Multimedia Tools and Applications (2024) 83:75995–76033
https://doi.org/10.1007/s11042-024-18559-4

End-to-end pseudo relevance feedback based vertical web search queries recommendation

Tajmir Khan[1] · Umer Rashid[1] · Abdur Rehman Khan[2]

Received: 9 April 2023 / Revised: 26 October 2023 / Accepted: 29 January 2024 /
Published online: 21 February 2024

Abstract

Nowadays, the web has emerged as an enormous multimedia data resource. Social media platforms are becoming the mass producers of user-generated multimedia content. Web search engines usually organize media-specific information, such as text, images, video, etc., in specialized data repositories (verticals), providing easy access to multimedia content. Web search engines have become efficient in retrieving data across various distinct verticals. However, users still need help formulating queries to retrieve the relevant multimedia results from verticals. Novice users often issue short-length ambiguous queries due to a lack of domain knowledge or query formulation experience, resulting in the retrieval of irrelevant results. The query formulation itself is a time-consuming process for the users. We presented an end-to-end deep-learning automatic query recommendation approach to address the associated issues. The proposed method autonomously extracts the domain knowledge using pseudo-relevance feedback, transforms it into a unified text-to-text summary, and assists users in generating non-ambiguous and well-balanced query recommendations. The proposed system employs Google's real-time dataset and is compared to the Google search engine. The evaluation consists of empirical and usability perspectives. The empirical evaluation of shorter query formulation and reformulation time obtained 89% accuracy scores in automated query recommendation. The usability testing (N=37) reveals 85.4% usefulness & ease-of-use, and "A" category proposed system usability.

Keywords Deep learning · Query recommendation · Multimedia information retrieval · Natural language processing · Vertical web search · Pseudo relevance feedback

1 Introduction

In the early days, content on the web was limited in quantity, having fewer modalities of information associated with them [1]. The ability of the web and the internet to easily connect people across the world has resulted in a surging usage of social media websites [2]. As a

✉ Umer Rashid
umerrashid@qau.edu.pk

[1] Department of Computer Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan

[2] Department of Computer Science, National University of Modern Languages, Lahore 54000, Pakistan

result, the number of internet users reaches approximately 2.89 billion, which is 42.3% of the overall world human population [3]. Users often share text, audio, images, and video on social media platforms such as Facebook, Instagram, Twitter, YouTube, Flicker, etc., causing an immense growth of multimedia data [2]. In the early days, users generated approximately one petabyte of data daily [4]. In 2012, it increased to approximately 2.5 exabytes, and in the next year, it was 4.4 zettabytes approximately. A recent study in 2020 confirmed 40 zettabytes of multimedia content on the web [5].

Presently, users interact with the web to satisfy their information needs [6]. The information gap of a user intrigues this information-seeking journey [7]. However, the exponential multimedia growth creates challenges in retrieving relevant information [8]. Web search engines are the main gateway for accessing proliferating information [8]. The search engines take in the user information needs in the form of queries that contain specific keywords related to the required topics [7]. Subsequently, the search engines retrieve and present the search results matching the keywords in the user queries with minimal response time [6]. Therefore, search engines have become an entry point to quickly find relevant multimedia information due to their ability to retrieve the appropriate results from immense piles of multimedia content [6, 9].

However, the retrieval of relevant information mainly depends on the quality of the user query issued to a search engine [7]. Users of web search engines are often reported to need more prior domain knowledge and search expertise [7]. Subsequently, the users may issue generic queries consisting of a few keywords that challenge the query-matching algorithms in the retrieval of the immense information [3]. The users find navigating ample information space difficult [7]. As a result, recent users' information-seeking behavior reports state that 65% of the users reformulate queries, and among them, 78% removed one word, and the 91% of the users added at least one term in query modification [10, 11]. However, the complexity of the search task also increases query complexity up to 27.61% [12].

To ease this problem, the web search engines provide query recommendations to formulate or refine their queries and assist users in effective query formulation [3]. The literature defines ambiguous queries as a primary reason for unsuccessful information retrieval since the expression of information needs is inadequate, and the quality of the submitted query is essential to retrieve the relevant information [3, 13]. However, query recommendation in terms of "query auto-completion" or recommendation of semantically relevant queries may better satisfy the users' information needs [3]. The research presented in this article focuses on the latter type of query recommendation.

The conventional query auto-completion systems rely upon historical query logs to suggest potential query completion candidates [14]. The users requiring expertise in domain knowledge are less satisfied with the auto-completion paradigm due to their unfamiliarity with the vocabulary terms [14]. The post-query recommendation systems identify the user intent based on their initial query [15]. This implicit contextual information captured by the algorithms is often referred to as Pseudo Relevance Feedback (PRF) [15]. Few studies investigate the effectiveness of query formulation assistance provided by the existing state-of-the-art search engines on the queries generated via the PRF [3]. The query formulation algorithms and deep learning approaches can be employed in web search engines to recommend advanced and semantically related queries in retrieval [3].

Hence, the recent research focuses on enhanced user interaction with the information. These include intuitively visualizing the search results [1], re-modeling the user interaction with the post-retrieved search results [7], and user usability analysis [16]. However, at present, the effectiveness of a search engine depends on the quality of the search queries issued. Contrarily, the users' information needs are becoming exploratory and vague, with unclear search goals [7]. Therefore, there exists a research gap that unveils the user factors in query formulation and the effect of query-assisting approaches in enhancing the user searching experience. While the existing approaches evaluate the proposed approach from the empirical perspectives [7], there is a need to perform detailed user behavioral and usability analysis via standard usability protocols.

1.1 Problem statement

Due to the exponential multimedia growth on the web, users face the retrieval of relevant information on web search engines challenge [8]. The search engines retrieve and present the search results by matching the keywords in the user queries [6]. However, a successful retrieval depends on the relevant query formulation that encapsulates domain knowledge and retrieval expertise [3]. Contrarily, the users' information needs are becoming exploratory. The search goals often are vague and undefined. Consequently, users issue short-type and ill-defined search queries, resulting in either retrieval of irrelevant or overly generic search results. Hence, there exists a need for an approach that assists users in formulating well-articulated queries to enhance search retrieval performance and gain domain knowledge, without added effort at the users' end.

1.2 Research objectives

Recent research indicates that users face difficulty in finding relevant information during informational search tasks via the existing vertical web search engines. The existing techniques need to emphasize presenting techniques that extend the capabilities of the existing state-of-the-art general-purpose web search engines. Previous approaches employed content recommendation techniques [17], visualization systems [1], and data restructuring [7] to improve the results retrieval. However, the existing web search engines can retrieve relevant and precise information in response to user queries if a well-articulated query is issued. However, queries inadequately convey significant aspects of the searchers' intent, which degrades the search engines' retrieval performance. Therefore, our primary objective is to assist the users by recommending well-balanced and high-impact queries to retrieve relevant information and complete the search task with minimum effort. In this research, we:

- Investigated a query recommendation approach to generate high-impact queries for general-purpose vertical web search engines.
- Performed a comparative analysis of system usability and searching efforts in the proposed approach and the existing general-purpose vertical web search engines.
- Analyzed key similarities and differences in various query recommendation techniques.

1.3 Research contributions

This research deploys state-of-the-art deep learning techniques using a PRF approach to get the user's implicit intent, aid the user in well-articulated query formulation, and retrieve useful research results. Specifically, our research contributions include the following:

- conceptualizing and implementing PRF-based architectural approach that encapsulates generic and domain-independent, state-of-the-art deep learning neural models.
- Analyzing and reporting comparative usability and behavioral analysis of the proposed approach and the existing general-purpose vertical web search engine.

- Aggregating and analyzing key similarities and differences in the existing and proposed query recommendation techniques.

This research proposed an approach to modify the user queries in vertical web search using PRF. Primarily, we generated queries to assist novice and expert users in query formulation using top-k documents as PRF. It is a well-known method for addressing query intention mismatching problems between the retrieved results and users' information needs. Therefore, one of the most effective ways is to bridge the gap between user query intent and actual results [18]. We applied deep learning memorization and query generation models to generate high-impact user queries. Our investigation reveals that the system-generated queries show a high resemblance with queries entered by the domain experts. Finally, to further assist users, we provide a complete query table to select the query easily and retrieve the relevant information for a large corpus in the vertical web search engines. The proposed solution augments the existing search engine's capabilities and therefore requires no particular search expertise or training at the user's end.

The proposed approach is further evaluated empirically in terms of generated query accuracy and qualitatively to measure users' searching efforts, information gain, and system usability. Our empirical experimental results with the baseline system (Google search engine) obtain above 80% query recommendation accuracy. The within-subjects usability analysis (N=37) via standard usability instruments unveils more than twice as much reduced searching efforts than the traditional search paradigms, and 88% of perceived usefulness by the users when compared with the existing search engine. To the best of our knowledge, prior research has yet to be evaluated against existing search engines to determine the effectiveness of the suggested query recommendation mechanism.

The rest of the discussion is organized as follows. Section 2 provides the literature review. Section 3 presents our proposed approach. Section 4 explains the instantiation details of the proposed approach. Section 5 discusses the evaluation of the proposed approach. Section 6 provides in-depth insight into the obtained results, and finally, Section 7 concludes our discussion and highlights future research directions.

2 Literature review

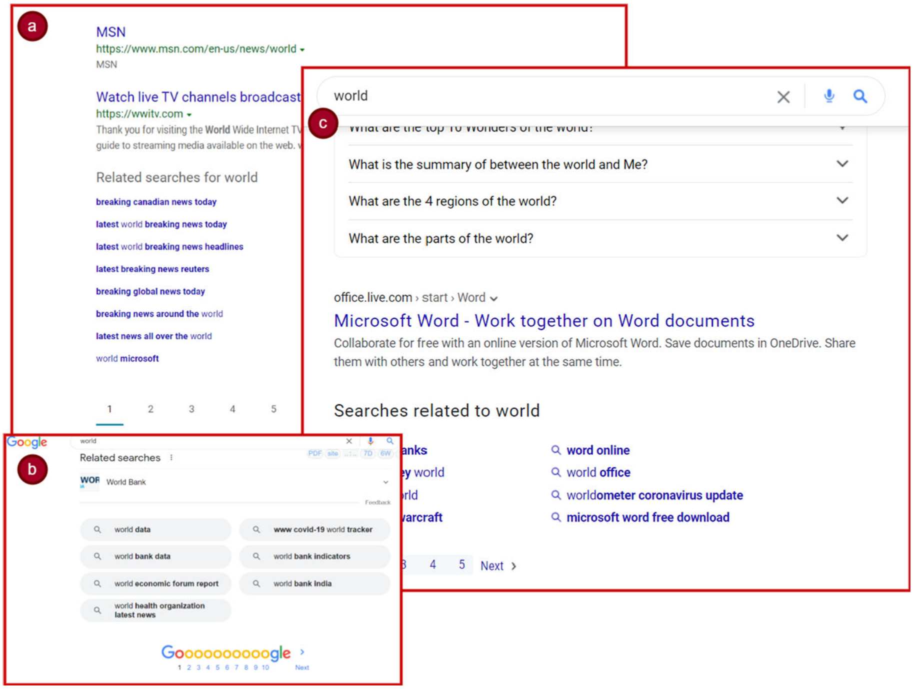
The vertical web search engines are emerging as retrieval gateways to retrieve relevant information from extensive multimedia sources that are further categorized into distinct verticals (i.e., web, image, video, news, etc.) [1, 19–21]. However, multimedia document retrieval over the web is still challenging since vertical web search engines are known to suffer significantly in the case of short, ambiguous, and ill-defined query expressions [20, 22]. The users often express dissatisfaction with the retrieved multimedia content [23]. To ease this problem, researchers have devised numerous approaches that aid users in retrieving relevant information by enhancing the ill-defined query. To further improve the efficacy, the query enhancement approaches often incorporate users' relevance feedback to recommend well-articulated queries that may help in the retrieval of precise information [24]. The detailed discussion on the query enhancement, relevance feedback, and query recommendation approaches, along with the state-of-the-art techniques, are discussed in the subsequent subsections.

2.1 Query enhancement

The information seekers usually express information needs through textual keywords over vertical web search engines [25]. The query representations are subsequently submitted to the search engines to rank the list of retrieved relevant results [25]. The users' ill-defined, short, and ambiguous retrieval expressions jeopardize the entire retrieval process [20, 26]. In query enhancement, the query modification and recommendation techniques address the query expression shortcomings [27]. The former automatically adds extra keywords and phrases with the existing queries to narrow down the search scope [27]. The latter refine query terms after the inputs of initial queries by the users [28]. Ahmad et al. presented an architecture using the standard seq-2-seq model to retain words in query reformulation [29]. They encode information in search actions and context to facilitate document ranking and query suggestion tasks. Chen et al. used recurrent neural networks to capture queries in the current user session and model the short-term context to predict future queries [30]. Ahmad et al. used multi-task neural session relevance to predict users' result clicks and future queries [31]. They used RNN with an attention mechanism to model users' search behavior effectively, while the attention mechanism captured user preferences. The present state-of-the-art web search engines currently adopt a similar approach. Figure 1 depicts query expression and recommendation over vertical web search engines. Mainly, at the end of the search engine result page, a list of recommended queries is provided by mainstream search engines such as Bing Google (Fig. 1(a)), Google (Fig. 1(b)), Yahoo! (Fig. 1(c)), etc. These query terms often incorporate additional keywords from various domains to help the users specify their search intent to retrieve precise results.

2.2 Relevance feedback

The explicit relevance feedback is usually captured by considering the user's mentioned preferences and selected documents [26]. Jeffery et al. developed a decision-theoretic framework that collects user feedback to improve the query results [32]. Balakrishnan et al. used a combination of users' comments, ratings, and referrals as indicators to retrieve the relevant results [33]. Their work transformed the user query into triplets (object, attribute, and value) to clarify ambiguous queries. Jayarathna et al. devised a technique to extract user relevance feedback to highlight the relevant text on the selected pages [34]. They profiled the users' interests based on aggregating implicit and semi-explicit user interest data across multiple everyday applications. Alternatively, the implicit feedback obtains the user interest and preference by observing users' seeking and interaction behavior [26]. Xu et al. used users' eye movement via an eye-tracking camera to obtain explicit feedback from the users [35]. The notion was to acquire user attention times and predict the user's attention through data mining over a new online item. Su et al. inferred the relative importance of information mappings, type, and contextual similarity, to better align with the searcher's intent [36]. They transformed queries into knowledge graphs that can convert logic, natural, and exemplary queries. The graph query is subsequently used as an intermediate to support question answering. Stai et al. proposed an approach based on information type and context on the graph. The former to infer the parent-child relationship between two concepts (e.g., video and multimedia), and the latter to imply the concept similarity (e.g., building and real estate), respectively [37]. The PRF uses top-ranked document keywords to boost information retrieval performance, reduce vocabulary mismatch, and update the query model [15, 38]. Zamani et al. retrieved top-*k* documents from an initial query and considered them pseudo-relevant documents [38]. They



MSN
https://www.msn.com/en-us/news/world
MSN

Watch live TV channels broadcast
https://wwitv.com
Thank you for visiting the World Wide Internet TV
guide to streaming media available on the web.

