ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/385656754

# Evaluating the summarization algorithms of Ask.com, Bing, Google and Yahoo: search snippets-result conformance

**Article** *in* Performance Measurement and Metrics · November 2024

DOI: 10.1108/PMM-02-2024-0012

CITATIONS
0

READS
30

**3 authors:**

Mahdi Zeynali-Tazehkandi
University of Tabriz
**30** PUBLICATIONS **27** CITATIONS
SEE PROFILE

Mohsen Nowkarizi
Ferdowsi University of Mashhad
**100** PUBLICATIONS **357** CITATIONS
SEE PROFILE

Yeganeh Zhian-Feizabadi
Ferdowsi University of Mashhad
**1** PUBLICATION **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Mahdi Zeynali-Tazehkandi on 30 August 2025.

The user has requested enhancement of the downloaded file.

Performance Measurement and Metrics

AQ: 1

# Evaluating the summarization algorithms of Ask.com, Bing, Google and Yahoo: search snippets-result conformance

AQ: 2 Mahdi Zeynali-Tazehkandi, Mohsen Nowkarizi and
AQ: 3 Yeganeh Zhian-Feizabadi
AQ: 4 *Ferdowsi University of Mashhad, Mashhad, Iran*

Received 17 February 2024
Revised 11 August 2024
Accepted 31 October 2024

AQ: 5

## Abstract

**Purpose –** Objective: The objective of the research was to evaluate the performance of the summarization algorithms utilized by Ask.com, Bing, Google and Yahoo.

**Design/methodology/approach –** This study evaluated the summarization algorithms employed by search engines. In this regard, a set of 30 search terms was carefully constructed using schema ontology. These terms were subsequently employed to interrogate the selected search engines. A random selection of three results from each set of ten retrieved results was then extracted for thorough evaluation. The evaluation process involved quantifying the degree of relevance deviation between the search snippets and the corresponding full-text results. This analysis served to effectively measure the effectiveness of the search engines' summarization algorithms.

**Findings –** The research findings indicated a substantial discrepancy between the relevance of search snippets and the corresponding full-text search results. Furthermore, the findings underscored notable differences among the summarization algorithms utilized by Ask.com, Bing, Google and Yahoo. Additionally, the study found that the abstracting quality of Ask.com, Bing, Google and Yahoo was 43.4, 57.2, 51.8 and 47.8%, respectively.

**Practical implications –** Users are recommended to employ Bing and Google for conducting search queries. The developers of Ask.com and Yahoo are urged to refine their summarization algorithms.

**Originality/value –** This study constitutes the first-ever attempt to assess the summarization capabilities of Ask.com, Bing, Google and Yahoo.

**Keywords** Information search and retrieval, Web search engines, Performance evaluation, Web summaries, Search snippets

**Paper type** Research article

## 1. Introduction

AQ: 7 In the contemporary digital age, the World Wide Web has emerged as an indispensable resource for individuals to fulfill their professional responsibilities, make informed purchasing decisions, and enrich their leisure time. Chen *et al.* (2014) aptly highlight the pervasiveness of web resources, revealing their dominance over traditional information sources like libraries. With an estimated 6 billion indexable pages (Gulli and Signorini, 2005; Kunder, 2022), the World Wide Web presents users with an overwhelming volume of information, thereby necessitating the development of effective tools to navigate this vast digital landscape. In this regard, search engines have emerged as indispensable mediators, bridging the gap between the expansive information available on the internet and its users (Zhang *et al.*, 2019). The escalating number of search engine users (Kumar and Pavithra, 2010) underscores the crucial role of these platforms in facilitating access to relevant and up-to-date information, empowering individuals to make informed decisions and enhance their overall well-being. Additionally, studies have consistently demonstrated that users exhibit a higher degree of trust in search engines compared to traditional media or social networking platforms (Bonart *et al.*, 2020). These findings underscore the critical importance of identifying the most appropriate search engine for individual users. Researchers have accordingly employed diverse evaluation

*Conflict of interest:* The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Performance Measurement and Metrics
© Emerald Publishing Limited
e-ISSN: 1758-6925
p-ISSN: 1467-8047
DOI 10.1108/PMM-02-2024-0012

approaches, resulting in specialized subfields such as the evaluation of the search capabilities (Papenmeier *et al.*, 2021), the coverage and overlap of results (Tsay *et al.*, 2017; Mitra and Awekar, 2017), the retrieval effectiveness (Zhang *et al.*, 2019; Gul *et al.*, 2020) and search snippets (Hu *et al.*, 2019; Bilal and Huang, 2019) within the evaluation domain.

Search snippets, also known as document summarization (Tombros and Sanderson, 1998), and descriptions-result conformance (Lewandowski, 2008), are brief excerpts of webpage content that appear below links in search engine results pages (SERPs) (Gomez-Nieto *et al.*, 2014). These snippets offer users a concise overview of the webpage's content, enabling them to quickly determine its relevance to their search query.

The generation of search snippets is a form of extractive document summarization. In this process, relevant sentences or sentence fragments are selected from the webpage based on their correlation with the search terms (Turpin *et al.*, 2007). This ensures that the snippets accurately represent the webpage's content and provide users with the information they need to make informed decisions about whether to click through.

Search snippets play a crucial role in shaping user behavior and directing web traffic. They are the first element that users encounter in SERPs; so, their effectiveness in conveying the core message of a webpage can significantly impact click-through rates (Hu *et al.*, 2019). In essence, search snippets act as digital storefronts, showcasing the concise overview of a webpage's content and enticing users to explore further. By effectively communicating the webpage's value proposition, snippets can effectively convert potential visitors into engaged readers. As a result, these snippets should provide an accurate and comprehensive representation of the underlying content (Xue *et al.*, 2008).

Despite extensive searches of prominent electronic databases, including Emerald Publishing, Sage Publishing, Taylor and Francis Online, and Springer, as well as the scholarly search engine Google Scholar, it was observed that limited research has been conducted on the efficacy of search engine snippets. This highlights the necessity for further investigation into the effectiveness of summarization algorithms and the search snippets generated by these platforms. As a result, the effectiveness of search snippets provided by international search engines remains unclear, and a comprehensive examination and evaluation of these snippets is absent from the current body of research. To address this gap, the current research aims to compare the summarization algorithms employed by Ask.com, Bing, Google, and Yahoo search engines. To align with the research objectives, the following hypotheses have been formulated:

The main hypothesis

(1) There is a significant difference between the effectiveness of summarization algorithms employed by Ask.com, Bing, Google and Yahoo.

### *1.1 Sub-hypotheses*

(1) There is a significant difference between the relevance of the search snippets and the full text of the retrieved results on ASK.COM.

(2) There is a significant difference between the relevance of search snippets and the full text of results retrieved on Bing.

(3) There is a significant difference between the relevance of the search snippets and the full text of the retrieved results on Google.

(4) There is a significant difference between the relevance of search snippets and the full text of results retrieved on Yahoo.

## 2. Literature review

Various studies have been conducted on the evaluation of search engines such as MacFarlane (2007), Dai and Davison (2011) and Al-Eroud *et al.* (2011). This article provides a review of some of these key studies.

### *2.1 The search engine capabilities*

A crucial aspect of evaluating search engines lies in examining their user interfaces. They serve as the intermediary between humans and machines, constituting the visible and tangible facet of the search engine with which users interact directly. Each search engine offers distinct search capabilities and features. Accordingly, users formulate their query terms based on these capabilities. By enhancing their user interfaces with additional features, search engines can enable users to formulate more comprehensive and effective queries (Zhang and Fei, 2010; Papenmeier *et al.*, 2021). Hence, the user interface and suggested search terms provided by search engines are of paramount significance and have been recognized as a critical research area within search engine evaluation. On this point, Gossen *et al.* (2013) compared the user interfaces and functionalities of 12 children's search engines, including OneK, Yahoo Kids, AskKids, and Google. Their examination focused on key aspects such as page length, page size, spell checker availability, result presentation, result quantity, font size, and result highlighting. The findings indicated that children's specialized search engines did not exhibit a significant superiority to Google. Furthermore, all of the engines exhibited areas for improvement, particularly in catering to the developmental capacities of children.

In a 2016 study, Fattahi *et al.* investigated whether expanding search queries with keywords suggested by Google significantly improved the relevance of retrieved documents. To this end, 60 college students were enlisted, each of whom was instructed to select their favorite subject and conduct a Google search. The students then repeated the search, incorporating the keywords suggested by Google. The findings demonstrated the efficacy of Google's search engine recommender system in enhancing search results relevance.

