al.* [2018] - Thiago Akio Nakamuraa, Pedro H. Calais, Davi de Castro Reis, André Paim Lemos; 22 Dec 2018; An anatomy for neural search engines; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0020025518309952
530. *Gupta & Sharma* [2014] - Trilok Gupta, Archana Sharma; Apr 2014; Anatomy of Web Search Engines; https://ijarcce.com/wp-content/uploads/2012/03/IJARCCE3A-s-trilok-Anatomy-of-Web-Search.pdf
531. *Olaf Kopp* [2022]-2 - Olaf Kopp, 29 Jun 2022; How does Google understands search terms by search query processing?; https://www.kopp-online-marketing.com/search-query-processing
532. *Kruse et al.* [2005] - Peter M. Kruse, Andre Naujoks, D. Rösner, M. Kunze; 2005; Clever Search: A WordNet Based Wrapper for Internet Search Engines; https://arxiv.org/ftp/cs/papers/0501/0501086.pdf
533. *Osinski & Weiss* [2005] - S. Osinski, Dawid Weiss; 2005; A concept-driven algorithm for clustering search results; https://www.biz.uiowa.edu/faculty/nstreet/01439479.pdf
534. *Gaurkhede & Dhotre* [2014] - Ashwini R. Gaurkhede, D. Dhotre; 2014; A survey on user specific search using organized search history; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?arnumber=6981154
535. *Pradhan et al.* [2014] - Narendra Pradhan, K. Pandey, R. Sahu; 2014; A Web Search Engine-Based Approach to Measure Semantic Similarity between Words; https://www.researchgate.net/publication/224176330_A_Web_Search_Engine-Based_Approach_to_Measure_Semantic_Similarity_between_Words
536. *Gerasimov et al.* [2014] - Nikita Gerasimov, Maxim Mozgovoy, Alexey Lagunov; 23 Oct 2014; Semantic sentence structure search engine; https://www.researchgate.net/publication/297601983_Semantic_sentence_structure_search_engine

537. *Sokolan et al.* [2010] - Patrick Sokolan, Dennis Doherty, Claude Duguay, William Radcliffe, Virgil Bourassa, Tammara King, John Sheppard; 26 Apr 2010; Search engine data structure; https://patents.google.com/patent/US8375021B2/en
538. *Woo & Gu* [2007] - Choi Soon Woo, Lee Byung Gu; 02 Oct 2007; Techniques on storage structure and storing method for indexed keyword information using database from search engine with maintaining its search speed; https://typeset.io/papers/techniques-on-storage-structure-and-storing-method-for-3o1nhiulol
539. *Kong et al.* [2016] - Jing Kong, Alex Scott, and Georg M. Goerg; 2016; Improving semantic topic clustering for search queries with word co-occurrence and bigraph co-clustering; https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45569.pdf
540. *Rachel Handley* [2023] - Rachel Handley; 17 Jul 2023; Topic Clusters for SEO: What They Are & How to Create Them; https://www.semrush.com/blog/topic-clusters/
541. *Colton Miller* [quattr.com] - Keyword Clustering Explained; https://www.quattr.com/optimize-content/why-focus-on-keyword-clusters
542. *Ustinovskiy et al.* [2013] - Yury Ustinovskiy, Pavel Serdyukov; 27 Oct 2013; Personalization of web-search using short-term browsing context; https://www.researchgate.net/publication/266654563_Personalization_of_web-search_using_short-term_browsing_context
543. *Ronny Lempel* [2017] - Ronny Lempel; 27 Aug 2017; Personalization is a Two-Way Street; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3109859.3109869
544. *Kaingade et al.* [2015] - Rasika M. Kaingade, Hemant A. Tirmare; 28 Sep 2015; Personalization of Web Search based on privacy protected and auto-constructed user profile; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=7275711
545. *Timos Sellis* [2011] - Timos Sellis; 28 Sep 2011; Personalization in web search and data management; https://typeset.io/papers/personalization-in-web-search-and-data-management-40i05zs33c
546. *Masha Maksimava* [2017] - 22 Aug 2017; Google's Personalized Search Explained; https://www.link-assistant.com/news/personalized-search.html
547. *Nick Barney* [2022] - Nov 2022; search engine results page (SERP); https://www.techtarget.com/whatis/definition/search-engine-results-page-SERP
548. *Thais Montjean* [2025] – 7 Jan 2025; Personalization in 2025: Redefining customer-centric digital marketing; https://dotdigital.com/blog/personalization-in-digital-marketing/
549. *Patel et al.* [2015] - Patel Jay, Pinal Shah, Kamlesh Makvana, Parth Shah; 2015; Review on web search personalization through semantic data; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=7226083
550. *Stamou et al.* [2008] - Sofia Stamou, Alexandros Ntoulas; 2008; Search personalization through query and page topical analysis; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s11257-008-9056-y
551. *Arya et al.* [2014] - Himani Arya, Jaytrilok Choudhary, D. Tomar; 2014; A Survey on Techniques for Personalization of Web Search;

https://www.academia.edu/58592220/A_Survey_on_Techniques_for_Personalization_of_Web_Search?uc-sb-sw=16163231

552. *Shafiq et al.* [2015] - M. Omair Shafiq, Reda Alhajj, Jon G. Rokne; 2015; On personalizing Web search using social network analysis; https://www.researchgate.net/publication/275219347_On_personalizing_Web_search_using_social_network_analysis
553. *Osann-1* [2018] - Jr. Robert Osann; 07 Nov 2018; Internet search results annotation for missing search terms; https://patents.google.com/patent/US10324995B2/en
554. *Osann-3* [2020] - Jr. Robert Osann; 13 Feb 2020; Internet Search Results Annotation for Search Term Elements Present or Absent in Referenced Webpages and Descendant Webpages; https://typeset.io/papers/internet-search-results-annotation-for-search-term-elements-1zkzbwb2fa
555. *Liu et al.* [2013] - Sean Yaoxing Liu, Sagar Kamdar, Rishi Khaitan, Frances B. Haugen; 19 Feb 2013; Processing annotations for a search result; https://typeset.io/papers/processing-annotations-for-a-search-result-4rrkme05mg
556. *Osann-2* [2018] - Jr. Robert Osann; 25 Nov 2018; Internet search results annotation for missing search terms in referenced webpages and descendant webpages; https://typeset.io/papers/internet-search-results-annotation-for-missing-search-terms-3e6y0d7cg6
557. *Kim et al.* [2011] - Lawrence H. Kim, Seungphil Oh, Satish Kumar Musukula; 30 Dec 2011; Presenting search results annotated with domain endorsements; https://patents.google.com/patent/US9009129B2/en
558. *Liu et al.* [2020] - Wei Liu, Xiaofeng Meng, W. Meng, D. Raghu, V. Reddy, C. Jacob, A. Arasu; 2020; An Automatic Annotation Technique for Web Search Results; https://www.academia.edu/71045288/An_Automatic_Annotation_Technique_for_Web_Search_Results?uc-sb-sw=60452772
559. *Sriramoju* [2014] - Shoban Babu Sriramoju; 2014; An Application for Annotating Web SearchResults; https://www.researchgate.net/publication/318489670_An_Application_for_Annotating_Web_Search_Results
560. *Y. P. Jeba* [2013] - Y. P. Jeba; 2013; A SURVEY ON ANNOTATING SEARCH RESULTS FROM WEB DATABASES; https://www.semanticscholar.org/paper/A-SURVEY-ON-ANNOTATING-SEARCH-RESULTS-FROM-WEB-Jeba/fa4ed66c0b897201d2d2c3fdc18ace09671d3bed
561. *Lu et al.* [2013] - Yiyao Lu, Hai He, Hongkun Zhao, W. Meng, Clement T. Yu; 2013; Annotating Search Results from Web Databases; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=5989804
562. *Shao et al.* [2019] - Yunqiu Shao, Yiqun Liu, Fan Zhang, Min Zhang, Shaoping Ma; 2019; On Annotation Methodologies for Image Search Evaluation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3309994
563. *Rui et al.* [2007] - Xiaoguang Rui, Nenghai Yu, Taifeng Wang, Mingjing Li; 2007; A Search-Based Web Image Annotation Method; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=4284735

564. *Biradar & Kumar* [2006] - B. Biradar, B. Kumar; 2006; Internet Search Engines: A Comparative Study and Evaluation Methodology; https://www.semanticscholar.org/paper/Internet-Search-Engines%3A-A-Comparative-Study-and-Biradar-Kumar/4eb3e7f6bc3418030ea5337e1e23b7774214e399
565. *Bar-Ilan et al.* [2006] - Judit Bar-Ilan, Mazlita Mat-Hassan, Mark Levene; 2006; Methods for comparing rankings of search engine results; https://arxiv.org/pdf/cs/0505039.pdf
566. *Bar-Ilan* [2005] - Judit Bar-Ilan; 2005; Comparing rankings of search results on the Web; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0306457305000312
567. *Kumar & Pavithra* [2010] - B. Kumar, S. Pavithra; 2010; Evaluating the searching capabilities of search engines and metasearch engines: a comparative study; https://nopr.niscpr.res.in/bitstream/123456789/9745/4/ALIS%2057%282%29%2087-97.pdf
568. *Aryshandy et al.* [2021] - Gita Aryshandy, Yudi Fernando, Rayina Triningsih Dharmaputra, R. Ikhsan, Ika Sari; 2021; How Does Search Engine Optimization Affect Outcomes of Electronic Marketing Strategy?; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=9649450
569. *Egea & Menendez* [2006] - J. M. O. Egea, M. Menéndez; 2006; Global Marketing on the Internet; https://www.semanticscholar.org/paper/Global-Marketing-on-the-Internet-Egea-Men%C3%A9ndez/a846eac255900d293476c29479d403df91d5fb59
570. *Nathan* [2009] - C. S. Nathan; 2009; E-Marketing Methods and Application For Successful Marketing in Global Marketing Scenario; https://indianjournalofcapitalmarkets.com/index.php/ijom/article/view/36898
571. *V. Raja* [2012] - The Study of E-Commerce Service Systems in Global Viral Marketing Strategy; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2190787
572. *Bhandari & Bansal* [2018] - R. Bhandari, A. Bansal; 2018; Impact of Search Engine Optimization as a Marketing Tool; https://journals.sagepub.com/doi/abs/10.1177/2278682117754016?journalCode=brja
573. *Berman & Katona* [2012] - Ron Berman, Z. Katona; 2012; The Role of Search Engine Optimization in Search Marketing; https://www-jstor-org.ressources-electroniques.univ-lille.fr/stable/24545014?
574. *Bevendorff et al.* [2022] - Janek Bevendorff, Matti Wiegmann; 2022; THE IMPACT OF ONLINE AFFILIATE MARKETING ON WEB SEARCH; https://downloads.webis.de/publications/papers/bevendorff_2022c.pdf
575. *Terrance et al.* [2018] - Arokia Ramya Terrance, Shruti Shrivastava, Asmita Kumari; 2018; Importance of Search Engine Marketing in the Digital World; https://annals-csis.org/Volume_14/drp/pdf/24.pdf
576. *John Simon* [2001] - Steven John Simon; 01 Jan 2001; The Impact of Global Electronic Commerce on the General Agreement on Trade in Services (GATS); https://www.tandfonline.com/doi/abs/10.1080/1097198X.2001.10856292
577. *Kaur & Pathak* [2015] - Parminder Kaur, Ashutosh Pathak, Karamjeet Kaur; 01 Feb 2015; E-Marketing- A Global Perspective; https://www.ijera.com/papers/Vol5_issue2/Part%20-%205/L50205116124.pdf
578. *Rajamanickam* [2020] - Dr.M. Rajamanickam; 31 Mar 2020; Digital Marketing - A Global Perspective; https://www.semanticscholar.org/paper/Digital-Marketing-A-Global-Perspective-Rajamanickam/32cc66b43b105c2c3e5704a968470f85f8419f89

579. *Ferrera & Kessedjian* [2019] - Cécile Ferrera, Eowyne Kessedjian; 13 Mar 2019; Evolution of E-commerce and Global Marketing; https://www.econstor.eu/bitstream/10419/194869/1/19113-78049-1-10-20190316.pdf
580. *Hoque & Munir* [2025] - 100+ Search Engine Marketing (SEM) Statistics: Trends and Insights for 2025; https://www.agencyhandy.com/search-engine-marketing-statistics/
581. *Dave Chaffey* [2024] - Dave Chaffey; 01 Feb 2024; Search engine marketing statistics 2024; https://www.smartinsights.com/search-engine-marketing/search-engine-statistics/
582. *Semrush Team* [2023] - Aug 10, 2023; Top 106 SEO Statistics; https://www.semrush.com/blog/seo-statistics/
583. *Strahl et al.* [2021] - Jonathan Strahl, Jaakko Peltonen, Patrik Floréen; 26 Aug 2021; Directing and Combining Multiple Queries for Exploratory Search by Visual Interactive Intent Modeling; page 535-548; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-030-85613-7_34
584. *Raff et al.* [2014] - Paul Raff, Xiao Huang, Yichao Cai, Rui Ma, C. Aidan Crook, An Yan; 26 Mar 2014; Client intent in integrated search environment; https://patents.google.com/patent/US20170039269A1/fr
585. *Ruotsalo et al.* [2018] - Tuukka Ruotsalo, Jaakko Peltonen, Manuel J. A. Eugster, Dorota Glowacka, Patrik Floréen, Petri Myllymäki, Giulio Jacucci, Samuel Kaski; 03 Oct 2018; Interactive Intent Modeling for Exploratory Search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3231593
586. *Taraknath et al.* [2017] - Woddi Taraknath, Edwards Jonathan, Woundy Stephen, Ramanathan Venkatesan; 09 Feb 2017; Ranking of Search Results Based on Customer Intent; https://patents.google.com/patent/US20170039578A1/en
587. *Xiao et al.* [2011] - Dongqing Xiao, Haoliang Qi, Jingbin Gao, Zhongyuan Han, Muyun Yang, Sheng Li; 01 Jan 2011; HIT 2 Joint NLP Lab at the NTCIR-9 Intent Task; https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings9/NTCIR/16-NTCIR9-INTENT-XiaoD.pdf
588. *Xia et al.* [2013] - Yunqing Xia, Xiaoshi Zhong, Guoyu Tang, Junjun Wang, Qiang Zhou; 2013; Ranking Search Intents Underlying a Query; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-642-38824-8_23
589. *Mitsui et al.* [2017] - M. Mitsui, Jiqun Liu, N. Belkin, C. Shah; 2017; Predicting Information Seeking Intentions from Search Behaviors; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3077136.3080737
590. *Wu et al.* [2010] - Dayong Wu, Yu Zhang, Shiqi Zhao, Ting Liu; 2010; Identification of Web Query Intent Based on Query Text and Web Knowledge; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=5635695
591. *Maji et al.* [2020] - Subhadeep Maji, Priyank Patel, Bharat Thakarar, Mohit Kumar, Krishna Azad Tripathi; 14 Apr 2020; A Regularised Intent Model for Discovering Multiple Intents in E-Commerce Tail Queries; Page 665; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-030-45439-5_43
592. *Gupta et al.* [2023] - Neha Gupta, Shanmukha C. Guttula, Hima Patel; 08 Jul 2023; Multi-Intent Detection in User Provided Annotations for Programming by Examples Systems; https://arxiv.org/pdf/2307.03966.pdf
593. *Burchell et al.* [2020] - Laurie Burchell, Jie Chi, Tom Hosking1, Nina Markl, Bonnie Webber; 01 Dec 2020; Querent Intent in Multi-Sentence Questions; https://aclanthology.org/2020.law-1.13.pdf

594. *Hashemi & Zamani* [2022] - Helia Hashemi, Hamed Zamani; 17 Oct 2022; Stochastic Optimization of Text Set Generation for Learning Multiple Query Intent Representations; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3511808.3557666
595. *Mohammadi et al.* [2020] - Samin Mohammadi, Mathieu Chapon, Arthur Fremond; 2020; Query Intent Detection from the SEO Perspective; https://arxiv.org/pdf/2006.09119.pdf
596. *Mendoza & Zamora* [2009] - Marcelo Mendoza, Juan Zamora; 2009; Identifying the Intent of a User Query Using Support Vector Machines; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-642-03784-9_13
597. *Jouis et al.* [2012] - C. Jouis, Ismaïl Biskri, J. Ganascia, M. Roux; 2012; Next Generation Search Engines: Advanced Models for Information Retrieval; https://www.researchgate.net/publication/232746304_Next_Generation_Search_Engines_Advanced_Models_for_Information_Retrieval
598. *Yang et al.* [2000] - Qiang Yang, Hai-Feng Wang, Ji-Rong Wen, Gao Zhang, Ye Lu, Kai-Fu Lee; 2000; Towards a Next-Generation Search Engine; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/3-540-44533-1_5
599. *Wei-Ying Ma* [2005] - From relevance to intelligence: toward next generation web search; http://www.ccrc.nthu.edu.tw/news/2005/941108/Wei-Ying%20Ma.pdf
600. *Ma et al.* [2004] - Wei-Ying Ma, HongJiang Zhang, Hsiao-Wuen Hon; 2004; Towards Next Generation Web Information Retrieval; https://link.springer.com/chapter/10.1007/978-3-540-30480-7_3
601. *Li Zheng* [2014] - 01 Jan 2014; Towards next generation vertical search engines; https://typeset.io/papers/towards-next-generation-vertical-search-engines-3qts4yq7h9
602. *Serrano et al.* [2017] - Martin Serrano, Michael Boniface, Monique Calisti; Jun 2017; Next Generation Internet Research and Experimentation; https://www.researchgate.net/publication/322399974_Next_generation_internet_research_and_experimentation
603. *Widyantoro* [2007] - Dwi H. Widyantoro; 17 Jun 2007; Toward the Development of Next Generation Search Engine; https://typeset.io/papers/toward-the-development-of-next-generation-search-engine-5fc366dfp0
604. *Steven Miller* [2016] - Steven M. Miller; 01 Jan 2016; Preparing the next generation for the cognitive era; https://content.iospress.com/download/information-services-and-use/isu804?id=information-services-and-use%2Fisu804
605. *Devarakonda et al.* [2011] - Ranjeet Devarakonda, Leslie A. Hook, Giri Palanisamy, James M. Green; 01 Jan 2011; Next-Generation Search Engines for Information Retrieval; https://www.osti.gov/biblio/1011015
606. *Danny Goodwin* [2023] - 7 Sep 2023; Survey: 51% of Gen Z women prefer TikTok, not Google, for search; https://searchengineland.com/gen-z-tiktok-google-search-survey-431345
607. *Matt G. Southern* [2024] - January 15, 2024; TikTok Gains Traction As A Search Engine Among Gen Z [STUDY]; https://www.searchenginejournal.com/tiktok-gains-traction-as-a-search-engine-among-gen-z-study/505633/

608. *Neal Sivadas* [2023] - 31 Aug 2023; Why Gen Z Uses TikTok as a Search Engine, and Why You Should Too; https://www.linkedin.com/pulse/why-gen-z-uses-tiktok-search-engine-you-should-too-neal-sivadas
609. [246]*Kitty Bates* [2002] - 16 Dec 2022; Scrolling or Searching? TikTok; Gen Z's Search Engine; https://mrs.digital/blog/tiktok-gen-z-search-engine/
610. *House-of-communication* [2023] - 19 Sep 2023; How to target Gen Z via search; https://www.house-of-communication.com/uk/en/newsroom/2023/09/how-to-target-gen-z-via-search.html
611. *Hebenthal et al.* [2010] - Douglas C. Hebenthal, Cesare John Saretto, Kathleen Mulcahy, James E. Allard; 13 Oct 2010; Following online social behavior to enhance search experience; https://patents.google.com/patent/US20120095976A1/en
612. *Gupta & Dixit* [2021] - Ashlesha Gupta, Ashutosh Dixit; 08 Aug 2021; A novel user trend-based priority assigner and URL scheduler for dynamic incremental crawling; https://onlinelibrary.wiley.com/doi/epdf/10.1002/cpe.6555
613. *Nilsson & Alanko* [2018] - Rebecca Nilsson, Christa Alanko; 01 Jan 2018; STREAMLINE THE SEARCH ENGINE MARKETING STRATEGY : Generational Driven Search Behavior on Google; https://www.diva-portal.org/smash/get/diva2:1234136/FULLTEXT01.pdf
614. *Andrei Broder* [2007] - Andrei Z. Broder; 02 Apr 2007; The next generation web search and the demise of the classic IR model; https://www.researchgate.net/publication/221397487_The_Next_Generation_Web_Search_and_the_Demise_of_the_Classic_IR_Model
615. *Petrova & Trunin* [2020] - D. Petrova, P. Trunin; 01 Jan 2020; Revealing the mood of economic agents based on search queries; https://econpapers.repec.org/article/risapltrx/0400.htm
616. *Nicholas et al.* [2011] - D. Nicholas, I. Rowlands, David J. Clark, P. Williams; 2011; Google Generation II: web behaviour experiments with the BBC; https://www.emerald.com/insight/content/doi/10.1108/00012531111103768/full/html?skipTracking=true
617. *Jon Clark* [2022] - 3 Oct 2022; 19 advanced Google search operators you need to know; https://searchengineland.com/advanced-google-search-operators-388355
618. *Edwin Toonen* [2023] - 7 Dec 2023; Mastering Google Search: Using search operators for SEO; https://yoast.com/search-operators/
619. *Matteo Duo* [2023] - Matteo Duò, 14 Dec 2023; Google Search Operators: 40 Commands to Know (Improve Research, Competitive Analysis, and SEO); https://kinsta.com/blog/google-search-operators/
620. *Marcin Pastuszek* [2023] - 22 May 2023; Google Advanced Search Operators: Your Ultimate Guide to Enhanced Google Searches; https://seodity.com/blog/google-advanced-search-operators
621. *Carr & Zamansky* [2018] - Terence M. Carr, Leo Zamansky; 11 Dec 2018; Enhancement of search results; https://typeset.io/papers/enhancement-of-search-results-4wa7ch0vag
622. *Hubert et al.* [2011] - Gilles Hubert, Guillaume Cabanac, Christian Sallaberry, Damien Palacio; 2011; Query Operators Shown Beneficial for Improving Search Results; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-642-24469-8_14