Related searches for world
breaking canadian news today
latest world breaking news today
latest world breaking news headlines
latest breaking news reuters
breaking global news today
breaking news around the world
latest news all over the world
world microsoft

Related searches
World Bank
world data
www covid-19 world tracker
world bank data
world bank indicators
world economic forum report
world bank India
world health organization latest news

What are the top 10 wonders of the world?
What is the summary of between the world and Me?
What are the 4 regions of the world?
What are the parts of the world?

office.live.com › start › Word
Microsoft Word - Work together on Word documents
Collaborate for free with an online version of Microsoft Word. Save documents in OneDrive. Share them with others and work together at the same time.

Searches related to world
anks
y world
rld
arcraft
word online
world office
worldometer coronavirus update
microsoft word free download

Fig. 1 Screenshot of query recommendation on (a) Bing, (b) Google, and (c) Yahoo! search engine accessed on January 2022

used matrix factorization techniques to predict the weight of query terms for relevant item retrieval. The deep learning approaches have recently enhanced the traditional approaches. Almasri et al. created a corpus from retrieved top-k documents and formed a deep learning vector compromising terms similarity relationships using cosine measure [39]. They used deep learning vectors to capture similar query terms in sizeable unstructured text. It facilitated query terms expansion for retrieval of relevant information. Similarly, Keikha et al. mapped each query onto a set of related Wikipedia articles that collectively represent the search query's semantics in query expansion [40].

2.3 Query recommendation

The users have intentions to access relevant information from the internet. However, the search engine may need to quickly assess the searchers' intent [41]. The Query recommendation helps users refine new queries based on previous queries. The recommended queries relevant to the user's past queries improve the retrieval process and satisfy the user information needs [24, 30, 41–45]. Query recommendation removes the query ambiguities; however, a few users may need to become more familiar with relevant vocabulary terminology and cannot formulate effective queries. The query recommendation provides relevant and succinct queries to users [46]. The web search engines present the related queries to guide the users in search sessions by employing the $seq-2-seq$ model and query-aware attention mechanisms [44]. Li et al. suggested initial queries as a part of the user's search behavior. They further proposed a personalized query recommendation model based on the knowledge graph and metadata to recommend queries [47]. Alternatively, Zhang et al. [48], Chen et al. [28], and Cai et al. [49] exploited the diversification concept in query recommendation by covering the user's divergent exploration intent in exploration feedback. Mustar et al. discussed and compared the transformer-based model recommendation, and recurrent neural network-based model in query recommendations [50]. Bodigut et al. employed a deep Reinforcement Learning model and session-based user feedback to generate high-quality related queries [51]. They trained a Deep Reinforcement Learning model based on long-term session-based user feedback, syntactic relatedness, and estimated naturalness of generated query to predict the user's following query. The Existing research has used neural models along with the PRF to enhance the performance of query recommendations [15].

2.4 State-of-the-art

Wang et al. combined relevance feedback and semantic matching. They computed the relevance matching by weighting the query and document at the lexical level using the traditional BM25 method and chose n-top most documents in re-ranking [18]. Yu et al. employed a binary classification model based on Bidirectional Encoder Representations from Transformers (BERT) to establish semantic relevance between the query and the documents. They retrieved top-k documents from a dense retrieval model and built a PRF query encoder using the BERT encoder to refine the query representation [52]. Yu et al. also proposed a graph-based embedding method for the PRF using intra-node attention to contextualize the query according to each document [53]. Keikha et al. performed PRF using unsupervised and supervised term selection on the Wikipedia dataset. In unsupervised learning, they segmented the query based on decreasing permutations and retrieved the results matching the most prolonged permuted pattern using various ranking algorithms. Supervised learning trained a classifier to apply the most influential ranking algorithm given the query and the retrieved document set [40]. Valcarce et al. proposed a matrix decomposition problem involving the inter-term similarity matrix computation to expand the original query. Afterward, they used linear least squares regression with regularisation to solve the decomposition problem [54].

2.5 Issues and motivation

The explicit and implicit relevance feedback mechanisms require additional users' efforts (as responses) in searching and analysis of complex interaction scenarios, respectively. The query modification may restrict the continuity of users' search activities. Alternatively, the PRF may improve the overall retrieval effectiveness without sacrificing individual query performance [55]. Users have intentions to access relevant information from the internet. However, the vertical web search engine encounters challenges in assessing the users' intentions in the traditional query-response interaction paradigm [56]. The vertical web search engines usually auto-correct and modify the user queries by analyzing the users' query history [41]. Furthermore, the users often need help incorporating domain-relevant terminologies in their queries. As a result, query recommendation paradigms may provide irrelevant and less precise queries to the users [46].

Recent studies have shown that information-seeking activity requires integration of the seeker perspective as the PRF and the 5W's (what, why, when, where, and how) questions in query recommendations [57]. However, a query recommendation model based on 5W's is yet to be studied, and most of the existing approaches are trained on specific datasets [40]. Additionally, query recommendation models are usually compared with offline systems instantiated over static datasets. Additionally, the previous approaches are not extensively evaluated on the state-of-the-art vertical web search engines [40]. In this research, our objective is to investigate a deep learning-based neural network architecture and 5W's scheme of query formulation to obtain satisfactory query recommendations over vertical web search engines.

3 Query recommendation approach

We proposed a novel query recommendation approach by exploiting the PRF and a deep learning technique to capture users' intent and automatically predict the relevant queries for the users in online search activities over the vertical web search engine (Google). The following sections discuss the approach preliminaries, formalization, architecture, and instantiation.

3.1 Preliminaries

We propose a PRF mechanism to recommend queries in vertical web searches. Our approach targets the retrieval of the multimedia data that is the most relevant to the user queries. The users may give queries, and our approach modifies the ambiguous queries iteratively in an interactive way. Firstly, the user provides a pivot (start-up) query in the form of keywords (Fig. 2(a)). The textual query is passed to a vertical web search engine. Top-k documents from each vertical are selected to extract the user intent from the documents (Fig. 2(b)). Afterward, we extracted the metadata from the retrieved contextual information, performed text pre-processing by employing a Latent Semantic Analyzer (LSA), and generated the PRF summarized report (Fig. 2(c)). The PRF summarized report is further passed to a 5W Text-to-Text Transfer Transformer (T5) query generator deep neural model, which outputs a pool of potential candidate queries based on 5W (e.g., Who, What, Where, When, and Why) questions (Fig. 2(d)). The pool of candidate queries is fed to the Bidirectional Encoder Representations from Transformers (BERT) using the Siamese network text transformer neural model to extract the contextual information and perform semantic similarity matching with the contextually transformed user queries (Fig. 2(e)). Finally, a list of high-impact system-generated queries is shown to the user in descending order of their similarities against a pivot query (Fig. 2(f)). This research uses pre-trained[1] deep learning models. Hence, enabling the proposed approach to be domain-independent and eliminating the requirement of complex hyper-parameters tuning.

3.2 Formalization

Let C be the set of query formulation, context retrieval, PRF generator, context extractor, query generator, and query suggestion components, given as $C = \{Q_f, C_r, P_g, C_e, Q_s, Q_r\}$. The Q_f inputs a user query q and forms a mapping with $Q_f \rightarrow C_r \rightarrow f_m(q)$. Let M be the set of retrieved multimedia verticals $M = \{W, I, V\}$, where W, I, and V are the set of web, image, and video verticals retrieved using q. Associated with each vertical is the metadata given as $\forall \{W\} \exists \{t, u, d\}$. The t denotes the title, u denotes the URL, and d denotes the description. For image and video vertical, this additionally includes $\forall \{I, V\} \exists \{t, u, d, p\}$, where p includes thumbnail preview. We retrieve n results based on the initial context query

[1] Pre-trained model employed in this research uses the default python library configuration.

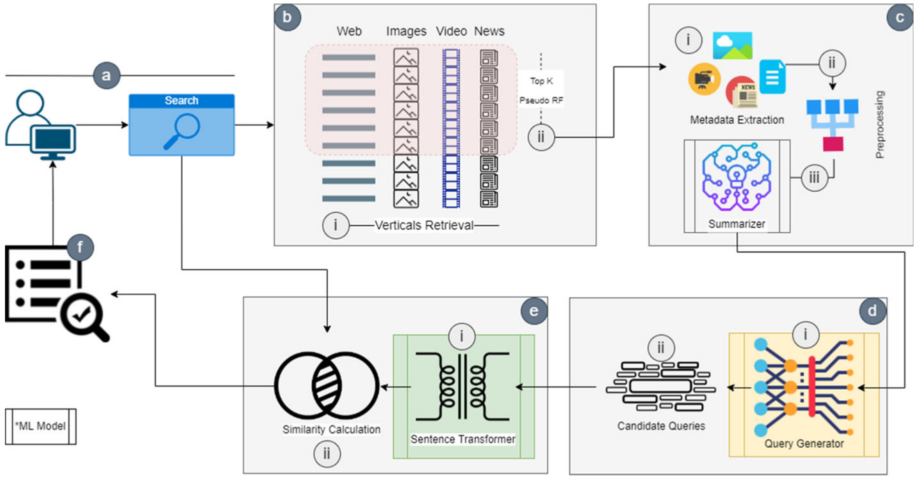


Fig. 2 Preliminaries to modify the user queries in vertical web search using the PRF

q, given as M = $\sum_{i=1}^{n}\{W_i, I_i, V_i\}$. To set initial context for P_g, we extracted top-k elements from the set M, given as the context set $C = \sum_{i=1}^{k}\{W_i^{\{t,d\}}, I_i^{\{t,d\}}, V_i^{\{t,d\}}\}$. Let $k = m$, where m is the number of retrieved results from a particular vertical. $\forall \,|\, c \,|\, \varepsilon \, C \geq |\alpha|$, we removed special characters, embedded URLs, and encoded HTML characters if the length of c is greater than a defined threshold $\alpha\varepsilon\mathbb{N}$, then we apply deep learning algorithm on the textual content in C_e to generate a set of the PRF summary S, where $|S| = 3m$. The S was then transformed into 386 dimensional sentence embedding via BERT model $\vec{E} = \{\forall s\varepsilon S|0 \leq s \leq 1 \wedge |s| = 386\}$ and passed to $Q_e \rightarrow f(E)$, generating a set of candidate queries Q_c. Afterward, we performed a pairwise similarity operation given as a dictionary mapping $Q_r = D : \{sim(q, \forall Q_c) \wedge \{d_i \in \mathbb{N} : d_i \geq d_{i+1}\}\}$, where sim is cosine similarity measure, and D is the similarity score sorted in descending order.

3.3 Architecture

The proposed architecture consists of six main components. These are; query formulation, context retrieval, relevance feedback generator, query generator, context extractor, and query suggestion components (Fig. 3). Mainly, we employed component object model notation to represent the architecture of the proposed approach. The boxes represent components, the boxes with two small rectangles represent modules, and the simple rectangular boxes represent interfaces. The dashed arrow lines show the dependency. The simple circle demonstrates intra-connectivity (among different modules), and the circle with the moon symbol indicates inter-connectivity (between an interface and a module). The starting state consists of only the outgoing edge from the circle symbol, and the end state denotes an undirected line with a circle and a dot symbol.

3.3.1 Query formulation

The query formulation component inputs the query terms from the user as keywords or phrases to start a new search session. The query keywords are dispatched to the search

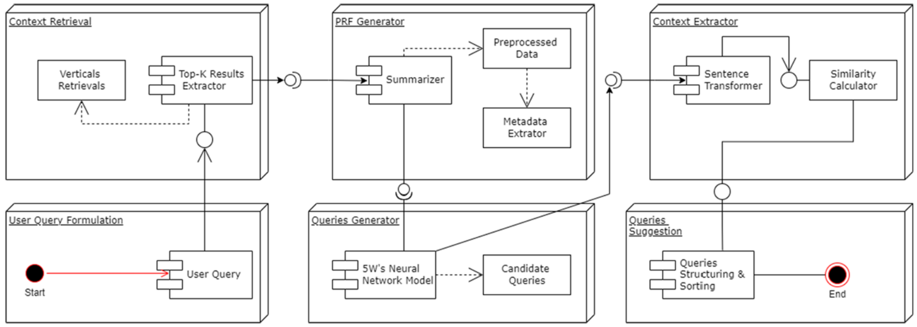


Fig. 3 Component-based proposed architectural diagram outlining components and their relationship

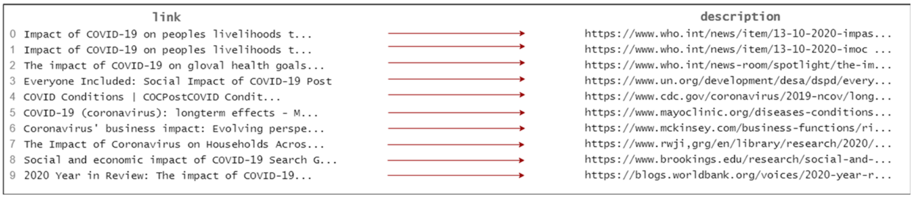
engine to retrieve the search results in real time. As opposed to generalizing the user intent via historical user-generated logs data, we used the user's initial intent as a PRF to start the user search journey. Therefore, this component is a starting point for the proposed architecture's data flow.

3.3.2 Relevance feedback generation

The relevance feedback generation component captures the user intends as a pivot user query given at start-up time. Particularly, against the user-issued query, the search results are crawled from the Google search engine in real-time from multiple verticals (image, video, and web). The retrieved results are stored as in-memory tables. The associated metadata is organized in a separate memory table for individual verticals. The rows in the memory table comprise the results, and the columns represent the features related to each result, such as title, description, thumbnail, etc., as shown in Fig. 4. The relevance feedback component considers the top-k items from each vertical as a piece of relevant information as PRF. This contextual information is then passed to the PRF generator component to extract users' intent.