In a subsequent study, Papenmeier *et al.* (2021) evaluated four user interfaces with a participant pool of 139 native English speaking adults aged 18–40. These digital natives resided in the United States, the United Kingdom, and Ireland. The results indicated that enhancing search capabilities and incorporating assistants such as chatbots could result in a significant increase in the number of product features mentioned (by 84%) and the use of natural language formulas (by 139%), compared to the standard search bar interface. In an earlier study, Aqle *et al.* (2022) developed a user interface known as "InteractSE" and conducted a comparative analysis with Google's user interface. Sixteen participants were asked to compare the two interfaces for complex search activities. The findings revealed that the designed user interface improved users' comprehension and exhibited superior quality compared to Google's.

### *2.2 Coverage and overlap of results*

Search engine coverage refers to the comprehensiveness and depth of the content indexed by a search engine within a particular subject domain. Overlap, an associated concept, implies the presence of a specific information resource in two or more collections or different databases (Egghe and Goovaerts, 2007). Of paramount importance is the fact that a search engine's coverage signifies the vastness of content it encompasses, providing users with immense accessibility. Consequently, individuals are keen to discern which search engine indexes the greatest number of web pages and thus offers the most comprehensive coverage. This topic has been extensively addressed within the field of search engine evaluation, where coverage and overlap are meticulously examined (Tsay *et al.*, 2017). They evaluated the coverage and overlap of two search engines, Google Scholar and Microsoft Academic, and two aggregate institutional repositories (OAIster and OpenDOAR), and two physics-related open sources (arXiv. organ and Astrophysics Data System). The results showed that Google Scholar,

Astrophysics Data System, Microsoft Academic, OpenDOAR, OAIster, arXiv.org, respectively, have covered 96.8, 96.5, 91.6, 48, 45 and 41% of physics-related resources. In a 2017 study, Mitra and Awekar observed that two popular web search engines, Google and Bing, exhibited a high degree of agreement in their search results. Their subsequent investigation, using 2,500 queries, revealed a noticeably lower overlap between these engines, with each engine consistently producing a unique subset of search results. Also, Loan and Sheikh (2018) assessed the precision of Google Scholar through the application of bibliometric analysis. Their findings revealed that Google Scholar effectively retrieves a range of resources, primarily encompassing books and journal articles; however, the coverage of other formats, such as reports and working papers, remained significantly limited. In their study, Rahim *et al.* (2019) extracted search terms from 76 open-access research articles indexed in the Web of Science within the field of Library and Information Science. The researchers utilized the Krejcie and Morgan sample size table to determine a sample of 170 search terms. Subsequently, this initial list was subsequently refined to yield 20 suitable queries. In order to assess the coverage of the search terms, the researchers entered each term into Yahoo and Bing. The retrieved results were then analyzed based on the number of duplicate links and dead links among the top 20 results. The findings demonstrated that Google consistently occupied the top ranking position, followed sequentially by Yahoo and Bing.

### *2.3 The retrieval effectiveness of results*

Comparative evaluation of a search engine's retrieval algorithm and ranking methodology constitutes a significant dimension of search engine assessment. Upon initiating a search query, the search engine commences a retrieval process from its vast database to identify relevant results. Beyond the paramount importance of retrieving pertinent documents, it is equally crucial to assess the ranking mechanism employed, ensuring that the most relevant documents are positioned at the forefront of the search results. In a study aimed at assessing the performance of Google and Bing, Lewandowski (2015) developed the Relevance Assessment Tool (RAT). Drawing upon a dataset of 1,000 questions extracted from a German search engine, Lewandowski's analysis revealed that Google consistently outperformed Bing, achieving a relevance score of 95% compared to Bing's 76%. Sadeghi and Vegas (2017) assessed the retrieval performance of Google in retrieving Persian documents. They employed 100 search topics and evaluated precision as their primary metric. The findings revealed that Google's precision ranged from approximately 45 to 72%.

Uyar and Aliyu (2015) employed a set of 60 queries centered on the nomenclature of individuals and flora. They calculated the precision of the top 100 search results. The findings revealed that Google achieved a precision of 48%, while Bing managed a mere 33%, thereby establishing Google's superior retrieval effectiveness.

In their 2016 study, Ali and Gul elicited 15 queries from postgraduate students of Library and Information Science. These queries were then submitted to both search engines, and the first 30 results were analyzed. The findings revealed that Google consistently outperformed Yahoo in terms of both precision and recall.

In a research, Balabantaray (2017) used 20 randomly selected queries from the most popular queries on each search engine such as Google, Yahoo, Bing, Ask.com, and AOL. He examined the first ten results retrieved by each engine. The results indicated that Google consistently outperformed the other search engines in terms of search capabilities and precision, retrieving the most relevant and informative results for all 20 queries.

Azizan *et al.* (2018) evaluted Google, Yahoo, DuckDuckGo, and Bing – in the domain of durian-related information retrieval. To identify relevant search terms, the researchers meticulously reviewed queries posed in online forums, groups, and social media platforms pertaining to durian. Subsequently, in consultation with subject matter experts, they selected eight pertinent keywords. Upon retrieving the search results, the first twenty entries were carefully evaluated and their precision scores were calculated. The findings revealed that

Yahoo exhibited superior performance when considering the top ten results, while Google demonstrated a stronger overall effectiveness.

To assess the performance of Google, Bing, and Ask.com in retrieving relevant search results, Wu *et al.* (2019) conducted a study involving eight graduate students. Each student was assigned the task of searching for ten queries obtained from text retrieval conferences held between 2009 and 2012. The students evaluated the first 20 search results per query as relevant or unrelated, using new browsers without search history to prevent algorithm bias. The results of the study indicated that Bing performed comparably to Google, outperforming Ask.com in terms of relevance and accuracy of search results.

Gul *et al.* (2020) employed precision and recall to assess the retrieval effectiveness of Google, Yahoo, and Bing in the life sciences and biomedicine domains. They initially identified 14 navigational queries based on the most highly cited publications in these fields. The first 20 results were evaluated. The findings demonstrated that Google, Yahoo, and Bing consistently exhibited superior precision and recall scores, respectively.

Zeynali Tazehkandi and Nowkarizi (2021) compared Google to three Iranian search engines (Parsijoo, Yooz, and Rismoon) using precision, recall, and normalized cumulative gain to assess Persian information retrieval. The study involved selecting 32 relevant Persian topics via stratified sampling. Simulated work tasks were then designed, and participants were asked to read each task, enter their information needs into the four search engines, and record the relevant results. The findings revealed that Google is more effective in retrieving accurate and relevant Persian information than the local search engines.

CheshmehSohrabi and Sadati (2022) evaluated the four general search engines including Google, Yahoo, Duck Duck Go, and Bing, and three specialized image search engines (Flickr, PicSearch, and GettyImages). According to the subject title of the Library of Congress, they designed 15 search terms. They examined the first 20 results and calculated two measures of recall and precision. The results showed that general search engines perform better than others.

Hussain *et al.* (2023) investigated the retrieval performance of Bing Images, Google Images, and Yahoo Image. The authors performed a comprehensive analysis of the engines' ability to effectively retrieve relevant images for two search queries: "COVID-19" and "Tourism". Google Images and Bing Images outperformed Yahoo Image in precision, recall, F-Measure, and Fallout Ratio.

Hussain *et al.* (2023) conducted an analysis of the top 10 images of 2016 to assess the effectiveness of Google Images, Yahoo Image Search, and Picsearch. They evaluated each engine's ability to retrieve relevant images based on established criteria. The first 50 results were used to calculate precision and recall. The findings revealed that Yahoo Image Search consistently outperformed the other two engines in retrieving relevant images.

Hajian Hoseinabadi and CheshmehSohrabi (2023) introduced a new metric to evaluate search engine quality based on relevance, ranking, novelty, coverage, and duplicate content. They compared Yahoo, Google, DuckDuckGo, and Bing. They devised a set of 40 search queries and analyzed the top 20 results. Their findings indicated that Google consistently outperformed the other search engines based on their proposed metric.

### *2.4 Summarization algorithms of search engines*

When users conduct a search in search engines, they are often presented with a query-based summary (Gomez-Nieto *et al.*, 2014) or search snippets (Hu *et al.*, 2019), which can be considered as an abstract of the retrieved document. These summaries influence users' decisions about whether to delve into the full text or not (Marcos *et al.*, 2015). Therefore, the effectiveness of the provided summaries or search snippets is crucial. Some studies suggest search engines can use enticing search snippets to steer users towards specific websites. Building upon this observation, Turpin *et al.* (2007) proposed algorithms and data structures for an efficient query-based summarization system called "engine snippets." They conducted experiments using a Sun Fire V210 Server, measuring RAM storage, average time to generate

10 snippets, and collection size of the wt50 g corpus. Their method, utilizing the zlib compression library, yielded a 58% reduction in snippet generation time compared to a baseline method.

