*International Journal of Engineering & Technology*, 7 (4.35) (2018) 276-281

**International Journal of Engineering & Technology**

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***



# Evaluating Search Results in Exploratory Search

**Mohammed Najah Mahdi[1*], Abdul Rahim Ahmad[2], Roslan Ismail[3]**

[1,2]*Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Selangor, Malaysia.*
[1,2]*College of Computer Science & Information Technology, Universiti Tenaga Nasional, Selangor, Malaysia.*
**Corresponding author E-mail: najah.mahdi@uniten.edu.my*

**Abstract**

The volume of information available on the World Wide Web is quite significant. This plethora of information has, to a large extent constantly challenged researchers in seeking for ways of making the information easily accessible to end users in as convenient ways as possible. Characteristically, ensuring the accessibility of this large amount of information involves three all-encompassing processes of retrieval, organization and presentation. The World Wide Web offers a platform for sharing information from large database repositories globally. However, information needs to be searched with specialized tools commonly referred to as search engines. While a number of search engines does presently exist, most of these search engines are noted for their inability to retrieve information usable to the end user. It is therefore critical for the results generated from search engines to be intelligently organized for the optimal usefulness of the information to the searcher. Exploratory web technologies is capable of filing this gap. Therefore, this paper reviews exploratory search as a mechanism for conducting result-oriented search. It also reviews the ways of evaluating the search results obtained from an exploratory search.

***Keywords***: *Information Retrieval; Search Engine; Exploratory Search; Faceted Search Engine*.

## 1. Introduction

We presently live in a world of effective immeasurable data and information. Alost everyone in the world owns and carries a device such as a laptop and mobile phone, which allows them instantaneous access to a summation of almost the totality of human knowledge[1]. As such, the concept of information and knowledge has changed significantly, which has made access less cumbersome. Hence, mechanisms directed towards obtaining and retaining reliable information and knowledge as and when needed has become crucial. The modern search engine is the route into which a vast amount of data is accessed. As such, Information Retrieval (IR) is conceived with the aim of finding the most relevant information in order to accurately and speedily respond to the query of a user, and to present that information back to the user for further use based on prioritized results of the search query[2]. Within the context of IR, a user is more concerned about the precision of a search. That is, having preference over a limited number of results that closely matches the needs of the searcher as opposed to countless number of results containing the answers but hidden in irrelevant documents[3]. The procedures involved in trying to locate a certain piece of information in an assemblage of documents refers to a query[4, 5], and is most often undertaken within materials, usually documents of unstructured nature such as text that are within large collections similar to those typically found in computers.

## 2. Review Methodology

The review evaluating search results for this ES was developed by following the guidelines as proposed by[6]. The steps in this evaluating are shown in Table 1.

**Table 1:** Phases and detailed evaluating search results for this ES process

| Phase | Detailed steps |
|---|---|
| Planning | Identify the need for ES |
| Conducting | • Carry out a comprehensive search for primary studies.<br>• Assess and record the criteria of included studies.<br>• Classify data needed to answer the research questions.<br>• Extract data from each included study.<br>• Summarize and synthesise study results.<br>• Interpret results to determine their applicability. |
| Documenting | Write up study as a report. |

## 3. Research Background

The WWW has become a universal repository of human knowledge and culture. Given that millions of Internet users have created hundreds of billions of documents that compose the largest repository of human knowledge in history, finding information on the web has become quite challenging and often requires submitting queries to a search engine[7]. SEs are one of the most popular tools on the Web and are designed to help users find useful information that could contain text, pictures, or videos[8]. Users can type a few terms (words) into the engine and it will return a list of documents related to this query.

Simonini and Zhu [9] Visualize huge amount of data, based on a Bayesian suggestion algorithm and the widely used enterprise search platform Solr. This article presents a new approach to visu-

**Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**

alize huge amount of data, based on a Bayesian suggestion algorithm and the widely used enterprise search platform Solr. The proposal approach is a dynamic and visual tool for big data based on an extension of the popular open source enterprise search engine Apache Solr.

Kules and Shneiderman [10] described the main points of the search results based on a strong coupling of categories and the search results page. They established a link between two direct searchers to check how fast and capable they can identify good results by defining categories that employ unknown terms and by assessing how the search results are related to their queries.

### 3.1. Information Retrieval and Searching the Web

One of the most critical tools in the process of retrieving information is searching. Search engines eases and increases the possibility of browsing hypertext documents on the web. As such, and in recent times, traditional methods of storing information such as libraries have become irrelevant[11]. Additionally, it is common knowledge that previously, users need to know the exact location of the data they have needed of before proceeding to search. This is based on the premise that pages on the internet are usually not as orderly as books on shelves of a library[12]. More so, the data that users search for are not cataloged in one location, and knowing the URL of the expected data is not a guarantee that such data can or will be found. In certain instances, a forward address is usually made available, but the page would have been moved, or was available briefly[13, 14].Hence, it has become imperative to develop a common and easy method of searching for and retrieving information.