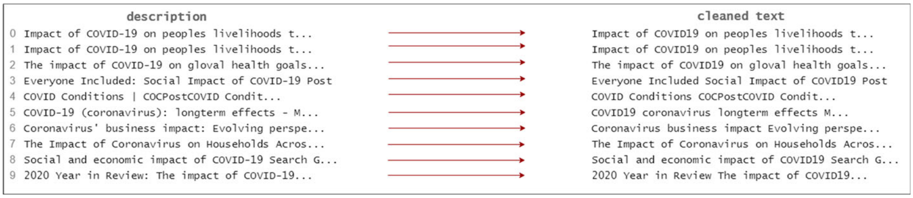
3.3.3 Relevance feedback extraction

The relevance feedback extraction component uses the search results in the web, image, and video verticals and generates a summarized PRF report. This summarized report is stored as in-memory tables. The associated metadata is organized in a separate memory table for individual verticals. The rows in the memory table comprise the results, and the columns represent the features related to each result, such as title, description, thumbnail,



```
                      link                                                          description
0 Impact of COVID-19 on peoples livelihoods t...      ───→      https://www.who.int/news/item/13-10-2020-impas...
1 Impact of COVID-19 on peoples livelihoods t...      ───→      https://www.who.int/news/item/13-10-2020-imoc ...
2 The impact of COVID-19 on gloval health goals...    ───→      https://www.who.int/news-room/spotlight/the-im...
3 Everyone Included: Social Impact of COVID-19 Post   ───→      https://www.un.org/development/desa/dspd/every...
4 COVID Conditions | COCPostCOVID Condit...           ───→      https://www.cdc.gov/coronavirus/2019-ncov/long...
5 COVID-19 (coronavirus): longterm effects - M...     ───→      https://www.mayoclinic.org/diseases-conditions...
6 Coronavirus' business impact: Evolving perspe...    ───→      https://www.mckinsey.com/business-functions/ri...
7 The Impact of Coronavirus on Households Acros...    ───→      https://www.rwji,grg/en/library/research/2020/...
8 Social and economic impact of COVID-19 Search G...  ───→      https://www.brookings.edu/research/social-and-...
9 2020 Year in Review: The impact of COVID-19...      ───→      https://blogs.worldbank.org/voices/2020-year-r...
```

Fig. 4 The verticals storage memory table. The rows represent each search results and the columns represent features associated with each result



	description		cleaned text
0	Impact of COVID-19 on peoples livelihoods t...	→	Impact of COVID19 on peoples livelihoods t...
1	Impact of COVID-19 on peoples livelihoods t...	→	Impact of COVID19 on peoples livelihoods t...
2	The impact of COVID-19 on gloval health goals...	→	The impact of COVID19 on gloval health goals...
3	Everyone Included: Social Impact of COVID-19 Post	→	Everyone Included Social Impact of COVID19 Post
4	COVID Conditions | COCPostCOVID Condit...	→	COVID Conditions COCPostCOVID Condit...
5	COVID-19 (coronavirus): longterm effects - M...	→	COVID19 coronavirus longterm effects M...
6	Coronavirus' business impact: Evolving perspe...	→	Coronavirus business impact Evolving perspe...
7	The Impact of Coronavirus on Households Acros...	→	The Impact of Coronavirus on Households Acros...
8	Social and economic impact of COVID-19 Search G...	→	Social and economic impact of COVID19 Search G...
9	2020 Year in Review: The impact of COVID-19...	→	2020 Year in Review The impact of COVID19...

Fig. 5 The verticals storage memory table after pre-processing via relevance feedback extractor component

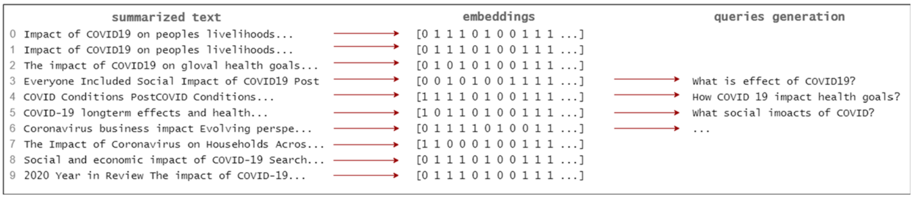
etc., as shown in Fig. 4. Afterwards, multiple operations are performed over the data stored in memory tables. Firstly, the data cleansing operation is performed on the data stored in memory tables. Specifically, special characters, URLs, and encoded HTML characters are removed to ensure the exploitation of system-generated keywords as near-natural language substitutes. The cleaned text is associated with the document in the memory tables, as shown in Fig. 5. The token length of the cleaned text is validated and summarized via the Latent Semantic Analyzer (LSA). This detailed summarized report eliminates the redundancy in the information and encapsulates all the necessary contextual information concisely to facilitate near-human-like query recommendations.

3.3.4 Query generation

The query generation component inputs the summarized documents and maps the summarized words to embedding sequences. These embeddings are further passed to the questions generation model that outputs an arrangement of words in the querying format. These queries follow the pattern of Harold Lasswell's well-known "5W" (who, when, what, where, why the) communication model [58]. We used a T5 sequence-to-sequence pretrained transformer-based deep learning questions generation model. Specifically, the summarized PRF report is broken into atomic words called tokens and mapped to an embedding sequence. These embeddings are passed to the feed-forward neural network and output a text sequence of words in the querying format. These queries follow the pattern of Harold Lasswell's well-known "5W" (who, when, what, where, why) model of communication [58]. When adequately augmented with domain-specific keywords, such questions allow a search engine to effectively set the context for retrieving relevant information (e.g., entity, place, date, etc.), which is otherwise challenging to assemble manually. This output is stored in the form of vectors, as shown in Fig. 6.

3.3.5 Context extraction

Finding the most relevant queries for recommendation to the user after generating a pool of candidate queries is challenging. The context extractor component retrieves the 5W-generated



	summarized text		embeddings		queries generation
0	Impact of COVID19 on peoples livelihoods...	→	[0 1 1 1 0 1 0 0 1 1 1 ...]		
1	Impact of COVID19 on peoples livelihoods...	→	[0 1 1 1 0 1 0 0 1 1 1 ...]		
2	The impact of COVID19 on gloval health goals...	→	[0 1 0 1 0 1 0 0 1 1 1 ...]		
3	Everyone Included Social Impact of COVID19 Post	→	[0 0 1 0 1 0 0 1 1 1 1 ...]	→	What is effect of COVID19?
4	COVID Conditions PostCOVID Conditions...	→	[1 1 1 1 0 1 0 0 1 1 1 ...]	→	How COVID 19 impact health goals?
5	COVID-19 longterm effects and health...	→	[1 0 1 1 0 1 0 0 1 1 1 ...]	→	What social imoacts of COVID?
6	Coronavirus business impact Evolving perspe...	→	[0 1 1 1 1 0 1 0 0 1 1 ...]	→	...
7	The Impact of Coronavirus on Households Acros...	→	[1 1 0 0 0 1 0 0 1 1 1 ...]		
8	Social and economic impact of COVID-19 Search...	→	[0 1 1 1 0 1 0 0 1 1 1 ...]		
9	2020 Year in Review The impact of COVID-19...	→	[0 1 1 1 0 1 0 0 1 1 1 ...]		

Fig. 6 Queries generation after applying T5 model via query generation component

queries and computes their similarities with the base query (user's start-up query). Instead of using generic similarity measures such as cosine similarity, we used a semantics-based similarity measure on each candidate query and the user's initial intent to find the semantic relatedness of the candidate query. Hence, the context extractor component retrieves the generated queries from the vector and computes their similarities with the pivot query via the sentence transformer model. The sentence transformer translates the 5W-generated queries into a 384-dimensional feature space embedding that encapsulates inherited textual context and semantics. Based on the lower 384-dimensional dense vector space, we performed a pairwise cosine similarity measure on each candidate query against the user's initial intent to find the semantic relatedness of the candidate query and discard the irrelevant queries if their similarity score is below a certain threshold. In this case, we took 0.50 as a threshold value. The queries below the 0.50 similarity score are eliminated since they denote more irrelevancy than relevancy [7]. Finally, the similarity associated with each query is stored in a dictionary, as shown in Fig. 7.

3.3.6 Query recommendation

The query recommendation component sorts the potential query recommendations in descending order of their similarity scores computed by the context extractor component. The most relevant query is therefore placed in the top-most position. These queries represent the users' interests according to their initial intent. Users can choose one of the recommended queries to retrieve the most relevant multimedia results. These queries represent the user's interests according to their initial intent. A user can choose one of the recommended queries to retrieve the most relevant multimedia results. The queries recommender dictionary reduces the burden on the users by assisting in the formulation of well-balanced queries. The queries recommender dictionary reduces the burden on the users by assisting in formulating well-balanced queries. Table 1 provides the query statistics generated by top-100 search results in each vertical, equally divided into a result set of 10. Overall, an average of 69 questions were generated from the aggregated web, image, and news verticals, with an average of 21 questions per vertical. The query terms used to calculate the statistics are the same as in Table 3.

4 Instantiation

The proposed architecture is instantiated by considering the empirical and usability perspectives. The practical perspective includes developing a back-end model to support the research objectives. The usability is leveraged by instantiating a traditional web search engine-like user interface that provides user interaction with an inherited back-end model. Furthermore, a detailed system walk-through explains the back-end model's linkage with the corresponding

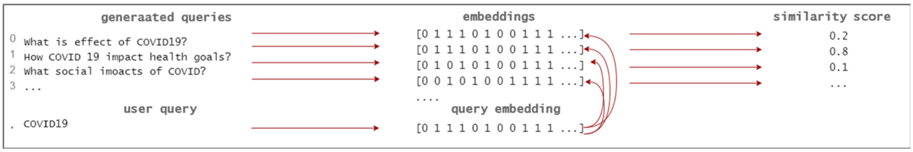


Fig. 7 Calculating queries semantic similarity scores in context extraction component via sentence embedding

Table 1 System suggested queries recommendation statistics against top-100 queries divided equally into the set (S) of 10

Query	Vertical	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Total/Average/Std.
Q1	text	4	3	2	2	3	3	4	3	4	2	30/3/0.8
	image	1	2	2	2	2	2	2	2	3	2	20/2/0.4
	video	2	1	2	3	1	3	2	2	2	2	20/2/0.6
Total/Average		7/2.3	6/2	6/2	7/2.3	6/2	8/2.7	8/2.7	7/2.3	9/3	6/2	70/23.3/1.3
Q2	text	3	1	3	3	2	3	2	3	2	2	24/2.4/0.7
	image	2	2	2	1	1	2	2	2	2	2	18/1.8/0.4
	video	2	2	2	1	1	2	2	3	1	2	18/1.8/0.6
Total/Average		7/2.3	5/1.7	7/2.3	5/1.7	4/1.3	7/2.3	6/2	8/2.7	5/1.7	6/2	60/20/1.2
Q3	text	2	3	3	3	2	4	2	3	3	2	27/2.7/0.6
	image	2	2	2	2	2	2	3	1	1	1	18/1.8/0.6
	video	2	2	1	2	3	2	2	2	2	2	20/2/0.4
Total/Average		6/2	7/2.3	7/2.3	7/2.3	7/2.3	8/2.7	7/2.3	6/2	6/2	5/1.7	65/21.7/0.8
Q4	text	2	2	2	3	3	3	3	3	3	3	27/2.7/0.5
	image	2	2	2	2	1	2	2	2	2	2	19/1.9/0.3
	video	2	2	1	2	2	3	1	3	2	2	20/2/0.6
Total/Average		6/2	6/2	5/1.7	7/2.3	6/2	8/2.7	6/2	7/2.3	7/2.3	7/2.3	66/22/0.8
Q5	text	2	2	3	2	3	3	3	3	3	3	27/2.7/0.5
	image	2	2	3	2	2	2	2	2	2	2	21/2.1/0.3
	video	1	2	2	2	3	2	3	3	3	3	24/2.4/0.7

Table 1 continued

Query	Vertical	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Total/Average/Std.
Total/Average		5/1.7	6/2	7/2.3	6/2	8/2.7	7/2.3	8/2.7	8/2.7	8/2.7	8/2.7	72/24/1.0
Q6	text	2	2	3	3	3	3	3	3	3	3	28/2.8/0.4
	image	2	1	3	2	2	1	2	2	2	2	19/1.9/0.5
	video	2	2	2	1	2	2	2	2	2	1	18/1.8/0.4
Total/Average		6/2	5/1.7	8/2.7	6/2	7/2.3	6/2	7/2.3	7/2.3	7/2.3	6/2	55/18.3/0.8
Q7	text	2	2	2	1	2	2	2	2	2	2	19/1.9/0.3
	image	2	2	3	3	3	3	3	3	3	3	28/2.8/0.4
	video	2	2	1	2	1	2	3	1	2	3	19/1.9/0.7
Total/Average		6/2	6/2	6/2	6/2	6/2	7/2.3	8/2.7	6/2	7/2.3	8/2.7	66/22/0.8
Q8	text	3	2	2	3	2	3	3	3	3	3	27/2.7/0.5
	image	2	1	2	2	2	2	2	2	2	2	19/1.9/0.3
	video	2	1	2	2	2	2	2	1	1	1	16/1.6/0.5
Total/Average		7/2.3	4/1.3	6/2	7/2.3	6/2	7/2.3	7/2.3	6/2	6/2	6/2	62/20.7/0.9
Q9	text	2	2	2	3	3	3	3	3	3	3	27/2.7/0.5
	image	1	1	1	1	1	2	2	1	1	1	12/1.2/0.4
	video	2	1	2	1	2	2	2	2	2	2	18/1.8/0.4
Total/Average		5/1.7	4/1.3	5/1.7	5/1.7	6/2	7/2.3	7/2.3	6/2	6/2	6/2	57/19/0.9
Q10	text	3	2	2	2	3	3	3	2	2	2	24/2.4/0.5
	image	2	2	2	1	1	2	1	2	3	2	18/1.8/0.6
	video	2	2	2	2	1	2	1	1	3	2	18/1.8/0.6
Total/Average		7/2.3	6/2	6/2	5/1.7	5/1.7	7/2.3	5/1.7	5/1.7	8/2.7	6/2	60/20/1.2/1.0
Grand Total / Mean Average / Mean Std.												**69.3/21.1/0.1**

instantiated search user interface. A detailed discussion of each is provided in the subsequent sections.

4.1 Tool development

We implemented a full-fledged tool to suggest the queries for the Google search results retrieved from three distinct verticals, i.e., web, image, and video. The tool is primarily implemented in Python 3.7 programming language and Django framework[2]. Django offers a comprehensive solution based on a software architecture and layered design convention. In this approach, the model layer handles the architectural components' implementation details summarized in Table 2. The template layer deals with web interface design using HTML, CSS, and JavaScript, further elaborated in the subsequent subsection. The view layer acts as an intermediary, taking user inputs from the template and passing them to the model for processing.

4.2 User interface & usage scenario

The proposed end-to-end tool deploys a user-friendly conventional search engine-like interface design that primarily provides iterative query recommendations over Google search results. The interface design primarily provides iterative query recommendations for the Google search results retrieved from multiple verticals. The objective is to satisfy users' information needs over vertical web search engines by recommending various queries in an iterative query recommendation search session. We illustrate a usage scenario of user interaction with the proposed automatic query recommendation mechanism in Fig. 8.