623. *Haas et al.* [2011] - Kevin Haas, P. Mika, P. Tarjan, Roi Blanco; 2011; Enhanced results for web search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2009916.2010014
624. *Lucas & Topi* [2002] - Wendy T. Lucas, Heikki Topi; 2002; Form and function: The impact of query term and operator usage on Web search results; https://asistdl-onlinelibrary-wiley-com.ressources-electroniques.univ-lille.fr/doi/full/10.1002/asi.10013
625. *Eastman & Jansen* [2003] - C. Eastman, B. Jansen; 2003; Coverage, relevance, and ranking: The impact of query operators on Web search engine results; https://faculty.ist.psu.edu/jjansen/academic/pubs/tois_2003.pdf
626. *Jovicic* [2023] - Aleksandra Beka Jovicic; 25 Sep 2023; What Is Keyword Intent & How to Align Keywords with User Needs; https://www.semrush.com/blog/keyword-intent/
627. *wordstream-2* - Keyword Intent – The Secret to Attracting the Right Traffic; https://www.wordstream.com/keyword-intent
628. *Nysveen & Pedersen* [2005] - Herbjørn Nysveen, Per E. Pedersen; 2005; Search mode and purchase intention in online shopping behaviour; https://www.inderscience.com/info/inarticle.php?artid=8103
629. *Shim et al.* [2001] - Simon S. Y. Shim, Mary Ann Eastlick, Sherry L Lotz, Patricia Warrington; 2001; An online prepurchase intentions model: The role of intention to search; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0022435901000513
630. *Kim & Lee* [2004] - Jae-Il Kim, Heechung Lee, H. Kim; 2004; Factors Affecting Online Search Intention and Online Purchase Intention; https://s-space.snu.ac.kr/bitstream/10371/1809/1/SJBv10n2_027.pdf
631. *Guo & Agichtein* [2010] - Qi Guo, Eugene Agichtein; 2010; Exploring searcher interactions for distinguishing types of commercial intent; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1772690.1772827
632. *Park et al.* [2010] - Kinam Park, Taemin Lee, Soonyoung Jung, Sangyep Nam, Heuiseok Lim, 2010; Extracting Search Intentions from Web Search Logs; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=5581286
633. *Tsukuda et al.* [2013] - Kosetsu Tsukuda, Tetsuya Sakai, Zhicheng Dou, Katsumi Tanaka; 2013; Estimating Intent Types for Search Result Diversification; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-642-45068-6_3
634. *Ahmadvand et al.* [2020] - A. Ahmadvand, S. Kallumadi, F. Javed, Eugene Agichtein; 2020; JointMap: Joint Query Intent Understanding For Modeling Intent Hierarchies in E-commerce Search; https://arxiv.org/pdf/2005.13783.pdf
635. *Garigliotti & Balog* [2018] - Darío Garigliotti, K. Balog; 2018; IntentsKB: A Knowledge Base of Entity-Oriented Search Intents; https://krisztianbalog.com/files/cikm2018-intents.pdf
636. *Sia et al.* [2012] - Ka Cheung Sia, Kyrylo Tropin, Bhuvan Middha, Paul N. Bennett, Krysta M. Svore; 07 Mar 2012; Search results by mapping associated with disparate taxonomies; https://patents.google.com/patent/US20130238608A1/en
637. *Dou & Guo* [2020] - Zhicheng Dou, Jiafeng Guo; 01 Jan 2020; Query Intent Understanding; https://www.researchgate.net/publication/347290102_Query_Intent_Understanding

638. *Ghanekar et al.* [2010] - Joy Ghanekar, Jerry Cheng, Edward Stanley Ott, Marc Davis; 09 Feb 2010; Entity-based search results and clusters on maps; https://typeset.io/papers/entity-based-search-results-and-clusters-on-maps-3d4mqez7rs
639. *Su et al.* [2021] - Zhan Su, Zhicheng Dou, Yutao Zhu, Xubo Qin, Ji-Rong Wen; 11 Jul 2021; Modeling Intent Graph for Search Result Diversification; Page-746; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3404835.3462872
640. *Jiang Zong* [2015] - Jiang Zong-l; 2015; Automatic Identification of Query Intent by Using Multiple Features; https://www.semanticscholar.org/paper/Automatic-Identification-of-Query-Intent-by-Using-Zong-l/3ad6b66e27efb2069badabc3e9b4269e52105a7d
641. *LN-Search-Engine* [2023] - 12 Dec 2023; What are the most effective ways to handle misspellings in search queries?; https://www.linkedin.com/advice/0/what-most-effective-ways-handle-misspellings-search-mlknc
642. *MacLeod* [2021] - Edith MacLeod; 31 Mar, 2021; Google's new approach to misspellings using deep learning and NLP; https://www.wordtracker.com/blog/keyword-research/google-outlines-new-approach-to-misspellings-in-search-using-deep-learning
643. *Zhang et al.* [2020] - Ruqing Zhang, Jiafeng Guo, Yixing Fan, Yanyan Lan, Xueqi Cheng; 19 Oct 2020; Query Understanding via Intent Description Generation; https://jiafengguo.github.io/2020/2020-Query%20Understanding%20via%20Intent%20Description%20Generation.pdf
644. *Roy et al.* [2015] - Rishiraj Saha Roy, Rahul Katare, Niloy Ganguly, Srivatsan Laxman; 2015; Discovering and understanding word level user intent in Web search queries; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3199173
645. *Donato* [2011] - D. Donato; 2011; Toward a deeper understanding of user intent and query expressiveness; https://www.semanticscholar.org/paper/Toward-a-deeper-understanding-of-user-intent-and-Donato/6961997b9b4661b8ef203d41e3c905d41524a52f
646. *Duan & Hsu* [2011] - Huizhong Duan, B. Hsu; 2011; Online spelling correction for query completion; https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/WWW11-OnlineSpellingCorrection.pdf
647. *Cao et al.* [2009] - Huanhuan Cao, D. Hu, Dou Shen, Daxin Jiang, Jian-Tao Sun, Enhong Chen; 2009; Context-aware query classification; https://www.researchgate.net/publication/221300584_Context-Aware_Query_Classification
648. *Das et al.* [2013] - Ariyam Das, C. Mandal, C. Reade; 2013; Determining the User Intent Behind Web Search Queries by Learning from Past User Interactions with Search Results; https://cse.iitkgp.ac.in/~chitta/pubs/comad13-queryGoal.pdf
649. *Liu et al.* [2020] - Xiaowei Liu, Weiwei Guo, Huiji Gao, Bo Long; 2020; Deep Search Query Intent Understanding; https://arxiv.org/pdf/2008.06759.pdf
650. *Hafernik & Jansen* [2013] - C. Hafernik, B. Jansen; 2013; Understanding the specificity of web search queries; https://faculty.ist.psu.edu/jjansen/academic/jansen_specifity.pdf

651. *Athukorala et al.* [2014] - Kumaripaba Athukorala, Antti Oulasvirta, D. Glowacka, Jilles Vreeken; 2014; Narrow or Broad?: Estimating Subjective Specificity in Exploratory Search; https://vreeken.eu/pubs/2014/foraging-atukorala,oulasvirta,glowacka,vreeken,jaccuci.pdf
652. *Wang et al.* [2009] - Xuanhui Wang, Deepayan Chakrabarti, Kunal Punera; 2009; Mining broad latent query aspects from search sessions; https://www.researchgate.net/publication/220271991_Mining_broad_latent_query_aspects_from_search_sessions
653. *Freund & Toms* [2005] - L. Freund, Elaine Toms; 2005; Understanding the brevity of Web queries; https://www.researchgate.net/publication/247986434_Understanding_the_brevity_of_Web_queries
654. *Yoon & Pantel* [2018] - Sungro Yoon, Soo-Min Pantel, Opeyemi Akanji, Vivek Kumar; 20 Mar 2018; Categorical exploration facilitation responsive to broad search queries; https://typeset.io/papers/categorical-exploration-facilitation-responsive-to-broad-7paoszgdzn
655. *Sperr Jr.* [2018] - Edwin V. Sperr Jr.; 23 Sep 2018; Mind the Gap: Identifying What Is Missed When Searching Only the Broad Scope with PubMed Clinical Queries; https://www.biorxiv.org/content/10.1101/423111v1
656. *Bemmelen & Plough* [2008] - John van Bemmelen, Schering-Plough; 2008; Applying SMQs to Adverse Event Data; https://www.lexjansen.com/phuse/2008/tu/TU05.pdf
657. *Brian Dean* [2022] - 13 Oct 2022; Search Intent; https://backlinko.com/hub/seo/search-intent
658. *Janiszewski* [1998] - Chris Janiszewski; 1998; The Influence of Display Characteristics on Visual Exploratory Search Behavior Get access Arrow; https://academic.oup.com/jcr/article-abstract/25/3/290/1795657?redirectedFrom=fulltext&login=false
659. *Beck et al.* [2014] - François Beck, Jean-Baptiste Richard, Viêt Nguyen-Thanh, Ilaria Montagni, Isabelle Parizot, Emilie Renahy; 13 May 2014; Use of the internet as a health information resource among French young adults: results from a nationally representative survey; https://www.researchgate.net/publication/262304346_Use_of_the_Internet_as_a_Health_Information_Resource_Among_French_Young_Adults_Results_From_a_Nationally_Representative_Survey
660. *Figueiredo & Paiva* [2015] - Dijanice Alves Figueiredo, Eliane Bezerra Paiva; 14 Apr 2015; Estudo do Comportamento Informacional dos usuários da Médiathèque Simone de Beauvoir da Aliança Francesa João Pessoa; https://typeset.io/papers/estudo-do-comportamento-informacional-dos-usuarios-da-3zmyixg2p3
661. *Florence* [2005] - Florence Gonod Boissin; 01 Sep 2005; Information-seeking behaviour and use of the Internet by French general practitioners: a qualitative study; https://pubmed.ncbi.nlm.nih.gov/16109145/
662. *Ducrot et al.* [2021] - Pauline Ducrot, Ilaria Montagni, Viêt Nguyen Thanh, Anne-Juliette Serry, Jean-Baptiste Richard; 14 Apr 2021; Evolution of Online Health-Related Information Seeking in France From 2010 to 2017: Results From Nationally Representative Surveys; https://typeset.io/papers/evolution-of-online-health-related-information-seeking-in-4sv80gsm66
663. *Woolford* [2015] - Cecile Woolford; Jan 2015; FRENCH FAMILIES IN THE TOURISM INDUSTRY: THE INFLUENCE OF SOCIAL MEDIA AND INTERNET ON THEIR INFORMATION SEARCH BEHAVIOUR

AND THEIR DECISIONMAKING CRITERIA; https://esource.dbs.ie/server/api/core/bitstreams/fd7ba70c-ac98-404d-9ccc-268acf9d8713/content

664. *Rolland & Freeman* [2010] - Slyvie Rolland, Ina Freeman; March 2010; A new measure of e-service quality in France; https://www-emerald-com.ressources-electroniques.univ-lille.fr/insight/content/doi/10.1108/09590551011052106/full/pdf?title=a-new-measure-of-eservice-quality-in-france
665. *Jolivet & Turon* [2019] - Grégory Jolivet, Hélène Turon; 2019; Consumer Search Costs and Preferences on the Internet; https://research-information.bris.ac.uk/ws/portalfiles/portal/150990544/joliron_reresubmit.pdf
666. *Muller & Bolger* [1985] - T. Muller, Christopher Bolger; 1985; SEARCH BEHAVIOUR OF FRENCH AND ENGLISH CANADIANS IN AUTOMOBILE PURCHASE; https://doi.org/10.1108/eb008288
667. *Bressolles & Durrieu* [2010] - Grégory Bressolles, F. Durrieu; 2010; A typology of online buyers for French wine web sites based on electronic service quality dimensions; https://www.researchgate.net/publication/235279196_A_typology_of_online_buyers_for_French_wine_web_sites_based_on_electronic_service_quality_dimensions
668. *Kai Lemmetty* [2022] - Online shopping trends in France to grow your business; 2022; https://www.gaasly.com/blog/online-shopping-trends-in-france
669. *Manon Fischer* [2023] - 2 Aug 2023; 8 consumer habits to know before expanding your brand to France; https://www.hive.app/blog/consumer-habits-france-expansion
670. *Renahy et al.* [2007] - Emilie Renahy, Isabelle Parizot, Pierre Chauvin; 11 Oct 2007; WHIST: a web-based survey on health information seeking on Internet in France, 2007; https://www.researchgate.net/publication/23162848_WHIST_A_web-based_survey_on_health_information_seeking_on_Internet_in_France_2007
671. *Renahy et al.* [2010] - Emilie Renahy, Isabelle Parizot, Pierre Chauvin; 19 Mar 2010; Determinants of the frequency of online health information seeking: results of a web-based survey conducted in France in 2007; https://typeset.io/papers/determinants-of-the-frequency-of-online-health-information-4w5jj1dtlv
672. *Weaver et al.* [2010] - James B. Weaver, Darren Mays, Stephanie Sargent Weaver, Gary L. Hopkins, Doğan Eroğlu, Jay M. Bernhardt; Aug 2010; Health Information–Seeking Behaviors, Health Indicators, and Health Risks; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2901281/
673. *Jia et al.* [2021] - Xiaoyun Jia, Yan Pang, Liangni Sally Liu; 16 Dec 2021; Online Health Information Seeking Behavior: A Systematic Review; https://www.mdpi.com/2227-9032/9/12/1740
674. *Boyce et al.* [2022] - LeAnn Boyce, Ahasan Harun, Gayle Prybutok, Victor R Prybutok; Apr 2022; Exploring the factors in information seeking behavior: a perspective from multinational COPD online forums; https://pubmed.ncbi.nlm.nih.gov/34293130/
675. *Zhang et al.* [2019] - Runtong Zhang, Xinyi Lu, Wen Wu, Xiaopu Shang; May 2019; Why do patients follow physicians' advice? The influence of patients' regulatory focus on adherence: an empirical study in China; https://pubmed.ncbi.nlm.nih.gov/31077196/

676. *Suenaga & Vicente* [2021] - Hiroaki Suenaga & Maria Rosalía Vicente; Sep 2021; Online and offline health information seeking and the demand for physician services; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s10198-021-01352-7
677. Liu ZW. [2022] - Research Progress and Model Construction for Online Health Information Seeking Behavior; Front Res Metr Anal. 2022 Feb 11;6:706164; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8873569/
678. *Rashid & Wasif* [2016] – Junaid Rashid, Muhammad Wasif Nisar; 10 Oct 2016; A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines; https://www.mecs-press.org/ijitcs/ijitcs-v8-n10/IJITCS-V8-N10-10.pdf
679. *Joakim* [2020] - Joakim Edlund; 2020; Cognitive Search Engine Optimization; https://www.diva-portal.org/smash/get/diva2:1470637/FULLTEXT01.pdf
680. *Kiseleva et al* [2015] - Julia Kiseleva, Jaap Kamps, Vadim Nikulin, Nikita Makarov; Oct 2015; Behavioral Dynamics from the SERP's Perspective: What are Failed SERPs and How to Fix Them?; https://www.researchgate.net/publication/281967311_Behavioral_Dynamics_from_the_SERP%27s_PerspectiveWhat_are_Failed_SERPs_and_How_to_Fix_Them
681. *Gong & Cosma* [2023] - Yan Gong, Georgina Cosma; Nov 2023; BOON: A NEURAL SEARCH ENGINE FOR CROSS-MODAL INFORMATION RETRIEVAL; http://arxiv.org/abs/2307.14240
682. *Philipp & White* [2017] - George Philipp, Ryen W. White; Dec 2017; Interactions between Health Searchers and Search Engines; http://arxiv.org/abs/1712.03622
683. *Hannak et al.* [2013] - Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, Christo Wilson; 2013; Measuring personalization of web search; https://cbw.sh/static/pdf/fp039-hannak.pdf
684. *Yao et al.* [2021] - JING YAO, ZHICHENG DOU, JUN XU, JI-RONG WEN; May 2021; RLPS: A Reinforcement Learning–Based Framework for Personalized Search; http://playbigdata.ruc.edu.cn/dou/publication/2021_TOIS_RLPer.pdf
685. *Almeida* [2017] - Fernando Almeida; 2017; Concept and Dimensions of Web 4.0; https://www.researchgate.net/publication/321366810_Concept_and_Dimensions_of_Web_40
686. *Asaad Khaleel* [2021] - Asaad Khaleel Ibrahim; 2021; Evolution of the Web: from Web 1.0 to 4.0; https://journal.qubahan.com/index.php/qaj/article/view/75/49
687. *França et al.* [2021] - R. P. França, Ana Carolina Borges Monteiro, Rangel Arthur, Y. Iano; 2021; An Overview of Web 2.0 and Its Technologies and Their Impact in the Modern Era; https://ouci.dntb.gov.ua/en/works/4E0zz0K7/
688. *Dom et al.* [2005] - B. Dom, K. Bharat, A. Broder, Marc Najork, Jan O. Pedersen, Yoshinobu Tonomu; 2005; How search engines shape the web; https://www.ra.ethz.ch/cdstore/www2005/docs/p879.pdf
689. *Martasari* [2023] - Gacelia Weny Martasari; 25 Jun 2023; Impact of Industrial Technology 4.0 In Improving Service Quality and Customer Experience on E-Commerce Platforms: Literature Review; https://typeset.io/papers/impact-of-industrial-technology-4-0-in-improving-service-1v1j497o

690. *Rajkumar et al.* [2012] - N. Rajkumar, B. Gohin, Viji Vinod; 01 Jan 2012; Search engine: Intelligent web service search; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/document/6719089
691. *Srikantaiah et al.* [2012] - Srikantaiah K C, Srikanth P L, Tejaswi, Shaila K, Venugopal K. R, L M Patnaik; 01 Jan 2012; Ranking Search Engine Result Pages based on Trustworthiness of Websites; https://arxiv.org/ftp/arxiv/papers/1209/1209.5244.pdf
692. *Agile District* [2022] - 30 Sep 2022; Web 4.0 Explained – A Brief!; https://www.linkedin.com/pulse/web-40-explained-brief-agiledistrict
693. *Strivemindz.com* [2023] - 26 Oct 2023; What Is Web 4.0: An Overview of the Latest Evolution of the Internet; https://www.strivemindz.com/blog/what-is-web-4-0/
694. *BairesDev.com* - Web 4.0? It's Time To Talk About It; https://www.bairesdev.com/blog/web-4-0-its-time-to-talk-about-it/
695. *Szeredi et al.* [2014] - Pter Szeredi, Gergely Lukcsy, Tams Benk; 2014; The Semantic Web Explained: The Technology and Mathematics behind Web 3.0; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.5555/2635961
696. *Sheth & Thirunarayan* [2012] - A. Sheth, K. Thirunarayan; 2012; Semantics Empowered Web 3.0: Managing Enterprise, Social, Sensor, and Cloud-based Data and Services for Advanced Applications; https://corescholar.libraries.wright.edu/books/70/
697. *Raghavan* [2023] - Rithesh Raghavan; 01 Jun 2023; WEB 1.0 VS WEB 2.0 VS WEB 3.0 VS WEB 4.0 VS WEB 5.0 – EVOLUTION OF THE WORLD WIDE WEB; https://acodez.in/evolution-of-the-world-wide-web/
698. *cronuts-digital* - Difference between web 1.0 web 2.0 and web 3.0 web 4.0; https://cronuts.digital/en/diferencias-entre-web-10-web-20-web-30-web40/
699. *Aghaei et.al* [2012] - Aghaei et.al; Jan 2012; EVOLUTION OF THE WORLD WIDE WEB: FROM WEB 1.0 TO WEB 4.0; https://airccse.org/journal/ijwest/papers/3112ijwest01.pdf
700. *Kusum* [2025] – Kusum Saini; 31 Jul 2025; Web 1.0, 2.0, 3.0, & 4.0: A Detailed Guide; https://www.simplilearn.com/what-is-web-1-0-web-2-0-and-web-3-0-with-their-difference-article
701. *Lin et al.* [2022] - Yi-Lan Lin, Zhipeng Gao, Yaofeng Tu, Hongyang Du, Dusit Niyato, Jiawen Kang, Hui Yang; 30 Nov 2022; A Blockchain-based Semantic Exchange Framework for Web 3.0 toward Participatory Economy; https://arxiv.org/pdf/2211.16662.pdf
702. *Rettinger et al.* [2018] - Achim Rettinger, Stefan Zander, Maribel Acosta, York Sure-Vetter; 01 Jan 2018; Semantic Technologies: Enabler for Knowledge 4.0; https://www.researchgate.net/publication/323395886_Semantic_Technologies_Enabler_for_Knowledge_40
703. *Daniel Harris* [2008] - Daniel Harris; 2008; Web 2.0 Evolution into The Intelligent Web 3.0: 100 Most Asked Questions on Transformation, Ubiquitous Connectivity, Network Computing, Open Technologies, ... Databases and Intelligent Applications; https://pdfroom.com/books/web-20-evolution-into-the-intelligent-web-30-100-most-asked-questions-on-transformation-ubiquitous-connectivity-network-computing-open-technologies-databases-and-intelligent-applications/e1j5KvZZ2Kr

