Evaluation of Search Engines Performance on Standardize Queries

Abubakar Kamal, Sarah Bukhari*, Umar Khattab, and Ayesha Khan
Bahauddin Zakariya University, Multan

ABSTRACT The rapid expansion of web content has intensified the need for efficient and accurate search engines capable of responding effectively to diverse user queries. The study aims to compare and evaluate the performance of five major search engines (Google, Bing, Yahoo, Baidu, and DuckDuckGo) in terms of their retrieval speed and relevance across different query types: navigational, informational, and transactional. A set of standardized queries was designed, categorized, and submitted to each search engine under controlled experimental conditions. The number of relevant documents retrieved and the retrieval speed were recorded and analyzed. The results showed that Bing outperformed others in navigational queries, Google dominated informational queries, and Yahoo surprisingly excelled in transactional searches. DuckDuckGo demonstrated competitive performance, especially for navigational tasks. While, Baidu showed strength mainly in informational queries. The findings confirmed that no single search engine is superior across all query types and that search engines optimize differently depending on user intent. This research highlights the importance of query-type-specific evaluation for a better understanding of search engine performance and user satisfaction.

INDEX TERMS Baidu, Bing, DuckDuckGo, Google, Yahoo, queries, search engine

1-Introduction

In the digital age, search engines are crucial resources for information gathering because they act as entry points to the enormous amount of knowledge accessible via the Internet and the World Wide Web. Many search engines have appeared over time; all attempting to meet user demands with unique features and customized functionalities [1]. The assessment of these search engines, meanwhile, has not kept up with their rapid advancement. The evaluation of web search engines is essential for two reasons: it influences the development of search algorithms and aids web users in selecting the appropriate search engine [1]. Traditional search engines are beneficial for locating information online and have become more intelligent with time. However, they have the drawback of not understanding the definitions of the words and expressions used on the web pages and how they relate to one another. These days the technology for keyword-based searches has reached a standstill. About 25% of web searchers do not find adequate results in the first batch of URLs returned, partly because of the web's daily 60-terabyte growth in size [2], [3].

Our keyword-based search engines cite the quantity of web content surpasses technological advancements. To draw in search engines, user's/web pages may contain hundreds of keywords. However, they promote the keywords, rather than providing the content related to them. Semantics (the study of meaning in language) is used in semantic search to generate highly relevant search results, rather than relying on ranking algorithms like Google Page Rank to predict relevancy [2]. The goal is typically to deliver the requested information to the user and not to force them to sift through a list of unrelated keyword results [4].

These days, there are many search engines, but none of them can provide a comprehensive and inclusive expression of the Web. There are several retrieval methods available today which use precision, relative recall, dead links, duplicate links, and unique links to assess how well search engines perform [1]. The most popular search engines used are Google, Yahoo, Bing, Baidu, and DuckDuckGo. Their volume of search ratio is 84.16%, 7.61%, 4.40%, 1.68%, and 2.28%, respectively [5]. The World Wide Web, according to the researchers, is at least 11. It has five billion pages. (Dr Aditi Sharan) On the other hand there is a much larger and deeper web that is concealed within databases with an estimated 3 trillion pages that search engines do not index [3].

This study focuses on how the top 5 search engines are different from each other in terms of retrieval speed and how they response to different types of queries, such as navigational, transactional, and informational queries. The process used to choose and classify nine representative search queries representing informational, transactional, and navigational types is explained in detail in the following section. Additionally, it describes the controlled experimental setting that was used to test each search engine, thus ensuring uniformity across browser devices and internet conditions. The section also describes the performance metrics—such as retrieval speed and response effectiveness—that are used to compare and

*Correspondence Author: drsarahbukhari@gmail.com

objectively evaluate Google, Bing, Yahoo, Baidu, and DuckDuckGo.

2-Literature Review

The World Wide Web's exponential growth has increased the significance of search engines as the main resources to find information. The way that the top five search engines (Google, Bing, Yahoo, Baidu and DuckDuckGo) respond to various user query types, such as transactional, informational and navigational queries, as well as their usability and retrieval speed, varies greatly [6].