As it emerges from Fig. 8, a user instantiates a search session by giving the textual query "Impact of covid 19" in the search bar of the designed user interface (Fig. 8(a)). The initial query is dispatched to the Google search engine (Fig. 8(b)). In response, web, image, and video verticals are retrieved from Google in real-time (Fig. 8(c & d)). The web search results are, by default, presented on the main page of the tools' interface in distinct verticals. In addition, the main page also displays the list of relevant suggested queries in a separate query table (Fig. 8(e)).

The user selects the query "What is covid 19 impact on societies and economies?" and top-ranked relevant search results are presented in a separate Google search engine results page (SERP) to the users (Fig. 8(b)). We want to clarify that each query contains a Google icon in front of it. When the user clicks on this icon, the query goes to the Google interface, and retrieved results are displayed on the Google SERP. The user can also view the updated results based on the selected query from the query table on the tools' main page (Fig. 8(e)).

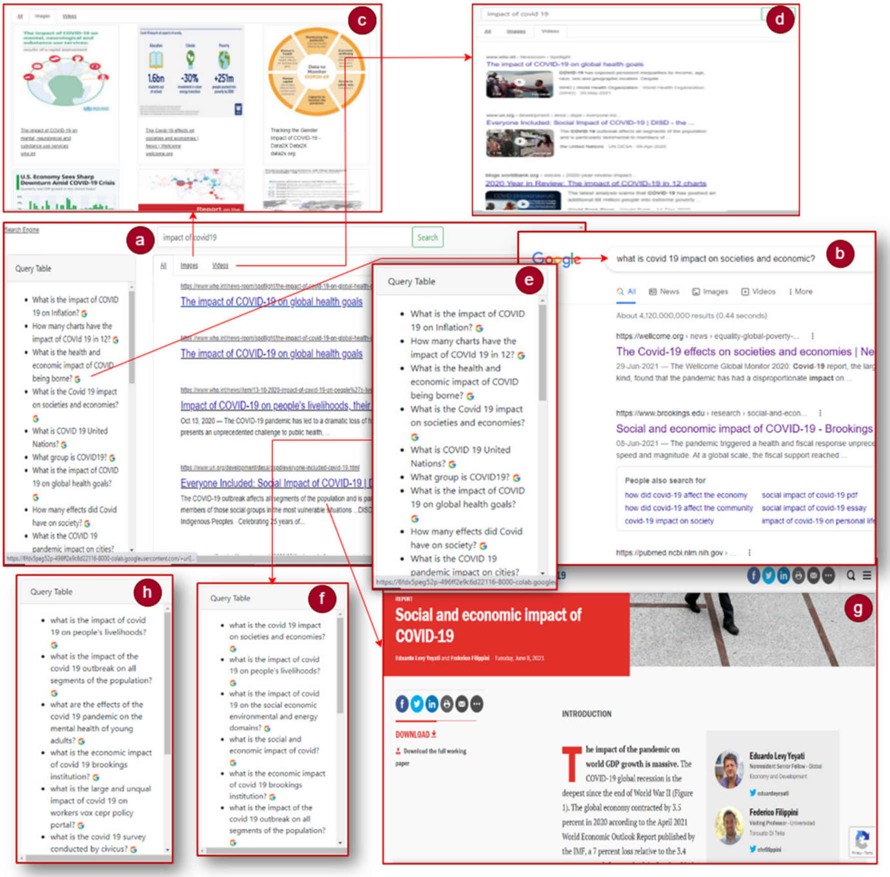
As the user clicks "What is covid 19 impact on societies and economy?" the results are presented in the Google SERP, the user selects the document of their choice, and the selected document is displayed in a separate browser window (Fig. 8(g)). As the user selects a query from the set of suggested queries (Fig. 8(f)), the query table is updated automatically based on the PRF computed over the updated set of the document retrieved via the newly selected user query from the query table (Fig. 8(h)). The procedure continues, and the system refreshes the query table based on the user selection of a query from the query table iteratively until the satisfaction of the user's information needs.

[2] https://www.djangoproject.com/

Table 2 Employed model architecture, input shape, output, and pretraining dataset

Model name	Architecture	Input shape	Output	Training dataset
Sumy[a]	Latent semantic analyzer	results x terms	Summarized sentences	–
T5[b]	Seq-2-Seq transformers	summarized sentences	5W's questions generation	SQuADv1 [59]
MiniLM-L6-v2[c]	Sentence transformers	5W's questions generation	384-D semantic embedding	STS [60]

[a] https://pypi.org/project/sumy/
[b] https://github.com/patil-suraj/question_generation
[c] https://huggingface.co/sentence-transformers/paraphrase-MiniLM-L6-v2



Search Engine

impact of covid19

Query Table

- What is the impact of COVID 19 on Inflation?
- How many charts have the impact of COVId 19 in 12?
- What is the health and economic impact of COVID being borne?
- What is the Covid 19 impact on societies and economies?
- What is COVID 19 United Nations?
- What group is COVID19?
- What is the impact of COVID 19 on global health goals?
- How many effects did Covid have on society?
- What is the COVID 19 pandemic impact on cities?

The impact of COVID-19 on global health goals

The impact of COVID-19 on global health goals

Impact of COVID-19 on people's livelihoods, their

Oct 13, 2020 — The COVID-19 pandemic has led to a dramatic loss of h presents an unprecedented challenge to public health, ...

https://www.un.org/development/desa/dspd/everyone-included-covid-19.html

Everyone Included: Social Impact of COVID-19 | D

The COVID-19 outbreak affects all segments of the population and is pa members of those social groups in the most vulnerable situations ...DISD Indigenous Peoples · Celebrating 25 years of...

what is covid 19 impact on societies and economic?

About 4,120,000,000 results (0.44 seconds)

https://wellcome.org › news › equality-global-poverty-...

The Covid-19 effects on societies and economies | Ne

29-Jun-2021 — The Wellcome Global Monitor 2020: **Covid-19** report, the larg kind, found that the pandemic has had a disproportionate **impact** on ...

https://www.brookings.edu › research › social-and-econ...

Social and economic impact of COVID-19 - Brookings

08-Jun-2021 — The pandemic triggered a health and fiscal response unprece speed and magnitude. At a global scale, the fiscal support reached ...

People also search for

how did covid-19 affect the economy	social impact of covid-19 pdf
how did covid-19 affect the community	social impact of covid-19 essay
covid-19 impact on society	impact of covid-19 on personal life

https://pubmed.ncbi.nlm.nih.gov › ...

U.S. Economy Sees Sharp Downturn Amid COVID-19 Crisis

Tracking the Gender Impact of COVID-19 - Data2X Data2X data2x.org

impact of covid 19

The impact of COVID-19 on global health goals

Everyone Included: Social Impact of COVID-19 | DISD - the ...

2020 Year in Review: The impact of COVID-19 in 12 charts

Query Table

- What is the impact of COVID 19 on Inflation?
- How many charts have the impact of COVId 19 in 12?
- What is the health and economic impact of COVID being borne?
- What is the Covid 19 impact on societies and economies?
- What is COVID 19 United Nations?
- What group is COVID19?
- What is the impact of COVID 19 on global health goals?
- How many effects did Covid have on society?
- What is the COVID 19 pandemic impact on cities?

Query Table

- what is the covid 19 impact on societies and economies?
- what is the impact of covid 19 on people's livelihoods?
- what is the impact of covid 19 on the social economic environmental and energy domains?
- what is the social and economic impact of covid?
- what is the economic impact of covid 19 brookings institution?
- what is the impact of the covid 19 outbreak on all segments of the population?

REPORT

Social and economic impact of COVID-19

Eduardo Levy Yeyati and Federico Filippini · Tuesday, June 8, 2021

DOWNLOAD

Download the full working paper

INTRODUCTION

The impact of the pandemic on world GDP growth is massive. The COVID-19 global recession is the deepest since the end of World War II (Figure 1). The global economy contracted by 3.5 percent in 2020 according to the April 2021 World Economic Outlook Report published by the IMF, a 7 percent loss relative to the 3.4

Eduardo Levy Yeyati
Nonresident Senior Fellow - Global Economy and Development
eduardoyeyati

Federico Filippini
Visiting Professor - Universidad Torcuato Di Tella
efefilippini

Query Table

- what is the impact of covid 19 on people's livelihoods?
- what is the impact of the covid 19 outbreak on all segments of the population?
- what are the effects of the covid 19 pandemic on the mental health of young adults?
- what is the economic impact of covid 19 brookings institution?
- what is the large and unqual impact of covid 19 on workers vox cepr policy portal?
- what is the covid 19 survey conducted by civicus?

Fig. 8 Usage scenario via the proposed query recommendation mechanism depicting the user query, verticals, query table, results page, and accessing of information from the actual sources

5 Evaluation

We measured the proposed query recommendation approach using empirical, usability, and behavioral perspectives. The empirical analysis is conducted via accuracy measure. We employed and measured standard system usability and satisfaction instruments in usability evaluation. The user behavior analysis is performed by analyzing users' micro-level (e.g., clicks, keystrokes, time, etc.) interaction and information-seeking activity. The obtained results are further compared with the Google search engine. A detailed discussion of each aspect of the evaluation is provided in the following sections.

5.1 Dataset & benchmarks

The query recommendation system can be employed over the choice of various datasets. These include Text REtrieval Conference (TREC) [15, 18, 40, 52–54], MECANEX's platform storage unit [37], MS MARCO passage training data [52, 53], Associated Press collection [54], WT10G collection [54], GOV2 collection [54], and user search-query logs [51], etc. Similarly, the proposed queries recommendation system was bench-marked and compared against a range of generic systems developed based on algorithms; for example, Term frequency-inverse documents frequency, BM25, neural information retrieval models [15, 33], etc. The Google search engine can also be used as a baseline [33].

We chose Google as our baseline benchmark system and used its real-time dataset, which served twofold purposes. Firstly, the real-time dataset allowed the users to interact with the search results retrieved from Google in real time. Secondly, it allowed in-depth evaluation of the baseline system. Unlike previous approaches that used generic baseline systems and offline datasets, this type of evaluation setup created a realistic comparative environment, uncovering inherent strengths and weaknesses of the existing widely used state-of-the-art web search engine and allowing direct comparability with the proposed system.

5.2 Query accuracy scores & word length

Empirical evaluation of a search system instantiated over a real dataset is challenging since the data is retrieved in real-time and is often unlabeled [7]. In this scenario, a human evaluator expert must label the data and establish a ground truth [7]. Therefore, the association of ground truth with the system-recommended queries in real-time to evaluate the accuracy of the proposed approach is complex [7]. Existing research requires recruiting two human experts to eliminate the subjectivity of the experts and biases in the obtained accuracy scores [7]. Hence, we hired two human experts to address the challenge of manually annotating the recommended queries as relevant and irrelevant to establish ground truth relevancy in such a scenario. The human evaluator experts were graduate degree holders. The first human expert had 12+ years of librarianship experience, and the second had under-graduate level teaching experience of 5+ years. The former human evaluator expert evaluated the queries' relevancy from the technical perspective, whereas the latter human evaluator expert evaluated the queries' relevancy from the general users' perspective.

Firstly, the system-recommended queries are sorted in descending order of similarity scores. The queries were generated via the initial sample of ill-defined intent of the users (pivot queries) presented in Table 3. The human evaluator experts encoded their response to each system-recommended query in the binary, where 1 denoted relevancy and 0 represented irrelevancy of the recommended query to the user intent. The existing studies mostly use top-10 results for measuring precision [61].

Table 3 List of initial ill-defined queries are provided to the system as PRF to retrieve results and generate system-suggested queries

No.	Query
Q_1	White cat
Q_2	Information retrieval state of the art
Q_3	Amazing birds
Q_4	Bugatti
Q_5	The political issues
Q_6	Challenges in machine learning
Q_7	Role of technology in science
Q_8	Impact of covid
Q_9	Height of Asian people
Q_{10}	Road accident

The human evaluator experts were given the top-20 queries surpassing the defined cosine similarity score threshold $\alpha > 0.5$ to establish the query relevancy further using accuracy measures. We ignored recommended queries having similarities less than α since we considered them irrelevant [7]. From the obtained truth table, we calculated the accumulative accuracy score in the iteration of 5-steps, as shown in Table 4.

Accuracy is the fraction of relevant items to the overall retrieved items. The formulation of accuracy scores is calculated as $\sum_{Q=1}^{n}\sum_{RQ=1}^{k}\frac{r}{k} \therefore r = \{0|1\}$. The Q denotes the query topics presented in Table 3, which are dispatched to the search engine to retrieve the result set. The RQ denotes the recommended queries generated by the proposed system based on the retrieved dataset. For each query Q, we calculated the accumulative accuracy of the system-generated query based on the relevant feedback provided by the evaluator experts. Finally, we calculated the mean average accuracy score by averaging the experts' overall accuracy scores.

The overall 89% mean average accuracy score is achieved. The average and cumulative accuracy scores for each top-k query against all queries in Table 3 are presented in Fig. 9. Based on ground truth values obtained from the human expert, we measured the relevance of the system-recommended queries by a human expert in terms of the accuracy scores. The accuracy scores were computed using an expert's fraction of relevant recommended queries to the overall system recommended queries. We calculated the cumulative precision score of top-20 system-generated queries to measure our approach's effectiveness.

In addition to measuring system-recommended queries' accuracy, we also measured the word length in the recommended queries. This length is categorized into stopwords, non-stopwords, and the total number of words that may exist in the query. The non-stopwords indicate the essential keywords in a query, whereas the stopwords denote less influential query keywords [62]. The stopwords are the words that convey no special meaning (e.g., is, the, an, of, etc.). The non-stop words are independently meaningful, such as world, football, machine, etc. The total number of terms is a sum of stopwords and non-stopwords present in the query. Overall, a mean average length of top-20 queries is recorded that is 3.69, 6.55, and 10.25 words for stopwords, non-stopwords, and total words, respectively. According to the existing literature, an optimal balance query length should be between $5 - 10$ words [63]. Therefore, the proposed system recommended optimally balanced queries to users. The detailed query length scores are elaborated in Table 5.