Lewandowski (2008) compared the relevance of search snippets with the full text of retrieved results from Google, MSN, Yahoo, Ask.com, and Seekport. He recruited 40 participants from the University of Dusseldorf, representing a diverse range of educational backgrounds and experiences. Participants were asked to evaluate the relevance of both the snippets and the full text of the results.

Xue *et al*. (2008) proposed WebSIS, an interactive search system using image snippets. They compared text, image, and combined snippets based on precision and time efficiency. Forty-one users tested 29 queries using three snippet types. Combining image and text snippets improved user search and query refinement.

In a two-phase study involving 276 participants, Teevan *et al*. (2009) evaluated the effectiveness of three different Web result representations – text snippets, visual snippets, and thumbnails – in supporting search and re-visitation tasks. Their findings indicated that text snippets were particularly effective for identifying and accessing new Web pages that were unfamiliar to the participants. Thumbnails, on the other hand, emerged as a valuable tool for aiding in the re-discovery of previously visited Web pages.

Marcos *et al.* (2015) examined the impact of snippet richness (plain or rich), snippet position (top-ranked or bottom-ranked), and snippet type (Author, Google Plus, Google Places, Multimedia, or Review) on user behavior during web searching. Two experiments were conducted: an offline study (60 participants, eye tracking) and an online study (110 participants, mouse tracking). The results indicated that snippet richness plays a more significant role in user attention, particularly for bottom-ranked results. Additionally, Multimedia snippets were found to be the most visually appealing and noticeable among the various snippet types studied.

In a recent study, Kim *et al.* (2017) compared the snippet size of web search result between desktop and mobile screen. They recruited 24 participants with varying backgrounds from inside and outside a university campus. Each participant completed 12 tasks, including six tasks for each task type, and two of the six tasks included the same type of snippet length. The findings suggest that users with long snippets on mobile devices exhibit longer search times with no better search accuracy for informational tasks. The overall findings suggest that, unlike desktop users, mobile users are best served by snippets of two to three lines.

Finally, the rise of technology and search engines highlights the need to assess their performance. Consequently, numerous studies have been conducted to examine various aspects of search engines, including their user interfaces (Papenmeier *et al.*, 2021), the quality of recommender systems, the coverage and overlap of search results (Rahim *et al.*, 2019), and the efficiency of document retrieval and ranking algorithms (Gul *et al.*, 2020). Additionally, some research has focused on the advertising aspects of search engines, highlighting their potential to influence user behavior through the display of attractive search snippets. Scholars such as Lewandowski (2008), Marcos *et al.* (2015) and Hu *et al.* (2019) have delved into this topic from diverse perspectives, recognizing that search snippets serve as gateways that may prompt users to explore the full text of relevant documents. However, despite this recognition, only a handful of studies, such as Lewandowski (2008), have explicitly evaluated the effectiveness of search snippets in achieving this goal. This indicates a significant gap in the research landscape of search engine evaluation. However, Table 1 has been prepared for a quick overview of the studies conducted in this area. T1

## 3. Methodology

This research falls within the domain of applied evaluative research, particularly performance measurement (Connaway and Powell, 2016), more detailed information about the research design and method is provided below:

**Table 1.** The overview of previous studies

| Row | Categories | Studies |
|---|---|---|
| 1 | The search engine capabilities | Gossen *et al.* (2013), Fattahi *et al.* (2016), Papenmeier *et al.* (2021), and Aqle *et al.* (2022) |
| 2 | Coverage and overlap of results | Tsay *et al.* (2017), Mitra and Awekar (2017), Loan and Sheikh (2018), and Rahim *et al.* (2019) |
| 3 | The retrieval effectiveness of results | Uyar and Aliyu (2015), Balabantaray (2017), Wu *et al.* (2019), and Gul *et al.* (2020) |
| 4 | Summarization algorithms of search engines | Turpin *et al.* (2007), Lewandowski (2008), Marcos *et al.* (2015), and Kim *et al.* (2017) |

**Source(s):** Authors' own work

*3.1 Selection of the search engines*

T2 To select search engines, two criteria including previous studies (Table 2) and search engine
T3 ranking websites (Table 3) were used, both of which are referred to below.

In addition to scrutinizing search engines featured in prior studies, the evaluation encompassed the ranking of each search engine across Search engine watch, Search engine journal, Net market share, Stanventures, and Ebizmba. In essence, securing the top spot on these websites served as an additional criterion for inclusion, as documented in Table 3.

Finally, based on the previous studies (Table 2) and their global rankings on search engine evaluation websites (Table 3), four international search engines, namely Ask.com, Bing, Google, and Yahoo, were chosen for a comprehensive assessment of their search snippet effectiveness.

*3.2 Selection of the keywords*

In this investigation, the schema design, which can be viewed as an instance of an ontology, served as the foundation for selecting keywords. Drawing upon prior research conducted by Lewandowski (2008) and Balabantaray (2017), 30 keywords were developed based on the schema ontology and then examined using the aforementioned search engines by one of the researchers.

*3.3 Searching keywords on the search engines*

According to Wu *et al.* (2019) for eliminating the influence of browser variations and search history on the results, all searches were conducted by one researcher using a newly installed instance of the Google Chrome browser. Additionally, in accordance with Balabantaray (2017), each keyword across all search engines was queried on a single day, preventing the potential impact of newly added web resources or search engine index updates on the research findings.

*3.3 Selection of the results*

Empirical evidence from previous studies such as Çakir *et al.* (2008), Fendley and Kidambi (2010) and Yalun and Bitirim (2013) suggests that over 80% of users typically only examine the first ten search results. Consequently, in this study, the primary focus for evaluative purposes was the first ten results. Subsequently, utilizing a random selection process, three results from the initial ten were chosen for further examination. Finally, the entire text of the selected results was retained.

*3.4 Selection of the evaluators*

To assess the relevance of the results, a panel of university [1] students was assembled. Based on Krejcie and Morgan's sampling table (Connaway and Powell, 2016) and the prior studies (Zeynali Tazehkandi and Nowkarizi, 2021), a stratified random sampling method was employed to select 360 students from across all academic disciplines.

**Table 2.** The search engines evaluated in recent studies

| Row | The studies | The evaluated search engines |
|---|---|---|
| 1 | Lewandowski (2008) | Google, MSN, Yahoo, Ask.com and Seekport |
| 2 | Anastasiu *et al.* (2013) | Google and Yahoo |
| 3 | Gossen *et al.* (2013) | Google, Yahoo Kid, onek, ASK Kid and so on |
| 4 | Orduña-Malea *et al.* (2014) | Google |
| 5 | Qumsiyeh and Ng (2014) | Google, Yahoo and Bing |
| 6 | Lewandowski (2015) | Google and Bing |
| 7 | Uyar and Aliyu (2015) | Google and Bing |
| 8 | Ali and Gul (2016) | Google and Yahoo |
| 9 | Fattahi *et al.* (2016) | Google |
| 10 | Tsay *et al.* (2017) | Google Scholar and Microsoft Academic |
| 11 | Laura and Me (2017) | SICH |
| 12 | Sadeghi and Vegas (2017) | Google |
| 13 | Balabantaray (2017) | Google, Yahoo, Bing, Ask.com, and AOL |
| 14 | Mitra and Awekar (2017) | Google and Bing |
| 15 | Uyar and Karapinar (2017) | Google and Bing |
| 16 | Loan and Sheikh (2018) | Google Scholar |
| 17 | Azizan *et al.* (2018) | Google, Yahoo, Duck Duck Go and Bing |
| 18 | Bilal and Huang (2019) | Google and Bing |
| 19 | Rahim *et al.* (2019) | Yahoo and Bing |
| 20 | Wu *et al.* (2019) | Google, Bing, and Ask.com |
| 21 | Lewandowski and Sünkler (2019) | Google |
| 22 | Zavadski and Toepfl (2019) | Google and Yandex |
| 23 | Niforatos *et al.* (2019) | Google |
| 24 | Anuyah *et al.* (2020) | Google, Bing, Kidrex and Kidzsearch |
| 25 | Gul *et al.* (2020) | Google, Yahoo, and Bing |
| 26 | Zeynali Tazehkandi and Nowkarizi (2021) | Google, Parsijoo, Yooz and Rismoon |
| 27 | Manouchehri *et al.* (2021) | Google |
| 28 | Aqle *et al.* (2022) | Google |
| 29 | CheshmehSohrabi and Sadati (2022) | Google, Yahoo, DuckDuckGo, and Bing |
| 30 | Hussain *et al.* (2023) | Bing Images, Google Images and Yahoo Image |
| 31 | Hussain *et al.* (2023) | Google Images, Yahoo Image Search, Picsearch |
| 32 | Hajian Hoseinabadi and CheshmehSohrabi (2023) | Yahoo, Google, DuckDuckGo and Bing |