In a randomized controlled trial comparing Google, Yahoo, and Bing, [7] found that Google outperformed both Yahoo and Bing in usability when handling ambiguous (primarily informational) queries. Google received the highest System Usability Scale (SUS) score of 81.11, followed by Yahoo (65.42) and Bing (55.56). The study found that especially for navigational and informational queries, Google was able to provide more precise and contextually relevant results more quickly, thanks to its sophisticated algorithms and customized search mechanisms [8]. Users also reported a smoother Google interface and fewer navigation steps, which may indicate higher retrieval efficiency. Bing and Yahoo, while still functional, often lagged in interpreting short or ambiguous queries due to less sophisticated natural language processing.

China's top search engine Baidu has changed by putting a lot of emphasis on localization, mobile integration, and AI-enhanced services, according to [9]. Since its server network is dispersed and aligned with Chinese language structures, Baidu provides a fast retrieval speed within its domestic infrastructure, frequently outperforming competitors locally. Concerns regarding transactional queries where sponsored content might supersede relevance (affecting accuracy and user trust) are raised by the search engine's reliance on paid bidding rankings for query result placement. Baidu also incorporates multimedia search and Baidu Pedia, two tools for informational searches; however, there are complaints about the quality of the content and the abundance of advertisements, which impair usability and perceived objectivity [9].

DuckDuckGo is a privacy-focused substitute that doesn't track user data or customize search results, according to [10]. This greatly improves privacy and neutrality, but it restricts its capacity to provide customized results for navigational or context-dependent informational queries [11]. Zero-click information bangs (!) for direct site searches etc. A. Fast answers for facts weather and computations, as well as! Amazon (for Amazon queries) make it easier to complete transactional and informational queries. In addition, DuckDuckGo employs infinite scroll rather than paginated search results which lessens cognitive load and speed up retrieval for users who don't like flipping between pages. Further, its tools for developers (e.g., technical documentation lookups, URL encoders, and password generators) offer benefits beyond standard search, especially for specialized user groups [12].

Google is often recognized for its superior precision and vast coverage. It uses advanced algorithms including PageRank, which rates pages based on their relevance and popularity. Studies show that Google outperforms competitors in handling diverse query types and delivering user-centric features, such as package tracking and public data analysis [10], [13].

Yahoo's performance has been evaluated as second only to Google in terms of precision. In fact, it offers better coverage in some advanced search scenarios. However, it retrieves a higher number of dead links than its competitors, impacting its reliability [1], [14].

Bing demonstrates superior performance in retrieving fewer dead and duplicate links as compared to Yahoo, especially when handling complex queries. Its indexing techniques ensure high recall rates, though precision remains a challenge in certain cases [13], [15].

DuckDuckGo is known for its privacy-centric approach. It integrates unique features such as zero-click information and customizable search environments. It does not track users, which appeals to privacy-conscious audiences, although its precision and recall lag behind Google and Bing [13], [16].

3-Methodology

A comparative experimental research design is used in this study to systematically assess the performance differences between the top five search engines: Google, Bing, Yahoo, Baidu, and DuckDuckGo. Two important performance metrics are the focus of the comparison.

- **Retrieval Speed:** The amount of time it takes for the entire set of search results to appear after a query is submitted.
- **Response Effectiveness:** The capacity to appropriately respond to multiple user query types including transactional, informational, and navigational queries.

An experimental design is used to warrant objective, quantifiable measures under controlled settings and to allow for a fair and unbiased evaluation of the respective search engine's functions.