5.3 Usability evaluation

A usability study is a research method used to assess the effectiveness of a proposed system observing real users' interactions and collecting behavioral data. We followed the protocols of standard usability measures during user evaluation. We chose Google as a baseline, which holds several purposes. Firstly, recent studies show Google as the most preferred search engine by users [64]. Secondly, existing users are proficient in using Google for daily search searching [7]. Hence, we chose Google to determine the effectiveness of our proposed approach. To eliminate biases, we performed multiple standard protocols including a within-subjects study design and alternating system evaluation order to eliminate the order-induced learning effect. Secondly, we used the same dataset in both systems. Furthermore, the search results were retrieved from Google in real-time against user-issued queries to eliminate dataset-induced biases [65]. Moreover, the users of a new system are often reported to be resilient to accept a new system [66]. Hence, the direct comparison between Google and the proposed system is deemed to be a rigorous testing protocol to evaluate the actual effectiveness of the

Table 4 Query recommendation table showing cumulative accuracy scores obtained from the human experts (E) for each top-k query generated against the initial ill-defined user query

Top-k	#	Q_1		Q_2		Q_3		Q_4		Q_5		Q_6		Q_7		Q_8		Q_9		Q_{10}	
		E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2
5	RQ_1	1	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1
	RQ_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	RQ_3	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	RQ_4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	RQ_5	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
	Avg. Accuracy	1	1	0.8	1	0.8	1	1	1	1	1	0.8	1	1	0.6	1	1	1	1	1	1
10	RQ_6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	RQ_7	1	1	0	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
	RQ_8	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
	RQ_9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
	RQ_10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
	Avg. Accuracy	1	1	0.8	1	0.9	1	0.8	1	0.8	1	0.9	1	1	0.8	1	1	0.9	1	1	1
15	RQ_11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
	RQ_12	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1
	RQ_13	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
	RQ_14	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1
	RQ_15	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Avg. Accuracy	1	1	0.73	0.93	0.93	1	0.73	1	0.73	1	0.87	1	1	0.87	1	1	0.66	1	1	1
20	RQ_{16}	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1
	RQ_{17}	1	1	0	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
	RQ_{18}	1	1	1	0	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1
	RQ_{19}	1	0	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1
	RQ_{20}	1	1	0	1	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1
	Avg. Accuracy	1	0.95	0.65	0.85	0.9	1	0.55	1	0.55	1	0.9	1	0.95	0.85	1	1	0.65	1	0.95	1
Mean Avg. Accuracy		$E_1 = 0.81 E_2 = 0.97$																			
Overall Mean Avg. Accuracy		0.89																			

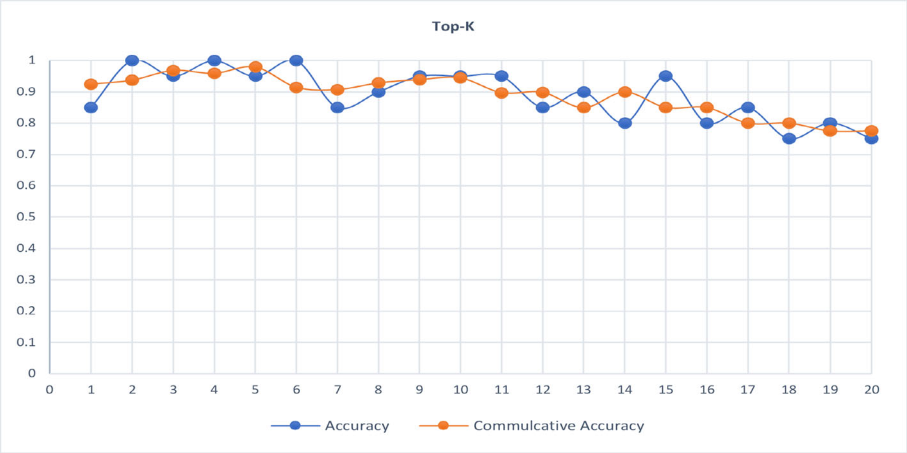


Fig. 9 Recommended queries average and cumulative accuracy chart against queries evaluated by the experts at each top-k position

proposed system. A detailed discussion of each protocol of usability evaluation employed in this research including selection of users, evaluation apparatus and tasks, study conduction procedure, and choice of usability measuring instruments are discussed in the subsequent subsections.

5.3.1 Users

A minimum sample size of $12 - 14$ users is required to analyze the performance metric success rate and uncover usability issues [67]. We recruited 37 users, including 72.9% males and 27.1% females, to evaluate the proposed system. The users were recruited via an open advertisement, had no prior interest in this research, and were English proficient. The users were selected based on their willingness to perform the evaluation and were not incentivized to participate. The selected sample of users belongs to diverse backgrounds to generalize the implication of the proposed approach. The users were mainly students (43.24% users), freelancers (10.81% users), developers (2.70% users), teachers (16.21% users), lab assistants (2.70% users), professional (e.g., engineer, clerk, and technical assistants) employees working in industries (24.32% users). All the users were information literate and could express their information needs and find relevant information from the web via search engines [68]. They could identify, find, evaluate, and use information effectively. The minimum, maximum, and average age of users are 20, 47, and 27.7 years, respectively.

The users were organized via a within-subjects study design. This organization allows a user to utilize both, the proposed and the baseline system. Hence the users are able to more precisely determine which system adapted effectively to their needs and hence is more widely used and well-known for users' comparative analysis [1, 69]. The learning effect was countered by randomly alternating the order of the system a user evaluated.

5.3.2 Apparatus

The experimentation was conducted on a desktop computer containing Intel(R) Core(TM) $i7 - 6700$ CPU @ 3.40GHz processor, 8GB DDR3 RAM, and 64-bit Windows-10 operating

Table 5 Recommended query words length including stopwords, non-stopwords, and total words

#	Stop words/Non-Stopwords/Total Words									
	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}
RQ_1	4/4/8	4/4/8	6/13/19	0/6/6	3/5/8	2/4/6	5/5/10	3/6/9	4/6/10	3/6/9
RQ_2	5/5/10	5/5/10	5/6/11	3/5/8	4/6/10	3/6/9	2/6/8	4/7/11	4/5/9	2/5/7
RQ_3	5/8/13	4/5/9	2/8/10	2/4/6	8/10/18	4/10/14	2/6/8	3/8/11	4/5/9	3/7/10
RQ_4	2/5/7	3/2/5	4/10/14	1/4/5	4/6/10	3/7/10	0/2/1	3/9/12	4/6/10	5/9/14
RQ_5	1/6/8	5/1/6	2/7/9	3/5/8	6/7/13	4/7/11	2/4/6	5/7/12	3/7/10	3/7/10
RQ_6	3/6/9	3/3/6	5/5/10	4/7/11	4/8/12	2/7/9	5/7/12	5/7/12	1/4/5	5/6/11
RQ_7	3/7/10	6/3/9	7/7/14	3/4/7	4/5/9	4/6/10	6/8/14	3/5/7	2/5/7	6/12/18
RQ_8	4/6/10	7/4/11	3/7/10	4/7/11	3/6/9	4/6/10	2/6/8	3/8/11	3/4/7	5/8/13
RQ_9	1/4/5	5/1/6	3/4/7	3/6/9	4/6/10	4/8/12	4/6/10	1/5/6	3/6/9	2/12/14
RQ_{10}	1/6/7	7/1/8	5/7/12	3/4/7	5/6/11	5/10/15	5/6/11	3/9/12	5/5/10	4/9/13
RQ_{11}	3/7/10	4/3/7	0/15/15	2/8/10	3/4/7	5/10/15	3/3/6	4/6/10	5/5/10	4/7/11
RQ_{12}	2/5/7	4/2/6	7/6/13	3/6/9	0/4/4	3/9/12	9/8/17	4/6/10	4/6/10	3/7/10
RQ_{13}	2/13/15	1/2/3	5/7/12	3/5/8	0/9/9	3/8/11	7/8/15	4/7/11	3/5/8	6/8/14
RQ_{14}	1/9/10	6/1/7	0/10/10	4/5/9	8/8/16	8/6/14	5/6/11	3/8/11	5/8/13	5/11/16
RQ_{15}	2/6/8	5/2/7	2/6/8	4/7/11	3/19/22	8/8/16	2/8/10	9/10/19	3/5/8	5/11/16
RQ_{16}	1/11/12	2/1/3	5/5/10	7/5/12	7/8/15	3/7/10	1/6/7	4/9/13	3/6/9	2/3/5
RQ_{17}	7/6/13	8/7/15	3/5/8	5/5/10	5/9/14	1/10/11	4/6/10	3/7/9	3/6/9	1/14/15
RQ_{18}	5/7/12	4/5/9	1/6/7	2/7/9	4/5/9	2/7/9	3/7/10	4/9/13	4/6/10	5/8/13
RQ_{19}	0/13/12	3/0/3	3/8/11	0/6/6	3/6/9	1/9/10	7/8/15	8/7/15	10/15/25	3/7/10
RQ_{20}	1/7/8	4/1/5	3/7/10	1/6/7	5/6/11	4/7/11	5/7/12	6/9/15	2/9/11	3/5/8
Avg. Length	2.65/7.2/7.2	4.5/2.65/2.65	3.55/7.45/7.45	2.85/5.6/5.6	4.15/7.15/7.15	3.65/7.6/7.6	3.95/6.15/6.15	4.1/7.45/7.45	3.75/6.2/6.2	3.75/8.1/8.1
Mean Avg. Length		3.69/6.55/10.25								

system. The system was connected to a 19 inches widescreen LCD monitor with a $1366x786$ pixels resolution. Google Chrome web browser facilitated the user's interaction with the proposed tool via a standard keyboard and mouse. We recorded user sessions using a free screen recording software[3] to analyze users' information-seeking and interaction behavior.

5.3.3 Tasks

In total, 4 topics were designed. The user had to select the most unfamiliar topic from the available list of topics and search for information about the chosen topic to eliminate user bias induced by prior knowledge. Each topic contained a set of identically structured questions using which a user had to explore the search space and synthesize the information from various sources [70]. We further defined 6 sub-tasks about the focused topic within each topic. These sub-tasks were designed using a combination of less complex lookup-based (who, when, and where, etc.) and more complex exploratory-based (what, how, and why, etc.) tasks [22]. The user was required to search each question and synthesize their answer in a written form from the first 4 tasks. The last 2 questions instructed the user to download a picture or watch a video clip about the topic. These instructions were comprehensive and designed ambiguously to examine the level of assistance provided to the user in the system-recommended queries. Table 6 shows these 4 topics and planned sub-tasks.

The users were asked to select the least familiar task for the proposed and baseline system. We assumed that the users' less familiarity with the selected topics might induce ill-defined queries in the search activities. The users' least knowledge about the selected topic may cause extended exploration activities. In the proposed system, the $Task_1$, $Task_2$. $Task_3$, and $Task_4$ were chose by 21.6%, 16.2%, 16.2%, and 45.9% of the users, respectively. For the baseline system, 21.6% of the users chose $Task_1$, 21.6% of the users chose $Task_2$, 32.4% of the users chose $Task_3$, and 24.5% of the users chose $Task_4$ for evaluation.

5.3.4 Instruments

In this research, we are primarily interested in evaluating the systems' usability, ease-of-use and usefulness, and users' perceived system effectiveness. The usability of a system is verified using a standard System Usability Scale (SUS) instrument [71]. The SUS contains ten questions, each consisting of a 5-point Likert scale, where each item is scored between the values 1 and 5. One is subtracted from the user response from the odd-numbered items, and for the even-numbered items, the maximum value (i.e., 5) is subtracted from the user responses. The scale values are in 0 to 4 intervals, with four being the most positive response. Finally, the converted responses are summed for all the items and multiplied by 2.5. This conversion outputs the range of possible values from 0 to 100.

The user ease-of-use and usefulness are measured using the After Scenario Questionnaire (ASQ) [72]. The ASQ also consists of 5-points Likert scale items where 1 denotes minimum and 5 denotes maximum user satisfaction. This score can be represented as an average raw score and additionally in terms of percentage. To measure the subjective effectiveness of the system-recommended queries on the proposed and the baseline system, we designed a custom 5-points (1 minimum and 5 maximum) Likert scale-based questionnaire consisting of 11 items (Table 7). This scale was measured categorically and further validated using the Content Validity Index (CVI) technique proposed in literature [73]. To calculate CVI, we hired two human experts from the statistical domain with graduate statistics degrees and

[3] https://screenrec.com/

Table 6 Search tasks employed to evaluate the query recommendation in the proposed approach and Google search engine

Task 1: War in Afghanistan.	**Task 2: Covid 19.**
- What is the name of the war in Afghanistan?	- When did the covid19 pandemic start in Pakistan?
- What were the causes of the Afghanistan war?	- How covid-19 spread?
- Who were the war rivals in Afghanistan?	- What are the SOPs to prevent the covid19?
- What are the names of current famous warlords in Afghanistan?	- What are the symptoms of covid-19?
- Give me insight related to bomb blast on children in Afghanistan.	- Download an image of the prime minister of Pakistan during vaccination.
- Can I get the clips of oath-taking ceremony of Taliban cabinet.	- Watch the video while lunching the covid19 vaccines of PakVac.
Task 3: Wildlife in Africa.	**Task 4: Pandora papers.**
- Which Is the biggest forest in Africa?	- When the Pandora papers were leaked?
- What is the greatest threat to the wildlife of Africa?	- What is Pandora paper?
- Which animals are endangered in Africa?	- How many documents are there in the Pandora Papers?
- Name the Well-known animals of African wildlife?	- Who leaked the Pandora papers?
- Download the images of top five hunted animals in Africa.	- Download image of Pandora paper memes.
- Watch videos of lions attacking crocodiles.	- Clips of Pandora papers.

data analysis expertise. The CVI measures the clarity and relevance of the items designed concerning the questionnaire objective to be achieved. Each item is scored between 1 to 4, where 1 denotes strong disagreement and 4 denotes strong agreement in relevancy and

Table 7 Recommended queries effectiveness measurement questionnaire

Q. No.	Item
1	I could easily find the {feature}?
2	The recommended {feature} has been produced by the system as per my query response.
3	The {feature}, produced by the proposed system is helpful for me.
4	The{feature} saved my time.
5	I got my relevant materials after choosing an option from the {feature}.
6	The {feature} covers all my query aspects.
7	The {feature} presents my domain of interest.
8	The {feature} helped me to understand the topic.
9	The {feature} contains the queries related to my task.
10	I am sure that this {system} suggested queries meet my requirements, which otherwise would not be met by the other system.
11	The {feature} in {system} added a suitable term to my query.

The tag {system} denotes the current system in use (baseline/proposed), and the {feature} denotes the "query table" for the proposed system and "related queries" for the baseline system clarity. Based on the obtained feedback, the CVI (S-CVI) score is calculated by averaging the number of agreements. Afterward, Universal Agreement (UA) score is calculated based on the binary agreement scores of the experts. According to the scoring scale discussed in literature [73], average S-CVI and UA scores should be 0.8 and 0.9 or higher, respectively. We achieved an average of 0.95 S-CVI and 0.90 UA scores, falling in the well-designed questionnaire category. The validity scores are presented in Table 8.

Hence, the usability instruments employed in this research are designed to incorporate the user factors via standard instruments. This eliminates the bias factor in excerpt feedback from the users and summarizes the key findings. Hence, the usability excerpts usability metrics we rather observed via user-generated log analysis e.g., number of clicks, heatmaps, query issuing behavior, etc., to empirically observe the user feedback.