**Source(s):** Authors' own work

**Table 3.** Ranking of search tools based on the websites

| Rank | Search engine watch[1] | Search engine journal[2] | Net market share[3] | Stanventures[4] | Ebizmba[5] |
|---|---|---|---|---|---|
| 1 | Google | Google | Google | Google | Google |
| 2 | Bing | Bing | Bing | Bing | Bing |
| 3 | DuckDuckGo | Yandex | Baidu | Yahoo | Yahoo |
| 4 | Qoura | Cc search | Yahoo | Baidu | Baidu |
| 5 | Dogpile | Swisscows | Yandex | DuckDuckGo | ASK.COM |
| 6 | Vimeo | DuckDuckGo | DuckDuckGo | Yandex | AOL |
| 7 | Yandex | StartPage | ASK.COM | ASK.COM | DuckDuckGo |
| 8 | Boardreader | Search Encrypt | Naver | Naver | WolframAlpha |
| 9 | WolframAlpha | Gibiru | Ecosia | AOL | Yandex |
| 10 | StartPage | One search | AOL | Seznam | WebCrawler |

**Note(s):** [1]https://www.searchenginewatch.com, [2]https://www.searchenginejournal.com, [3]https://netmarketshare.com, [4]https://www.stanventures.com, [5]http://www.ebizmba.com

**Source(s):** Authors' own work

*3.5 Research forms*

Following the established guidelines for questionnaire construction (Pickard, 2013; Connaway and Powell, 2016), the forms commenced with a concise instruction guide on navigating the questionnaire. This was followed by the presentation of search snippets. Subsequently, participants were instructed to read the provided search snippets and assess the level of motivation they generated on a Likert scale, indicating their inclination to delve into the full text of the search results. At the conclusion of the questionnaire, participants were requested to provide demographic information about themselves.

*3.6 Calculating the effectiveness of summarization algorithm (search snippets)*

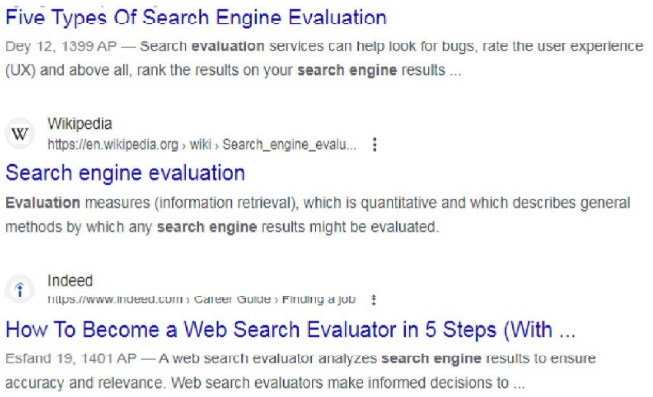
Search snippets, also referred to as document summarization (Turpin *et al.*, 2007) or document descriptions (Lewandowski, 2008), are concise summaries of webpages that provide readers with a comprehensive understanding of the full text. This aligns with the traditional definition of an abstract (Lancaster, 1991; Meadow *et al.*, 2007). Search snippets, akin to miniature versions of documents, serve as succinct summaries of web pages. These well-crafted excerpts empower readers to swiftly and accurately grasp the essence of a webpage, enabling them to discern whether or not they require delving into the full text (Hartley, 2008; Hu *et al.*, 2019). In this regard, when users initiate a search query, search engines promptly present users with
F1 summaries and descriptions of documents, constituting query-based abstracts. Figure 1 showcases three search snippets prepared by Google.

Figure 1 presents the search snippets of three search results, titled "Five Types of Search Engine Evaluation," "Search engine evaluation," and "How to Become a Web Search Evaluator in 5 Steps." Assume the user has clicked on the second result and proceeded to read the entire webpage
F2 based on the webpage's summary. The full text of the second result is displayed in Figure 2.

The extent to which a search engine summary (Figure 1) accurately reflects the full text of a web page serves as a measure of the effectiveness of the summarization process. In this regard, Lewandowski (2008) proposed a formula (Equation 1) for assessing this effectiveness.

$$Q_{sa} = N_S - \left(\frac{\Sigma|R_{ss} - R_f|}{N_p}\right) \quad (1)$$

In his study, Lewandowski (2008) introduced the concept of "description-result conformance" ($Q_{sa}$),which can be interpreted as the effectiveness of abstracting or summarizing documents. In this research,we adopt Lewandowski's definition and refer to "$Q_{sa}$" as the effectiveness of the summarization algorithm. Furthermore,we define "$N_s$" as the number of scale's




AQ: 8 **Figure 1.** Example of three search engine results snippets from Google. Source: Authors' own work

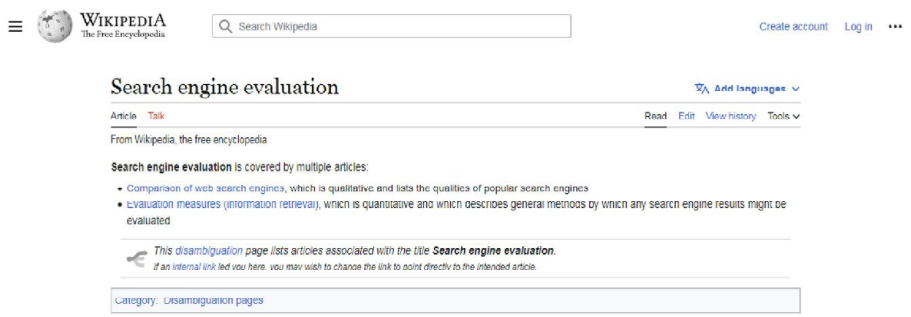



**Figure 2.** Full text of the second result entitled "Search engine evaluation". Source: Authors' own work

points,"$R_{ss}$" as the relevance of the search snippets,"$R_f$" as the relevance of the full text results,and "$N_p$" as the number of pairs. Using these definitions, we were able to determine the effectiveness of search engines in abstracting or summarizing documents.

### *3.7 Validity and reliability*

The validity of the findings was supported by the consensus of experts and the consistency with previous research (Lewandowski, 2008). Furthermore, during the implementation phases, the research form, search terms, and related aspects were subjected to a review by five experts. Based on the feedback received, necessary modifications were implemented. Subsequently, to assess the comprehensibility of the research forms for students, the forms were distributed to a group of 20 students for evaluation. They unanimously affirmed the forms' clarity and ease of understanding. Consequently, the forms were disseminated among the target population. To evaluate reliability, 30 search forms were administered to a cohort of participants (30 individuals) for completion twice, with an interval of two weeks between administrations. The correlation between the test and retest phases was determined to be 0.72, confirming the forms' reliability.

## 4. Results

### *4.1 The first sub-hypothesis*

There is a significant difference between the relevance of the search snippets and the full text of the retrieved results on ASK.com

Prior to testing the hypothesis, the normality of the data was assessed using the Kolmogorov-Smirnov test. The test statistic yielded a value of 0.01, which is less than the significance level of 0.05. This indicates that the null hypothesis of normality should not be accepted, suggesting that the data distribution deviates from normality. Consequently, the Sign test, a non-parametric test suitable for non-normal data, was employed to examine the hypothesis. The results of this test are presented in Table 4. T4

**Table 4.** Comparison of relevance between search snippets and full-text results in Ask.com

| Variable | Number | Negative differences | Positive differences | Ties | Test statistics (*Z*) | *p*-value |
|---|---|---|---|---|---|---|
| Value | 90 | 20 | 46 | 24 | −3.077 | 0.002 |

**Source(s):** Authors' own work

As indicated in Table 4, the *p*-value (0.002) is less than the significance level of 0.05. This indicates that there is a statistically significant difference in the relevance of search snippets and the entire text of results retrieved on ASK.COM. To provide further insights into the differences between pairs of descriptors and the entire results, Figure 3 is presented.

*4.2 The second sub-hypothesis*

There is a significant difference between the relevance of search snippets and the full text of results retrieved in Bing.

To assess the appropriate statistical test for evaluating the hypothesis, it was imperative to determine the normality of the data distribution. This was achieved through the application of the Kolmogorov-Smirnov test. Given the test statistic's value of (0.01) falling below the significance level of (0.05), the null hypothesis was rejected. This implies that the data distribution deviates from normality. Subsequently, to investigate the significance of the disparity between the relevance of search snippets and full text results in Bing, the sign test was employed. The outcomes are summarized in Table 5.

Since the significance level (0.004) falls below the conventional threshold of 0.05 (as indicated in Table 5), the null hypothesis cannot be upheld. This implies that a statistically significant difference exists between the relevance of search snippets and the full text of search results retrieved in Bing. Figure 4 provides further insights into the extent of this relevance disparity between search snippets and complete search results.