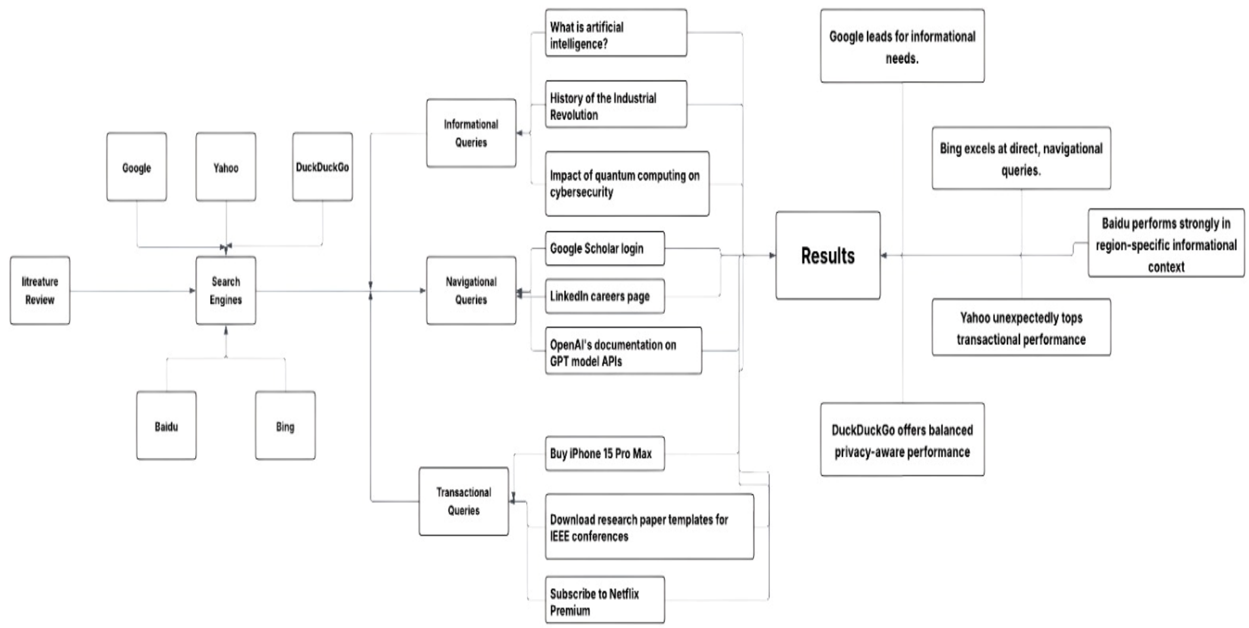

Figure 1 Research Design

3.1-Query Type Definition and Selection

In this study, queries were categorized into three types following industry standards.

- **Navigational Queries**: Queries intended to locate a specific website or page (e.g., "Facebook login", "Harvard University official site").
- **Informational Queries**: Queries intended to acquire knowledge about a topic without necessarily performing an action (e.g., "causes of global warming", "symptoms of vitamin D deficiency").
- **Transactional Queries**: Queries intended to perform an action, such as purchasing or downloading (e.g., "buy iPhone 15 online", "download MS Office trial version")

A total of 9 queries were selected including

- 3 navigational queries
- 3 informational queries
- 3 transactional queries

Table 1. Types of Query for Search Engine

Query Type	**Queries**
Navigational	
Q1	Google Scholar login
Q2	LinkedIn careers page
Q3	OpenAI's documentation on GPT model APIs
Informational	
Q1	What is artificial intelligence?
Q2	History of the Industrial Revolution
Q3	Impact of quantum computing on cybersecurity

Transactional	
Q1	Buy iPhone 15 Pro Max
Q2	Download research paper templates for IEEE conferences
Q3	Subscribe to Netflix Premium

The queries were carefully selected to represent real-world search behavior patterns, ensuring relevance across different user intents. Efforts were made to maintain linguistic simplicity and avoid bias toward any specific search engine's indexing strategy.

3.2- Data Collection Procedure

Each query was entered into Google, Yahoo, Bing, Baidu, and DuckDuckGo individually under identical conditions (device, browser, internet speed, and location). For each query submission, the number of relevant documents appearing in the top results was counted. A document was considered relevant if it correctly satisfied the user's intent.
Measurements were repeated three times to minimize random errors and average values were taken.
The results were organized into tables for each query type.

- Table II: Navigational Queries
- Table III: Informational Queries
- Table IV: Transactional Queries

Each table records the

- Number of relevant documents retrieved for Q1, Q2, and Q3 per search engine.
- Total number of relevant documents across the three queries.
- Average precision (%) calculated based on the maximum possible relevant documents.

Evaluation Metrics
The following metrics were used for analysis.

- **Number of Relevant Documents**: Total documents correctly matching the user query.
- **Average Retrieval Percentage**: Average proportion of relevant documents retrieved, calculated as

Average (%) = (Maximum possible documents / Total relevant documents) × 100
Comparative Performance: Evaluated by comparing average precision percentages across all engines and query types.

3.3-Data Analysis Strategy

After data collection

- For every type of search, the total number of documents that each search engine returned was tabulated.
- Average precision was calculated and compared.
- Performance trends were analyzed across different query types to evaluate
  - Which engine performs better for navigational queries?
  - Which engine excels at informational queries?
  - Which engine is more suitable for transactional queries?
- Descriptive statistics and visual graphs (bar charts) were used to make the comparison clear and intuitive.

3.4-Experimental Environment and Setup

To maintain **experimental consistency**, all search engine tests were conducted under the following controlled conditions.