5.3.5 Procedure

After designing the study protocol, the evaluation started by obtaining the demographic background details from the users. Afterward, to further eliminate biases in results, the user selected two topics from the list of four, one for the proposed system and the other for the baseline system. The user had to choose the least familiar topic from the given topics and search for information or answers about the selected topic. The users were then provided the systems to evaluate in the alternating order to eliminate the users' prior learn-ability effect and biases induced due to the task order [69]. To ensure uniformity, the task exploration time limit was set to a maximum of 15 minutes. Finally, using standard usability instruments, we measured users' felt difficulty, interest, knowledge gain, and usability components. We recorded all the interaction information using screen recording software to analyze their behavioral characteristics, such as the number of clicks, keystrokes, time spent, etc.

5.3.6 Usability results

The usability evaluation results showed that the users reported both systems (proposed and baseline) as highly usable. The users scored the baseline (Google) system 78 out of 100 overall

Table 8 Questionnaire content validity scores obtained from the human experts

Items	Clarity		Relevance				
	Expert 1	Expert 2	Expert 1	Expert 2	Expert in Agreement	I-CVI	UA
1	4	4	4	4	2	1	1
2	4	4	4	4	2	1	1
3	4	4	4	4	2	1	1
4	4	4	4	4	2	1	1
5	4	4	4	4	2	1	1
6	4	4	3	4	2	1	1
7	4	4	2	4	1	0.5	0
8	4	4	4	4	2	1	1
9	4	4	4	4	2	1	1
10	4	4	4	4	2	1	1
11	4	4	3	4	2	1	1
Average scores	**1**	**1**	**0.9**	**1**	**0.9**	**S-CVI = 0.95**	**S-CVI UA = 0.90**

system usability. However, users in the proposed system were marginally more satisfied in terms of usability, giving an overall usability score of 80 out of 100. Figure 10 demonstrates the result of the SUS questionnaire. According to the interpretation of the achieved results presented in the study [71], the baseline system achieves "B+," and the proposed approach achieves "A-" grade.

The overall user satisfaction is determined through the standard After-Scenario Questionnaire (ASQ). We chose the questions in ASQ related to usefulness and ease-of-use. A higher score denotes better usefulness and ease-of-use. Figure 10 shows the obtained result via both systems. According to the results, most users perceived the proposed tool as approximately 10% more valuable than the baseline system. Similarly, the proposed system obtained 16% more usage in query formulation than the baseline system.

The effectiveness of the recommended queries given by both systems (proposed and baseline) is measured using a custom-defined 11-items questionnaire consisting of a 5-points Likert scale. We changed the terms in the questionnaire presented during the evaluation of the proposed and baseline systems to avoid the users' memorization effect. Specifically, we named the query suggestion component in the proposed system the "query table". We called the query suggestion component in the baseline system the "related queries". This aspect is further highlighted in Table 7. Most users agreed with the assistance provided by the system-recommended queries in the proposed system (Fig. 11). The users mainly reported neutral to a mild agreement on system recommended query usefulness in the baseline system (Fig. 12).

5.4 User behavior evaluation

To the best of our knowledge, there was a lack of PRF studies that analyzed the users' behavioral interaction with the tool. In the subsequent sections, we discuss users' behavioral interaction in-depth, including users' time analysis (time spent during exploration), query log analysis (# of queries issued and reformulations), and the level of information gained after a search session. A detailed discussion on each is provided in subsequent sections.

5.4.1 Time analysis

The participant selected two least familiar tasks, one for each system. We measured the time spent in information exploration (from search task initiation to completion) via the proposed and baseline systems. The maximum 15 minute time limit is used in most usability-based studies in such types of exploration activities [74]. Hence, the content exploration time limit

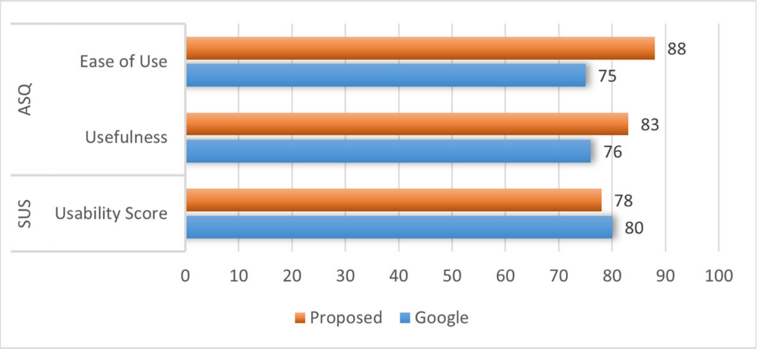


Fig. 10 Usability score of proposed and baseline systems using SUS instrument

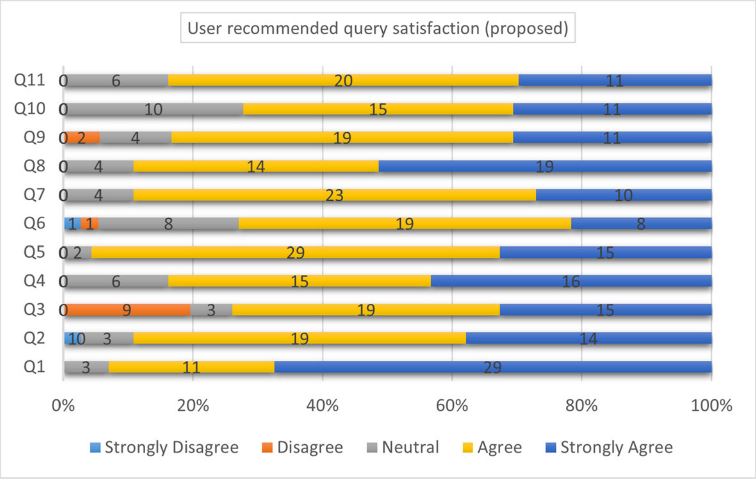


Fig. 11 System recommended queries results via the proposed system

was set to a maximum of 15 minutes. The lower completion time denotes the user information needs satisfaction in a shorter period. In contrast, the longer task completion time indicates that users must put more effort into satisfying their information needs. The average task completion time over proposed and baseline systems were 6 and 7.5 minutes, respectively (Fig. 13).

5.4.2 Log analysis

After expressing initial intent to the system, users of the proposed system were inclined to look at the list of recommended queries presented by the query table component, showing the query table's usefulness. The query table component provides ease in query formulation,

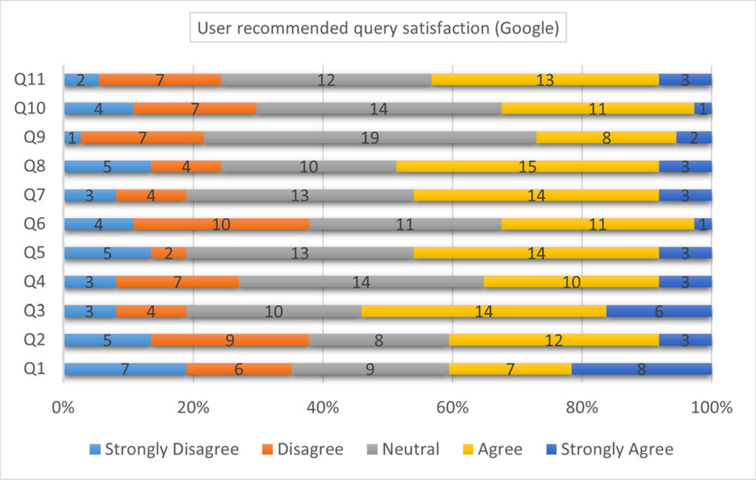


Fig. 12 System recommended queries results via the baseline system

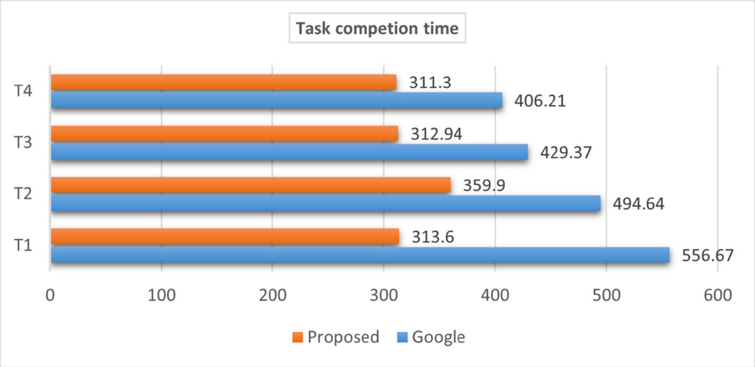
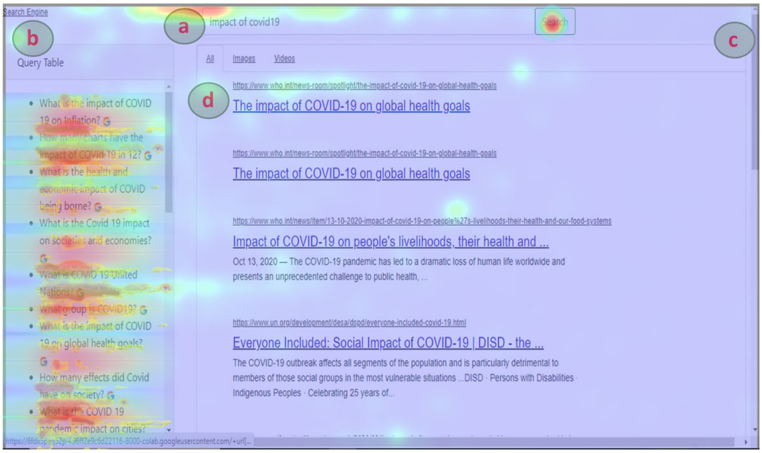


Fig. 13 Task completion time (in seconds) in proposed and baseline systems

vocabulary usage, and exploring the information via just one click, saving users' time. We generated a heat map based on the users' mouse clicks during the evaluation process to determine the query table's usage. The red color in the heat map shows the most mouse clicks, and the green color indicates the fewer user clicks. As evident from Fig. 14, users could effectively use the query table and find their desired information within the top three search results retrieved via the suggested queries.

The existing search engines are precision-oriented and therefore retrieve the most matching results in the initial query issued by the users. We measured to what extent the users could meet their informational needs by manually giving only one query. Hence, we named it a one-shot query. The existing state-of-the-art information retrieval systems reduce the manual need to enter the query by suggesting relevant queries to the user. The intuition behind a one-shot query is to determine the degree to which the suggested queries are usable to the users to eliminate the need for manual query reformulations. To the best of our knowledge, this aspect of suggested query usability needs to be evaluated in previous studies. Figure 15 depicts the one-shot query evaluation task-wise result produced from a total of 37 participants. In the



- What is the impact of COVID 19 on Inflation?
- How many charts have the impact of COVID 19 in 12?
- What is the health and economic impact of COVID being borne?
- What is the Covid 19 impact on societies and economies?
- What is COVID 19 United Nations?
- What group is COVID19?
- What is the impact of COVID 19 on global health goals?
- How many effects did Covid have on society?
- What is the COVID 19 pandemic impact on cities?

https://www.who.int/news-room/spotlight/the-impact-of-covid-19-on-global-health-goals
The impact of COVID-19 on global health goals

https://www.who.int/news-room/spotlight/the-impact-of-covid-19-on-global-health-goals
The impact of COVID-19 on global health goals

https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people%27s-livelihoods-their-health-and-our-food-systems
Impact of COVID-19 on people's livelihoods, their health and ...
Oct 13, 2020 — The COVID-19 pandemic has led to a dramatic loss of human life worldwide and presents an unprecedented challenge to public health, ...

https://www.un.org/development/desa/dspd/everyone-included-covid-19.html
Everyone Included: Social Impact of COVID-19 | DISD - the ...
The COVID-19 outbreak affects all segments of the population and is particularly detrimental to members of those social groups in the most vulnerable situations ...DISD · Persons with Disabilities · Indigenous Peoples · Celebrating 25 years of...

https://6fdx5pcg52p-496ff2e9c6d22116-8000-colab.googleusercontent.com/+url[...

Fig. 14 Heatmap of user interaction with the proposed system

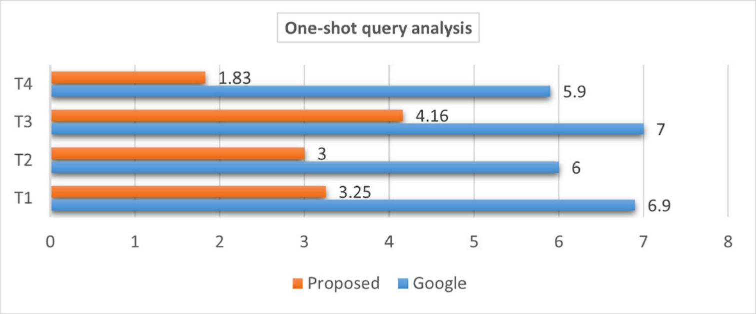


Fig. 15 Comparison of query formulation in proposed and baseline Systems

proposed system, users could complete a goal using only a single query because several relevant queries were generated automatically in a query table. Overall, on average, users needed to reformulate 2 manual queries to complete their goal on the proposed system. Contrarily, they needed to manually reformulate nearly 5 queries on the baseline system to achieve their goals. The analysis shows that the proposed system generated more relevant queries to the user tasks. Therefore, the proposed system could assist the user in meeting their goal twice as much as the baseline system.

In a traditional query reformulation mechanism, the users manually enter the keywords to search for required information. The users reformulate the queries manually by entering new or modifying existing keywords until their information needs are satisfied. It often happens when the user needs help determining which vocabulary terms to use to retrieve the most relevant information. Figure 16 demonstrates the users' average number of query reformulations. Overall, the users on the proposed system needed to reformulate two queries, and on the baseline system, nearly five queries to get the relevant results. Therefore, the proposed system users could complete the goal with more than twice the reduced need to reformulate queries.