704. *Gan et al.* [2023] - Wensheng Gan, Zhenqiang Ye, Shicheng Wan, P. Yu; 2023; Web 3.0: The Future of Internet; https://arxiv.org/pdf/2304.06032.pdf
705. *R-S-Kumar* [2022] - R. S. Kumar; 2022; AN OVERVIEW OF THE EXPECTED INFLUENCE OF WEB 3.0 ON e-COMMERCE AND ALLIED DOMAINS; https://www.trendytechjournals.com/ijtret/volume6/issue6-7.pdf
706. *Fan et al.* [2023] - Yuqing Fan, Tianyi Huang, Yiran Meng, Sheng-Hsien Cheng; 2023; The current opportunities and challenges of Web 3.0; https://arxiv.org/pdf/2306.03351.pdf
707. *William* [2022] - James William; 9 May 2022; Learn About The Upcoming Techniques Web 3.0 VS Web 4.0; https://blog.temok.com/web-30-vs-web-40/
708. *Gupta et al.* [2023] - Shikha Gupta, Dheeraj Patel, Jyotsna Rai; 30 Jun 2023; Scientometric review of Web 3.0; https://www.researchgate.net/publication/372037084_Scientometric_review_of_Web_30
709. *Vasiliu-Feltes* [2023] - Dr. Ingrid Vasiliu-Feltes; 23 Jan 2023; Leveraging Web 3.0 Architecture for Healthcare 4.0; https://telehealthandmedicinetoday.com/index.php/journal/article/view/398
710. *automationedge.com-1* [2023] - 18 Apr 2023; Top 7 Healthcare Trends to Watch for in 2025; https://automationedge.com/home-health-care-automation/blogs/top-7-healthcare-trends-to-watch-for-in-2025/
711. Yang et al. [2023] - Soorim Yang, Hyeong-jun Joo, Jaeho Kim; 23 Dec 2023; Metaverse search system: Architecture, challenges, and potential applications; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S2405959523001649
712. *Lung et al.* [2009] - KKF Lung, WC Tang, KYK Lee; 2009 - Method Of Sorting The Result Set Of A Search Engine; https://patents.google.com/patent/US20110161309A1/en
713. *Mmeje* [2024] - Chima Mmeje; Dec. 2024; What is a SERP Feature? Common Types And How To Win Them; https://moz.com/learn/seo/serp-features
714. *Wiki Syntax* - https://en.wikipedia.org/wiki/Syntax
715. *Wiki Phrase* - https://en.wikipedia.org/wiki/Phrase
716. *Oxfordlearnerdictionaries* - Dictionary meaning of Word Phrase; Oxford Learners Dictionaries; https://www.oxfordlearnersdictionaries.com/definition/english/phrase_1
717. *Shhub et al.* [2023] - Aya Shhub , Zaira Jimenez, and Michael Solis; 05 Jan 2023; A Synthesis of Reading Prosody: Evaluating Phrasing and Syntax Interventions; https://www.researchgate.net/publication/366895189_A_Synthesis_of_Reading_Prosody_Evaluating_Phrasing_and_Syntax_Interventions
718. *PageRank-Wiki* - https://en.wikipedia.org/wiki/PageRank
719. *Anna Gargioni* - Top Google Search Secrets that Will Make You Look Smarter; https://www.greengeeks.com/blog/top-google-search-secrets-that-will-make-you-look-smarter/
720. *Pcmag-GSOs* - Search Expression; https://www.pcmag.com/encyclopedia/term/search-expression
721. *Condron et al.* [2022] - SIDRA CONDRON, PATRICK MCGRANE, ZAC HARRIS; 26 Jan 2022; Google Advanced Search Operators: 50 + Google Search Commands; https://www.spyfu.com/blog/google-search-operators/

722. *Google-GSOs-1* - Overview of Google search operators; https://developers.google.com/search/docs/advanced/debug/search-operators/overview
723. *Google-GSOs-2* - Refine Google searches; https://support.google.com/websearch/answer/2466433?hl=en
724. *Moz-GSOs* – 62 Advanced Google Search Operators, Use Cases & Cheatsheet; https://moz.com/learn/seo/search-operators
725. *Technolush* [2019] - 09 Jul 2019; Advanced Google Search Tips; https://www.technolush.com/blog/advanced-google-search-tips
726. *Edosomwan* [2010] - Joseph Edosomwan, Taiwo O. Edosomwan; Oct 2010; Comparative analysis of some search engines; South African Journal of Science 106 (11/12); https://www.researchgate.net/publication/307685487_Comparative_analysis_of_some_search_engines
727. *Acharya et al.* [2008] - Kshitish Acharya, Greta Kasliwal & Haritha Haridas; 22 Jul 2008; A comparative analysis of 21 literature search engines; https://www.nature.com/articles/npre.2008.2101.1
728. *Garoufallou* [2012] - Emmanouel Garoufallou; 20 Apr 2012; Evaluating search engines: A comparative study between international and Greek SE by Greek librarians; https://www.emerald.com/insight/content/doi/10.1108/00330331211221837/full/html
729. *ML-Wiki* - Machine Learning; https://en.wikipedia.org/wiki/Machine_learning
730. *Dandekar* [2016] - Nikhil Dandekar; 07 Apr 2016; What are some uses of machine learning in search engines? ; https://medium.com/@nikhilbd/what-are-some-uses-of-machine-learning-in-search-engines-5770f534d46b
731. *Grigorov* [2019] - Dido Grigorov; 18 Nov 2019; 8 Machine Learning Principles and Models Used by Search Engines; https://serpact.com/8-machine-learning-principles-and-models-used-by-search-engines/
732. *Wroe* [2019] - Tiffani Wroe;14 Nov 2019; How Google Search (Probably) Uses Machine Learning; https://www.embeddedcomputing.com/technology/ai-machine-learning/how-google-search-probably-uses-machine-learning
733. *Wiggers* [2020] - Kyle Wiggers; 15 Oct 2020; Google details how it's using AI and machine learning to improve search; https://venturebeat.com/2020/10/15/google-details-how-its-using-ai-and-machine-learning-to-improve-search/
734. *Houmb et al.* [2006] - Shiv Hilde Houmb, Indrakshi Ray, & Indrajit Ray; 2006; Estimating the Relative Trustworthiness of Information Sources in Security Solution Evaluation; https://www.cs.colostate.edu/~iray/research/papers/iTrust06.pdf
735. *Wiki-OA* - Online Advertising - http://en.wikipedia.org/wiki/Online_advertising#Delivery_methods
736. *IAB-Report* [2017] - IAB internet advertising revenue report 2016 full year results; https://www.iab.com/wp-content/uploads/2016/04/IAB_Internet_Advertising_Revenue_Report_FY_2016.pdf
737. *Maynes & Everdell* [2014] - Rebecca Maynes & Ian Everdell; 2014; The evolution of Google Search results pages & their effects on user behaviour; http://seopaslaugos.com/wp-content/uploads/2014/The_Evolution_of_Google_Search_Results_Pages_and_Their_Effect_on_User_Behaviour.pdf

738. *De Silva & Alahakoon* [2022] - Daswin De Silva, Damminda Alahakoon; 2022; An artificial intelligence life cycle: From conception to production; https://www.cell.com/patterns/fulltext/S2666-3899(22)00074-5?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2666389922000745%3Fshowall%3Dtrue
739. *wordstream-1* - What Is a SERP? A Guide to Search Engine Results Pages; https://www.wordstream.com/serp
740. *Google.com-SERP* - Automatically generating and ranking results; https://www.google.com/search/howsearchworks/how-search-works/ranking-results/
741. *Jenny Chang* [2024] - 90 Google Search Statistics for 2024: Usage & User Behavior Data; https://financesonline.com/google-search-statistics/
742. *Kyle Byers* [2020] - 5 Apr 2020; What 3.25 Billion Website Visits Tell Us About the Internet's Top Traffic Sources; https://growthbadger.com/traffic-study/
743. *Olaf Kopp* [2024] - Olaf Kopp; 28 Jan 2024; Social media has become one of the most important gatekeepers for content; https://www.kopp-online-marketing.com/social-media-has-become-one-of-the-most-important-gatekeepers-for-content
744. *Rose & Levinson* [2004] - Daniel Rose and Danny Levinson; Apr 2004; Understanding User Goals in Web Search; http://www.medien.ifi.lmu.de/lehre/ws0910/mmi2/literature/UnderstandingUserGoalsInWebSearch_Rose.pdf
745. *Broder* [2002] - Andrei Broder; Sep 2002; A taxonomy of web search - IBM Research; https://dl.acm.org/doi/abs/10.1145/792550.792552
746. *A. K. Dani* [2008] - Amar Kumar Dani; May 2008; Automatic Identification of user goals in web search based on classification of click-through results; http://cse.iitkgp.ac.in/~abhij/facad/03UG/Report/03CS3014_Amar_Dani.pdf
747. *George Lawton-techtarget* - George Lawton; What is generative AI? Generative AI explained; https://www.techtarget.com/searchenterpriseai/definition/generative-AI
748. *Philip Resnik* [1995] - Philip Resnik; 28 Nov 1995; Using information content to evaluate semantic similarity in a taxonomy; https://arxiv.org/pdf/cmp-lg/9511007.pdf
749. *Dekang Lin* [1998] - Dekang Lin; 24 Jul 1998; An information-theoretic definition of similarity; https://www.semanticscholar.org/paper/An-Information-Theoretic-Definition-of-Similarity-Lin/cc0c3033ea7d4e19e1f5ac71934759507e126162
750. *Jiang & Conrath* [1997] - Jiang JJ, Conrath DW; 20 Sep 1997; Semantic similarity based on corpus statistics and lexical taxonomy; https://arxiv.org/pdf/cmp-lg/9709008.pdf
751. *Hirst & St-onge* [1995] - Graeme Hirst, David St-onge; Oct 1995; Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms; https://www.researchgate.net/publication/2735129_Lexical_Chains_as_Representations_of_Context_for_the_Detection_and_Correction_of_Malapropisms
752. *Banerjee S, Pedrsen* [2003] - Banerjee S, Pedrsen T; 2003; Extended gloss overlaps as a measure of semantic relatedness; In Proceedings of the eighteenth international joint conference on artificial intelligence (Acapulco, Mexico, August 9–15, 2003), IJCAI '03, pp 805–810; https://www.d.umn.edu/~tpederse/Pubs/ijcai03.pdf

753. *Patwardhan* [2003] - Patwardhan S; Aug 2003; Incorporating dictionary and corpus information into a context vector measure of semantic relatedness; https://www.academia.edu/327271/Incorporating_Dictionary_and_Corpus_Information_Into_a_Context_Vector_Measure_of_Semantic_Relatedness
754. *Khurram et al.* [2018] - Khurram Shahzad, Ifrah Pervaz, Rao Muhammad Adeel Nawab; Sep 2018; WordNet-based Semantic Similarity Measures for Process Model Matching; https://ceur-ws.org/Vol-2218/paper4.pdf
755. *funsimmat-sss* - Semantic similarity scores; https://funsimmat.bioinf.mpi-inf.mpg.de/help3.php#:~:text=Semantic%20similarity%20scores&text=Resnik's%20measure%20is%20a%20measure,UniProt%20database%20as%20its%20probability
756. *Chandrasekaran & Mago* [2021] - DHIVYA CHANDRASEKARAN, VIJAY MAGO; 30 Jan 2021; Evolution of Semantic Similarity- A Survey; https://arxiv.org/pdf/2004.13820.pdf
757. *Anna & Francesco* [2021] - ANNA FORMICA, FRANCESCO TAGLINO; 22 Jul 2021; An Enriched Information-Theoretic Definition of Semantic Similarity in a Taxonomy; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=9483871
758. *Han et al.* [2021] - Bin Han, Chirag Shah, and Daniel Saelid; 06 May 2021; Users Perception of Search Engine Biases and Satisfaction; https://arxiv.org/pdf/2105.02898.pdf
759. *Gao & Shah* [2020] - Ruoyuan Gao, Chirag Shah; 01 Jan 2020; Toward creating a fairer ranking in search engine results; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0306457319304121
760. *Ensan & Al-Obeidat* [2019] - Faezeh Ensan, Feras Al-Obeidat; 11 Aug 2019; Relevance-based entity selection for ad hoc retrieval; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0306457318306836
761. *Google SERP View* [2006] - 22 Mar 2006; New Google SERP Really Soon; http://googlesystem.blogspot.com/2006/03/new-google-serp-really-soon.html
762. *Deepawaliseotips* [2023] - 6 Jun 2023; Mastering Google Search Operators and Commands: A Comprehensive Guide; https://www.deepawaliseotips.com/mastering-google-search-operators-commands-comprehensive-guide/
763. *Khaled* [2007] - Khaled M. Nagi; Jan 2007; Bringing Information Retrieval Back To Database Management Systems; https://www.researchgate.net/publication/221057625_Bringing_Information_Retrieval_Back_To_Database_Management_Systems
764. *Ullah et al.* [2018] - Arif Ullah, Nazri Mohd Nawi, Edi Sutoyo, Muhammad Aamir; Nov 2018; Search Engine Optimization Algorithms for Page Ranking: Comparative Study; https://www.researchgate.net/publication/329605308_Search_Engine_Optimization_Algorithms_for_Page_Ranking_Comparative_Study
765. *Pundkar & Baheti* [2015] - Sumedh Pundkar, Kapil Baheti; Jun 2015; Semantic Search Engine using Ontologies; https://www.academia.edu/15099592/Semantic_Search_Engine_using_Ontologies

766. *Olaf Kopp-Intents* [2022] - 18 Jul 2022; How to use 12 micro intents for SEO and content journey mapping; https://searchengineland.com/micro-intents-seo-content-journey-mapping-386544
767. *Jansen & Spink* [2006] - Jansen BJ, Spink A; 2006; How are we searching the world wide web? A comparison of nine search engine transaction logs; Information Process Manag, 42(1):248–63; https://doi.org/10.1016/j.ipm.2004.10.007
768. *Mark Sanderson* [2008] - Mark Sanderson; Jul 2008; Ambiguous queries: test collections need more sense; 31st ACM SIGIR conference Singapore, ACM, New York, NY, USA, pp. 499–506; https://www.researchgate.net/publication/221301174_Ambiguous_queries_test_collections_need_more_sense
769. *Xu & Ye* [2018] - Jialu Xu, Feiyue Ye; Nov 2018; Query recommendation using hybrid query relevance; Future Internet; https://doi.org/10.3390/fi10110112.
770. *Santos et al.* [2015] - Santos RLT, Macdonald C, Ounis I.; Mar 2015; Search result diversification; https://doi.org/10.1561/1500000040.
771. *Bendersky et al.* [2011] - M. Bendersky, W. B. Croft, and Y. Diao; 9-12 Feb 2011; Quality-biased ranking of web documents; https://people.cs.umass.edu/~yanlei/publications/wsdm11.pdf
772. *Carpineto & Romano* [2012] - Carpineto C, Romano G.; Jan 2012; A survey of automatic query expansion in information retrieval; https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f90f1695f32d88e89773181d4b20e734ed686187
773. *Carpineto et al.* [2009] - C. Carpineto, S. Osiński, G. Romano, and D. Weiss, 2009. A survey of web clustering engines. ACM Computing Surveys. 41(3):1–38 ; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1541880.1541884
774. *Hearst & Pedersen* [1996] - Hearst MA, Pedersen JO; Aug 1996; Reexamining the cluster hypothesis: scatter/gather on retrieval results; Proceedings of ACM SIGIR conference New York; https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=30e6ff76eb5ba028128984c3b15eb38bf191c69d
775. *Ngo & Nguyen* [2004] - Ngo CL, Nguyen HS; 2004; A tolerance rough set approach to clustering Web search results; In: Proceedings of the 8th European conference on principles and practice of knowledge discovery in databases *(Pisa, Italy, September 20 – 24, 2004).* PKDD'04, Lecture Notes in Computer Science, 3202, Springer, pp 515–517; https://www.researchgate.net/publication/220699585_A_Tolerance_Rough_Set_Approach_to_Clustering_Web_Search_Results
776. *Zamir & Etzioni* [1998] - Zamir O, Etzioni O; 24-28 Aug 1998; Web document clustering: a feasibility demonstration; SIGIR'98, ACM, New York, NY, USA, pp. 46–54; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/290941.290956
777. *Alghamdi & Alfalqi* [2015] - Alghamdi R, Alfalqi K.; 2015; A survey of topic modeling in text mining; IJACSA, 6(1):147–53; https://thesai.org/Downloads/Volume6No1/Paper_21-A_Survey_of_Topic_Modeling_in_Text_Mining.pdf

778. *Goyal et al.* [2013] - Goyal P, Mehala N, Bansal A; Oct 2013; A robust approach for finding conceptually related queries using feature selection and tripartite graph structure; J Inform Sci. 39(5):575–92; https://journals-sagepub-com.ressources-electroniques.univ-lille.fr/doi/epub/10.1177/0165551513477819

779. *Fu & Yu* [2020] - Yu FU, You YU; 2020; Research on text representation method based on improved TF-IDF; In: Journal of physics: conference series. 1486. pp. 072032; https://iopscience.iop.org/article/10.1088/1742-6596/1486/7/072032/pdf

780. *Blei et al.* [2003] - David M. Blei, Andrew Y. Ng, Michael I. Jordan; 2003; Latent dirichlet allocation; In Journal of machine Learning research, 3(Jan):993–1022; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.5555/944919.944937

781. *Jelodar et al.* [2019] - Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, Zhao L; 2019; Latent dirichlet allocation (LDA) and Topic modeling: models, applications, a survey, Multimedia Tools Appl; 78(11):15169–211; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s11042-018-6894-4

782. *Tong & Zhang* [2016] - Tong Z, Zhang H; 21-22 May 2016; A text mining research based on LDA topic modelling; In: Proceedings of the sixth international conference on computer science, engineering and information technology (CCSEIT), Vienna, Austria, pp 21–22; https://www.researchgate.net/publication/303563965_A_Text_Mining_Research_Based_on_LDA_Topic_Modelling

783. *Potha & Stamatatos* [2018] - Potha N, Stamatatos E; 2018; Intrinsic author verification using topic modeling; In: Proceedings of the 10th hellenic conference on artificial intelligence (SETN'18), (Patras, Greece) (pp. 1–7); https://doi.org/10.1145/3200947.3201013

784. *Bruce Croft* [2019] - Bruce Croft W; 2019; The importance of interaction for information retrieval; In: 42nd international ACM SIGIR conference on research and development in information retrieval (SIGIR '19), (Paris, France, July 21–25, 2019), pp 1–2; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3331184.3331185

785. *Bae et al.* [2020] - Bae J, Helldin T, Riveiro M, Nowaczyk S, Mohamed-Rafik Bouguelia, Falkman G; 2020; Interactive clustering: a comprehensive review; ACM Comput. Surv. 53, 1, Article 1 (February 2020), 39 pages; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3340960

786. *Zamani et al.* [2020] - Zamani H, Dumais S, Craswell N, Bennett P, Lueck G; 2020; Generating clarifying questions for information retrieval; In: Proceedings of the 29th international conference on world wide web. (WWW'20), (Taipei, Taiwan, April 20–24, 2020). Association for Computing Machinery, New York, NY, USA pp. 418–428; https://www.microsoft.com/en-us/research/uploads/prod/2020/01/webconf-2020-camera-zamani-et-al.pdf

787. *Mehala & Goyal* [2011] - Mehala N, Goyal P ; 2011; Concept based query recommendation; In: Proceedings of the ninth Australasian data mining conference (Ballarat, Australia, Dec 01–02, 2011), AusDM'11, ACM, New York, NY, USA, 121, pp 69–78; https://crpit.scem.westernsydney.edu.au/confpapers/CRPITV121Goyal.pdf

788. *Dumais* [2005] - Dumais ST; 22 Sep 2005; Latent semantic indexing; Ann Rev Inform Sci Technol, 38(1):188–230; https://asistdl-onlinelibrary-wiley-com.ressources-electroniques.univ-lille.fr/doi/pdf/10.1002/aris.1440380105
789. *Liu et al.* [2014] - Liu Y, Wang C, Zhou K, Nie J, Zhang M, Ma S; 2014; From skimming to reading: a two-stage examination model for web search; In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management (Shanghai, China, November 3–7, 2014). CIKM'14, ACM, New York, NY, USA, pp. 233–240; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2661829.2661907
790. *Raghavan et al.* [1989] - Raghavan VV, Bollman P, Jung GS; 1989; A critical investigation of recall and precision as measures of retrieval system performance; ACM Trans Inform Syst. 1989, 7:205–29; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/65943.65945
791. *Pedrsen et al.* [2004] - Pedrsen T, Patwardhan S, Michelizzi J; Apr 2004; Wordnet: Similarity: measuring the relatedness of concepts; In: Proceedings of the Demonstration Papers HLT-NAACL, Boston, MA, USA, ACL, Stroudsburg, PA, USA, pp. 38–41; https://www.researchgate.net/publication/2903260_WordNetSimilarity_-_Measuring_the_Relatedness_of_Concepts
792. *Leacock et al.* [1998] - C. Leacock, Geore A Miller and M. Chodorow; 1998; Combining local context and WordNet similarity for word sense identification; In: WordNet: an electronic lexical database, The MIT Press, Cambridge, 11, pp. 265–283; https://www.researchgate.net/publication/200045856_Combining_Local_Context_and_WordNet_Similarity_for_Word_Sense_Identification
793. *Wu & Palmer* [1994] - Wu Z, Palmer M; 1994; Verb semantics and lexical selection; In: Proceedings of the 32nd Annual meeting of the association for computational linguistics (Las Cruces, New Mexico, USA, June 27–30, 1994). ACL'94, pp. 133–138; https://aclanthology.org/P94-1019.pdf
794. *Chin et al.* [2015] - Jessie Chin, Evan Anderson, Chieh-Li Chin, Wai-Tat Fu; 2015; Age differences in information search: An exploration-exploitation tradeoff model; https://www.researchgate.net/publication/278302235_Age_differences_in_information_search_An_exploration-exploitation_tradeoff_model
795. *Di Caprio et al.* [2022] - Debora Di Caprio, Francisco J. Santos-Arteaga, Madjid Tavana; 01 Apr 2022; An information retrieval benchmarking model of satisficing and impatient users' behavior in online search environments; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S095741742101650X
796. *Russo & Russo* [2020] - Luca Russo & Selena Russo; 13 Feb 2020; Search engines, cognitive biases and the man–computer interaction: a theoretical framework for empirical researches about cognitive biases in online search on health-related topics; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s11019-020-09940-9
797. *Yi Song* [2019] - Yi Song; 2019; The Research on User Interest Mining Based on Deep Learning; https://iopscience.iop.org/article/10.1088/1742-6596/1345/6/062041/pdf