- **Device**: A laptop running Windows 11 with 16 GB RAM, 256 NVMe, and an Intel Core i5 processor.
- **Browser**: Google Chrome (Version 122) in **incognito mode** to eliminate the effects of cache, cookies, and personalization.
- **Internet Connection:** Storm Fiber broadband connection with a stable upload speed of 19 Mbps and download speed of 30 Mbps.
- **Location**: All searches were conducted from the same physical location (Multan), minimizing geographic influence on search engine localization and content delivery.
- **Testing Times**: Searches were performed during three distinct periods of the day: morning (9 AM), afternoon (2 PM), and evening (7 PM) to account for possible server load fluctuations.

No VPNs, browser extensions, or third-party software were used during the testing phase to ensure the purity of connection and to observe default search engine behaviors.

This study breaks down the queries by types, such as informational, navigational, and transactional. Further, it displays the number of pertinent documents that each search engine returned for the first fifty documents that were returned in Table 2, 3, and 4.

4-Results

The findings from the experimental study evaluated the number of relevant documents retrieved by five search engines (Google, Yahoo, Bing, Baidu, and DuckDuckGo) for three types of queries: navigational, informational, and transactional. These are shown in the following tables.

TABLE II

Number of Relevant Documents Retrieved using Navigational QUERY

Query Number	Google	Yahoo	Bing	Baidu	DuckDuckGo
Q1	30	19	22	14	22
Q2	32	21	40	33	28
Q3	16	31	34	16	31
Total	78	71	76	63	81
Avg (%)	52	47.3	64.0	42.0	54.0

TABLE III

Number of Relevant Documents Retrieved using Informational QUERY

Query Number	Google	Yahoo	Bing	Baidu	DuckDuckGo
Q1	38	27	35	35	36
Q2	42	26	38	40	26
Q3	32	30	30	31	31
Total	112	83	103	106	93
Avg (%)	74.7	55.3	68.7	70.7	62.0

TABLE IV

Number of Relevant Documents Retrieved using Transactional QUERY

Query Number	Google	Yahoo	Bing	Baidu	DuckDuckGo
Q1	23	45	38	29	36
Q2	22	35	28	25	29
Q3	20	32	32	30	31
Total	65	112	98	84	96
Avg (%)	43.3	74.7	65.3	56.0	64.0

The analysis reveals key insights into the effectiveness of each search engine in different query types.

4.1-Performance on Navigational Queries

The navigational query results are presented in Table II. According to the results, Google retrieved 78 relevant documents, Yahoo retrieved 71, Bing retrieved 76, Baidu retrieved 63, and DuckDuckGo retrieved 81.

The average precision rates for navigational queries were

- Bing achieved the highest average with 64.0%.
- DuckDuckGo followed with 54.0%.
- Google obtained 52.0%.
- Yahoo achieved 47.3%.
- Baidu had the lowest performance, with 42.0%.

These results indicate that Bing and DuckDuckGo performed better in retrieving relevant documents for navigational queries than Google, Yahoo, and Baidu.

4.2-Performance on Informational Queries

The informational query results are presented in Table III.

Google retrieved 112 relevant documents, Yahoo 83, Bing 103, Baidu 106, and DuckDuckGo 93.

The average precision rates were

- Google recorded the highest average at 74.7%.
- Baidu followed closely with 70.7%.
- Bing achieved 68.7%.
- DuckDuckGo obtained 62.0%.
- Yahoo scored the lowest at 55.3%.

The findings suggested that Google was the most effective in handling informational queries, followed closely by Baidu and Bing, while Yahoo showed significantly lower performance in this category.

4.3-Performance on Transactional Queries

The results for transactional queries are presented in Table IV. Google retrieved 65 relevant documents, Yahoo 112, Bing 98, Baidu 84, and DuckDuckGo retrieved 96 documents.

The average precision rates for transactional queries were

- Yahoo achieved the highest precision with 74.7%.
- Bing followed with 65.3%.
- DuckDuckGo achieved 64.0%.
- Baidu obtained 56.0%.
- Google recorded the lowest with 43.3%.

Unlike the other query types, Yahoo outperformed all other search engines in transactional query retrieval, followed by Bing and DuckDuckGo, whereas Google had the lowest average precision in this category.