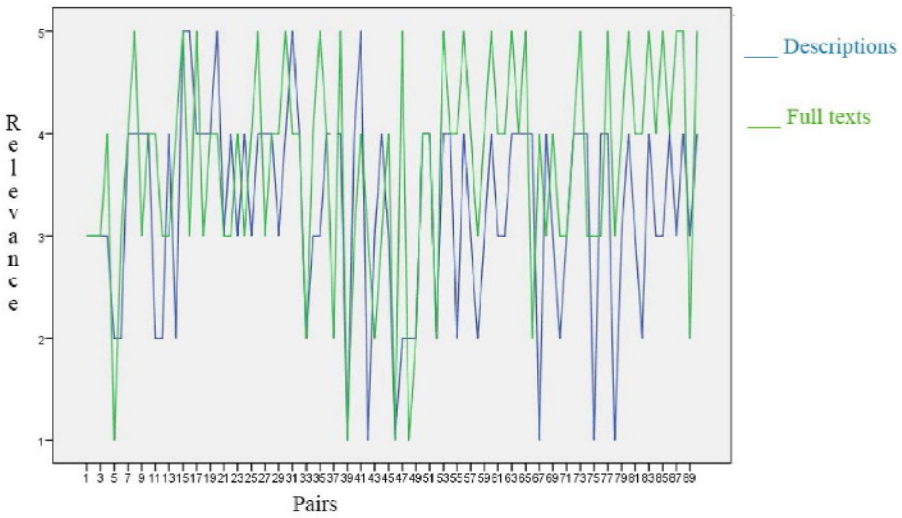



**Figure 3.** Comparison of relevance between the full text of search results and their search snippets in ASK.COM. Source: Authors' own work

**Table 5.** Sign test for the relevance difference between the search snippets and the full text of Bing's results

| Variables | Number | Negative differences | Positive differences | Ties | Test statistics (*Z*) | *p*-value |
|---|---|---|---|---|---|---|
| Values | 90 | 24 | 50 | 16 | −2.906 | 0.004 |

**Source(s):** Authors' own work

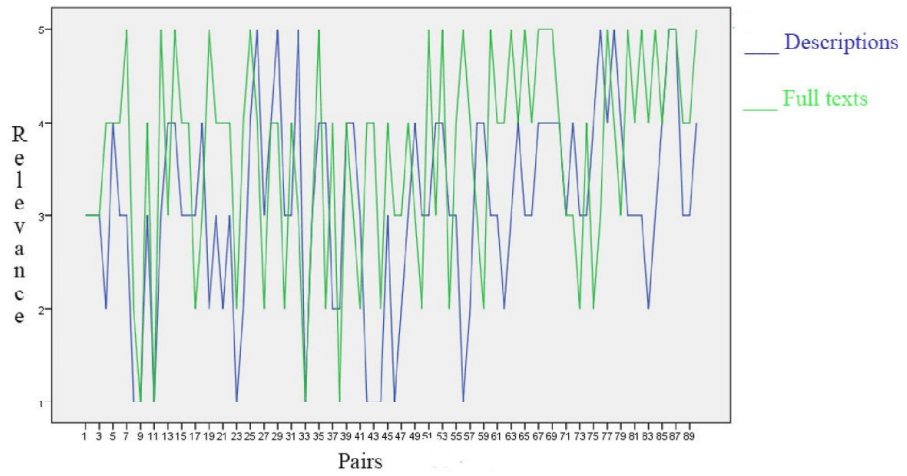



**Figure 4.** Relevance difference between the results' full text and their search snippets in Bing. Source: Authors' own work

*4.3 The third sub-hypothesis*

There is a significant difference between the relevance of the search snippets and the full text of the retrieved results in the Google search engine.

In view of the necessity to assess the normality of the data to identify the appropriate hypothesis test, the Kolmogorov-Smirnov test was employed. The statistical value obtained from this test (0.01) was found to be less than the conventional threshold (0.05), indicating that the data distribution deviates from normality. Consequently, the sign test was utilized to evaluate the proposed hypothesis, with the results summarized in Table 6. T6

As evidenced by the significance level of the test (0.001) being less than the conventional threshold of 0.05, the null hypothesis is rejected. This indicates a statistically significant difference between the relevance of search snippets and the full text of results retrieved in Google. To further elucidate this relevance disparity, Figure 5 presents a graphical representation of the pairwise comparisons between search snippets and full text results. F5

*4.4 The fourth sub-hypothesis*

There is a significant difference between the relevance of search snippets and the full text of results retrieved in Yahoo search engine.

To determine the appropriate statistical test for examining the hypothesized difference in relevance between search snippets and full text results in Yahoo, the normality of the data distribution was assessed using the Kolmogorov-Smirnov test. The test revealed that the data distribution deviates significantly from normality ($p$-value <0.05). Consequently, the non-parametric Sign test was employed to evaluate the statistical significance of the observed difference.

**Table 6.** Sign test for relevance difference between the search snippets and the full text of Google's results

| Variable | Number | Negative differences | Positive differences | Ties | Test statistics ($Z$) | $p$-value |
|---|---|---|---|---|---|---|
| Value | 90 | 13 | 52 | 25 | −4.788 | 0.001 |

**Source(s):** Authors' own work

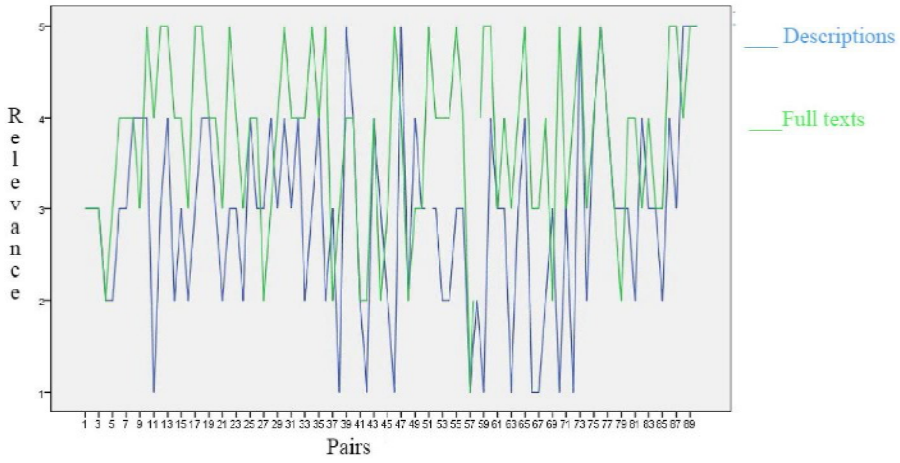



**Figure 5.** Relevance difference between the results' full text and their search snippets in Google. Source: Authors' own work

A statistically significant difference was observed between the relevance of search snippets and the full text of retrieved results in Yahoo, with a significance level of 0.021 (Table 7). This difference is further illustrated in Figure 6, which provides a clear visualization of the discrepancies between snippet and full-text relevance.

*4.5 The main hypothesis of the research*

There is a significant difference between the quality of abstracting of Ask.com, Bing, Google and Yahoo.

To determine the appropriate statistical test for the proposed hypothesis, it was essential to assess the normality of the data. Consequently, the distribution of the data was examined using the Kolmogorov-Smirnov test. Given that the test statistic yielded a value of 0.01, which is lower than the predetermined significance level of 0.05, the null hypothesis and the assertion that the data conforms to a normal distribution were rejected. Accordingly, Friedman's test was employed to evaluate the hypothesis, and the results are presented in Table 8.

Table 8 indicates the number of search keywords, the chi-square statistic, the degree of freedom, and the significance level. Because the test's significance level (0.003) is less than 5%, the null hypothesis and the assertion of equal abstracting quality among search engines are not supported. To put it another way, there is a meaningful disparity in the abstracting quality of search engines, and the average rank of each search engine is displayed in Table 9.