798. *Shi et al.* [2014] - Conglei Shi, Yingcai Wu, Shixia Liu, Hong Zhou, Huamin Qu; Dec 2014; LoyalTracker: Visualizing Loyalty Dynamics in Search Engines; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=6876038
799. *Linus Philip* [2015] - Lawrence, Linus Philip; 2015; What are the effects of different search engine interface designs on user behaviour?; https://essay.utwente.nl/68261/1/Lawrence_MA_MB.pdf
800. *Bitirim & Ertugrul* [2020] - Yıltan Bitirim, Duygu Çelik Ertuğrul; 31 Jul 2020; An analysis of user behaviors on the search engine results pages based on the demographic characteristics; https://itiis.org/digital-library/23717
801. *Azzopardi et al.* [2018] - Leif Azzopardi, Paul Thomas, Nick Craswell; 2018; Measuring the Utility of Search Engine Result Pages: An Information Foraging Based Measure; SIGIR'18, July 8-12, 2018, Ann Arbor, MI, USA; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/abs/10.1145/3209978.3210027
802. *Zanjani & Khadivi* [2014] - Mohammad Daryaie Zanjani, Shahram Khadivi; 05 Oct 2014; Predicting user click behaviour in search engine advertisements; https://www.researchgate.net/publication/288057237_Predicting_user_click_behaviour_in_search_engine_advertisements
803. *Lewandowski & Kammerer* [2020] - Dirk Lewandowski, Yvonne Kammerer; 23 Apr 2020; Factors influencing viewing behaviour on search engine results pages: a review of eye-tracking research; https://searchstudies.org/wp-content/uploads/2021/04/20191203_Eyetracking-Review_Preprint.pdf
804. *Chuklin et al.-1* [2015] - Aleksandr Chuklin, Ilya Markov, Maarten de Rijke; 2015; An Introduction to Click Models for Web Search; https://staff.fnwi.uva.nl/m.derijke/wp-content/papercite-data/pdf/chuklin-introduction-2015.pdf
805. *Schultheib* [2023] - Sebastian Schultheib; Jan 2023; How search engine marketing influences user knowledge gain; https://arxiv.org/ftp/arxiv/papers/2301/2301.10086.pdf
806. *Benjelloun et al.* [2020] - Omar Benjelloun, Shiyu Chen, Natasha Noy; 2-6 Nov 2020; Google Dataset Search by the Numbers;; The semantic web - ISWC 2020 Book, p 667-682; 19th International Semantic Web Conference, Athens, Greece; https://link-springer-com.ressources-electroniques.univ-lille.fr/book/10.1007/978-3-030-62466-8
807. *Renhe et al.* [2022] - Jiang Renhe, Chen Quanjun, Cai Zekun, Fan Zipei, Song Xuan, Tsubouchi Kota, Shibasaki Ryosuke; Feb 2022; Will you go where you search? A deep learning framework for estimating user search-and-go behavior; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S092523122031496X
808. *Tung et al.* [2019] - Vuong Tung, Saastamoinen Miamaria, Jacucci Giulio, Ruotsalo Tuukka; Nov 2019; Understanding user behavior in naturalistic information search tasks; https://web-p-ebscohost-com.ressources-electroniques.univ-lille.fr/ehost/pdfviewer/pdfviewer?vid=0&sid=7bbe630d-e187-4a74-986c-b9723ad31d38%40redis
809. *Schultheib & Lewandowski* [2021]-2 - Schultheib Sebastian, Lewandowski Dirk; Mar 2021; How users' knowledge of advertisements influences their viewing and selection behavior in search engines; https://web-p-ebscohost-

com.ressources-electroniques.univ-lille.fr/ehost/pdfviewer/pdfviewer?vid=0&sid=eca879ed-f030-420f-8678-3bebc80f1dac%40redis

810. *Makeev et al.* [2014] - Stanislav Makeev, Andrey Plakhov, and Pavel Serdyukov; 13-16 Apr 2014; Personalizing Aggregated Search, p 197-209; Advances in Information Retrieval Book; 36th European Conference on IR Research, ECIR 2014, Amsterdam; https://link-springer-com.ressources-electroniques.univ-lille.fr/book/10.1007/978-3-319-06028-6
811. *Medlar & Glowacka* [2020] - Medlar Alan, Głowacka Dorota; Feb 2020; Exploratory search: user behaviour and search engine adaptation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3352683.3352687
812. *Eldad & Elad* [2021] - Yechiam Eldad, Yom-Tov Elad; Jul 2021; Unique Internet Search Strategies of Individuals With Self-Stated Autism: Quantitative Analysis of Search Engine Users' Investigative Behaviors; Journal of medical Internet research, Vol.23 (7); https://www.readcube.com/articles/10.2196/23829
813. *Ashkan & Clarke* [2012] - Ashkan Azin; Clarke Charles L. A.; 29 Mar 2012; Impact of query intent and search context on clickthrough behavior in sponsored search; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s10115-012-0485-x
814. *Singer et al.* [2012] - Georg Singer, U. Norbisrath, D. Lewandowski; 7 Jun 2012; Ordinary search engine users carrying out complex search tasks; https://www.researchgate.net/publication/225279577_Ordinary_Search_Engine_Users_Carrying_Out_Complex_Search_Tasks
815. *Thomas et al.* [2014] - Paul Thomas, Alistair Moffat, P. Bailey, Falk Scholer; 26 Aug 2014; Modeling decision points in user search behavior; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/2637002.2637032
816. *Zerhoudi et al.* [2021] - Saber Zerhoudi, M. Granitzer, Jörg Schlötterer, C. Seifert; 13 Dec 2021; Query Change as a Contextual Markov Model for Simulating User Search Behaviour; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/3503162.3503165
817. *Yamin & Ramayah* [2013] - Fadhilah Mat Yamin, T. Ramayah; 2013; INTERFACING GOOGLE SEARCH ENGINE TO CAPTURE USER WEB SEARCH BEHAVIOR; https://academic-pub.org/ojs/index.php/ijecs/article/view/1111
818. *Priyadharshini et al.* [2018] - L. Priyadharshini, S. Florence, K. Prema, C. Kumari; 5 Feb 2018; Predicting implicit search behaviors using log analysis; https://www.sciencepubco.com/index.php/ijet/article/view/9582/3314
819. *Kazutoshi* [2016] - Kazutoshi Umemoto; 23 Mar 2016; A Study on Fine-Grained User Behavior Analysis in Web Search; https://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/215679/5/djohk00603.pdf
820. *Jiang et al.* [2014] - Jiepu Jiang, Daqing He, J. Allan; 3 Jul 2014; Searching, browsing, and clicking in a search session: changes in user behavior by task and over time; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2600428.2609633

821. *Kori et al.* [2015] - Shogo Kori, Yanjun Zhu, Koichi Yamaguchi, Satoru Takiguchi, Y. Takama; 1 Nov 2015; Analysis of user's behaviour based on search intentions for information retrieval using search engines; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=7407098
822. *Nasution* [2018] - M K M Nasution; 2018; Semantic interpretation of search engine resultant; https://iopscience.iop.org/article/10.1088/1757-899X/300/1/012053/pdf
823. *Yamin et al.* [2014] - Fadhilah Mat Yamin, T. Ramayah, Wan Hussain Wan Ishak, Siti Norezam Othman; 2014; Search Log Analysis Method to Uncover User Search Behaviour on Web Searching Environment; https://core.ac.uk/download/42979793.pdf
824. *Kopeinik et al.* [2023] - Simone Kopeinik, Martina Mara, Linda Ratz, Klara Krieg, Markus Schedl, Navid Rekabsaz; Apr 2023; Show me a "Male Nurse" How Gender Bias is Reflected in the Query Formulation of Search Engine Users; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3544548.3580863
825. *Shen & Martin* [2010] - Yun Shen, Trevor Martin; 2010; Intelligent User Search Behaviour Knowledge Discovery; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=5584867
826. *Zhang & Moffat* [2006] - Yuye Zhang, Alistair Moffat; 2006; Some Observations on User Search Behavior; https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=27e4dcd17c6af73477db3c909007fd1a5c2e783e
827. *Wolfram* [2013] - Dietmar Wolfram; 15 Oct 2013; A Query-Level Examination of End User Searching Behaviour on the Excite Search Engine; https://journals.library.ualberta.ca/ojs.cais-acsi.ca/index.php/cais-asci/article/view/30/26
828. *Jose & Lal* [2013] - Jeeva Jose, P. Sojan Lal; Jun 2013; Analysis of the Temporal Behaviour of Search Engine Crawlers at Web Sites; https://paper.researchbib.com/view/paper/4831
829. *Murtuza et al.* [2010] - Murtuza Kutub, R. Prachetaa, Mangesh Bedekar; 2010; User Web Search Behaviour; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=5698386
830. *Kaushik & Jones* [2021] - Abhishek Kaushik, Gareth J. F. Jones; 9 Apr 2021; Exploring Current User Web Search Behaviours in Analysis Tasks to be Supported in Conversational Search; https://arxiv.org/pdf/2104.04501.pdf
831. *Neelima & Rodda* [2016] - G. Neelima; Sireesha Rodda; 2016; Predicting user behavior through sessions using the web log mining; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=7449167
832. *Sebastian* [2019] - Dungs Sebastian; 13 Mar 2019; Modelling temporal patterns in user behaviour; https://duepublico2.uni-due.de/servlets/MCRFileNodeServlet/duepublico_derivate_00070044/Diss_Dungs.pdf
833. *Marchionini* [2006] - Marchionini G.; 2006; Exploratory search: from finding to understanding; Communications of the ACM 49, 4, 41–46; https://www.researchgate.net/publication/220422328_Marchionini_G_Exploratory_search_from_finding_to_understanding_Comm_ACM_494_41-46
834. *Zhang-1 et al.* [2021] - Fan Zhang, Yiqun Liu, Jiaxin Mao, Min Zhang, Shaoping Ma; 21 Feb 2021; User behavior modeling for Web search evaluation; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S2666651021000048

835. *Yao-2 et al.* [2021] - Jing Yao, Zhicheng Dou, Ruobing Xie, Yanxiong Lu, Zhiping Wang and Ji-Rong Wen; 30 Sep 2021; USER: A Unified Information Search and Recommendation Model based on Integrated Behavior Sequence; https://arxiv.org/pdf/2109.15012.pdf
836. *Khamsum et al.* [2014] - Kinley Khamsum, Tjondronegoro Dian, Partridge Helen, Edwards Sylvia; 2014; Modeling users' web search behavior and their cognitive styles; https://research-repository.griffith.edu.au/bitstream/handle/10072/62985/96431_1.pdf?sequence=1
837. *Hendahewa & Shah* [2015] - Chathra Hendahewa, Chirag Shah; Sep 2015; Implicit search feature based approach to assist users in exploratory search tasks; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0306457315000795
838. *Kucukyilmaz* [2021] - Tayfun Kucukyilmaz; May 2021; Exploiting temporal changes in query submission behavior for improving the search engine result cache performance; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0306457321000418
839. *Bai et al.* [2017] - XIAO BAI, IOANNIS ARAPAKIS, B. BARLA CAMBAZOGLU, ANA FREIRE; Mar 2017; Understanding and Leveraging the Impact of Response Latency on User Behaviour in Web Search; https://iarapakis.github.io/papers/TOIS17.pdf
840. *Chen et al.* [2021] - Jia Chen, Jiaxin Mao, Yiqun Liu, Fan Zhang, Min Zhang, Shaoping Ma; Apr 2021; Towards a Better Understanding of Query Reformulation Behavior in Web Search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/3442381.3450127
841. *Brian Dean* [2025] - Brian Dean; 18 Aug 2023; How People Use Google Search (New User Behavior Study); https://backlinko.com/google-user-behavior
842. *Zhitomirsky et al.* [2016] - Maayan Zhitomirsky-Geffet, Judit Bar-Ilan, Mark Levene; 18 Jul 2016; Testing the stability of "wisdom of crowds" judgments of search results over time and their similarity with the search engine rankings; https://www.researchgate.net/publication/303821723_Testing_the_stability_of_wisdom_of_crowds_judgments_of_search_results_over_time_and_their_similarity_with_the_search_engine_rankings
843. *Pawitra* [2015] - Chiravirakul Pawitra; 2015; Search Satisfaction - Choice Overload, Variety Seeking and Serendipity in Search Engine Use; https://purehost.bath.ac.uk/ws/portalfiles/portal/187961401/PawitraThesis_Hardbound.pdf
844. *Arora & Govilkar* [2016] - Nirali Arora, Sharvari Govilkar; 2016; Efficient Hybrid Ranking Algorithm for Search Engine; https://www.ijcsit.com/docs/Volume%207/vol7issue3/ijcsit20160703125.pdf
845. *Jonsson & Stahl* [2015] - G Jonsson, L Ståhl; 2015; Users' Perception and Evaluation of a Search Engine Result Page; https://www.diva-portal.org/smash/get/diva2:929087/FULLTEXT01.pdf
846. *Salehi et al.* [2020] - Sarah Salehi, Jia DU, Helen Ashman; Jan 2020; Personalized web search and user satisfaction: a user centred evaluation; https://www.researchgate.net/publication/347699307_Personalised_Web_search_and_user_satisfaction_a_user-centred_evaluation

847. *Alsaffar et al.* [2018] - Alsaffar Alsaffar, Lyn Pemberton, Karina Rodriguez Echavarria; Visual Search Behaviour amongst Arabic, English and Chinese Search Engine Users; https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2018.134

848. *Reyes-Menendez et al.* [2018] - Ana Reyes-Menendez, Jose Ramon Saura, Pedro R. Palos-Sanchez, Jose Alvarez-Garcia; 18 Sep 2018; Understanding User Behavioral Intention to Adopt a Search Engine that Promotes Sustainable Water Management; https://www.mdpi.com/2073-8994/10/11/584

849. *Su et al.* [2018] - Ning Su, Jiyin He, Yiqun Liu, Min Zhang, Shaoping Ma; Feb 2018; User Intent, Behaviour, and Perceived Satisfaction in Product Search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3159652.3159714

850. *Tsai & Wu* [2017] - Ming-Fong Tsai, Yi-Hong Wu; 17 Oct 2022; User intent prediction search engine system based on query analysis and image recognition technologies; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s11227-022-04874-w

851. *Lambert & Loiselle* [2007] - S. Lambert, C. Loiselle; 01 Oct 2007; Health Information—Seeking Behavior; https://www.researchgate.net/publication/5917676_Health_Information-Seeking_Behavior

852. *Jacobs et al.* [2017] - Wura Jacobs, Ann O. Amuta, Kwon Chan Jeon; 13 Mar 2017; Health information seeking in the digital age: An analysis of health information seeking behavior among US adults; https://www.tandfonline.com/doi/epdf/10.1080/23311886.2017.1302785?needAccess=true

853. *Zimmerman & Shaw* [2020] - Margaret S. Zimmerman, George Shaw Jr; 12 Feb 2020; Health information seeking behaviour: a concept analysis; https://onlinelibrary.wiley.com/doi/epdf/10.1111/hir.12287

854. *Jia-2 et al.* [2021] - Xiaoyun Jia, Yan Pang, Liangni Sally Liu; Dec 2021; Online Health Information Seeking Behavior: A Systematic Review; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8701665/pdf/healthcare-09-01740.pdf

855. *Mirzaei et al.* [2021] - Ardalan Mirzaei, Parisa Aslani, Edward Joseph Luca, Carl Richard Schneider; Jul 2021; Predictors of Health Information–Seeking Behavior: Systematic Literature Review and Network Analysis; https://www.jmir.org/2021/7/e21680/PDF

856. *Chang & Huang* [2020] - Cheng-Ching Chang, Mu-Hsuan Huang; Oct 2020; Antecedents predicting health information seeking: A systematic review and meta-analysis; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0268401219301082

857. *Schoenherr* [2022] - Jordan Richard Schoenherr; 2022; Vaccination, Values, and Health Information Search Behaviour in the New Folk Medicine; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/document/10227112

858. *Kim et al.* [2023] - Lyounghee Lian Kim, Yangsun Hong, Sumaira Abrar, C. P. Fitzgerald; 2023; Relationships Between Social Media Use, Exposure to Vaccine Misinformation and Online Health Information Seeking Behaviour; https://journals.sagepub.com/doi/abs/10.1177/09732586231166111

859. *Pang et al.* [2014] - Patrick Cheong-Iao Pang, Patrick Cheong-Iao Pang, Kathleen Gray, Shanton Chang, Shanton Chang, Jon Pearce, Jon Pearce, Karin Verspoor, Karin Verspoor; 2014; Online Health Information Seeking Behaviour: Understanding Different Search Approaches; https://core.ac.uk/download/pdf/301362845.pdf
860. *Song et al.* [2022] - Qiurong Song, Qiurong Song, Jiepu Jiang, Jiepu Jiang; 2022; How Misinformation Density Affects Health Information Search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3485447.3512141
861. *Kerstin Horch* [2021] - Kerstin Horch; 2021; Searching for health information on the Internet - Results from the KomPaS study; https://www.rki.de/EN/Content/Health_Monitoring/Health_Reporting/GBEDownloadsJ/FactSheets_en/JoHM_02_2021_Searching_for_information_KomPaS.pdf?__blob=publicationFile
862. *Neeragatti et al.* [2023] - Surya Neeragatti, Ranjit Kumar Dehury, Nishchala Sripathi; 2023; Determinants of Digital Health Information Search (DHIS) Behaviour: Extending UTAUT with healthcare behaviour constructs; https://journal.achsm.org.au/index.php/achsm/article/view/1685/993
863. *Liu & Zou* [2023] - Yan Liu, Peng Zou; 2023; Psychological Effect of Discrete Outbreak Events of COVID-19 on Health Information Search in China; https://pubmed.ncbi.nlm.nih.gov/36829338/
864. *Weimann-Saks* [2022] - Dana Weimann-Saks; Dec 2022; Young adults searching for Health-Related Information on the Internet; https://www.researchgate.net/publication/366612139_Young_adults_searching_for_Health-Related_Information_on_the_Internet
865. *Schoenherr & Thomson* [2020] - Jordan Richard Schoenherr, Robert Thomson; 2020; Health Information Seeking Behaviour, Risk Communication, and Mobility During COVID-19; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/stamp/stamp.jsp?tp=&arnumber=9462203
866. *Wikipedia-HBM –* Health Belief Model; https://en.wikipedia.org/wiki/Health_belief_model#cite_ref-1
867. *Rosenstock* [1974] - I. M. Rosenstock; 1974; The Health Belief Model and Preventive Health Behavior; Health Education Monographs, vol. 2, p. 354–386; https://www-jstor-org.ressources-electroniques.univ-lille.fr/stable/45240623
868. *Rusova et al.* [2022] - Jitka Rusova, Adela Michkova, Ondrej Pleskot, Ondrej Podeszwa, Eva Hlavackova; 2022; The Importance of Selected Criteria for Assessing the Credibility of Information on the Internet from the Perspective of Users Searching Health and Disease Information Websites; https://jecs.pl/index.php/jecs/article/view/1456
869. *Paunic et al.* [2023] - Milica Paunic, Claire Rim, Olla Hilal, Renee Nassar, Farwa Zaib, Kayla Touma, Mahmoud Hossami, Rhonda Abdel-Nabi, Roaa Hirmiz, Caroline Hamm; 2023; Engine failures: A critical analysis of current clinical trial websites' search engines; https://ascopubs.org/doi/pdf/10.1200/JCO.2023.41.16_suppl.e18537
870. *Liu & Xu* [2020] - Zhengyang Liu, Chenchen Xu; 2020; A Large Scale Study on Health Information Retrieval for Laypersons; https://www.semanticscholar.org/paper/A-Large-Scale-Study-on-Health-Information-Retrieval-Liu-Xu/314f7908d6a636e0e115668580e084dfd3797eb1