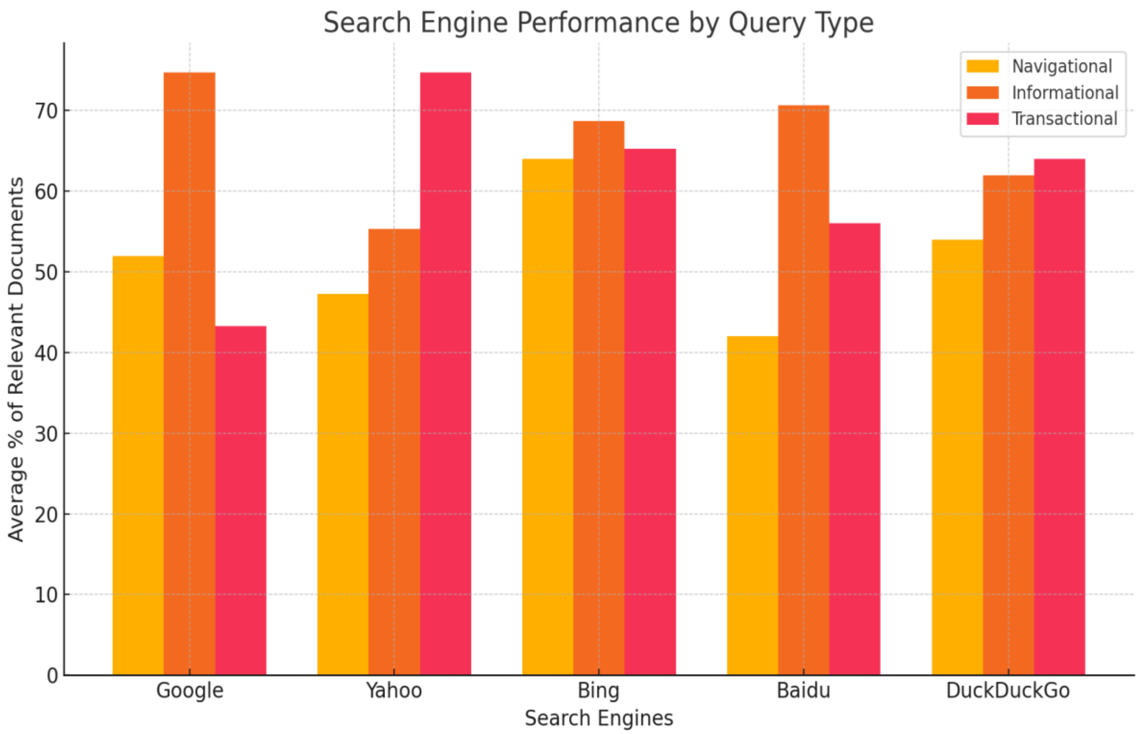

Figure 2 Performance by Query Type

Bing emerged as the most reliable search engine based on average retrieval percentage for various query types, with navigational queries showing the best results. For informational queries, Google still remained the best option, but Yahoo performed the best for transactional queries. The type of search query being conducted may influence the search engine selection, as each search engine exhibits strengths in particular domains. The findings suggest that for users who are primarily interested in navigational searches, Bing is the best option. Google is the best option for informational searches, while Yahoo offers the best results for transactional searches. As a result, the best search engine might change based on the specific user's purpose.

5-Discussion

The findings indicate notable variations in the capacity of the top five search engines—Google Bing Yahoo Baidu and DuckDuckGo—to return pertinent documents in response to navigational, informational, and transactional queries. The performance disparities found in this study are generally in line with the results from previous investigations [1]. Bing outperformed DuckDuckGo and Google by a small margin exhibiting the highest precision (64.0%) in case of navigational queries. DuckDuckGo's privacy-centric strategy makes its comparatively good performance noteworthy. Nevertheless, an earlier study by [17] contended that rather than creating highly customized or predictive navigation pathways, DuckDuckGo frequently depends on combining results from several sources. This lends credence to the idea that clean aggregation, rather than algorithmic prediction, is what makes DuckDuckGo so successful. With the highest precision (74 points) for informational queries Google continues to dominate the market. This result is highly consistent with the findings from several previous studies which highlighted that Google's sophisticated semantic search features, such as contextual analysis and knowledge graph integration allow it to better meet a wide range of informational needs than its rivals. In terms of informational query performance, Google was closely followed by Baidu and Bing in this research. Despite being regionally tailored for Chinese-language queries, Baidu's comparatively strong performance is consistent with earlier findings by [17]. The study pointed out that although some translation or indexing errors still exist in global contexts, regional search engines like Baidu show competitive results when assessed locally. A somewhat surprising pattern emerged in the context of transactional queries, with Yahoo outperforming the others with an average precision of 74.7%. In contrast to earlier researches, Yahoo was frequently rated as having lower overall effectiveness because it retrieved dead links. Yahoo's commercial partnerships and content aggregations has improved its visibility for searches related to transactions and e-commerce. Additionally, Bing and DuckDuckGo showed strong performance in transactional contexts, indicating that they successfully incorporate product pages and sponsored listings without noticeably sacrificing relevancy. Google's comparatively poor performance (43.3%) for transactional queries is consistent with earlier issues raised by [17], who observed that Google occasionally lessens its direct transactional focus by prioritizing wider information access.