The Table 9 presents the mean rank of relevance differences for search snippets and full text results across four search engines: Bing, Google, Yahoo, and Ask.com. A lower mean rank indicates a higher quality of abstracting, as it suggests that the search snippet more accurately

**Table 7.** Sign test for relevance difference between the search snippets and the full text of Yahoo's results

| Variable | Number | Negative differences | Positive differences | Ties | Test statistics ($Z$) | $p$-value |
|---|---|---|---|---|---|---|
| Value | 90 | 22 | 41 | 27 | −2.305 | 0.021 |

**Source(s):** Authors' own work

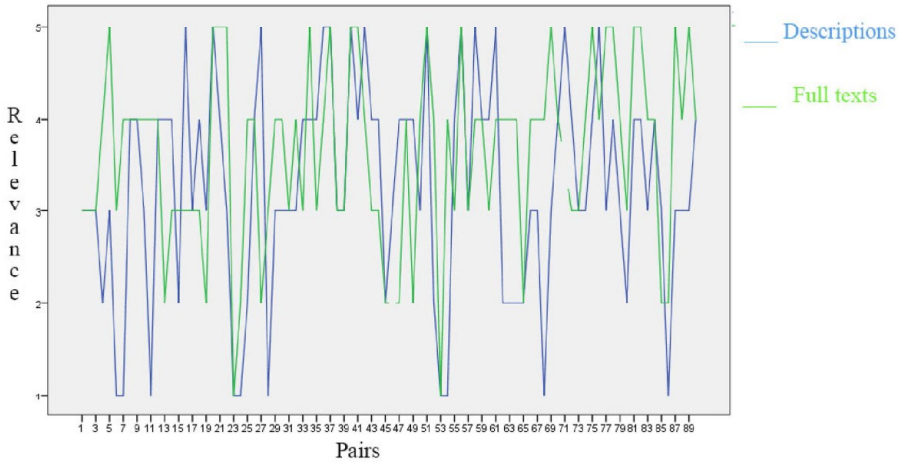



**Figure 6.** Relevance difference between the results' full text and their search snippets in Yahoo. Source: Authors' own work

**Table 8.** Friedman's test for the qualities of abstracts of search engines

| Variables | Number | Test statstics (Chi-square) | Df | *p*-value |
|---|---|---|---|---|
| Values | 90 | 13.92 | 3 | 0.003 |

**Source(s):** Authors' own work

**Table 9.** Mean rank of relevance differences of search snippets-full texts' results in the search engines

| Search engines | Ask.com | Bing | Google | Yahoo |
|---|---|---|---|---|
| Mean rank | 2.83 | 2.14 | 2.41 | 2.61 |

**Source(s):** Authors' own work

reflects the content of the full text document. Based on the data in Table 8, it can be concluded that Bing, Google, Yahoo, and Ask.com exhibit increasing levels of abstracting quality, respectively. To further illustrate this relationship, Figure 7 provides a graphical representation of the abstracting quality of each search engine. To exemplify the method of calculating search engine abstracting quality as formulated in Equation (1), a specific example is presented. For example, Table 9 illustrates a discrepancy of 2.83 in relevance scores between search snippets and full-text results on ASK.com. Considering that the effectiveness of summarization algorithms is evaluated using a 5-point scale, their influence on ASK.com's search results can be estimated by subtracting the observed difference (2.83) from the maximum effectiveness score (5). This calculation yields a value of 2.17. To express the quality of the summarization algorithm as a percentage, this value can be divided by the maximum score (5) and multiplied by 100. This results in an estimated impact of 43.4% (Figure 7). F7

In the study, 360 pairs of search snippet-full text results were analyzed. To provide further insights into the summarization algorithms employed by search engines in each pair, Figure 8 is presented. F8

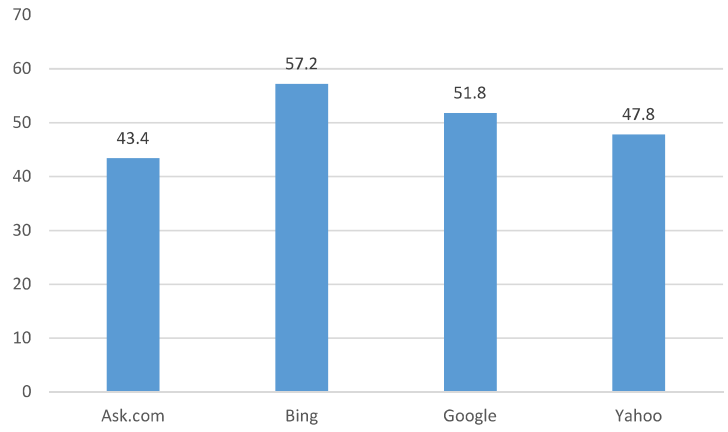



**Figure 7.** Effectiveness of summarization algorithms employed by Ask.com, Bing, Google and Yahoo. Source: Authors' own work

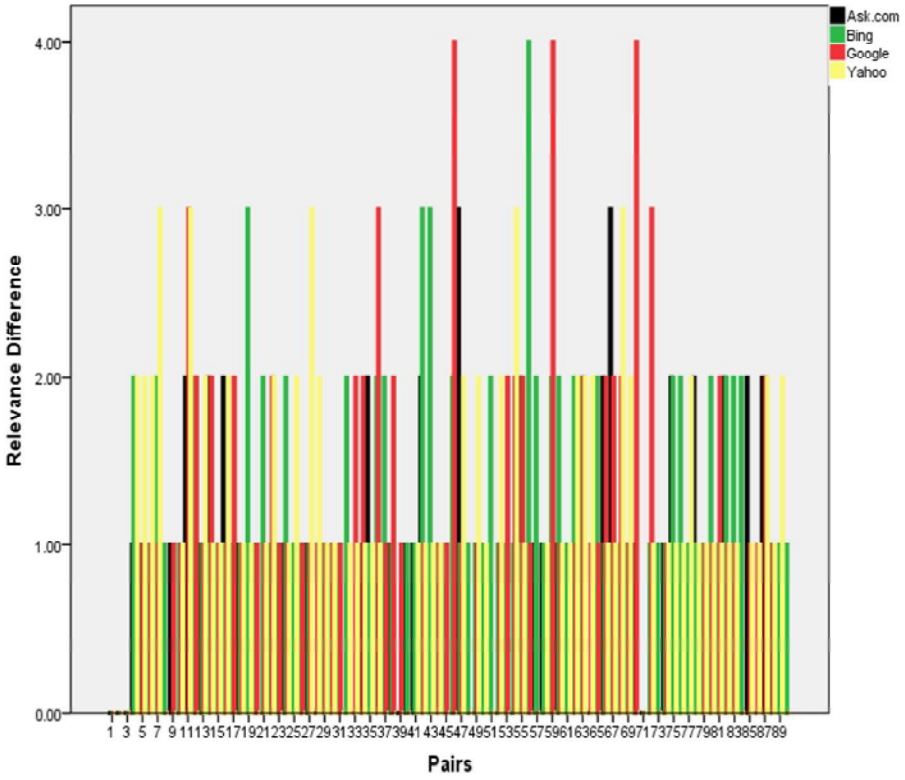



**Figure 8.** The relevance differences of search snippets-full texts results in Ask.com, Bing, Google and Yahoo. Source: Authors' own work

Figure 8 presents a visual representation of the variation in relevance between search snippets and the full text of retrieved search results for each unique search term. The black, green, red, and yellow bars correspond to the respective search engines: Ask.com, Bing, Google, and Yahoo. It is important to note that a 5-point Likert scale was employed in this research, implying that a difference score equates to a 20% discrepancy in relevance between the search snippets and the full text of the results.

As evident in Figure 8, out of 360 pairs of search snippet-full text evaluations, only four pairs exhibit an 80% relevance difference (Three pairs of Google and one pair of Bing). Additionally, only 12 pairs of search snippet-full text results demonstrate a 60% relevance disparity. In other words, it is highly improbable that the relevance difference between search snippets and the entire text of the results exceeds 40% across all four search engines. Approximately 40% of the pairs exhibit a 20% relevance difference, while nearly 25% of the pairs exhibit no relevance difference at all.

## 5. Discussion and conclusion

This study focuses on evaluating the summarization algorithms of four prominent international search engines: Ask.com, Bing, Google, and Yahoo. In the research, 360 pairs of search snippets and corresponding full-text results were evaluated by university students, revealing statistically significant differences across all four search engines considered. While Rahim *et al.* (2019) found that Google, Yahoo, and Bing exhibited superior performance, the current research yielded slightly divergent results. Azizan *et al.* (2018) reported that Google outperformed the others, while Bing emerged as the top performer in this study. This discrepancy could be attributed to the varying evaluation criteria employed in these studies.

The results of Wu *et al.* (2019) indicated comparable performance between Bing and Google, with Bing outperforming Ask.com, which aligns with our study's results. Additionally, our findings corroborate those of Ali and Gul (2016), as Google demonstrated superior performance over Yahoo. In contrast, Hajian Hoseinabadi and CheshmehSohrabi (2023) observed a better performance for Google, while Bing attained a higher score in our study. This discrepancy likely stems from differences in evaluation criteria. Overall, our findings align with those of previous studies, consistently demonstrating Google's superior performance among search engines.

As evident from Figure 7, the summarization effectiveness of Ask.com, Bing, Google, and Yahoo search engines is 43.4%, 57.2%, 51.8%, and 47.8%, respectively. Comparing the results of the research with others such as the effectiveness of Google is in the range of 50–57% (Zeynali Tazehkandi and Nowkarizi, 2021); the precision of Google, Yahoo, Bing, DuckDuckGo search engines is 50, 40, 48 and 44%, respectively (Azizan *et al.*, 2018); and the recall of Google, Bing and Yahoo search engines is 36, 31 and 33%, respectively (Gul *et al.*, 2020); So it can be concluded that the summarization algorithm of the search engines is similar to their retrieval algorithm, but according to Wu *et al.* (2019) the coverage of Ask.com, Google and Bing is 80, 90 and 90% respectively.