871. *Chiu et al.* [2022] - Yen-Lin Chiu, Chin-Chung Tsai, Jyh-Chong Liang; 02 Sep 2022; Laypeople's Online Health Information Search Strategies and Use for Health-Related Problems: Cross-sectional Survey; https://www.jmir.org/2022/9/e29609/PDF
872. *Htay et al.* [2022] - Mila Nu Nu Htay, Laurence Lloyd Parial, Ma. Carmen Tolabing, Kevin Dadaczynski, Orkan Okan, Angela Yee Man Leung, Tin Tin Su; Apr 2022; Digital health literacy, online information-seeking behaviour, and satisfaction of Covid-19 information among the university students of East and South-East Asia; https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0266276
873. *Ni et al.* [2024] - Zijun Ni, Lingli Zhu, Shuyan Li, Yuping Zhang, Ruiyi Zhao; Jan 2024; Characteristics and associated factors of health information-seeking behaviour among patients with inflammatory bowel disease in the digital era: a scoping review; https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-024-17758-w
874. *Ghezzi et al.* [2019] - Pietro Ghezzi, Peter G Bannister, Gonzalo Casino, Alessia Catalani, Michel Goldman, Jessica Morley, Marie Neunez, Andreu Prados, Mariarosaria Taddeo, Tania Vanzolini, Luciano Floridi; 02 Dec 2019; Online information of vaccines: information quality is an ethical responsibility of search engines; https://arxiv.org/ftp/arxiv/papers/1912/1912.00898.pdf
875. *Yan Zhang* [2013] - Yan Zhang; 2013; Consumer health information searching process in real life settings; https://asistdl-onlinelibrary-wiley-com.ressources-electroniques.univ-lille.fr/doi/pdf/10.1002/meet.14504901046
876. *Alfano et al.* [2020] - Marco Alfano, Biagio Lenzitti, Davide Taibi, Markus Helfert; 2020; Language Complexity in On-line Health Information Retrieval; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-030-52677-1_5
877. *Tuliao et al.* [2022] - Antover P. Tuliao, Natira Mullet, Lindsey G. Hawkins, Derek Holyoak, Marisa Weerts, Anthony Inyang; 13 Jan 2022; Searching for Mental Health Services: Search Strings and Information Acquisition; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s41347-021-00238-y
878. *Zhang-2 et al.* [2021] - Di Zhang, Shi Zhen, Hongchao Hu, Gang Kevin Han; Mar 2021; Classification of the Use of Online Health Information Channels and Variation in Motivations for Channel Selection: Cross-sectional Survey; https://www.jmir.org/2021/3/e24945/PDF
879. *Kim et al.* [2019] - Sung Un Kim, Ivana Martinovic, Snjezana Stanarevic Katavic; 22 Oct 2019; The use of mobile devices and applications for health information: A survey of Croatian students; https://journals-sagepub-com.ressources-electroniques.univ-lille.fr/doi/epub/10.1177/0961000619880937
880. *Li-2 et al.* [2023] - Zhiwei Li,Yinguang Fan, Hong Su, Zhiwei Xu, Hung Chak Ho, Hao Zheng, Junwen Tao, Yunquan Zhang, Kejia Hu, Mohammad Zahid Hossain, Qi Zhao, Cunrui Huang, Jian Cheng; 2023; The 2022 Summer record-breaking heatwave and health information-seeking behaviours: an infodemiology study in Mainland China; https://gh.bmj.com/content/bmjgh/8/9/e013231.full.pdf
881. *Every Author* [2023] - Every Author; 05 May 2023; 7 Types of Research Gaps in Literature Review; https://everyauthorpub.com/every-author-blog/f/7-types-of-research-gaps-in-literature-review
882. *Amanda Bolderston* [2008] - Amanda Bolderston; 2008; Writing an Effective Literature Review; https://www.academia.edu/20417534/Writing_an_Effective_Literature_Review

883. *Hamilton & Clare* [2004] - Helen Hamilton, Judith Clare; 2004; Reviewing the literature: making 'the literature' work for you; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S1322769608604370
884. *Ahn et al.* [2018] - Jae-Hyeon Ahn, Yoon-Soo Bae, Jaehyeon Ju, Wonseok Oh; Oct 2018; Attention Adjustment, Renewal, and Equilibrium Seeking in Online Search: An Eye-Tracking Approach; https://www.researchgate.net/publication/329743074_Attention_Adjustment_Renewal_and_Equilibrium_Seeking_in_Online_Search_An_Eye-Tracking_Approach
885. *Arguello* [2014] - Jaime Arguello, 13-16 Apr 2014; Predicting Search Task Difficulty; Advances in Information Retrieval Book, p 88-99; 36th European Conference on IR Research, ECIR 2014, Amsterdam; https://link-springer-com.ressources-electroniques.univ-lille.fr/book/10.1007/978-3-319-06028-6
886. *Croft et al.* [2010] - W. Bruce Croft, Donald Metzler, Trevor Strohman; 2010; Search Engines: Information Retrieval in Practice - Book; https://ciir.cs.umass.edu/downloads/SEIRiP.pdf
887. *Adita Joshi* [2022] - Adita Joshi; 10 Jan 2022; How to Choose the Right Research Methodology; https://researcher.life/blog/article/choose-research-methodology/
888. *Uttkarsha Bhosale* [2023] - Uttkarsha Bhosale; 7 Jul 2023; How to Choose Best Research Methodology for Your Study; https://www.enago.com/academy/choose-best-research-methodology/
889. *Kolkova* [2023] - Kateryna Kolkova; 10 Apr 2023; Research Methods vs Research Methodologies | Know the Difference; https://www.youtube.com/watch?v=BH3j2hFk0aY
890. *Newcastle* [2024] - 6 Mar 2024; Research Methods: What are research methods?; https://libguides.newcastle.edu.au/researchmethods
891. *MvsM-bizmanualz.com* - Methods Vs Methodology; https://www.bizmanualz.com/write-better-procedures/methods-vs-methodology.html
892. *Jarvelin & Kekalainen* [2002] - Kalervo Järvelin, Jaana Kekäläinen; 01 Oct 2002; Cumulated gain-based (CGB - DCG) evaluation of IR techniques; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/582415.582418
893. *Moffat & Zobel* [2008] - ALISTAIR MOFFAT, JUSTIN ZOBEL; Dec 2008; Rank-Biased Precision (RBP) for Measurement of Retrieval Effectiveness; https://people.eng.unimelb.edu.au/jzobel/fulltext/acmtois08.pdf
894. *Chapelle et al.* [2009] - Olivier Chapelle, Donald Metzler, Ya Zhang, Pierre Grinspan; Nov. 2009; Expected Reciprocal Rank (ERR) for Graded Relevance; https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7e3cf6492128f915112ca01dcb77c766129e65cb
895. *Chapelle et al.* [2011] - Olivier Chapelle, Shihao Ji, Ciya Liao, Emre Velipasaoglu, Larry Lai, Su-Lin Wu; 28 Apr 2011; Intent-based diversification of web search results: metrics and algorithms; https://titan.cs.gsu.edu/~sji/papers/err_ia.pdf
896. *Evangelopoulos et al.* [2016] - Xenophon Evangelopoulos, Victor Giannakouris-Salalidis, Lazaros Iliadis, Christos Makris, Yannis Plegas, Antonia Plerou, Spyros Sioutas; Jan 2016; Evaluating information retrieval using document

popularity: An implementation on MapReduce; Volume-51, P 16-23; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0952197616000270

897. *Yilmaz et al.* [2010] - Emine Yilmaz, Milad Shokouhi, Nick Craswell, Stephen Robertson; 26 Oct 2010; Expected Browsing Utility (EBU) for Web Search Evaluation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1871437.1871672
898. *Carterette* [2011] - Ben Carterette; July 2011; System Effectiveness, User Models, and User Utility: A Conceptual Framework for Investigation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2009916.2010037
899. *Smucker & Clarke-1* [2012] - Mark D. Smucker, Charles L. A. Clarke; Oct 2012; Modeling User Variance in Time-Biased Gain (TBG); https://plg.uwaterloo.ca/~claclark/smucker-clarke-hcir2012-user-variance.pdf
900. *Smucker & Clarke-2* [2012] - Mark D. Smucker, Charles L.A. Clarke; Aug 2012; Time-based calibration of effectiveness measures; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2348283.2348300
901. *Smucker & Clarke-3* [2012] - Mark D. Smucker, Charles L.A. Clarke; Oct 2012; Stochastic simulation of time-biased gain; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2396761.2398568
902. *Moffat et al.* [2012] - Alistair Moffat, Falk Scholer, Paul Thomas; Dec 2012; INSQ & SDCG - Models and Metrics: IR Evaluation as a User Process; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2407085.2407092
903. *Moffat et al.* [2015] - Alistair Moffat, Falk Scholer, Paul Thomas, Peter Bailey; Dec 2015; INST: An Adaptive Metric for Information Retrieval Evaluation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2838931.2838938
904. *Moffat & Alfan* [2020] - Alistair Moffat, Alfan Farizki Wicaksono; Feb 2020; Metrics, User Models, and Satisfaction; https://www.academia.edu/73832267/Metrics_User_Models_and_Satisfaction
905. *Moffat et al.* [2016] - Alistair Moffat, Falk Scholer, Paul Thomas, Peter Bailey; 2016; Incorporating User Expectations and Behavior into the Measurement of Search Effectiveness; https://people.eng.unimelb.edu.au/ammoffat/abstracts/mbst17acmtois.pdf
906. *Wicaksono* [2020] - Alfan F. Wicaksono; Dec 2020; Modelling Search Session; https://ir.cs.ui.ac.id/new/talks/slides/ir_nlp_talk_2_alfan2.pdf
907. *Chuklin et al.* [2013] - Aleksandr Chuklin, Pavel Serdyukov, Maarten de Rijke; 2013; Click Model-Based Information Retrieval Metrics (CMBM); https://www.researchgate.net/profile/Aleksandr-Chuklin/publication/309180427_slides_Click_model-based_information_retrieval_metrics/links/580350b908ae310e0d9df4c1/
908. *Chuklin et al.-2* [2015] - Aleksandr Chuklin, Ilya Markov, Maarten de Rijke; Click Models for Web Search - Book; https://clickmodels.weebly.com/uploads/5/2/2/5/52257029/mc2015-clickmodels.pdf
909. *Moffat et al.* [2019] - Leif Azzopardi, Paul Thomas, Alistair Moffat; Jul 2019; cwl_eval: An Evaluation Tool for Information Retrieval; https://www.microsoft.com/en-us/research/uploads/prod/2019/06/CWL_Eval_Sigir_Demo_Paper.pdf

910. *Moffat et al.* [2013] - Falk Scholer, Paul Thomas, Alistair Moffat; Oct 2013; Users Versus Models: What Observation Tells Us About Effectiveness Metrics; https://www.khoury.northeastern.edu/home/vip/teach/IRcourse/5_eval_userstudy/other_notes/users_vs_models_effectiveness_metrics.pdf
911. *Jiang & Allan* [2016] - Jiepu Jiang and James Allan; Mar 2016; Adaptive Effort for Search Evaluation Metrics; https://maroo.cs.umass.edu/getpdf.php?id=1201
912. *Chuklin & Rjike* [2016] - Aleksandr Chuklin, Maarten de Rijke; Oct 2016; Incorporating Clicks, Attention and Satisfaction (CAS) into a Search Engine Result Page Evaluation Model; https://www.researchgate.net/publication/307543608_Incorporating_Clicks_Attention_and_Satisfaction_into_a_Search_Engine_Result_Page_Evaluation_Model
913. *Luo et al.* [2017] - Cheng Luo, Yiqun Liu, Tetsuya Sakai, Fan Zhang, Min Zhang, Shaoping Ma; Aug 2017; Evaluating Mobile Search with Height-Biased Gain (HBG); http://www.thuir.cn/group/~mzhang/publications/sigir2017Luo-HBG.pdf
914. *Zhang et al.* [2017] - Fan Zhang, Yiqun Liu, Xin Li, Min Zhang, Yinghui Xu, Shaoping Ma; Aug 2017; Evaluating Web Search with a Bejeweled Player Model (BPM); https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3077136.3080841
915. *Jiang & Allan* [2017] - Jiepu Jiang, James Allan; Nov 2017; Adaptive Persistence for Search Effectiveness Measures; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3132847.3133033
916. *Thomas et al.* [2018] - Paul Thomas, Alistair Moffat, Peter Bailey, Falk Scholer, Nick Craswell; Dec 2018; Better effectiveness metrics for serps, cards, and rankings; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3291992.3292002
917. *Sakai & Dou* [2013] - Tetsuya Sakai, Zhicheng Dou; Jul 2013; Summaries, ranked retrieval and sessions: a unified framework for information access evaluation (U-measure); https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2484028.2484031
918. *Ferrante et al.* [2014] - Marco Ferrante, Nicola Ferro, Maria Maistro; Jul 2014; Injecting user models and time into precision via Markov chains; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/2600428.2609637
919. *Azzopardi et al.* [2020] - Leif Azzopardi, Ryen W. White, Paul Thomas, Nick Craswell; Mar 2020; Data-Driven Evaluation Metrics for Heterogeneous Search Engine Result Pages; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3343413.3377959
920. *Ferrante & Ferro* [2020] - Marco Ferrante, Nicola Ferro; Sep 2020; Exploiting Stopping Time to Evaluate Accumulated Relevance; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3409256.3409832
921. *Robertson* [2008] - Stephan Robertson; Jul 2008; A new interpretation of average precision; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1390334.1390453

922. *Clarke et al.* [2008] - Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, Ian MacKinnon; Jul 2008; Novelty and diversity in information retrieval evaluation; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1390334.1390446
923. *Agrawal et al.* [2009] - Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, Samuel Ieong; Feb 2009; Diversifying search results; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1498759.1498766
924. *Craswell et al.* [2008] - Nick Craswell, Onno Zoeter, Michael Taylor, Bill Ramsey; Feb 2008; An experimental comparison of click position-bias models; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1341531.1341545
925. *Sakai & Robertson* [2008] - Tetsuya Sakai, Stephen Robertson; Dec 2008; Modelling A User Population for Designing Information Retrieval Metrics; https://www.researchgate.net/profile/Stephen-Robertson-11/publication/228568551_Modelling_a_user_population_for_designing_information_retrieval_metrics/links/00b495346c7bef323f000000/Modelling-a-user-population-for-designing-information-retrieval-metrics.pdf
926. *Chapelle & Zhang* [2009] - Olivier Chapelle, Ya Zhang; Apr 2009; A dynamic bayesian network click model for web search ranking; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1526709.1526711
927. *Dupret & Piwowarski* [2008] - Georges E. Dupret, Benjamin Piwowarski; Jul 2008: A user browsing model to predict search engine click data from past observations; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1390334.1390392
928. *Guo et al.* [2009] - Fan Guo, Chao Liu, Yi Min Wang; Feb 2009; Efficient multiple-click models in web search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1498759.1498818
929. Li et al. [2009] - Jane Li, Scott Huffman, Akihito Tokuda; Jul 2009; Good abandonment in mobile and PC internet search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1571941.1571951
930. *Eric Charnov* [1976] - Eric L. Charnov; Apr 1976; Optimal Foraging, the Marginal Value Theorem; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/004058097690040X
931. *Pirolli & Card* [1999] - Peter Pirolli, Stuart K. Card; Oct 1999; Information foraging; https://www.researchgate.net/publication/229101074_Information_Foraging
932. *Diaz et al.* [2020] - Fernando Diaz, Bhaskar Mitra, Michael D. Ekstrand, Asia J. Biega, Ben Carterette; Oct 2020; Evaluating Stochastic Rankings with Expected Exposure; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3340531.3411962
933. *Breuer et al.* [2021] - Timo Breuer, Nicola Ferro, Maria Maistro, Philipp Schaer; Jan 2021; repro_eval: A Python Interface to Reproducibility Measures of System-oriented IR Experiments; https://www.researchgate.net/publication/348659215_repro_eval_A_Python_Interface_to_Reproducibility_Measures_of_System-oriented_IR_Experiments
934. *Giannakouris et al.* [2014] - Giannakouris Salalidis Victor, Plerou Antonia, Sioutas Spyros; Sep 2014; CSMR: A Scalable Algorithm for Text Clustering with Cosine Similarity and MapReduce;

https://www.researchgate.net/publication/268925574_CSMR_A_Scalable_Algorithm_for_Text_Clustering_with_Cosine_Similarity_and_MapReduce

935. *Xenophon et al.* [2014] - Xenophon Evangelopoulos, Christos Makris, Yannis Plegas; 2014; Reciprocal Rank Using Web Page Popularity; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-662-44722-2_13
936. *Moher et al.* [2009] - David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman; 21 Jul 2009; Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement; PRISMA Group: https://www.bmj.com/content/339/bmj.b2535
937. *Charlotte Strange* [2022] - Charlotte Strange; 8 Dec 2022; A reflection on choosing a thesis topic; https://edpsy.org.uk/blog/2022/a-reflection-on-choosing-a-thesis-topic/
938. *Reid* [2023] - Elizabeth Reid; 10 May 2023; Supercharging Search with generative AI; https://blog.google/products/search/generative-ai-search/
939. *A. Newell* [1963] - Allen Newell; Feb 1963; A guide to the general problem solver program gps-2-2 book; https://stacks.stanford.edu/file/druid:zk239tp3547/zk239tp3547.pdf
940. *Webdesignmuseum* – Google in 1998; https://www.webdesignmuseum.org/gallery/google-1998
941. *Clicknathan* [2014] - Google's Search Results Explained; https://clicknathan.com/web-design/googles-search-results-explained/
942. *Clifford* [2017] - How Google Ranks First Page Results and How You can Get There; https://www.locallogy.com/google-first-page
943. *Google-Support-2021* – Indexed but not showing in search results; https://support.google.com/webmasters/thread/115232862/indexed-but-not-showing-in-search-results?hl=en
944. *Danny Sullivan* Google [2023] - 25 biggest moments in Search, from helpful images to AI; https://blog.google/products/search/25-biggest-google-search-updates/
945. *Google Press* [2004] - Google Announces Desktop Search; https://googlepress.blogspot.com/2004/10/google-announces-desktop-search.html
946. *Google Search Central* - Introducing Rich Snippets; https://developers.google.com/search/blog/2009/05/introducing-rich-snippets
947. *similarweb-zero-click* - Zero-Click Searches; https://support.similarweb.com/hc/en-us/articles/360006488277-Zero-Click-Searches
948. *Google Timeline Wiki* - Timeline of Google Search; https://en.wikipedia.org/wiki/Timeline_of_Google_Search
949. *Roger Montti* [2022] - Why You Should Use Boolean Search Operators For SEO; https://www.searchenginejournal.com/how-to-use-boolean-search-operators-seo-application/7822/
950. *Eran Fisher* [2014] - Eran Fisher; 2015; BOOK: Internet and Emotions; https://www.academia.edu/8238627/BOOK_Internet_and_Emotions

951. *Miller & Moses* [2023] - Samuel Miller and Elijah Matthew Moses; 12 Jul 2023; Journey Through the Evolution of the Web: From Web 1.0 to Web 6.0 and Beyond; https://medium.com/@efficio/journey-through-the-evolution-of-the-web-from-web-1-0-to-web-6-0-and-beyond-7d87661ddf2a
952. *Yuan et al.* [2020] - Bei Yuan, Minhong Wang, Jeroen J. G. van Merriënboer, Xu Tao, Andre Kushniruk, Jun Peng; 01 Apr 2020; Investigating the Role of Cognitive Feedback in Practice-Oriented Learning for Clinical Diagnostics; https://typeset.io/papers/investigating-the-role-of-cognitive-feedback-in-practice-18uog1219g
953. *Alnatsheh* [2020] - Essam Alnatsheh; 19 Nov 2020: User Behavior for Neural Network-based Web Search Results Filtering; https://typeset.io/papers/user-behavior-for-neural-network-based-web-search-results-4rldqi2cow
954. *Xin-Fu* [2010] - Xin Fu; 01 Jan 2010; Towards a model of implicit feedback for Web search; https://onlinelibrary-wiley-com.ressources-electroniques.univ-lille.fr/doi/pdf/10.1002/asi.21198
955. *Gwizdka* [2009] - Jacek Gwizdka; 15 Dec 2009; Assessing Cognitive Load on Web Search Tasks; https://arxiv.org/ftp/arxiv/papers/1001/1001.1685.pdf
956. *Peterson et al.* [2022] - Ashley M. Peterson, Andrew C. High, Raven Maragh-Lloyd, Ryan G. Stoldt, Brian Ekdale; 20 Oct 2022; Trust in Online Search Results During Uncertain Times; https://www.researchgate.net/publication/365153516_Trust_in_Online_Search_Results_During_Uncertain_Times
957. *Mana* [2014] - Jordi Torras Mañá; 21 Nov 2014; Semantic search engine; https://patents.google.com/patent/US20160147878A1/en
958. *Zhao* [2008] - Zhao Jun-qin; 01 Jan 2008; Research on Intelligent Search Engine Based on Semantic Comprehension; https://typeset.io/papers/research-on-intelligent-search-engine-based-on-semantic-ziiyx7j1th
959. *Sembok et al.* [2012] - Tengku Mohd Tengku Sembok, Mohamad Fauzan Noordin, Roslina Othman, Rabiah Abdul Kadir; 01 Jan 2012; Logical-linguistic semantic in search engine; https://typeset.io/papers/logical-linguistic-semantic-in-search-engine-5cf9aufyku
960. *Santha & Jayakumar* [2019] - A. C. Santha Sheela, C. Jayakumar; 14 Mar 2019; Comparative Study of Syntactic Search Engine and Semantic Search Engine: A Survey; https://typeset.io/papers/comparative-study-of-syntactic-search-engine-and-semantic-1gou4qfyjp
961. *Dziak et al. GP* [2015] - Scott M. Dziak, Haitao Xia, Lu Lu; 30 Jun 2015; Loop adaptation control with pattern detection; https://patents.google.com/patent/US9553739B1/en
962. *Rahman & Seyal* [2014] - Mohd Noah A. Rahman, Siti Aminah Maidin Afzaal H. Seyal; 01 Jan 2014; Search Engine Development: Adaptation from Supervised Learning Methodology; https://www.academia.edu/8371950/Search_Engine_Development_Adaptation_from_Supervised_Learning_Methodology
963. *Lv & Zhai* [2009] - Yuanhua Lv, ChengXiang Zhai; Adaptive relevance feedback in information retrieval; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1645953.1645988
964. *Joachims & Radlinski* [2007] - Thorsten Joachims, Filip Radlinski; Search Engines that Learn from Implicit Feedback; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/document/4292009