### 3.2. Important Problems in IR

IR is mainly purposed to make information available to a user in order to aid the user in carrying out tasks in an efficient, effective and timely manner. Some of the tasks for which IR was set out for includes, amongst others, problems solving, decision making, effective preparation and equipping for performing future tasks and better assimilation of information needed for performance present or future tasks. Additionally, IR is done to address activities that are non-crucial such as getting optimal entertainment, inquisitiveness, trivialities or any other activities that makes one happy. Thus, any IR system should be able to address the challenges as illustrated below:

1. Document collection: This involves selecting the documents that are to be searched for. In order to find out huge parts of the Web vis-à-vis smaller groups of seed pages, the search engine does a Web crawl on the hyperlinked structure of web pages.
2. Information processing: When a new document is located, text, images, other multimedia content and meta-data are parsed. The data are thereafter utilized to ascertain if the content of what is being searched for is contained in the page.
3. Indexing: It is important to compress and store the massive amount of data obtained. However, these processes must be done efficiently in order to allow for frequency and continuity. It is commonplace that users are unwilling to wait for pages to load (a delay), as such, this index is capable of providing quick representations of large number of web pages to the search engine.
4. Query processing: Originally, when searching for the purpose of research, comprehensive descriptions of the specific information searched for were provided. However, current search engines are prompted with just two or three keywords, hence expanding the scope of enquiries using more words/phrases to help the query process is advised. Search engines (SE) can also use more and specific information such as location, history of search, etc., side-by-side the actual query.
5. Document Ranking: IR system takes required representations of documents and returns a number representing the likelihood that the document will meet the need of the searcher. A prioritised list of documents is then presented based on the number of returns. As such, a proper, empirical and theory-based retrieval models, or learned rank algorithms are brought to light.
6. Result Evaluation: In order that IR performs optimally, having a mechanism needed to assess its quality is essential. However, the core of this evaluation is in view of having certain ways of utilizing users in finding some samples of the true set of pertinent information.

Presently, a number of search tools are available that allows for quick and easy means of locating information on the web. However, two rudimentary methods are noted for organizing and locating information. The methods are, (1) the directories and (2) search engines[15]. Characteristically, in the two approaches, the information is contained in databases created manually or with the use of crawlers. Search request is responded to by the search tool to retrieve information from its already made database of indexed web pages[16]. Although there are many search engines, they are classifiable into Directories, Spider Based Search Engine and the new developed Exploratory Search (ES).

### 3.3. Directories

IR enables the use of directories in finding and organizing information. Directories are a ranked depiction of hyperlinks to web pages and presentations compartmentalized into topics and subtopics. Directories are categorized as either broad or specific, with web directories noted to be directories used for gathering diverse resources. Web directories are massive assemblages of links to sites organized in dissimilar classes. The sites in a Web directory are listed in some chronological order (mostly alphabetical), which eases the search and retrieval process for users. Though a number of Web directories offer some sort of search functionality, sort, search directories are profoundly not the same as search engines based on two factors[11]. First, humans do edit most directories. Secondly, crawlers do not automatically gather the corresponding URLs. They are submitted by the owners of the site. Characteristically, human editors customarily does review and classification of the Web pages and presentations added to a directory[17]. Interestingly, the core advantage of Web directories is that a human does the viewing and checking of the pages. This is done to reduce the possibility of web pages being classified in the wrong categories. However, the disadvantage of this approach is that the lists in web directories could go obsolete where no human is available to edit and update on an on-going basis. Notably, the most common all-purpose search directories are Google Directory and Umdum Directory[19].

### 3.4. Search Engine (SE)

SE as the second approach to locating and organizing information on Web pages such as Google, Alta vista etc. does the following:

1. Accepts query submission from user comprising of a word or phrase defining the exact information a user intends to access on the web.
2. Explores the database to match the query.
3. Assembles and returns a clickable list of URLs matching the query.
4. Receive revised query resubmission from user if any.

Characteristically, SE is a software package designed for searching of information on the World Wide Web. It searches for documents using definite keywords and in turn returns a list of the documents (usually referred to as hits), where the keywords are found. Basically, the information may be made up of web pages, pictures and related/unrelated classes of files[18, 20]. Some SE also mine data accessible in databases or open directories. Hence,

without a sophisticated SE it will be almost difficult to locate whatsoever on the Web without knowing a specific URL[21, 22].