A comparative analysis of the research findings with those of other studies reveals consistent patterns. Zeynali Tazehkandi and Nowkarizi (2021) reported an effectiveness range of 50–57% for Google, while Azizan *et al.* (2018) found precision scores of 50% for Google, 40% for Yahoo, 48% for Bing, and 44% for DuckDuckGo. Moreover, Gul *et al.* (2020) documented recall rates of 36% for Google, 31% for Bing, and 33% for Yahoo. These results suggest that search engines' abstracting algorithms similar to their retrieval algorithms. However, a different picture emerges when considering coverage metrics. Wu *et al.* (2019) demonstrated that Ask.com, Google, and Bing achieved coverage rates of 80%, 90%, and 90%, respectively. This indicates that search engines excel at indexing and retrieving relevant content, but their ability to accurately summarize and represent information may be less robust.

The search engines provide more comprehensive coverage of information than their effective summarizations, but the summarization algorithms of two search engines (Bing and

Google) are acceptable. All four search engines, Ask.com, Bing, Google, and Yahoo, scored more than 40% in the summarization effectiveness test. Bing scored more than 55%, while Google scored close to 55%, which suggests that they provide more accurate and informative search snippets. Based on these results, users are encouraged to use Bing and Google for their search engine queries. The designers of Ask.com and Yahoo are advised to improve the accuracy and comprehensiveness of their summarization algorithms to provide a better user experience.

**About the authors**

Mahdi Zeynali-Tazehkandi is a Ph. D. candidate in the Department of Library and Information Science (LIS), Ferdowsi University of Mashhad, Iran. Also, he is a reviewer in Sage Publishing, Springer Nature Group and Emerald Publishing. He has published some articles in scholarly journals. His main research interests are evaluation of information retrieval systems, science studies and education of LIS.

Dr Mohsen Nowkarizi is a professor in the Department of Library and Information Science (LIS), Ferdowsi University of Mashhad, Iran. He has authored some publications and articles in professional and scholarly publications. His main research interests are science studies, human-computer interaction and search engines.

Yeganeh Zhian-Feizabadi is a master student in the Department of Library and Information Science (LIS), Ferdowsi University of Mashhad, Iran. Her main research interests are Internet things and evaluation of information retrieval systems.

**Ethical approval**

All procedures (Human consent) in this study were conducted in accordance with the "University review board's (Ref/58476) approved protocols."

**Data availability statement**

The datasets used in this investigation are accessible upon reasonable request from the corresponding author.

**Note**

1. According to the guidelines of the journal, the name of the university has been removed to make the identity of the authors anonymous.

**References**

Al-Eroud, A.F., Al-Ramahi, M.A., Al-Kabi, M.N., Alsmadi, I.M. and Al-Shawakfa, E.M. (2011), "Evaluating Google queries based on language preferences", *Journal of Information Science*, Vol. 37 No. 3, pp. 282-292, doi: 10.1177/0165551511403383.

Ali, S. and Gul, S. (2016), "Search engine effectiveness using query classification: a study", *Online Information Review*, Vol. 40 No. 4, pp. 515-528, doi: 10.1108/OIR-07-2015-0243.

Anastasiu, D.C., Gao, B.J., Jiang, X. and Karypis, G. (2013), "A novel two-box search paradigm for query disambiguation", *World Wide Web*, Vol. 16 No. 1, pp. 1-29, doi: 10.1007/s11280-011-0154-0.

Anuyah, O., Milton, A., Green, M. and Pera, M.S. (2020), "An empirical analysis of search engines' response to web search queries associated with the classroom setting", *Aslib Journal of Information Management*, Vol. 72 No. 1, pp. 88-111, doi: 10.1108/AJIM-06-2019-0143.

Aqle, A., Al-Thani, D. and Jaoua, A. (2022), "Can search result summaries enhance the web search efficiency and experiences of the visually impaired users?", *Universal Access in the Information Society*, Vol. 21 No. 1, pp. 171-192, doi: 10.1007/s10209-020-00777-w.

Azizan, A., Bakar, Z.A., Rahman, N., Masrom, S. and Khairuddin, N. (2018), "A comparative evaluation of search engines on finding specific domain information on the web", *International Journal of Engineering and Technologies*, Vol. 7 No. 4, pp. 1-4, doi: 10.14419/ijet.v7i4.33.23471.

Balabantaray, R.C. (2017), "Evaluation of web search engine based on ranking of results and its features", *International Journal of Information and Communication Technology*, Vol. 10 No. 4, pp. 392-405, doi: 10.1504/IJICT.2017.084337.

Bilal, D. and Huang, L.-M. (2019), "Readability and word complexity of SERPs snippets and web pages on children's search queries", *Aslib Journal of Information Management*, Vol. 71 No. 2, pp. 241-259, doi: 10.1108/AJIM-05-2018-0124.

Bonart, M., Samokhina, A., Heisenberg, G. and Schaer, P. (2020), "An investigation of biases in web search engine query suggestions", *Online Information Review*, Vol. 44 No. 2, pp. 365-381, doi: 10.1108/OIR-11-2018-0341.

Çakir, E., Bahceci, H. and Bitirim, Y. (2008), "An evaluation of major image search engines on various query topics", in *2008 the Third International Conference on Internet Monitoring and Protection*, pp. 161-165, doi: 10.1109/ICIMP.2008.9.

Chen, Y., Jeon, G.Y. and Kim, Y.-M. (2014), "A day without a search engine: an experimental study of online and offline searches", *Experimental Economics*, Vol. 17 No. 4, pp. 512-536, doi: 10.1007/s10683-013-9381-9.

CheshmehSohrabi, M. and Sadati, E.A. (2022), "Performance evaluation of web search engines in image retrieval: an experimental study", *Information Development*, Vol. 38 No. 4, pp. 522-534, doi: 10.1177/02666669211010211.

Connaway, L.S. and Powell, R.R. (2016), *Basic Research Methods for Librarians*, ABC-CLIO. AQ: 9

Dai, N. and Davison, B.D. (2011), "Topic-sensitive search engine evaluation", *Online Information Review*, Vol. 35 No. 6, pp. 893-908, doi: 10.1108/14684521111193184.

Egghe, L. and Goovaerts, M. (2007), "A note on measuring overlap", *Journal of Information Science*, Vol. 33 No. 2, pp. 189-195, doi: 10.1177/0165551506075325.

Fattahi, R., Parirokh, M., Dayyani, M.H., Khosravi, A. and Zareivenovel, M. (2016), "Effectiveness of Google keyword suggestion on users' relevance judgment", *The Electronic Library*, Vol. 34 No. 2, pp. 302-314, doi: 10.1108/EL-03-2015-0035.

Fendley, R. and Kidambi, P. (2010), "Benchmarking Web-based image retrieval", in *Proceedings of the Sixteenth Americas Conference on Information Systems*, Lima, pp. 1446-1458.

Gomez-Nieto, E., Roman, F.S., Pagliosa, P., Casaca, W., Helou, E.S., Oliveira, M.C.F.d. and Nonato, L.G. (2014), "Similarity preserving snippet-based visualization of web search results", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20 No. 3, pp. 457-470, doi: 10.1109/tvcg.2013.242.

Gossen, T., Hempel, J. and Nürnberger, A. (2013), "Find it if you can: usability case study of search engines for young users", *Personal and Ubiquitous Computing*, Vol. 17 No. 8, pp. 1593-1603, doi: 10.1007/s00779-012-0523-4.

Gul, S., Ali, S. and Hussain, A. (2020), "Retrieval performance of Google, Yahoo and Bing for navigational queries in the field of 'life science and biomedicine'", *Data Technologies and Applications*, Vol. 54 No. 2, pp. 133-150, doi: 10.1108/DTA-05-2019-0083.

Gulli, A. and Signorini, A. (2005), "The indexable web is more than 11.5 billion pages", *Special Interest Tracks and Posters of the 14th International Conference on World Wide Web*, pp. 902-903, doi: 10.1145/1062745.1062789.

Hajian Hoseinabadi, A. and CheshmehSohrabi, M. (2023), "Proposing a new combined indicator for measuring search engine performance and evaluating Google, Yahoo, DuckDuckGo, and Bing search engines based on combined indicator", *Journal of Librarianship and Information Science*, Vol. 56 No. 1, pp. 178-197, doi: 10.1177/09610006221138579.

Hartley, J. (2008), *Academic Writing and Publishing: A Practical Handbook*, Routledge, London.