965. *Bendersky et al.* [2018] - Michael Bendersky, Xuanhui Wang, Marc Najork, Donald Metzler; 01 Jul 2018; Learning with Sparse and Biased Feedback for Personal Search; https://typeset.io/papers/learning-with-sparse-and-biased-feedback-for-personal-search-56yvqu5un6
966. *Mittal & Agarwal* [2019] - Saloni Mittal, Manoj K. Agarwal; 03 Aug 2019; Named Entity Recognition in Local Intent Web Search Queries; https://link-springer-com.ressources-electroniques.univ-lille.fr/chapter/10.1007/978-3-030-27615-7_31
967. *Padmasundari & Bangalore* [2018] - A Padmasundari, Srinivas Bangalore; 6 Sep 2018; Intent Discovery Through Unsupervised Semantic Text Clustering; https://web.archive.org/web/20190220231647/http://pdfs.semanticscholar.org/29a6/ecd69146a65463fbd622ca164f2569ab63cd.pdf
968. *Chimienti et al.* [2022] - Michela Chimienti, Ivan Danzi, Vincenzo Gattulli, Donato Impedovo, Giuseppe Pirlo, Davide Veneto; 2022; Behavioral Analysis for User Satisfaction; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/document/9999081
969. *Eugene et al.* [2019] - Agichtein Eugene, Brill Eric, Dumais Susan; 17 Jan 2019; Improving Web Search Ranking by Incorporating User Behavior Information; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/3308774.3308778
970. *Dandekar et al. GP* [2021] - Pranav Dandekar, Ashish Goel, Peter Lofgren, Fisher Matthew; 2 Dec 2021; Entity recognition from an image; https://patentimages.storage.googleapis.com/13/af/64/3281e3ce7ec0ba/US20210374386A1.pdf
971. *Kasturia et al.* [2021] - Vaibhav Kasturia, Marcel Gohsen, Matthias Hagen; 18 May 2021; Entity-Based Query Interpretation; https://arxiv.org/pdf/2105.08581v1
972. *Kota et al.* [2021] - Nagaraj Kota, Venkatesh Duppada, Ashvini Jindal, Mohit Wadhwa; 5 Nov 2021; Understanding Job Seeker Funnel for Search and Discovery Personalization; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3459637.3481959
973. *Chebil et al.* [2020] - Wiem Chebil, Mohammad Wedyan, Haiyan Lu, Omar Ghaleb Elshaweesh; 01 Apr 2020; Context-Aware Personalized Web Search Using Navigation History; https://typeset.io/papers/context-aware-personalized-web-search-using-navigation-3xq47c70sj
974. *Zhang et al.* [2021] - Chuxu Zhang, Huaxiu Yao, Lu Yu, Chao Huang, Dongjin Song, Haifeng Chen, Meng Jiang, Nitesh V. Chawla; 26 May 2021; Inductive Contextual Relation Learning for Personalization; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3450353
975. *Shen et al.* [2021] - Yilin Shen, Xiangyu Zeng, Yu Wang, Hongxia Jin; 17 Aug 2021; System and method for personalized natural language understanding; https://patentimages.storage.googleapis.com/e0/73/b7/16604fffd6a7b1/US11094317.pdf
976. *Solaiman & Bhargava* [2023] - KMA Solaiman, Bharat Bhargava; Feb 2023; Feature Centric Multi-modal Information Retrieval in Open World Environment (FemmIR);

https://www.researchgate.net/publication/368295794_Feature_Centric_Multi-modal_Information_Retrieval_in_Open_World_Environment_FemmIR

977. *Tomar & Bharti* [2023] - Deepak Tomar, Jyoti Bharti; 25 Jun 2023; Transformer based multimodal similarity search method for E-Commerce platforms; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/document/10183514
978. *Tautkute et al.* [2019] - Ivona Tautkute, Tomasz Trzcinski, Aleksander P. Skorupa, Krzysztof Lukasz, Krzysztof Marasek; 17 Jun 2019; DeepStyle: Multimodal Search Engine for Fashion and Interior Design; https://typeset.io/papers/deepstyle-multimodal-search-engine-for-fashion-and-interior-43pbi6sh34
979. *Pinho et al.* [2016] - Eduardo Pinho, Tiago Marques Godinho, Frederico Valente, Carlos Costa; 26 Aug 2016; A Multimodal Search Engine for Medical Imaging Studies; https://link-springer-com.ressources-electroniques.univ-lille.fr/article/10.1007/s10278-016-9903-z
980. *Shaowei Wang* [2015] - Shaowei Wang; 2015; Multimodal Code Search; https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=1118&context=etd_coll
981. *Duclos* [2019] - Vincent Duclos, 24 Oct 2019; Algorithmic futures: The life and death of Google Flu Trends; https://typeset.io/papers/algorithmic-futures-the-life-and-death-of-google-flu-trends-34n5xsnski
982. *Woloszko* [2020] - Nicolas Woloszko; 01 Dec 2020; Tracking activity in real time with Google Trends; https://one.oecd.org/document/ECO/WKP(2020)42/En/pdf
983. *Gillespie* [2016] - Tarleton Gillespie; 1 Feb 2016: #trendingistrending: when algorithms become culture; https://tarletongillespie.org/essays/Gillespie%20-%20trendingistrending%20PREPRINT.pdf
984. *Fang & Chen* [2016] - Ze-Han Fang, Chien Chin Chen; Aug 2016; A novel trend surveillance system using the information from web search engines; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0167923616300847
985. *Yadav et al.* [2010] - Divakar Yadav, A. K. Sharma, J. P. Gupta; 27 Dec 2010; Users search trends on WWW and their analysis; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/1963564.1963574
986. *Sullivan* [2007] - Danny Sullivan; 16 May 2007; Google Launches “Universal Search” & Blended Results; https://searchengineland.com/google-20-google-universal-search-11232
987. *Xiao-xing* [2008] - Gao Xiao-xing; 01 Jan 2008; A Personalized Information Retrieval Algorithm Based on Word-Correlativity Model; https://typeset.io/papers/a-personalized-information-retrieval-algorithm-based-on-word-4jmag3k3gr
988. *Golbus et al.* [2014] - Peter B. Golbus, Imed Zitouni, Jinyoung Kim, Ahmed Hassan, Fernando Diaz; Apr 2014; Contextual and dimensional relevance judgments for reusable SERP-level evaluation; https://typeset.io/papers/contextual-and-dimensional-relevance-judgments-for-reusable-42e2hy81n7
989. *Geetha Rani* [2013] - S. Geetha Rani; Jul 2013; A New Ranking Algorithm for Ranking Search Results of Search Engine based on Personalized User Profile; https://research.ijcaonline.org/volume74/number15/pxc3889961.pdf

990. *Rathod et al.* [2015] - Meenakshi Rathod, Reshma Y. Totare, Prajakta Raskar, Pooja Modak, Venuvanshi Bhujbal; Apr 2015; Personalized Web Search Protecting Privacy Using Greedy Algorithm; https://ijsrd.com/articles/IJSRDV3I1015.pdf
991. *Thakur & Kumari* [2019] - Jawahar Thakur, Anita Kumari; 2019; Semantic Web Search Engines: A Comparative Survey; https://typeset.io/papers/semantic-web-search-engines-a-comparative-survey-pa1z97wrq4
992. *El-gayar et al.* [2015] - M. M. El-gayar, N.Mekky, A. Atwan; 01 Jan 2015; Efficient Proposed Framework for Semantic Search Engine using New Semantic Ranking Algorithm; https://www.researchgate.net/publication/297734275_Efficient_Proposed_Framework_for_Semantic_Search_Engine_using_New_Semantic_Ranking_Algorithm
993. *Pandiarajan et al.* [2020] - Sudhakar Pandiarajan, V. M. Yazhmozhi, P. Praveen kumar; 22 Jun 2020; Semantic Search Engine Using Natural Language Processing; https://ieeexplore-ieee-org.ressources-electroniques.univ-lille.fr/document/9140266
994. *Alan Clark et al.* [2020] - David Alan Clark, Smith Justin N, Nowotarski Mark S; Nov 2020; Semantic analysis system for ranking search resultshttps://patentimages.storage.googleapis.com/d9/ac/a2/3107e92139e08d/US10846295.pdf
995. *Banerjee & Choudhary* [2017] - Aritra Banerjee, Shrey Choudhary; 07 Sep 2017; Advanced Page Rank Algorithm with Semantics, In Links, Out Links and Google Analytics; https://www.researchgate.net/publication/319643027_Advanced_Page_Rank_Algorithm_with_Semantics_In_Links_Out_Links_and_Google_Analytics
996. *More et al.* [2021] - Manju More, S. Sushanth Kumar, Mustafa Basthikodi, J. Thriveni, K. Venugopal; 2O21; Novel Semantic Approach Based Ranking Algorithm Framework For Advancing Search Engines; https://www.webology.org/data-cms/articles/20220206020408pmwevology%2018%20(3)%20-%2064.pdf
997. *seozoom.com* [2023] - 18 Jul 2023; Semantic Search, the key to modern SEO; https://www.seozoom.com/semantic-search-the-key-to-modern-seo/
998. *Silvers* [2024] - Jon Silvers; 29 Apr 2024; The definitive guide to semantic search engines; https://www.algolia.com/blog/ai/the-definitive-guide-to-semantic-search-engines/
999. *Weisz et al.* [2023] - Justin D. Weisz, Michael Muller, Jessica He, Stephanie Houde; 13 Jan 2023; Toward General Design Principles for Generative AI Applications; IBM Company Research; https://typeset.io/papers/toward-general-design-principles-for-generative-ai-2f17jdg7
1000. *Louridas & Ebert* [2023] - Panagiotis Louridas, Christof Ebert; 01 Jul 2023; Generative AI for Software Practitioners; https://www.researchgate.net/publication/372212988_Generative_AI_for_Software_Practitioners
1001. *G-AI-SAP* - What is generative AI?; https://www.sap.com/products/artificial-intelligence/what-is-generative-ai.html
1002. *G-AI-Wiki* - Generative artificial intelligence; https://en.wikipedia.org/wiki/Generative_artificial_intelligence
1003. *Caines et al.* [2023] - Andrew Caines, Luca Benedetto, Shiva Taslimipoor, Christopher J. Davis, Yuan Gao, Zheng Yuan, Mark Elliott, Russell Moore, Christopher Bryant, Marek Rei, Helen Yannakoudakis, Diane Nicholls, Paula

Buttery; Jul 2023; On the application of Large Language Models for language teaching and assessment technology; https://www.researchgate.net/publication/372415980_On_the_application_of_Large_Language_Models_for_language_teaching_and_assessment_technology

1004. *Cacciamani et al.* [2023] - Giovanni Cacciamani, Michael Eppler, Conner Ganjavi, Gary S Collins, Inderbir Gill; Jul 2023; Development of the ChatGPT, Generative Artificial Intelligence and Natural Large Language Models for Accountable Reporting and Use (CANGARU) Guidelines; https://www.researchgate.net/publication/372444895_Development_of_the_ChatGPT_Generative_Artificial_Intelligence_and_Natural_Large_Language_Models_for_Accountable_Reporting_and_Use_CANGARU_Guidelines
1005. *Ma et al.* [2023] - Rao Ma, Meng Qian, Potsawee Manakul, Mark J. F. Gales, Kate Knill; Jul 2023; Can Generative Large Language Models Perform ASR Error Correction?; https://arxiv.org/pdf/2307.04172v1
1006. *Douglas* [2023] - Michael R. Douglas; Oct 2023; Large Language Models; https://arxiv.org/pdf/2307.05782
1007. *Bariah et al.* [2023] - Lina Bariah, Qiyang Zhao, Hang Zou, Yu Tian, Faouzi Bader, Merouane Debbah; 23 Dec 2023; Large Language Models for Telecom: The Next Big Thing?; https://arxiv.org/pdf/2306.10249
1008. *LLM-Databricks* [2024] - 19 Mar 2024; Generative AI and large language models (LLMs) on Databricks; https://docs.databricks.com/en/generative-ai/generative-ai.html
1009. *Dee* [2023] - Catherine Dee; 9 Nov 2023; Large language models (LLMs) vs generative AI: what's the difference?; https://www.algolia.com/blog/ai/large-language-models-llms-vs-generative-ai-whats-the-difference/
1010. *Bell* [2023] - Elizabeth Bell; 21 Dec 2023; Generative AI vs. Large Language Models (LLMs): What's the Difference?; https://appian.com/blog/acp/process-automation/generative-ai-vs-large-language-models.html
1011. *Ziems et al.* [2023] - Noah Ziems, Wenhao Yu, Zhihan Zhang, Meng Jiang; 16 May 2023; Large Language Models are Built-in Autoregressive Search Engines; https://arxiv.org/pdf/2305.09612
1012. *Gilbert et al.* [2023] - Henry Gilbert, M Sandborn, Douglas C. Schmidt, Jesse Spencer-Smith, Jules White; 25 Apr 2023; Semantic Compression With Large Language Models; https://arxiv.org/pdf/2304.12512
1013. *Suri & Ziaee* [2023] - Gaurav Suri, Ali Ziaee; 08 May 2023; Do Large Language Models Show Decision Heuristics Similar to Humans? A Case Study Using GPT-3.5; https://arxiv.org/pdf/2305.04400
1014. *Zhang et al.* [2024] - Zhenning Zhang, Yunan Zhang, Suyu Ge, Guangwei Weng, Mridu Narang, Xia Song, Saurabh Tiwary; 22 Feb 2024; GenSERP: Large Language Models for Whole Page Presentation; https://arxiv.org/abs/2402.14301
1015. *Marketbrew-4* - Unleashing the Power of Large Language Models for SEO; https://marketbrew.ai/unleashing-the-power-of-large-language-models-for-seo
1016. *Shorten & Cardenas* [2023] - Connor Shorten, Erika Cardenas; 13 Jun 2023; Large Language Models and Search; https://weaviate.io/blog/llms-and-search
1017. *Wiki-GenZ* – Generation Z https://en.wikipedia.org/wiki/Generation_Z
1018. *Pathmonk* [2023] - Pathmonk; 17 Aug 2023; Artificial Intelligence Meets Conversion Optimization: The Future of CRO; https://www.linkedin.com/pulse/artificial-intelligence-meets-conversion-optimization-future/

1019. *Raykova* [2024] - Lora Raykova; 25 Mar 2024; Artificial Intelligence Optimization (AIO): New Way to Speed Up Your Site; https://uxify.com/blog/post/artificial-intelligence-optimization-website-speed
1020. *Ayuya* [2024] - Collins Ayuya; 17 Apr 2024; AI Model Optimization: 6 Key Techniques; https://www.eweek.com/artificial-intelligence/ai-model-optimization/
1021. *Zhou et al.* [2023] - Chi Zhou, He Li, LinLin Zhang, Yufei Ren; 07 Jun 2023; Optimal Recommendation Strategies for AI-Powered E-Commerce Platforms: A Study of Duopoly Manufacturers and Market Competition; https://typeset.io/papers/optimal-recommendation-strategies-for-ai-powered-e-commerce-2n6mhsbz
1022. *Hao et Hu* [2022] - Hongzhong Hao, Sheng Hu; 21 Jan 2022; Recommendation Optimization of Physical Education for Developing the Intelligence of Autistic Children following Intelligent Collaborative Filtering Algorithm; https://typeset.io/papers/recommendation-optimization-of-physical-education-for-3utyt0gz
1023. *Koulidis & Ahmed* [2023] - Alexis Koulidis, Shehab E. Ahmed; 02 Oct 2023; Advisory System for Drilling Optimization Using Artificial Intelligence and In-Cutter Sensing Data; https://typeset.io/papers/advisory-system-for-drilling-optimization-using-artificial-3jfnn0zw
1024. *Yuan* [2022] - Xiaoyan Yuan; Aug 2022; Research on Digital Media Recommendation Algorithm Based on Artificial Intelligence Technology; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/3568923.3568931
1025. *Somosi* [2022] - Zoltán Somosi; 01 Jan 2022; The relationship between seo and artificial intelligence; https://www.researchgate.net/publication/365724518_The_relationship_between_seo_and_artificial_intelligence
1026. *Ding et al.* [2019] - Bailu Ding, Sudipto Das, Ryan Marcus, Wentao Wu, Surajit Chaudhuri, Vivek Narasayya; Jun 2019; AI Meets AI: Leveraging Query Executions to Improve Index Recommendations; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/10.1145/3299869.3324957
1027. *Wang et al.* [2019] - Jia Wang, Xia Song, Qibing Jin, Dan Song; Aug 2019; Research on recommender algorithm optimization based on statistics and preference model; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/abs/10.1145/3357254.3357291
1028. *Tavernier* [2020] - Jean Joseph Tavernier; 7 Jul 2020; Artificial intelligence system for generating intent-aware recommendations; https://patentimages.storage.googleapis.com/fa/46/6d/d7c2fc85450c5e/US10706450.pdf
1029. *Wei et al.* [2021] - Bi Wei, Chen Zeyu, Wang Fan, Wu Zeheng, Shi Lei, He Jingzhou; 17 Jun 2021; Artificial intelligence-based search recommendation method and device; https://patents.google.com/patent/WO2021114624A1/en
1030. *Khanal* [2023] - Devesh Khanal; 22 Aug 2023; Does Ranking on Google Impact ChatGPT's Recommendations?; https://www.growandconvert.com/seo/google-rankings-vs-chatgpt-recommendations/
1031. *Ma et al.* [2022] - Siyong Ma, Puja Das, Sofia Maria Nikolakaki, Qifeng Chen, Humeyra Topcu Altintas; Sep 2022; Two-Layer Bandit Optimization for Recommendations; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/abs/10.1145/3523227.3547396
1032. *Bodhost* [2024] - 10 Oct 2024; AI-Powered Webpage Optimization: What It Is And How To Use It; https://www.bodhost.com/blog/ai-powered-webpage-optimization-what-it-is-and-how-to-use-it/

1033. *Berry* [2023] - Sarah Berry; 25 Oct 2023; SEO vs. PPC: Differences, Pros, Cons, & Use Cases; https://www.seo.com/blog/seo-vs-ppc/
1034. *Sendurur & Yildirim* [2015] - Emine Sendurur, Zahide Yildirim; Jan 2015; Students' Web Search Strategies With Different Task Types: An Eye-Tracking Study; https://www.researchgate.net/publication/272928125_Students'_Web_Search_Strategies_With_Different_Task_Types_An_Eye-Tracking_Study
1035. *Hienert et al.* [2018] - Daniel Hienert, Matthew Mitsui, Philipp Mayr, Chirag Shah, Nicholas J. Belkin; Mar 2018; The Role of the Task Topic in Web Search of Different Task Types; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3176349.3176382
1036. *Sikora et al.* [2005] - Cynthia A. Sikora, Linda A. Roberts, Bonnie M. Kudrick; 2005; TRENDS IN WEB SEARCHING: TASK-SPECIFIC CHARACTERISTICS; https://humanlinkllc.com/trends-in-web-searching-task-specific-characteristics/
1037. *Liu et al.* [2019] - Jiqun Liu, Matthew Mitsui, Nicholas J. Belkin, Chirag Shah; Mar 2019; Task, Information Seeking Intentions, and User Behavior: Toward A Multi-level Understanding of Web Search; https://dl-acm-org.ressources-electroniques.univ-lille.fr/doi/pdf/10.1145/3295750.3298922
1038. *Tegze* [2023] - Jan Tegze; Sep 2023; Filter Bubble: How It Affects Your Search?; https://newsletter.fullstackrecruiter.net/p/how-filter-bubble-affecting-your-search
1039. *Mueller-Bloch & Kranz* [2015] - Christoph Mueller-Bloch, Johann Kranz; Dec 2015; A framework for rigorously identifying research gaps in qualitative literature reviews; https://www.researchgate.net/publication/283271278_A_Framework_for_Rigorously_Identifying_Research_Gaps_in_Qualitative_Literature_Reviews
1040. *Brocke et al.* [2009] - Jan vom Brocke, Alexander Simons, Björn Niehaves, Anne Cleven; Jun 2009; Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process; https://www.researchgate.net/publication/259440652_Reconstructing_the_Giant_On_the_Importance_of_Rigour_in_Documenting_the_Literature_Search_Process
1041. *Kavuri & Milne* [2009] - Anil Kavuri, Alistair Milne; Jan 2019; Fintech and the Future of Financial Services: What Are the Research Gaps?; https://www.researchgate.net/publication/331244922_Fintech_and_the_Future_of_Financial_Services_What_Are_the_Research_Gaps
1042. *Tutors India* [2023] - Tutors India; Jul 2023; What are the different types of research gaps in the literature review?; https://www.tutorsindia.com/blog/what-are-the-different-types-of-research-gaps-in-the-literature-review/
1043. *Lacoste-Badie* [2014] - Sophie Lacoste-badie, Jan 2014; Attention portée à une publicité contenant un visage souriant chez les seniors : une étude eye-tracking; 13th IMTC, Venice; https://archives.marketing-trends-congress.com/2014/pages/PDF/050.pdf
1044. *Lacoste-Badie et al.* [2019] - Sophie Lacoste-Badie, M. Minvielle, O. Droulers; Jun 2019; Attention to food health warnings in children's advertising: a French perspective; https://shs.hal.science/halshs-02301956

1045. *Lacoste-Badie et al.* [2013] - Olivier Droulers, Mathieu Lajante, Éric Jamet, Sophie Lacoste-Badie, Morgane Minvielle; 2013; Effet de la complexité visuelle du packaging sur l'attention portée à la marque; https://www.cairn.info/revue-des-sciences-de-gestion-2013-3-page-39.htm&wt.src=pdf

1046. *Lacoste-Badie et al.* [2018] - Safaa Adil, Sophie Lacoste-Badie, Olivier Droulers; Feb 2018 - Face Presence and gaze Direction In Print Advertisements - how they Influence Consumer Responses - An Eye-tracking Study; https://www.researchgate.net/publication/323283511_Face_Presence_and_Gaze_Direction_In_Print_Advertisements_How_They_Influence_Consumer_Responses_An_Eye-Tracking_Study

1047. *Lacoste-Badie et al.* [2022] - S Lacoste-Badie, O Droulers, G Dossou, Karine Gallopel-Morvan; Aug 2022; Improving the effectiveness of pregnancy warning labels displayed on alcohol containers: a French eye-tracking study; https://hal.science/hal-03800591v2