We calculated the number of keystrokes (mouse clicks and keyboard presses) required to complete an exploratory search task. The total number of keystrokes denotes the micro-level efforts by the users to satisfy the information needs. The keystrokes meet these four tasks for both systems (proposed and baseline). Figure 17 shows the average keystrokes required to complete a complex search task. A minimum, maximum, and average of 35, 112, and 88 keystrokes were needed to satisfy the proposed system's information needs. On the contrary,

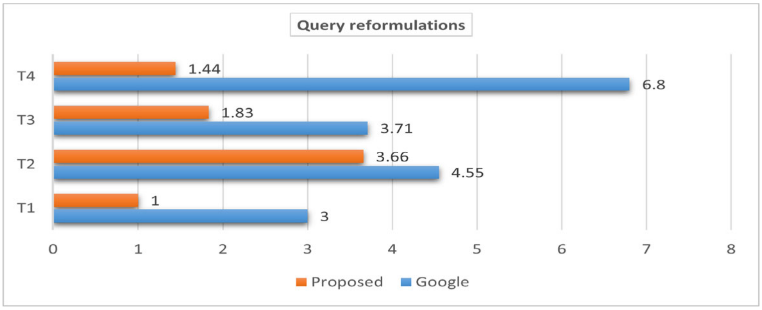


Fig. 16 Comparison of query reformulation in the proposed and baseline systems

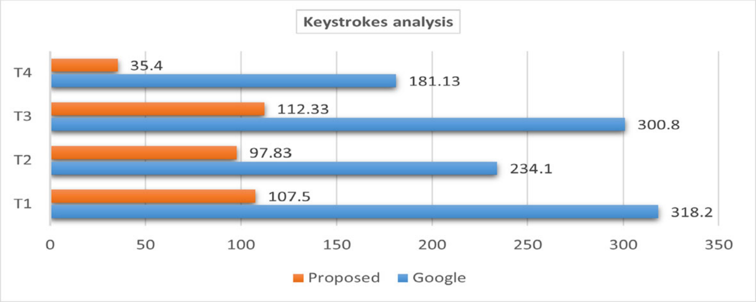


Fig. 17 Number of keystrokes comparison in proposed and baseline system

a minimum, maximum, and average of 98, 318, and 259 keystrokes were required to satisfy the information needs over the baseline system, respectively. The investigation reveals that users on the proposed system could complete the search task with 2.5 times fewer exploration efforts.

5.4.3 Cognitive analysis

We computed the task-based information gain through the average of all the information gained per task by the users, as shown in Fig. 18. Mainly, the same information gain result is produced for both systems. In tasks T_2, T_3, and T_4, users gained more information than Google. Contrarily, task T_1 performs better on the Google system than on our proposed system. Finally, it is concluded that all of these tasks have minimum differences concerning the information gained by the users.

Similarly, we calculated the average score of information gained by the users via our proposed and baseline system. We holistically analyzed the information gained by the users while performing search tasks. The system's average scores were used to calculate the information gain. The user-based information gain results show almost no significant difference between the proposed and the baseline system, as shown in Fig. 19. Therefore, it is concluded that there is an insignificant difference in the information gain of both systems with a minimum marginal difference.

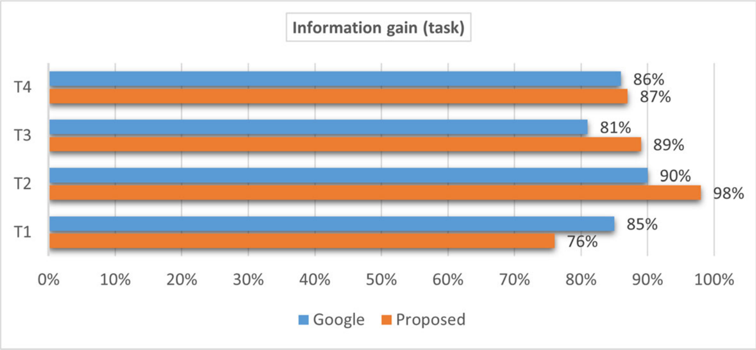


Fig. 18 Task-based information gain by the participants

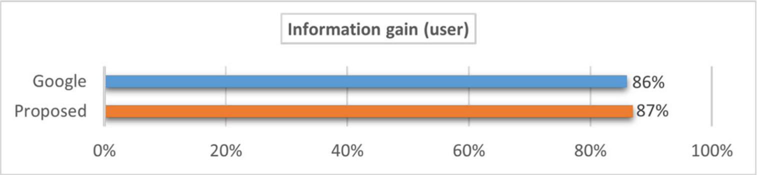


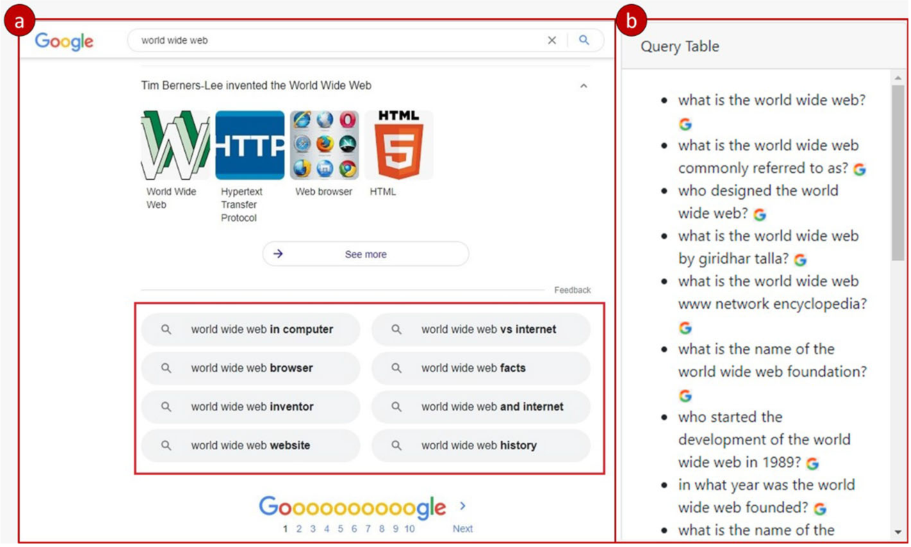
Fig. 19 User-based information gain by the participants

6 Results & discussion

In previous sections, we discussed the evaluation details and obtained results. This section analyzes the key similarities and differences in the obtained results. Therefore, the subsequent section provides a detailed discussion on a detailed comparison of the proposed approach with state-of-the-art PRF approaches, consolidating the previous techniques and highlighting the implications of PRF in information retrieval.

6.1 Query table vs. related queries

The Google search engine presents the related queries usually at the end of the search results (Fig. 20a). On the contrary, the existing usability research concerning user interfaces stresses the maximization of the visibility of interaction components [22]. Therefore the components most interacted with by the users, e.g., the query recommendation panel, should be placed on top of the screen or as a fixed panel on either the left or right side of the screen [22]. Following the defined standards, we opted to display the related query in the query table panel placed on the left side of the screen (Fig. 20b).



world wide web **in computer**
world wide web **vs internet**
world wide web **browser**
world wide web **facts**
world wide web **inventor**
world wide web **and internet**
world wide web **website**
world wide web **history**

- what is the world wide web?
- what is the world wide web commonly referred to as?
- who designed the world wide web?
- what is the world wide web by giridhar talla?
- what is the world wide web www network encyclopedia?
- what is the name of the world wide web foundation?
- who started the development of the world wide web in 1989?
- in what year was the world wide web founded?
- what is the name of the

Fig. 20 Screenshots of interface design (a) related queries in Google search engine accessed on January 2022, and (b) query recommendation table in our proposed approach

The query generation mechanism is also distinct in Google and the proposed approach. Google uses a recommendation approach that may utilize contextual parameters such as user query history and auto-correction mechanism in recommendations [75]. The latter generates queries by exploiting the user intent in summarized information associated with top-ranked results dispersed in different verticals. We critically analyzed and compared the significance of the query recommendation in the baseline system (Google) and our proposed approach via an 11-point instrument (Fig. 12). As it emerges from Figs. 11 and 12 that overall, 39% of the users were delighted with the proposed system, and only 9% were satisfied with the baseline system in terms of query recommendation paradigm.

6.2 Key similarities & differences

We found that the user gained the same information in the proposed and baseline systems. However, a significant difference was observed in the users' search efforts. Users on the baseline system spent 36% extra time to find relevant information and reformulate three additional queries. The searching efforts, e.g., keystrokes, were also reported to be 2.5 times higher in the baseline system. In terms of the overall system usability, both systems obtained similar scores. However, the proposed system achieved one higher grade in terms of usability. Subsequently, users found the proposed system more useful and easy to use than the baseline system. It could be due to the adoption of usability design principles in designing query table panels [22]. Furthermore, the PRF and the deep learning approach in query recommendation provide semantically related query recommendations.

6.3 Critical analysis & research credibility

The existing research holistically categorizes the relevance feedback approaches into implicit, explicit, and pseudo-relevance. The explicit relevance feedback approach requires efforts at the users' end, whereas getting the user intent from the implicit relevance feedback is challenging. We mainly used the PRF approach in this research. The PRF approach bridges implicit and explicit feedback that considers the user's initial intent and guides the rest of the search process. Therefore, an increasing trend in the PRF is observed in the existing research [15, 18, 51–54].

Initially, the feedback-based approaches were implemented by employing a generic statistical model such as a weighted graph [18, 33, 37]. With the proven machine and deep learning effectiveness, recent approaches primarily deploy the relevance feedback models based on machine, and deep learning [15, 18, 51–53]. Therefore, we deployed a combination of deep learning models to leverage effectiveness in PRF. The adopted procedure allowed us to semantically analyze the relationship between the user query intent and the recommended queries. In addition, existing research concentrated on calculating the accuracy scores of the proposed deep learning models on generic datasets. It focused on the actual usability of the results produced from the deep learning solution that was under-emphasized [15, 18, 33, 51–53]. Since the existing deep learning models are readily available to provide solutions [7], instead of designing and training the custom model, we emphasized deploying pre-trained deep learning models and calculating the usability of produced output from the users' perspective. We achieved 89% accuracy scores from Top-20 recommended queries.

Furthermore, most existing feedback approaches are implemented theoretically, and their effectiveness is obtained empirically (via precision, recall, accuracy scores, etc.) [15, 18, 33, 51–53]. Very few approaches exist that design a search user interface and evaluate the effectiveness of their approach from the usability and users' behavior perspectives [37]. Our implementation includes a standard-designed search user interface and deep-learning backend model. To the best of our knowledge, we are the first to analyze the usability issues comprehensively and understand users' behavioral characteristics in the PRF approaches (Table 9).

The usability and behavioral evaluation involve users' participation. We recruited the users via an open advertisement across the university departments and various industrial institutes to ensure the credibility of the results and overcome user-induced limitations. To eliminate task-based effects, we designed each task incorporating multiple sub-tasks based on convergent (simple) and divergent (complex) search tasks and randomly alternating the order in which users were assigned the system to evaluate. The difference in usability results was further reduced by employing standard usability (SUS) and user satisfaction (ASQ) scales. Moreover, the custom-designed questionnaire to get users' perspectives on the usability of the generated queries was validated (CVI). The empirical evaluation was rigorously evaluated against top-20 queries instead of the traditional top-10 queries evaluation procedure. To overcome bias in the labeling process, we recruited two human experts to compute the upper and lower boundaries of the accuracy scores. Moreover, the techniques used in this research were based on well-established libraries that can be used without prior domain knowledge. Therefore, we deem the validity of this study credible and reproducible.

6.4 Research implications & Study limitations

Web search engines are gateways to access immense, distributed, and heterogeneous resources on the web and attract a significant number of web users. They may collect large-scale search behavior log data [75]. Large-scale search behavior log data enabled substantial advances in search engines, i.e., the ability to identify relevant and irrelevant retrieval content, search results ranking, query modification, adaption, and recommendation [75]. The collected logs provided a rich picture of real people performing self-motivated searches [75]. However, researchers have demonstrated that logs alone should not be used as a direct source of relevancy judgments [75]. Instead, some implicit evidence of relevance requires careful interpretation [75].

The existing search engines can retrieve and rank relevant information within seconds [76]. Their effectiveness majorly depends on the quality of the query passed. Users' are often reported to issue generic and short-typed queries, which often results in the retrieval of an immense result set [7]. Therefore, to overcome this limitation of the web search engine, researchers devise query recommendation techniques to assist users in driving their search journey toward precise results set. The PRF, in our case, takes in the user's initial ambiguous query, retrieves the search results from different verticals, and performs question generation based on the 5W pattern on the retrieved result set. This process covers potential concepts in a particular search domain and allows users to choose the most relevant item. Since the users are effective in their memory to recall [7], this helps them identify the relevant information without added cognitive effort.

To enhance the queries' efficacy, we devised a PRF approach to extract and summarize the potential intentions from the retrieved results and recommend semantically similar queries that are well-articulated based on the 5W model. The proposed approach can be used on top of any search engine as a query recommendation mechanism to facilitate high-impact queries with user involvement. The empirical analysis has shown that the proposed system could produce relevant and well-balanced queries with 89% of the accuracy scores. However,

Table 9 Comparison of the proposed approach with the state-of-the-art approaches

Parameter	Technique	Balakrishnan et al. [33]	Li et al. [15]	Stai et al. [37]	Wang et al. [18]	Yu et al. [52]	Yu et al. [53]	Keikha et al. [40]	Valcarce et al. [54]	Bodigutla et al. [51]	Proposed
Feedback	Implicit			+						+	
	Explicit	+		+							
	Pseudo		+		+	+	+	+	+		+
Model	Deep Learning		+		+	+	+			+	+
	Machine Learning										+
	Statistical	+		+	+			+	+		
Ranking	Generic	+	+						+	+	
	Semantic					+		+			+
Dataset	TREC	+	+		+	+	+	+	+		
	Wikipedia							+			
	Real										+
Evaluation	Empirical	+	+	+	+	+	+	+	+	+	+
	Usability	+		+							+
	Behavioral			+							+
Accuracy/Precision	Top-6									+	+
	Top-10	+	+		+	+	+				+
	Top-15	+									+
	Top-20			+							+
	Top-100						+				
Implementation	Theoretical model	+	+	+	+	+	+	+	+	+	+
	Search user interface	+									+

in some rare and exceptional scenarios where a user types the query beyond the recognition ability of a web search engine, it can result in the retrieval of a very limited or empty result set, limiting the effectiveness of the PRF approaches. However, the key benefit of the PRF approach is no added cost at the users' end. Therefore, the proposed system can be a promising solution to ease query reformulation issues and reduce user cognitive efforts in vertical web search engines.

Presently, information retrieval systems are evaluated from the empirical and usability perspectives. An information retrieval system is usually evaluated empirically via accuracy [7], precision and recall [1], F-1 measure and sensitivity [16] scores. These scores are subject to the employed dataset, retrieval algorithm, and the results cut-off criteria (e.g., top-*k* results). The usability evaluation involves human participants. The subjective judgment of the human experts is measured via standard validated questionnaires [1]. Moreover, the implicit interaction of the users with the retrieval system is often evaluated via screen-recording [21], tracking scripts [77], and log analysis [77]. The usability evaluation is studied in a between-subjects or within-subjects study design [69]. Therefore, the difference in the stated evaluation technique, such as study design, choice of usability instruments (questionnaires), and the selection of users, is subject to the outcome of the usability results.

Overall, the proposed approach achieves "A-" grade compared to the baseline system's "B+" grade. Most users perceived the proposed tool as 10% more valuable and 16% more useful in query formulation than the baseline system. The increased usefulness could be attributed to the novel ability of the proposed approach to generate semantically similar and 5W's communication principle adhering queries within the user-defined intent. As a result, whilst the baseline system users remained neutral to mildly agreeing in overall query recommendation satisfaction, the proposed system was able to achieve agreeing to strongly agreeing query recommendation satisfaction. With each query session, the users were able to narrow down the search scope with the ability of the proposed system to generate well-articulated queries. Hence, the users completed the evaluation process 25% faster in the proposed system despite being newly introduced to the system with 1.5 to 2.5 times half the clicking efforts.