Hu, D., Jiang, S., Robertson, R.E. and Wilson, C. (2019), "Auditing the partisanship of Google search snippets", *The World Wide Web Conference*, San Francisco, CA, Association for Computing Machinery, pp. 693-704, doi: 10.1145/3308558.3313654.

Hussain, A., Gul, S., Nisa, N.T., Shueb, S., Gulzar, F. and Bano, S. (2023), "Image retrieval effectiveness of bing images, Google images and Yahoo image search in the scientific field of

AQ: 10

tourism and COVID-19", *Journal of Information Science*, 01655515231161560, doi: 10.1177/01655515231161560.

Kim, J., Thomas, P., Sankaranarayana, R., Gedeon, T. and Yoon, H.-J. (2017), "What snippet size is needed in mobile web search?", *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, Oslo, Norway, Association for Computing Machinery, pp. 97-106, doi: 10.1145/3020165.3020173.

Kumar, B. and Pavithra, S. (2010), "Evaluating the searching capabilities of search engines and metasearch engines: a comparative study", *Annals of Library and Information Studies*, Vol. 57 No. 2, pp. 87-97.

Kunder, M.d. (2022), "The size of the World Wide Web (The Internet)".

Lancaster, F.W. (1991), *Indexing and Abstracting in Theory and Practice*, Library Association London, London.

Laura, L. and Me, G. (2017), "Searching the Web for illegal content: the anatomy of a semantic search engine", *Soft Computing*, Vol. 21 No. 5, pp. 1245-1252, doi: 10.1007/s00500-015-1857-4.

Lewandowski, D. (2008), "The retrieval effectiveness of web search engines: considering results descriptions", *Journal of Documentation*, Vol. 64 No. 6, pp. 915-937, doi: 10.1108/00220410810912451.

Lewandowski, D. (2015), "Evaluating the retrieval effectiveness of web search engines using a representative query sample", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 9, pp. 1763-1775, doi: 10.1002/asi.23304.

Lewandowski, D. and Sünkler, S. (2019), "What does Google recommend when you want to compare insurance offerings?", *Aslib Journal of Information Management*, Vol. 71 No. 3, pp. 310-324, doi: 10.1108/AJIM-07-2018-0172.

Loan, F.A. and Sheikh, S. (2018), "Is Google scholar really scholarly?", *Library Hi Tech News*, Vol. 35 No. 3, pp. 7-9, doi: 10.1108/LHTN-11-2017-0078.

MacFarlane, A. (2007), "Evaluation of web search for the information practitioner", *ASLIB Proceedings*, Vol. 59 No. 4/5, pp. 352-366, doi: 10.1108/00012530710817573.

Manouchehri, S., Mirzabeigi, M. and Jowkar, T. (2021), "Evaluating the effectiveness of Farsi-English query production using ontology: a case of scientometric ontology", *Aslib Journal of Information Management*, Vol. 73 No. 3, pp. 386-405, doi: 10.1108/AJIM-08-2020-0268.

Marcos, M.-C., Gavin, F. and Arapakis, I. (2015), "Effect of snippets on user experience in web search", in *Proceedings of the XVI International Conference on Human Computer Interaction, Vilanova i la Geltru*, Spain, Association for Computing Machinery, p. Article 47, doi: 10.1145/2829875.2829916.

Meadow, C.T., Boyce, B.R., Kraft, D.H. and Barry, C. (2007), *Text Information Retrieval Systems*, Emerald Publishing.

Mitra, A. and Awekar, A. (2017), "On low overlap among search results of academic search engines", in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 823-824, doi: 10.1145/3041021.3054265.

Niforatos, J.D., Feinstein, M.M. and Pescatore, R.M. (2019), "Search engine queries as a metric of public interest in anesthesia", *Anaesthesia and Intensive Care*, Vol. 47 No. 3, pp. 302-304, doi: 10.1177/0310057X19842574.

Orduña-Malea, E., Luis Ortega, J. and Aguillo, F.I. (2014), "Influence of language and file type on the web visibility of top European universities", *Aslib Journal of Information Management*, Vol. 66 No. 1, pp. 96-116, doi: 10.1108/AJIM-02-2013-0018.

Papenmeier, A., Kern, D., Hienert, D., Sliwa, A., Aker, A. and Fuhr, N. (2021), "Starting conversations with search engines – interfaces that elicit natural language queries", in *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, Canberra, Association for Computing Machinery, pp. 261-265, doi: 10.1145/3406522.3446035.

Pickard, A.J. (2013), *Research Methods in Information*, Facet publishing, London.

Qumsiyeh, R. and Ng, Y.-K. (2014), "Assisting web search using query suggestion based on word similarity measure and query modification patterns", *World Wide Web*, Vol. 17 No. 5, pp. 1141-1160, doi: 10.1007/s11280-013-0235-3.

Rahim, I., Mushtaq, H. and Ahmad, S. (2019), "Evaluation of search engines using advanced search: comparative analysis of yahoo and bing", *Library Philosophy and Practice*, pp. 1-12.

Sadeghi, M. and Vegas, J. (2017), "How well does Google work with Persian documents?", *Journal of Information Science*, Vol. 43 No. 3, pp. 316-327, doi: 10.1177/0165551516640437.

Teevan, J., Cutrell, E., Fisher, D., Drucker, S.M., Ramos, G., André, P. and Hu, C. (2009), "Visual snippets: summarizing web pages for search and revisitation", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2023-2032, Boston, MA, Association for Computing Machinery, doi: 10.1145/1518701.1519008.

Tombros, A. and Sanderson, M. (1998), "Advantages of query biased summaries in information retrieval", *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, pp. 2-10, doi: 10.1145/290941.290947.

Tsay, M.-y., Wu, T.-l. and Tseng, L.-l. (2017), "Completeness and overlap in open access systems: search engines, aggregate institutional repositories and physics-related open sources", *PLoS One*, Vol. 12 No. 12, e0189751, doi: 10.1371/journal.pone.0189751.

Turpin, A., Tsegay, Y., Hawking, D. and Williams, H.E. (2007), "Fast generation of result snippets in web search", in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, Association for Computing Machinery, pp. 127-134, doi: 10.1145/1277741.1277766.

Uyar, A. and Aliyu, F.M. (2015), "Evaluating search features of Google knowledge graph and bing satori", *Online Information Review*, Vol. 39 No. 2, pp. 197-213, doi: 10.1108/OIR-10-2014-0257.

Uyar, A. and Karapinar, R. (2017), "Investigating the precision of Web image search engines for popular and less popular entities", *Journal of Information Science*, Vol. 43 No. 3, pp. 378-392, doi: 10.1177/0165551516642929.

Wu, S., Zhang, Z. and Xu, C. (2019), "Evaluating the effectiveness of Web search engines on results diversification", *Information Research: An International Electronic Journal*, Vol. 24 No. 1. AQ: 11

Xue, X.-B., Zhou, Z.-H. and Zhang, Z. (2008), "Improving Web search using image snippets", *ACM Transactions on Internet Technology*, Vol. 8 No. 4, pp. 1-28, doi: 10.1145/1391949.1391955.

Yalun, M.T. and Bitirim, Y. (2013), "An evaluation of Turkish retrieval performance of popular search engines for Internet and image search by using common lists", *Proceedings of the Conference on Digital Information and Communication Technology and its Applications*, Ostrava, pp. 148-153.

Zavadski, A. and Toepfl, F. (2019), "Querying the Internet as a mnemonic practice: how search engines mediate four types of past events in Russia", *Media, Culture and Society*, Vol. 41 No. 1, pp. 21-37, doi: 10.1177/0163443718764565.

Zeynali Tazehkandi, M. and Nowkarizi, M. (2021), "Evaluating the effectiveness of Google, Parsijoo, Rismoon, and Yooz to retrieve Persian documents", *Library Hi Tech*, Vol. 39 No. 1, pp. 166-189, doi: 10.1108/LHT-11-2019-0229.

Zhang, J. and Fei, W. (2010), "Search engines' responses to several search feature selections", *The International Information and Library Review*, Vol. 42 No. 3, pp. 212-225, doi: 10.1080/10572317.2010.10762866.

Zhang, J., Cai, X., Le, T., Fei, W. and Ma, F. (2019), "A study on effective measurement of search results from search engines", *Journal of Global Information Management*, Vol. 27 No. 1, pp. 196-221, doi: 10.4018/JGIM.2019010110.

**Corresponding author**
Mohsen Nowkarizi can be contacted at: mnowkarizi@um.ac.ir AQ: 6

For instructions on how to order reprints of this article, please visit our website:
**www.emeraldgrouppublishing.com/licensing/reprints.htm**
Or contact us for further details: **permissions@emeraldinsight.com**

View publication stats