1048. *Lacoste-Badie et al.* [2020] - Sophie Lacoste-Badie, Junwei Yu, Olivier Droulers; Jun 2020; Small change, big change – Increasing attention with product package variations; https://hal.science/hal-04445762v3

1049. *Lacoste-Badie et al.* [2021] - Junwei Yu, Olivier Droulers, Sophie Lacoste-Badie; Nov 2021; Why display motion on packaging? The effect of implied motion on consumer behaviour; https://univ-angers.hal.science/hal-04429472v2

1050. *Granka et al.* [2004] - Laura A. Granka, Thorsten Joachims, Geri Gay; April 2004; Eye-Tracking Analysis of User Behavior in WWW-Search; https://www.researchgate.net/publication/2917730_Eye-Tracking_Analysis_of_User_Behavior_in_WWW-Search

1051. *Nettleton & Gonzalez-Caro* [2012] - David F. Nettleton, Cristina González-Caro; Oct 2012; Analysis of User Behavior for Web Search Success using Eye Tracker data; https://www.researchgate.net/publication/255569127_Analysis_of_User_Behavior_for_Web_Search_Success_using_Eye_Tracker_data

1052. *Strzelecki* [2020] - Artur Strzelecki; Jun 2020; Eye-Tracking Studies of Web Search Engines A Systematic Literature Review; https://www.mdpi.com/2078-2489/11/6/300

1053. *Crockford & Agius* [2006] - Chris Crockford, Harry Agius; Apr 2006: An empirical investigation into user navigation of digital video using the VCR-like control set; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S1071581905001631

1054. *Fubel et al.* [2023] - Erik Fubel, Niclas Michael Groll, Patrick Gundlach, Qiwei Han, Maximilian Kaiser; May 2023; Beyond Rankings: Exploring the Impact of SERP Features on Organic Click-through Rates; https://arxiv.org/abs/2306.01785

1055. *D'Angeli et al.* [2022] - Gessica D'Angeli, Martina Mesce, Luca Cerniglia, Silvia Cimino; Dec 2022; Observational Tools Using Video Recordings of Parent-Child Feeding Interactions: A Narrative Review; https://www.researchgate.net/publication/366122876_Observational_Tools_Using_Video_Recordings_of_Parent-Child_Feeding_Interactions_A_Narrative_Review

1056. *Asan & Montague* [2014] - Onur Asan, Enid Montague; Mar 2014; Using video-based observation research methods in primary care health encounters to evaluate complex interactions; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4350928/
1057. *Haidet et al.* [2009] - Kim Kopenhaver Haidet, Judith Tate, Dana Divirgilio-Thomas, Ann Kolanowski, Mary Beth Happ; Aug 2009; Methods to Improve Reliability of Video Recorded Behavioral Data; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2713814/
1058. *Caldwell & Atwal* [2005] - Kay Caldwell, Anita Atwal; Feb 2005; Non-participant observation: using video tapes to collect data in nursing research; https://pubmed.ncbi.nlm.nih.gov/16416979/
1059. *Latvala et al.* [2000] - Eila Latvala, Päivi Vuokila-Oikkonen, Sirpa Janhonen; May 2000; Videotaped recording as a method of participant observation in psychiatric nursing; https://www.researchgate.net/publication/227869048_Videotaped_recording_as_a_method_of_participant_observation_in_psychiatric_nursing
1060. *Gross* [1991] - Deborah Gross; Oct 1991; Issues Related to Validity of Videotaped Observational Data; https://journals-sagepub-com.ressources-electroniques.univ-lille.fr/doi/10.1177/019394599101300511
1061. *Casey* [2006] - Dympna Casey; Jan 2006; Choosing an appropriate method of data collection; https://go.gale.com/ps/i.do?id=GALE%7CA212034177&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=13515578&p=AONE&sw=w&userGroupName=anon%7Ec7931e25&aty=open-web-entry
1062. *Roter et al.* [2000] - Debra Roter, Susan Larson, Gary S Fischer, James A. Tulsky; December 2000; Experts Practice What They Preach: A Descriptive Study of Best and Normative Practices in End-of-Life Discussions; https://www.researchgate.net/publication/12216392_Experts_Practice_What_They_Preach_A_Descriptive_Study_of_Best_and_Normative_Practices_in_End-of-Life_Discussions
1063. *Harrison Elder* [1999] - Jennifer Harrison Elder; Nov 1999; Videotaped behavioral observations: Enhancing validity and reliability; https://www-sciencedirect-com.ressources-electroniques.univ-lille.fr/science/article/pii/S0897189799802730?via%3Dihub
1064. *Janice* [1988] - Light Janice; 1988; Interaction involving individuals using augmentative and alternative communication systems: State of the art and future directions; https://psycnet.apa.org/record/1989-12440-001
1065. *Badenhorst* [2022] - Cecile Badenhorst; 2022; Research Conceptualization (PPS&Q and research proposals); https://www.youtube.com/watch?v=KNgPbPM2tW8
1066. *Chloé R* [2020] - Dec 2020; Top health searches on Google in France; https://www.alioze.com/en/top-health-searches-google-france/
1067. *SEMrush* [2024] - Jul 2024; Top websites in France (Healthcare Industry); https://www.semrush.com/trending-websites/fr/healthcare
1068. *Demoule* [2021] - Justine Demoule; Aug 2021; [Infographic] Telemedicine : the state of play in France; https://www.apizee.com/telemedicine-the-state-of-play-in-france.php
1069. *Rahmawati et al.* [2022] - Diana Rahmawati, Koko Joni, Muhammad Iqbal Maulana, Rosida Vivin Nahari, Achmad Fiqhi Ibadillah, Heri Setiawan, Adi Kurniawan Saputro; Dec. 2022; The design of facial expression

detection system to determine the level of customer satisfaction using K-Nearest Neighbor Method; MATEC Web of Conferences V372; https://doi.org/10.1051/matecconf/202237206002

1070. Bayle-Iniguez [2025] – Anne Bayle-Iniguez; Sep. 2025; 55 % des Français ont déjà utilisé un moteur de recherche pour s'informer sur leurs symptômes; https://www.lequotidiendumedecin.fr/sante-societe/e-sante/55-des-francais-ont-deja-utilise-un-moteur-de-recherche-pour-sinformer-sur-leurs-symptomes#
1071. Chen et al. [2020] - Nan Chen, Xiangdong Su, Tongyang Liu, Qizhi Hao, Ming Wei ; Jul. 2020; A benchmark dataset and case study for Chinese medical question intent classification; https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1122-3
1072. Wang et al. [2023] - Yu Wang, Yuan Wang, Zhenwan Peng, Feifan Zhang, Luyao Zhou, Fei Yang; Aug. 2023; Medical text classification based on the discriminative pre-training model and prompt-tuning; https://pmc.ncbi.nlm.nih.gov/articles/PMC10408339/
1073. Haye & Harikumar [2025] - Faryab Haye, Vineeth Harikumar; Aug.2025; Learn how Amazon Health Services improved discovery in Amazon search using AWS ML and gen AI; https://aws.amazon.com/blogs/machine-learning/learn-how-amazon-health-services-improved-discovery-in-amazon-search-using-aws-ml-and-gen-ai/
1074. Huan et al. [2022] - Hai Huan, Zelin Guo,Tingting Cai,Zichen He; Mar. 2022; A text classification method based on a convolutional and bidirectional long short-term memory model; https://www.tandfonline.com/doi/full/10.1080/09540091.2022.2098926
1075. Rajib Rana [2016] - Gated Recurrent Unit (GRU) for Emotion Classification from Noisy Speech; https://arxiv.org/abs/1612.07778
1076. Zhang-2 et al. [2019] - Jingren Zhang, Fang'ai Liu, Weizhi Xu, Hui Yu; Nov. 2019; Feature Fusion Text Classification Model Combining CNN and BiGRU with Multi-Attention Mechanism; https://www.researchgate.net/publication/337238917_Feature_Fusion_Text_Classification_Model_Combining_CNN_and_BiGRU_with_Multi-Attention_Mechanism
1077. Jinbao et al. [2022] - Teng Jinbao, Kong Weiwei, Chang Yidan, Tian Qiaoxin, Shi Chenyuan, Li Long; Feb 2022; Text Classification Method Based on BiGRU-Attention and CNN Hybrid Model; https://dl.acm.org/doi/10.1145/3488933.3488970
1078. Zhang et al. [2022] - Yuanyuan Zhang, Yu Chen, Shengkang Yu, Xiaoqin Gu, Mengqiong Song, Yu Peng, Jianxia Chen, Qi Liu; Jul. 2022; Bi-GRU Relation Extraction Model Based on Keywords Attention; https://direct.mit.edu/dint/article/4/3/552/109785/Bi-GRU-Relation-Extraction-Model-Based-on-Keywords
1079. Wang et al. [2024] - Youwei Wang, Lizhou Feng, Ao Liu, Weiqi Wang, Yudong Hou; 2024; Dual BiGRU-CNN-based sentiment classification method combining global and local attention; https://link.springer.com/article/10.1007/s11227-023-05558-9
1080. Arif et al. [2024] - Arif Mahmood, Ramsha Imran, Syed Mustafa Haider Rizvi; Oct. 2024; Text Classification using Graph Convolutional Networks: A Comprehensive Survey; https://arxiv.org/html/2410.09399v1
1081. Yu et al. [2021] - Qing Yu, Ziyin Wang, Kaiwen Jiang; Jan. 2021; Research on Text Classification Based on BERT-BiGRU Model;

https://www.researchgate.net/publication/348594810_Research_on_Text_Classification_Based_on_BERT-BiGRU_Model

1082. Pires et al. [2019] - Telmo Pires, Eva Schlinger, Dan Garrette; Jul. 2019; How Multilingual is Multilingual BERT?; https://aclanthology.org/P19-1493/

1083. Martin et al. [2020] - Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric Villemonte de la Clergerie, Djamé Seddah, Benoît Sagot; May 2020; CamemBERT: a Tasty French Language Model; https://arxiv.org/abs/1911.03894

1084. Lee et al. [2021] - Tae-Hoon Lee, Young-Min Kim, Eunji Jeong, Seon-Ok Na; Aug.2021; Query Intent Detection for Medical Advice: Training Data Construction and Intent Classification; https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10594626

1085. McGee [2012] - Matt McGee; Aug. 2012; Yes, Bing Has Human Search Quality Raters & Here's How They Judge Web Pages; https://searchengineland.com/bing-search-quality-rating-guidelines-130592

1086. Alafari et al. (2025) - Fatmah Alafari, Maha Driss, Asma Cherif; May 2025; Advances in natural language processing for healthcare: A comprehensive review of techniques, applications, and future directions; https://www.sciencedirect.com/science/article/abs/pii/S1574013725000024?

1087. Dura et al. (2022) - Basile Dura, Charline Jean, Xavier Tannier, Alice Calliger, Romain Bey, Antoine Neuraz, Rémi Flicoteaux; Jul 2022; Learning structures of the French clinical language: development and validation of word embedding models using 21 million clinical reports from electronic health records; https://arxiv.org/abs/2207.12940?

1088. Dynomant et al. (2019) - Emeric Dynomant, Romain Lelong, Badisse Dahamna, Clément Massonnaud, Gaétan Kerdelhué, Julien Grosjean, Stéphane Canu, Stefan J Darmoni; Jul 2019; Word Embedding for the French Natural Language in Health Care: Comparative Study; https://medinform.jmir.org/2019/3/e12310/PDF

1089. Bazoge et al. (2023) - Adrien Bazoge, Emmanuel Morin, Béatrice Daille, Pierre-Antoine Gourraud; Dec 2023; Applying Natural Language Processing to Textual Data From Clinical Data Warehouses: Systematic Review; https://medinform.jmir.org/2023/1/e42477/PDF

1090. Neveol et al. (2018) - Aurélie Névéol, Hercules Dalianis, Sumithra Velupillai, Guergana Savova & Pierre Zweigenbaum; Mar 2018; Clinical Natural Language Processing in languages other than English: opportunities and challenges; https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-018-0179-8?

1091. Azzouzi et al. (2024) - Mohamed El Azzouzi, Gouenou Coatrieux, Reda Bellafqira, Denis Delamarre, Christine Riou, Naima Oubenali, Sandie Cabon, Marc Cuggia & Guillaume Bouzillé; Feb 2024; Automatic de-identification of French electronic health records: a cost-effective approach exploiting distant supervision and deep learning models; https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02422-5

1092. Gerardin et al. (2022) - Christel Gérardin, Perceval Wajsbürt, Pascal Vaillant, Ali Bellamine, Fabrice Carrat, Xavier Tannier; Jun 2022; Multilabel classification of medical concepts for patient clinical profile identification; Artificial Intelligence in Medicine, vol. 128, issue C; https://dl.acm.org/doi/abs/10.1016/j.artmed.2022.102311

1093. Zhang et al. (2022) - Yiming Zhang, Ying Weng, Jonathan Lund; Jan 2022; Applications of Explainable Artificial Intelligence in Diagnosis and Surgery; https://pmc.ncbi.nlm.nih.gov/articles/PMC8870992/
1094. Jansen & Booth (2010) - Jansen, B. J., & Booth, D. L.; 2010; Determining the user intent of web search engine queries; https://www.researchgate.net/publication/221023370_Determining_the_user_intent_of_web_search_engine_queries
1095. Garla & Brandt (2012) - Vijay N Garla & Cynthia Brandt; Oct 2012; Semantic similarity in the biomedical domain: an evaluation across knowledge sources; https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-261
1096. Kim et al. (2023) - Hwayeong Kim, Jiwoong Lee, Sangwoo Moon, Sangil Kim, Taehyeong Kim, Sang Wook Jin, Jung Lim Kim, Jonghoon Shin, Seung Uk Lee, Geunsoo Jang, Yuanmeng Hu & Jeong Rye Park; Jul 2023; Visual field prediction using a deep bidirectional gated recurrent unit network model; https://www.nature.com/articles/s41598-023-37360-1
1097. Broder (2002) - Andrei Broder; Sep. 2002; A taxonomy of web search; https://dl.acm.org/doi/10.1145/792550.792552
1098. Rose & Levinson (2004) - Daniel Rose, Danny Levinson; Apr 2004; Understanding user goals in web search; https://www.researchgate.net/publication/2947641_Understanding_User_Goals_in_Web_Search
1099. Amatriain (2020) - Xavier Amatriain; Nov. 2018; NLP & Healthcare: Understanding the Language of Médicine; https://medium.com/curai-tech/nlp-healthcare-understanding-the-language-of-medicine-e9917bbf49e7
1100. White & Horvitz (2009) - White Ryen W., Eric Horvitz; Sep. 2009; Cyberchondria: studies of the escalation of medical concerns in web search; http://ryenwhite.com/papers/WhiteTOIS2009.pdf
1101. Dernoncourt et al., 2017 - Dernoncourt F., Lee J. Y., Uzuner, O., Szolovits, P.; 2017; De-identification of patient notes with recurrent neural networks; https://pubmed.ncbi.nlm.nih.gov/28040687/
1102. Pedersen et al. (2007) - Pedersen T., Pakhomov S. V., Patwardhan S., Chute, C. G.; 2007; Measures of semantic similarity and relatedness in the biomedical domain; https://pubmed.ncbi.nlm.nih.gov/16875881/
1103. Peng et al. (2019) - Peng Y., Yan S., Lu, Z.; 2019; Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets; https://aclanthology.org/W19-5006/
1104. Spasic & Nenadic (2020) - Irena Spasic, Goran Nenadic; Mar 2020; Clinical Text Data in Machine Learning: Systematic Review; https://pmc.ncbi.nlm.nih.gov/articles/PMC7157505/
1105. Alonso & Contreras (2016) - Alonso I., Contreras D.; Feb 2016; Evaluation of semantic similarity metrics applied to the automatic retrieval of medical documents: An UMLS approach; https://repositorio.comillas.edu/xmlui/handle/11531/15474
1106. Amann et al. (2020) - Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I. Madai; Nov. 2020; Explainability for artificial intelligence in healthcare: a multidisciplinary perspective; https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6

1107. Yeung et al. (2024) - Joshua Au Yeung, Anthony Shek, Thomas Searle, Zeljko Kraljevic, Vlad Dinu, Mart Ratas, Mohammad Al-Agil, Aleksandra Foy, Barbara Rafferty, Vitaliy Oliynyk & James T. Teo; Nov. 2024; Natural language processing data services for healthcare providers. https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-024-02713-x
1108. Soganciogl u et al. (2017) - Gizem Soganciogl u, Hakime Öztürk, Arzucan Özgür; Jul 2017; BIOSSES: a semantic sentence similarity estimation system for the biomedical domain; https://pubmed.ncbi.nlm.nih.gov/28881973/
1109. Wieland-Jorna et al. (2024) - Yvonne Wieland-Jorna, Daan van Kooten, Robert A Verheij, Yvonne de Man, Anneke L Francke, Mariska G Oosterveld-Vlug; Jul 2024; Natural language processing systems for extracting information from electronic health records about activities of daily living. A systematic review. https://academic.oup.com/jamiaopen/article/7/2/ooae044/7681769
1110. Kannan & Li (2017) - P.K. Kannan, Hongshuang Alice Li; Mar 2017; Digital marketing: A framework, review and research agenda; https://www.sciencedirect.com/science/article/abs/pii/S0167811616301550
1111. Sbaffi & Rowley (2017) - Laura Sbaffi, Jennifer Rowley; Jun 2017; Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research; https://www.jmir.org/2017/6/e218/
1112. Bradely (1997) - Andrew P. Bradely; Jul 1997; The use of the area under the ROC curve in the evaluation of machine learning algorithms; https://www.sciencedirect.com/science/article/abs/pii/S0031320396001422
1113. Hanely & McNeil (1982) - J A Hanley, B J McNeil; Apr. 1982; The meaning and use of the area under a receiver operating characteristic (ROC) curve; https://pubs.rsna.org/doi/10.1148/radiology.143.1.7063747
1114. Fawcett (2006) - Tom Fawcett; Jun. 2006; An introduction to ROC analysis; https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X
1115. Sebastiani (2002) - Fabrizio Sebastiani; Mar. 2002; Machine learning in automated text categorization; https://dl.acm.org/doi/10.1145/505282.505283
1116. Yang & Liu (1999) - Yiming Yang, Xin Liu; Aug. 1999; A re-examination of text categorization methods; https://dl.acm.org/doi/10.1145/312624.312647
1117. Davis & Goadrich (2006) - Jesse Davis, Mark Goadrich; Jun. 2006; The relationship between Precision-Recall and ROC curves; https://dl.acm.org/doi/10.1145/1143844.1143874
1118. Chicco & Jurman (2020) - Davide Chicco, Giuseppe Jurman; Jan. 2020; The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation; https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7
1119. Hossain et al. (2023) - Hossain E., Rana R., Higgins N., Soar J., Barua P. D., Pisani A. R., Turner K.; Mar 2023; Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review; https://www.scienceDirect.com/science/article/abs/pii/S0010482523001142
1120. Essi & Pedro (2024) - Essi Pietilä, Pedro A. Moreno-Sánchez; Jan 2024; When an Explanation is not Enough: An Overview of Evaluation Metrics of Explainable AI Systems in the Healthcare Domain; https://www.researchgate.net/publication/377107192_When_an_Explanation_is_not_Enough_An_Overview_of_Evaluation_Metrics_of_Explainable_AI_Systems_in_the_Healthcare_Domain

1121. Sadeghi et al. (2024) - Zahra Sadeghi, Roohallah Alizadehsani, Mehmet Akif CIFCI, Samina Kausar, Rizwan Rehman, Priyakshi Mahanta, Pranjal Kumar Bora, Ammar Almasri, Rami S. Alkhawaldeh, Sadiq Hussain, Bilal Alatas, Afshin Shoeibi, Hossein Moosaei, Milan Hladík, Saeid Nahavandi, Panos M. Pardalos; Aug. 2024; A review of Explainable Artificial Intelligence in healthcare; https://www.sciencedirect.com/science/article/pii/S0045790624002982

1122. Davis F.D. (1989) - Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology; MIS Quarterly - Vol. 13, No. 3 (Sep., 1989), pp. 319-340 (22 pages); Published By: Management Information Systems Research Center, University of Minnesota; https://www.jstor.org/stable/249008

1123. Tennant et al. [2015] - Bethany Tennant, Michael Stellefson, Virginia Dodd, Beth Chaney, Don Chaney, Samantha Paige, Julia Alber; Mar. 2015; eHealth Literacy and Web 2.0 Health Information Seeking Behaviors Among Baby Boomers and Older Adults; https://www.jmir.org/2015/3/e70/

Thesis Content Index

Sr.	Unit	Sr.	Chapter	Sr.	Pillar Topic	Sr.	Topics Cluster	Pages
					Thesis Layout, Text Formatting and Navigation Guidelines			3-4
					Acknowledgement			5-6
					Abstract			7-8
					Introduction			9-18
1	Foreground Literature Review (FLR) - Dynamics of Web Search Marketing Psychology							19-249
		1	Chronology of Information Creation and Presentation towards SERP					20-37
				1	Chronology of G-AI Evolving - From AI to G-AI			20-23
				2	Chronology of SEMLAs Evolving - Teaching & Training SEs			23-27
				3	Chronology of Search Engines Evolving - From Web 1.0 to Web 6.0			27-30
				4	Chronology of SERP Interface Evolving - From Click-Search to Zero-Click Search to Conversational Search			30-34
				5	Chronology of SERP Features Evolving - From Read-only to Conversational			34-37
					Discussion			37
		2	Psychological Literature Review in the context of SERP					38-104
				1	Art of Phrasing of Search User and Search Engine *(A two way Neural & Semantic Understanding)*			**38-47**
						1	Morphology, Syntax and Semantics of a Search Query	38-39
						2	Inferring Linguistic Efficiency of a Search Query by SEs	39-41
						3	Framing Neuro-physiological evidences between Query-SERP-Searcher	41-43
						4	Learning of Search Engines from Searcher' Cognitive Feedbacks	43-46
							Discussion	46-47
				2	Natural Language Ontology *(Search Query and User Intent inferring by Search Engines)*			**47-57**
						1	Decoding Search Query using NLP Components	47-48
						2	Inferring Search Intent using Natural Language Understanding	48-52
						3	Creation of SERP Features using NLP Ontologies	52-57
							Discussion	57
				3	Topicality *(Search Context and Cognitive Accuracy)*			**58-69**
						1	Domain and Topic Knowledge of Search User retrieves better results	59-61
						2	User' Literary Skills precise Topic Search Terms	61-62
						3	Search Engine's Topical Ontology Methods	62-66
						4	Cognitive Evaluation of Topicality	66-68
							Discussion	68-69
				4	Semantic Web *(Predicting search user expectations from SERP - Mechanism of Web 3.0 & Web 4.0)*			**70-78**