All things considered, the results indicate that no single search engine is inherently better for every kind of query. Search engines have different optimization objectives that no single engine can fully satisfy. These include transactional conversion, efficiency navigation, speed, and informational depth. Moreover, the variation in search engine performance also reflects more general findings from testing model studies. The statement from [17] Several models (including TMMi TPI and ISO/IEC standards) are required in the field of software and system testing in 2019 in order to address various organizational priorities.

In contrast to earlier works, this research provides a more recent and varied analysis than [18], which concentrated on a smaller group of search engines and assessed them using a smaller query set, and [19] which mainly classified queries without in-depth performance measurement across a diverse set of engines. This study illustrates how search engine capabilities are changing by combining a wider range of engines and by comparing performance for different query types. Performance variations also imply that query intent has a significant impact on user satisfaction and search success, highlighting the necessity of context-aware engine selection and ongoing engine algorithm evaluation. Similarly, one metric or scenario is not sufficient to capture overall quality in search engine evaluation. Similar to how different testing models highlight different quality attributes in software systems, each engine seems to have developed certain strengths. For example, Google appears to have developed strengths in information retrieval, Bing in navigation efficiency, and Yahoo in transactional services.

6-Conclusion and Future Work

Five popular search engines (Google, Bing, Yahoo, Baidu, and DuckDuckGo) were compared in this study based on how well they handled transactional, informational, and navigational queries. The findings results showed that for all query types, no single search engine performed better than others. Bing exhibited notable strength in navigational searches, probably because of integration with AI-based enhancements, while Google dominated in informational queries, thanks to its sophisticated indexing and ranking algorithms. Oddly, Yahoo performed exceptionally well in transactional queries, indicating that performance in commerce-driven contexts is still influenced by focused indexing and legacy partnerships. Baidu stood out in region-specific informational queries and reaffirmed its dominance in the Chinese-language web ecosystem. Whereas, DuckDuckGo offered a balanced mid-range performance with noteworthy reliability and privacy features. These results demonstrate the dynamic and context-dependent nature of search engine effectiveness, confirming, expanding upon, and occasionally challenging earlier research.

A deeper understanding of multilingual and multicultural search behavior may be possible for future research by enlarging the dataset to include a greater variety of queries from various languages and domains. Reduced subjectivity and increased evaluation scalability can be achieved by integrating automated relevance judgment via ML classifiers or crowdsourcing. A more comprehensive understanding of search engine utility in the real world might also be possible with additional research into click behavior, user satisfaction, time-to-result, and mobile versus desktop performance. Finally, incorporating query expansion strategies [18] may show how engine precision is affected by semantic enrichment for various query intents.

References

[1] I. H. Akhoon, I. Rahim, H. Mushtaq, and S. Ahmad, *Evaluation of Search Engines Using Advanced Search: Comparative Analysis of Yahoo and Bing*. Library Philosophy and Practice; Lincoln, 2019.

[2] K.-L. Liu, A. Santoso, C. Yu, and W. Meng, "Discovering the representative of a search engine," in *Proc. 10th Int. Conf. Inf. Knowl. Manag. (CIKM)*, New York, NY, USA, Oct.