7 Conclusion and future work

Web users need help with query formulation to explore the search results. Formulating a well-defined and balanced query is necessary for the user to retrieve the relevant information. We proposed a query recommendation approach based on pseudo-relevance feedback and deep learning algorithms; the objective was to facilitate the users in balanced query formulation. Mainly, we provided a query table that allows sophisticated query formulation via standard deep learning algorithms. The pseudo-relevance feedback was extracted and summarized by exploiting the retrieved contextual information.

A pool of potential candidate queries was generated by employing 5W's principles and a deep neural network learning model. The pool of candidates was further transformed into a neural model for semantics information extraction. Finally, a list of high-impact queries was generated in descending order of similarity. We evaluated our proposed query recommendation approach from the empirical, usability, and behavioral aspects. We involved 37 participants in the evaluation and compared experimental results with the Google search engine as a baseline system. The empirical results showed 89% accuracy with well well-balanced average query length of 10 words. The usability evaluation surpassed the baseline system by achieving an "A"-grade and a robust query recommendation satisfaction response. The behavioral analysis showed that the users could complete complex search tasks with minimum query reformulation efforts.

In the future, we will investigate diversification via users' personalized information in the query recommendation. In addition, we intend to explore the multi-modal information, i.e., acoustic and visual, along with textual in query recommendation. We aim to focus on user-seeking behavior in multimedia query recommendations over the vertical web search engines.

Declarations

Competing of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Rashid U, Bhatti MA (2017) A framework to explore results in multiple media information aggregated search. Multimed Tools Appl 76(24):25787–25826
2. Pouyanfar S, Yang Y, Chen S-C, Shyu M-L, Iyengar S (2018) Multimedia big data analytics: a survey. ACM Comput Surv (CSUR) 51(1):1–34
3. Vidinli IB, Ozcan R (2016) New query suggestion framework and algorithms: a case study for an educational search engine. Inf Process Manage 52(5):733–752. https://doi.org/10.1016/j.ipm.2016.02.001
4. Kuzi S, Zhai C, Tian Y, Tang H (2020) Figexplorer: a system for retrieval and exploration of figures from collections of research articles. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, pp 2133–2136
5. Oussous A, Benjelloun F-Z, Lahcen AA, Belfkih S (2018) Big data technologies: a survey. J King Saud Univ- Comput Inf Sci 30(4):431–448
6. Pámies-Estrems D, Castellá-Roca J, Viejo A (2016) Working at the web search engine side to generate privacy-preserving user profiles. Expert Syst Appl 64:523–535. https://doi.org/10.1016/j.eswa.2016.08.033
7. Khan AR, Rashid U, Saleem K, Ahmed A (2021) An architecture for non-linear discovery of aggregated multimedia document web search results. PeerJ Comput Sci 7:449
8. Tan SS-L, Goonawardene N (2017) Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 19(1):9
9. Kathuria M, Nagpal C, Duhan N (2016) Journey of web search engines: milestones, challenges & innovations. Int J Inf Technol Comput Sci 12:47–58
10. Jiang J, Ni C (2016) What affects word changes in query reformulation during a task-based search session. In: Proceedings of the 2016 ACM on conference on human information interaction and retrieval, pp 111–120
11. Toms EG, O'Brien H, Mackenzie T, Jordan C, Freund L, Toze S, Dawe E, Macnutt A (2007) Task effects on interactive search: the query factor. In: International workshop of the initiative for the evaluation of XML Retrieval, pp 359–372 . Springer
12. Bilal D, Gwizdka J (2018) Children's query types and reformulations in google search. Inf Process Manage 54(6):1022–1041
13. Lin S-C, Yang J-H, Nogueira R, Tsai M-F, Wang C-J, Lin J (2020) Query reformulation using query history for passage retrieval in conversational search. arXiv:2005.02230
14. Maxwell D, Bailey P, Hawking D (2017) Large-scale generative query autocompletion. In: Proceedings of the 22nd australasian document computing symposium, pp 1–8
15. Li C, Sun Y, He B, Wang L, Hui K, Yates A, Sun L, Xu J (2018) Nprf: a neural pseudo relevance feedback framework for ad-hoc information retrieval. arXiv:1810.12936
16. Rashid U, Javid A, Khan AR, Liu L, Ahmed A, Khalid O, Saleem K, Meraj S, Iqbal U, Nawaz R (2022) A hybrid mask rcnn-based tool to localize dental cavities from real-time mixed photographic images. PeerJ Comput Sc 8:888
17. Rahman MM, Abdullah NA (2018) A personalized group-based recommendation approach for web search in e-learning. IEEE Access 6:34166–34178

18. Wang J, Pan M, He T, Huang X, Wang X, Tu X (2020) A pseudo-relevance feedback framework combining relevance matching and semantic matching for information retrieval. Inf Process Manage 57(6):102342
19. Rashid U, Viviani M, Pasi G (2016) A graph-based approach for visualizing and exploring a multimedia search result space. Inf Sci 370:303–322
20. Song W, Liang JZ, Cao XL, Park SC (2014) An effective query recommendation approach using semantic strategies for intelligent information retrieval. Expert Syst Appl 41(2):366–372. https://doi.org/10.1016/j.eswa.2013.07.052
21. Rashid U, Saleem K, Ahmed A (2021) Mirre approach: nonlinear and multimodal exploration of mir aggregated search results. Multimed Tools Appl 80(13):20217–20253
22. Russell-Rose T, Tate T (2012) Designing the Search Experience: the Information Architecture of Discovery. Newnes
23. Kofler C, Larson M, Hanjalic A (2016) User intent in multimedia search: a survey of the state of the art and future challenges. ACM Comput Surv (CSUR) 49(2):1–37
24. Liao Z, Song Y, Zhou D (2020) Query suggestion. In: Query understanding for search engines, pp 171–203. Springer
25. Kumar M, Bindal A, Gautam R, Bhatia R (2018) Keyword query based focused web crawler. Procedia Comput Sci 125:584–590
26. Ooi J, Ma X, Qin H, Liew SC (2015) A survey of query expansion, query suggestion and query refinement techniques. In: 2015 4th International conference on software engineering and computer systems (ICSECS), pp 112–117. IEEE
27. Azad HK, Deepak A (2019) Query expansion techniques for information retrieval: a survey. Inf Process Manage 56(5):1698–1735
28. Chen W, Cai F, Chen H, de Rijke M (2017) Personalized query suggestion diversification. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, pp 817–820
29. Ahmad WU, Chang K-W, Wang H (2019) Context attentive document ranking and query suggestion. In: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, pp 385–394
30. Chen W, Cai F, Chen H, de Rijke M (2018) Attention-based hierarchical neural query suggestion. In: The 41st international ACM SIGIR conference on research & development in information retrieval, pp 1093–1096
31. Ahmad W.U, Chang K.-W, Wang H (2018) Multi-task learning for document ranking and query suggestion. In: International conference on learning representations
32. Jeffery S.R, Franklin M.J, Halevy AY (2008) Pay-as-you-go user feedback for dataspace systems. In: Proceedings of the 2008 ACM SIGMOD international conference on management of data, pp 847–860
33. Balakrishnan V, Ahmadi K, Ravana SD (2015) Improving retrieval relevance using users' explicit feedback. Aslib Journal of Information Management
34. Jayarathna S, Patra A, Shipman F (2015) Unified relevance feedback for multi-application user interest modeling. In: Proceedings of the 15th ACM/IEEE-CS joint conference on digital libraries, pp 129–138
35. Xu S, Jiang H, Lau FC (2008) Personalized online document, image and video recommendation via commodity eye-tracking. In: Proceedings of the 2008 ACM conference on recommender systems, pp 83–90
36. Su Y, Yang S, Sun H, Srivatsa M, Kase S, Vanni M, Yan X (2015) Exploiting relevance feedback in knowledge graph search. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1135–1144
37. Stai E, Kafetzoglou S, Tsiropoulou EE, Papavassiliou S (2018) A holistic approach for personalization, relevance feedback & recommendation in enriched multimedia content. Multimed Tools Appl 77(1):283–326
38. Zamani H, Dadashkarimi J, Shakery A, Croft WB (2016) Pseudo-relevance feedback based on matrix factorization. In: Proceedings of the 25th ACM international on conference on information and knowledge management, pp 1483–1492
39. ALMasri M, Berrut C, Chevallet J-P (2016) A comparison of deep learning based query expansion with pseudo-relevance feedback and mutual information. In: European conference on information retrieval, pp 709–715. Springer
40. Keikha A, Ensan F, Bagheri E (2018) Query expansion using pseudo relevance feedback on wikipedia. J Intell Inf Syst 50(3):455–478
41. Jiang J-Y, Wang W (2018) Rin: reformulation inference network for context-aware query suggestion. In: Proceedings of the 27th ACM international conference on information and knowledge management, pp 197–206

42. Chen W, Cai F, Chen H, De Rijke M (2020) Personalized query suggestion diversification in information retrieval. Front Comput Sci 14(3):1–14
43. Ding H, Zhang S, Garigliotti D, Balog K (2018) Generating high-quality query suggestion candidates for task-based search. In: European conference on information retrieval, pp 625–631. Springer
44. Dehghani M, Rothe S, Alfonseca E, Fleury P (2017) Learning to attend, copy, and generate for session-based query suggestion. In: Proceedings of the 2017 ACM on conference on information and knowledge management, pp1747–1756
45. Sordoni A, Bengio Y, Vahabi H, Lioma C, Grue Simonsen J, Nie J-Y (2015) A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. In: Proceedings of the 24th ACM international on conference on information and knowledge management, pp 553–562
46. Shokouhi M (2013) Learning to personalize query auto-completion. In: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval, pp 103–112
47. Li X, Chen Y, Pettit B, Rijke MD (2019) Personalised reranking of paper recommendations using paper content and user behavior. ACM Trans Inf Syst (TOIS) 37(3):1–23
48. Zhang X, Jiang X, Qin J (2020) Time-aware query suggestion diversification for temporally ambiguous queries. The Electronic Library
49. Cai F, Reinanda R, Rijke MD (2016) Diversifying query auto-completion. ACM Trans Inf Syst (TOIS) 34(4):1–33
50. Mustar A, Lamprier S, Piwowarski B (2021) On the study of transformers for query suggestion. ACM Trans Inf Syst (TOIS) 40(1):1–27
51. Bodigutla PK (2021) High quality related search query suggestions using deep reinforcement learning. arXiv:2108.04452
52. Yu H, Xiong C, Callan J (2021) Improving query representations for dense retrieval with pseudo relevance feedback. arXiv:2108.13454
53. Yu H, Dai Z, Callan J (2021) Pgt: pseudo relevance feedback using a graph-based transformer. arXiv:2101.07918
54. Valcarce D, Parapar J, Barreiro Á (2018) Lime: linear methods for pseudo-relevance feedback. In: Proceedings of the 33rd annual ACM symposium on applied computing, pp 678–687
55. Lv Y, Zhai C, Chen W (2011) A boosting approach to improving pseudo-relevance feedback. In: Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval, pp 165–174
56. White RW, Roth RA (2009) Exploratory search: beyond the query-response paradigm. Synth Lect Inf Concepts Retr Serv 1(1):1–98
57. Atwood R, Dervin B (1981) Challenges to sociocultural predictors of information seeking: a text of race versus situation movement state. Ann Int Commun Assoc 5(1):549–569. https://doi.org/10.1080/23808985.1981.11923862
58. Wenxiu P (2015) Analysis of new media communication based on lasswell's "5w" model. J Educ Soc Res 5(3):245–245
59. McCarley J, Chakravarti R, Sil A (2019) Structured pruning of a bert-based question answering model. arXiv:1910.06360
60. Chandrasekaran D, Mago V (2021) Evolution of semantic similarity-a survey. ACM Comput Surv (CSUR) 54(2):1–37
61. Cutrell E, Guan Z (2007) What are you looking for? an eye-tracking study of information usage in web search. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 407–416
62. Thorleuchter D, den Poel DV, Prinzie A (2010) Mining ideas from textual information. Expert Syst Appl 37(10):7182–7188. https://doi.org/10.1016/j.eswa.2010.04.013
63. Chang Y, Ounis I, Kim M (2006) Query reformulation using automatically generated query concepts from a document space. Inf Process Manage 42(2):453–468
64. Khan A.R, Rashid U (2021) A relational aggregated disjoint multimedia search results approach using semantics. In: 2021 International conference on artificial intelligence (ICAI), pp 62–67. https://doi.org/10.1109/ICAI52203.2021.9445229
65. Khan AR, Rashid U, Ahmed N (2022) An explanatory study on user behavior in discovering aggregated multimedia web content. IEEE Access 10:56316–56330. https://doi.org/10.1109/ACCESS.2022.3177597
66. Shekhar A, Marsden N (2018) Cognitive walkthrough of a learning management system with gendered personas. In: Proceedings of the 4th conference on gender & IT, pp 191–198
67. Alroobaea R, Mayhew PJ (2014) How many participants are really enough for usability studies? In: 2014 Science and information conference, pp 48–56. IEEE
68. Marcum JW (2002) Rethinking Inf Lit Libr Q 72(1):1–26
69. Taramigkou M, Apostolou D, Mentzas G (2017) Supporting creativity through the interactive exploratory search paradigm. Int J Hum Comput Interact 33(2):94–114

70. Li Y, Belkin NJ (2008) A faceted approach to conceptualizing tasks in information seeking. Inf Process Manage 44(6):1822–1837
71. Lewis JR, Sauro J (2018) Item benchmarks for the system usability scale. Journal of Usability Studies 13(3)
72. Lewis JR (1991) Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq. ACM Sigchi Bulletin 23(1):78–81
73. Shi J, Mo X, Sun Z (2012) Content validity index in scale development. Zhong nan da xue xue bao. Yi xue ban= Journal of Central South University. Med Sci 37(2): 152–155
74. Brown A, Evans M, Jay C, Glancy M, Jones R, Harper S (2014) Hci over multiple screens. In: CHI'14 extended abstracts on human factors in computing systems, pp 665–674
75. Kim JY, Teevan J, Craswell N (2016) Explicit in situ user feedback for web search results. In: Proceedings of the 39th international acm sigir conference on research and development in information retrieval, pp 829–832
76. Tablan V, Bontcheva K, Roberts I, Cunningham H (2015) Mimir: an open-source semantic search framework for interactive information seeking and discovery. J Web Semant 30:52–68
77. Huurdeman H, Kamps J, Wilson ML (2019) The multi-stage experience: the simulated work task approach to studying information seeking stages. CEUR Workshop Proceedings

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.