				1	Semantic Web Approaches	69-71
				2	Semantic Web Algorithms	72-73
				3	Semantic Web Evolution	73-78
					Discussion	78
			5	SERP Personalization *(Tangible Behaviour that drives the Personalization)*		**78-89**
				1	SERP Personalization Factors	79-81
				2	SERP User Satisfaction Parameters	81-87
				3	SERP Personalization Algorithms	87-89
					Discussion	90
			6	Search Skills (Persistent Intangible Search Behaviour reduces Intent Ambiguity)		**90-103**
				1	Experience - Information Search Mental Models	91-93
				2	Expertise - Linguistic Nuances and Contextual Refining	93-100
				3	Exposure - Web Sources & Algorithms Tracking	100-102
					Discussion	102-103
			7	Thesis Research Variables		**103-104**
		3	Technical Literature Review in the context of SERP			105-150
			1	Data Science Technologies used in generation of SERPs		**105-110**
			2	Grand perspective of Artificial Intelligence and its specificity for SERPs		**110-113**
			3	Search Engine Machine Learning *(SEML)* and Neural Networks		**114-125**
				1	ML SERP gets base from Neural Network	114-115
				2	ML State-of-Art System for SERP	115-117
				3	Purposes of SEML for SERP	117-121
				4	ML Technology and Algorithmic Models for SERP	121-123
				5	Difference between AI and ML	123-125
			4	Deep Learning in SERP Production		**125-132**
				1	SERP features and mechanisms backed by DL	125-128
				2	DL simulates the learning process of human brain	128-129
				3	G-AI models development using DL	129-130
				4	DL and ML relation with each other	130-132
			5	Generative-AI and LLMs are the core of the SERP		**132-145**
				1	General G-AI Models and Technologies	134-136
				2	G-AI and Large Language Models	136-137
				3	G-AI' LLMs use in SERPs Generation	137-139
				4	LLM' Chat GPT and SERP Connect	139-145
			6	MetaSearch Engine and SERP		**144-150**
				1	Metaverse Virtual Indexing	145-146
				2	Metaverse Virtual Environment	146-147
				3	Integration of Virtual Assets of Metaverse into SERP	147-150
		4	Web Search Dynamics			151-249
			1	Web Search User at Mind SERP Model of Google		**151-158**

						1	Classification of Indexed web-products	153-155
						2	Anatomy of a Search Engine	155-156
						3	Search Engines Comparative and Competitive Analysis	156-158
				2	Three Stakeholders of SERP			**158-195**
						1	Web Content Creation by Content Creators for all stakeholders	160-163
						2	Off-Search Content Classification by Search Engine for Self	163-171
						3	On-Search Request Readying by Search Engine for Search User	**171-195**
							A - Search Query Parsing and Deciphering	171-174
							B - Search Intent Identification and Mapping	174-190
							C - SERP Fetching	190-195
				3	SERP Interface and Navigation			**195-203**
				4	Web Search Intelligence			**203-236**
						1	Search Engine's Search-As-Learning	203-206
						2	Defining Search Persona from Searcher' Search Pattern	206-208
						3	Web Content Annotation by Annotators for Search Engine	208-225
						4	Web Search Expressions and Operators	225-229
						5	SERP Personalization, SEO Challenges and AIRO	230-237
				5	Web Search Market Statistics and Trends			**236-245**
						1	Key Web Search Market Statistics of 2020s	238-243
						2	Key Web Search Market Statistics of 2010s	243
						3	Key factors behind growing influence of SERP Marketing	243-245
				6	SERPs for Gen-Z			**245-249**
2	Research Framework							250-345
		5	Synthesized Literature Review (SLR)					251-301
				1	Knowledge Gaps and Further Scopes - From overall SERP Perspective			251-276
				2	Knowledge Gaps and Further Scopes - From Health SERP Perspective			276-289
				3	54 Research Hypotheses Derivatives based on SLR			290-300
				4	Eight Research Hypothesis Derivatives based on FLR			300-301
		6	Research Scope					302-311
				1	Research Scope Validation: 44 Research Audience and Scope Selection			302-305
				2	Hypotheses Formation: 10 Potentials			305-309
				3	Research Motivation and Goals			309-311
				4	Evolution of Research Question			311
		7	R5Ms *(Research Model, Methodology, Method, Metrics and Modalities)* Framework					312-339
				1	Five R5Ms Testable Frameworks			312-314
				2	Relevant R5Ms Literature Review (R5MLR)			314-330
				3	Conceptualization of the Research Stattement			330-331

				4	R5Ms Components of the Research Problem to Test		331-333
				5	Research Problem Contemporary Literature Review 1 (RPCLR-1)		333-338
				6	Probable Research Biases		338-339
		8	Bibliographic Data Analysis				340-345
				1	Qualitative and Quantitative (Q&Q) Analysis of Bibliography		340-343
				2	Research Thesis Shaping-up Funnel		344-345
				3	PRISMA Validation of FLR, SLR, R5MLR, RPCLR and ALR		344-345
3	Research Test Ontology						346-411
		9	Test-Participants Profiling, Data Taxonomy, Research Model and Results Validation				347-405
				1	Psycho-Cognito Profiling and Hypotheses Evaluation		**347-361**
				2	Data Taxonomy - Collection Method, Labelling, Description, Token Analysis and Text Samples Similarity Analysis		**361-374**
				3	Python ML Model and Results Validation		**374-405**
						1 Search Intents Classification Methods and Results	374-379
						2 RNN BiGRU SML Search Intents Classification Model & RPCLR-2	379-405
		10	Thesis Wrap-up				406-411
				1	Results Discussion		406-408
				2	Results Conclusion		408-409
				3	Limitations of this thesis		409-411
				4	Future Scopes and Recommendations		411
					Appendix-1: SERP Anatomy Screenshots (SAS)		412-429
					Appendix-2: Query Text Generation Prompts and YouTube API		430-431
					Appendix-3: Prompts used for Data Labelling		432-437
					Appendix-4: Google Survey Questions		438-439
					Appendix-5: FLR and SLR based One-Liners Testable Factslist		440-447
					Appendix-6: VOSviewer Maps and Connected Papers Analysis		448-453
					Appendix-7: Acronyms and Abbreviations		454-459
					Appendix-8: Thesis Keywords - Detailed List		460-463
					Appendix-9: Borrowed/Sourced Graphics & Visuals		464-503
					Bibliography		504-586
					Thesis Index		587-590
					Graphics-Visuals (Charts, Diagrams, Tables, Screenshots)		591-594
					Declaration		595
					Abstract		596-597

Graphics-Visuals

(Charts, Diagrams, Tables, Screenshots)

Sr.	Charts	Page
1	Chart-4: Chronology of G-AI Evolution	23
2	Chart-5: Chronology of SEMLAs Evolving - Teaching & Training Ses	27
3	Chart-6: Chronology of Search Engines Evolving	30
4	Chart-7: Chronology of SERP Interface Evolving	34
5	Chart-8: Chronology of SERP Features Evolving	37
6	Chart-11: Sub-variables of Variable 1 'Art-of-Phrasing'	46
7	Chart-12: Sub-variables of Variable 2 'Lingual-Ontologies'	57
8	Chart-15: Sub-variables of Variable 3 'Topicality'	68
9	Chart-20: Sub-variables of Variable 6 'Semantic Web'	78
10	Chart-22: Sub-variables of Variable 5 'SERP Personalization'	89
11	Chart-26: Sub-variables of Variable 4 'Search Skills'	102
12	Chart-28: Data Science and SERP	110
13	Chart-30: Artificial Intelligence and SERP	113
14	Chart-31: SEML and SERP	125
15	Chart-32: Deep Learning and SERP	132
16	Chart-33: Chat-GPT and SERP	145
17	Chart-35: MetaSearch Engine and SERP	149
18	Chart-36: Web Search User at Mind SERP Model of Google	151
19	Chart-43: Search User at Mind Model of Google: A Synthesis	158
20	Chart-44: Three Stakeholders of SERP Model	159
21	Chart-45: Organic Search Filter	166
22	Chart-58: Three Stakeholders of SERP	195
23	Chart-59: SERP Interface Blocks and Navigation Menu	196
24	Chart-63: Web Search Intelligence	236
25	Chart-72: The Searcher: Search Expressions and Operators	249
26	Chart-89: Hypotheses Matching to Psychological Model Variables	308
27	Chart-91: SERP Search Path Mind-Map	332
28	Chart-92: R5Ms Framework of the Research Problem	335
29	Chart-95: Research Thesis Shaping-up Funnel	344
30	Chart-96: PRISMA Validation of FLR, SLR, R5MLR, RPCLR & ALR	345
31	Chart-97: Caricature of Chapters and Inter-Chapters Syncing	347
32	Chart-98: Heatmap of Health Categories vs. SERP Behaviour Dimensions	360
33	Chart-99: Distribution of cosine similarities for Standard Intent Category	372
34	Chart-100: Distribution of cosine similarities for Health Intent Category	373
Sr.	**Tables**	**Page**
1	Table-1: Three Stakeholders of SERP	10

2	Table-2: SERP Generation Catalysts	10-11
3	Table-4: Five Types of LR	15-16
4	Table-5: Inclusion and Exclusion criteria for Literature Review	16
5	Table-9: Six variables of this thesis	103
6	Table-10: Difference between AI and ML	124-125
7	Table-11: Three Processing Stages of a SERP Making	153
8	Table-12: Comparative Analysis of Search Engines	156-157
9	Table-13: Competitive Analysis of Search Engines	157-158
10	Table-15: Categorical classification of Content Creators and their Content	163
11	Table-16: Search Intent Combinations	182-184
12	Table-17: What “Keyword” and “a SERP Feature” indicate “Which” intent?	216-217
13	Table-18: Intent Likeliness versus Satisfaction behind a searcher query	217
14	Table-19: Value of a SERP against a search query	218
15	Table-20: Keyword Type mapping to Website Goal	218
16	Table-21: Knowledge Gaps and Future Scopes - Domain Knowledge Pursuance and Perceived Relevance	252-253
17	Table-22: Knowledge Gaps and Future Scopes - Biasing in Searcher’s Behaviour	254-255
18	Table-23: Knowledge Gaps and Future Scopes - Search Query Scope and SERP Utility	258-259
19	Table-24: Knowledge Gaps and Future Scopes - Result Ranking and Search Behaviour	260-261
20	Table-25: Knowledge Gaps and Future Scopes - User Gain Expectancy	263-264
21	Table-26: Knowledge Gaps and Future Scopes - Query Drifting and User Interest Decay	268-269
22	Table-27: Knowledge Gaps and Future Scopes - Clicking Intent Intensity	271-272
23	Table-28: Knowledge Gaps and Future Scopes - SERP Personalization Impact	272-273
24	Table-29: Knowledge Gaps and Future Scopes - Semantic Clustering and IR	275-276
25	Table-30: Knowledge Gaps and Future Scopes - Pre-Web HIS Outcomes and Methods	279
26	Table-31: Knowledge Gaps and Future Scopes - HIS in Web Era	282
27	Table-32: Knowledge Gaps and Future Scopes - Web HISB Characteristics	284-285
28	Table-33: Knowledge Gaps and Future Scopes - DHL Impact on HISB	288-289
29	Table-34: Research Hypothesis Formation	290-300
30	Table-35: Research Hypothesis Derivatives based on FLR	300-

		301
31	Table-36: Number of probable Research Gaps and audience selected	304-305
32	Table-37: Hypotheses derived from citations referred	306-307
33	Table-38: Research impact on Thesis Title	311
34	Table-40: Comparative assessment of five feasible research methods	313-314
35	Table-41: R3M for 37 Metrics for Search User Behaviour Measurement	316-323
36	Table-42: Comparison of model-based evaluation metrics-I	324-326
37	Table-43: Comparison of model-based evaluation metrics-II	326-328
38	Table-44: Comparison of model-based evaluation metrics-III	328-330
39	Table-46: Probable Research Biases of this thesis	338-339
40	Table-47: Quantitative Analysis of Bibliography	340
41	Table-48: Qualitative Analysis of Bibliography	341-342
42	Table-53: Summary of Hypotheses Validation	355
43	Table-54: Healthcare Categories and Keywords	356
44	Table-55: Search Behaviour Intensity Score	358-359
45	Table-56: Hypotheses Analysis based on Intensity Scores	360-361
46	Table-57: 4 Video Test Sets for Data Collection	363
47	Table-58: Data Collection Instructions	364-365
48	Table-59: Search Prompts Engineering of Scenarios for Video Recording Test Sets	365-366
49	Table-60: Data Creation Instructions	366-367
50	Table-61: AI Simulated Data using ChatGPT, Mistral and Perplexity	367-368
51	Table-62: Descriptive Statistics of Token Count for Standard Search Intent Category	370-371
52	Table-63: Descriptive Statistics of Token Count for Health Search Intent Category	371
53	Table-64: Descriptive Statistics of Token Count for Health Search Intent Category	372
54	Table-65: Descriptive Statistics of Token Count for Health Search Intent Category	373
55	Table-66: Supervised ML FeedForward NN Model Results Set-1	376-378
56	Table-67: Supervised ML FeedForward NN Model Results Set-2	378
57	Table-69: BiGRU Benchamark Standard Search Intents	387
58	Table-70: BiGRU Benchamark Health Search Intents	387

59	Table-71: Value Addition of RNN BiGRU Supervised ML Model	403
Sr.	**Diagrams**	**Page**
1	Diagram-4: Web Search Process Model	38
2	Diagram-8: Psychological model of SERP Behaviour	104
3	Diagram-10: Search Algorithms in Universe of DS-AI-ML-NN-DL-GAI-LLM	150
4	Diagram-16: Supervised ML FeedForward NN Model Pipeline	376
5	Diagram-17: RNN BiGRU SML Model Central Code	393
Sr.	**Screenshots**	**Page**
1	Screenshot-1: SERP obtained in Incognito Window of Google Chrome	119
2	Screenshot-2: SERP obtained in General Window of Google Chrome	120
3	Screenshot-3: We queried [Animal Protection] but got result for [Animal Welfare]	121
4	Screenshot-4: SERP obtained in General Window of Google Chrome	121
5	Screenshot-5: ChatGPT as another SERP Feature	142
6	Screenshot-6: Google SERP with language filter	165
7	Screenshot-7: SERP View Dec 2023	196

Declaration

Simplicity is a great virtue, but it requires hard work to achieve it and education to appreciate it. And to make matters worse: complexity sells better, Edsger W. Dijkstra [1984] believing this quote, I hereby declare that this Doctoral Research Thesis is the result of our own independent work and investigation, except where otherwise stated and duly acknowledged. This thesis has not been submitted previously, either in part or in full, for the award of any degree or diploma at this or any other academic institution. I affirm that all sources of information, data, figures, and concepts borrowed from published or unpublished works have been cited in accordance with academic integrity standards. Any ideas, theories, or research findings that are not ours have been properly referenced.

I accept full responsibility for the content of this thesis and the ethical conduct of the research undertaken; thanks to the learnings of my Thesis Director &Co-director, and 'Research Integrity in Scientific Professions Course' which is run by Bordeaux University, France and fecilitated to us as a training through our ADUM platform by our LUMEN Laboratory and Ecole Doctorale.

Lille, 25 / 09 / 2025 Phool Kumar

Étude explicative des facteurs psychologiques à l'origine du comportement de navigation des internautes français en matière de recherche d'informations sur la santé sur les SERP dans le cadre de G-AI

Résumé: Cette thèse présente une étude explicative et exploratoire des facteurs psychologiques, cognitifs, de compréhension, de pertinence, d'intention, de personnalisation et émotionnels qui sous-tendent le comportement de recherche d'informations de santé des utilisateurs francophones (HISB) sur les pages de résultats des moteurs de recherche (SERP). L'accent a été mis sur le développement d'un profilage psycho-cognito de nos participants au test, puis sur la classification des intentions de recherche en deux catégories. La recherche adopte une approche mixte, intégrant des tests de suivi du comportement de recherche des utilisateurs par vidéographie et une analyse de la perception basée sur les commentaires. Au total, 160 participants francophones ont créé 2 884 requêtes de recherche uniques liées à la santé, selon 4 scénarios de test avec une moyenne de 4 requêtes par scénario, ce qui a permis de générer 627 enregistrements vidéo valides. Ces enregistrements vidéo ont été transcrits et divisés en 9 330 échantillons de texte valides et logiques, soit l'équivalent de 815 430 jetons de données humaines réelles, ainsi que la création d'un autre ensemble de commentaires d'utilisateurs YouTube. En parallèle, 165 872 échantillons de texte simulés par l'IA (17 773 761 jetons) ont été générés à l'aide d'API (Application Programming Interface) de grands modèles de langage (LLM) ; en utilisant des invites textuelles basées sur des règles sous des capacités et des caractéristiques d'utilisation contrôlées. Les données humaines réelles et les données générées par l'IA ont été analysées à l'aide d'un modèle d'apprentissage automatique supervisé basé sur un réseau neuronal récurrent (RNN) personnalisé et de la méthode BiGRU (Bidirectional Gated Recurrent Unit) développée en Python. Cette étude assimilée du traitement cognitif, de l'engagement émotionnel et de la compréhension sémantique offre également des perspectives cruciales sur les expériences de recherche futuristes médiées par l'IA et la prise de décision marketing dans le contexte des soins de santé français. Cette recherche contribue aux domaines croisés du marketing numérique, de l'informatique de santé grand public et de l'interaction homme-machine en proposant un cadre comportemental raffiné pour la recherche influencée par l'IA. Elle fait également progresser le discours méthodologique sur la manière de comparer systématiquement la cognition humaine à l'intelligence générative dans ce contexte.

Keywords: *Comportement de recherche d'informations sur la santé en français, IA générative dans la recherche d'informations sur la santé sur les moteurs de recherche, classification des intentions de recherche, psychologie du marketing de recherche sur le Web, modélisation RNN-BiGRU, analyse prédictive par apprentissage automatique supervisé, analyse de similarité sémantique, IA explicable dans le marketing de la santé en français.*

Unité de recherche: Lumen Lab. (Axe MTD), IAE Lille, 104 avenue du Peuple Belge - 59043 LILLE
Ecole doctorale: Ecole doctorale des sciences juridiques, politiques et de gestion, n° 74, 1 place Déliot, 59000 Lille, edsjpg@univ-lille.fr, http://edsjpg.univ-lille.fr; Université de Lille, 42 rue Paul Duez, 59000 Lille

Explanatory study of the psychological factors behind French Web Searcher's Health Information Search Navigation Behaviour on SERP in purview of G-AI

Abstract: This thesis presents an **explanatory and exploratory investigation** into the psychological, cognitive, understanding, relevance, intent, personalization and emotional factors underlying French-speaking User's Health Information Search Behaviour (HISB) on Search Engine Result Pages (SERPs) with prime focus on developing Psycho-Cognito Profiling of our test-participants and then doing dual-category Search Intents Classification. The research adopts a mixed-methods design, incorporating testing of user's search behavioral tracking through videography, and commentary-based perception analysis. A total of 160 French speaking participants created 2,884 unique health-related search prompts; under 4-test scenario with average 4 prompts per scenario; using which generating 627 valid video recordings were created. These video recordings were transcribed and divided into 9330 valid and logical text samples, equivalent of 815,430 tokens of real-human data; along with creating another set of YouTube user comments. In parallel, 165, 872 AI-simulated text samples *(17, 773, 761 tokens)* were generated using APIs *(Application Programming Interface)* of Large Language Models (LLMs); using rule-based **text-prompts** under controlled capacities and usage features. Both **real-human data** and **AI-generated-data** were analyzed using a custom-built **Recurrent Neural Network (RNN)** based Supervised Machine Learning Model and **BiGRU** *(Bidirectional Gated Recurrent Unit)* method developed in Python to present an assimilated study of cognitive processing, emotional engagement, and semantic understanding alongside offering critical insights into futuristic AI-mediated search experiences and marketing decision making in French healthcare contexts. This research contributes to the intersecting fields of digital marketing, consumer health informatics, and human-computer interaction by proposing a refined behavioral framework for AI-influenced search. It also advances methodological discourse on how to systematically compare human cognition with generative intelligence in the context.

Keywords: *French Health Information Search Behaviour, Generative AI in Health Information Search on Search Engines, Search Intent Classification, Web Search Marketing Psychology, RNN-BiGRU Modeling, Supervised Machine Learning Predictive Analytics, Semantic Similarity Analysis, Explainable AI in French Health Marketing.*

Research unit: Lumen Lab. (Axe MTD), IAE Lille, 104 avenue du Peuple Belge - 59043 LILLE
Doctoral school: Ecole doctorale des sciences juridiques, politiques et de gestion, n° 74, 1 place Déliot, 59000 Lille, edsjpg@univ-lille.fr, http://edsjpg.univ-lille.fr; Université de Lille, 42 rue Paul Duez, 59000 Lille