2001, pp. 577–579, doi: https://doi.org/10.1145/502585.502696.
[3] J. Singh and S. Aditi, "A comparative study between keyword and semantic based search engines," Paper presented at International Conference on Cloud, Big Data and Trust, Nov. 13-15, 2013.
[4] L. Wang, J. Wang, M. Wang, Y. Li, Y. Liang, and D. Xu, "Using Internet search engines to obtain medical information: A comparative study," *J. Med. Internet Res.*, vol. 14, no. 3, May 2012, Art. no. 74, doi: https://doi.org/10.2196/jmir.1943.
[5] M. Morris, J. Teevan, and K. Panovich, "A comparison of information seeking using search engines and social networks," in *Proc. Int. AAAI Conf. Web Social Media (ICWSM)*, May 2010, pp. 291–294, doi: https://doi.org/10.1609/icwsm.v4i1.14069.
[6] L. Hsu and Z. Walter, "Search engine or content website? A local information seeking classification model based on consumer characteristics and website perceptions," *Int. J. Hum.-Comput. Interact.*, vol. 31, no. 4, pp. 263–276, Apr. 2015, doi: https://doi.org/10.1080/10447318.2014.999741.
[7] W. Nel, L. De Wet, and R. Schall, "Randomized controlled trial of the usability of major search engines (Google, Yahoo!, and Bing) when using ambiguous search queries," in *Proc. 4th Int. Conf. Comput.-Hum. Interact. Res. Appl. (CHIRA)*, 2020, pp. 152–161, doi: https://doi.org/10.5220/0010133601520161.
[8] A. Oeldorf-Hirsch, B. Hecht, M. R. Morris, J. Teevan, and D. Gergle, "To search or to ask," in *Proc. ACM Conf. Comput.-Supported Coop. Work Social Comput. (CSCW)*, Feb. 2014, pp. 16–27, doi: https://doi.org/10.1145/2531602.2531706.
[9] Y. Pan, "Critical analysis of Baidu's business strategy," *High. Bus. Econ. Manag.*, vol. 24, pp. 2247–2252, Jan. 2024, doi: https://doi.org/10.54097/vrhv2q88.
[10] V. S. Parsania and F. K. K. Kamani, "A comparative analysis: DuckDuckGo vs. Google search engine," *GRD J. Eng.*, vol. 2, no. 1, pp. 12–17, 2016.
[11] J. Morato, S. Sanchez-Cuadrado, and S. Navajas, "Evaluating retrieval and ranking strategies on the dark web: A focus on Tor search engines," *Inf. Discov. Deliv.*, Apr. 2025, doi: https://doi.org/10.1108/IDD-07-2024-0096.
[12] S. S., S. B., S. M., S. K., K. G., and A. R., "Natural language processing techniques for information retrieval: Enhancing search engines with semantic understanding," *ITM Web Conf.*, vol. 76, Mar. 2025, Art. no. 05013, doi: https://doi.org/10.1051/itmconf/20257605013.
[13] A. Giddens and E. Kuşdil, *Modernliğin Sonuçları*. Istanbul, Türkiye: Ayrıntı Yayınları, 2004.
[14] C. D. Schultz, "Informational, transactional, and navigational need of information: Relevance of search intention in search engine advertising," *Inf. Retrieval J.*, vol. 23, no. 2, pp. 117–135, Apr. 2020, doi: https://doi.org/10.1007/s10791-019-09368-7.
[15] S. Gul, S. Ali, and A. Hussain, "Retrieval performance of Google, Yahoo and Bing for navigational queries in the field of life science and biomedicine," *Data Technol. Appl.*, vol. 54, no. 2, pp. 133–150, June 2020, doi: https://doi.org/10.1108/DTA-05-2019-0083.
[16] N. Yagci, S. Sünkler, H. Häußler, and D. Lewandowski, "A comparison of source distribution and result overlap in web search engines," *Proc. Assoc. Inf. Sci. Technol.*, vol. 59, no. 1, pp. 346–357, Oct. 2022, doi: https://doi.org/10.1002/pra2.758.
[17] K. Hrabovská, B. Rossi, and T. Pitner, "Software testing process models benefits & drawbacks: A systematic literature review," *arXiv:1901.01450*, Jan. 2019, doi: https://doi.org/10.48550/arXiv.1901.01450.
[18] H. K. Azad and A. Deepak, "Query expansion techniques for information retrieval: A survey," *Inf. Process. Manag.*, vol. 56, no. 5, pp. 1698–1735, Sep. 2019, doi: https://doi.org/10.1016/j.ipm.2019.05.009.
[19] S. Ali and S. Gul, "Search engine effectiveness using query classification: A study," *Online Inf. Rev.*, vol. 40, no. 4, pp. 515–528, Aug. 2016, doi: https://doi.org/10.1108/OIR-07-2015-0243.