### 3.5. Web Search Engine

Three types of SEs are basically known: SE is that crawler based, SE that is human powered SE and Hybrid SE[23] as depicted in table 2 below.

**Table 2:** Types of SEs

| SE Types | Example and Description |
|---|---|
| Crawler-Based SE | Example: Google. Here,<br>Crawlers will visit a site, read the information, meta tag and download the documents by following any hyper-links<br>Returns all information to the central repository and index[24] the data. |
| Human-powered SE | Example: Google directory, Yahoo directory, Open directory.<br>Suitably acknowledged as Web directories.<br>More common due to the complex value of links submitted by humans which are indexed and categorized.<br>The information that is submitted is only indexed. |
| Hybrid SE | Example: MSN Search<br>Overlap between crawler-based and human-powered directories.<br>In search, both crawler human powered results are featured.<br>It favours one type of listings over another[14]. |

### 3.6. Exploratory SE

Current commercial SEs use query and answer. In exploratory SE, the user issues a query and gets a reply - a set of possibly applicable documents. The ES process is noted to be a better information search mechanism. This position is predicated on the notion that it seeks information in an unrestricted, insistent, and multi-faceted context. The ES process is also noted to be opportunistic, iterative, and multi-tactical and normally used in scientific investigations, learning, and decision-making. Exploratory strategies are used in all processes of seeking for information that reflect the choices and experiences of seeker.

A number of search activities that distinguishes the ES process from lookup search. It is also noted that the ES process is very important in learning and investigating diverse activities. Basically, when users' information needs are clearly defined, lookup is plausibly a satisfactory process of locating information as required. However, when the needs of users are not categorically spelt out, lookup quest might be essential but may not be adequate for users seeking for information for the purpose of learning and/or research. Table 3 shows the activities in ES. They are overlapping because there is a general interplay between them, and some activities which might have been embedded in others.

ES is a representation of a paradigm move from the investigative approach of query-document matching to direct guidance at all stages of seeking for information. In ES, a user submits a tentative query with the aim of getting useful documents. The user then further explores the environment to better understand how to exploit it, selectively seek and unreceptively obtain clues about their next line of search actions [25]. ES is noted to be a specialization of information exploration, a comprehensive class of undertakings for searching for novel information in a definite conceptual area; as in exploratory data analysis [26].

**Table 3:** ES Classification Activities

| ES Activities | Detail |
|---|---|
| Lookup | Under Lookup fact retrieval, the following are the contents: known item search, navigation, transaction, verifications and question/answering. |
| Learn | Under Learn, the following are found: knowledge acquisition, comprehension/interpretation, comparison, aggregation/integration and socialization. |
| Investigate | Under Investigate, the contents are, accretion, analysis, exclusion/negation, synthesis, evaluation, discovery, planning and forecasting, and transformation. |

Marchionini suggests that only the latter two activities constitute ES as illustrated this in Figure 3. This is an indication that a dissimilar notion for processing data, other than retrieving a set of important materials through a SE should be used. The advantage of this is that it shields searchers from the classical search where they are forced to browse through a long list of clues, which is ineffective, and, as they are unable to precisely define the content of their query. Hence, the use of the ES is proposed as a more effective mechanism for searching. Marchionini quoted the following about ES[27]:

"ES has the potential to give a more complete overview of a topic based on less specific queries. It also allows the user to discover previously unknown facts and to identify relationships within a topic of interest".

### 3.7. Exploratory Search Systems (ESS)

Exploratory Search Systems (ESS) is a search system that uses ES. It uses a new technology and interface paradigm that facilitates larger interactions with search systems. Visualization systems, document clustering browsing systems and intelligent content summarization systems are the examples of ESS. Basically, ESS are mainly set up to provide a wide-range of mechanisms to users to expedite their search process of research and learning. ESSs are also set up to guide searchers in their pursuit of researching unfamiliar grounds [28]. By way of various levels of exposures to information, people learn to explore compendiums and objects that can be found in a data bank. As such, one of the functions of the ESSs, which involves appropriately categorizing information in an understandable and exploratory fashion by users becomes paramount[29]. ESSs help users to engage in browsing that make the most of how much information gain is gotten, how well they decide on which navigational routes to take, and how well they comprehend the information they access. Additionally, through interface structures such as dynamic queries[30], ESSs is capable of helping users see the instantaneous effect of their decisions. ESSs does not just return a lone document or respond to just a query, it rather significantly stimulates searchers cognitive change by way of learning and understanding. Dynamic queries used in ESS is useful for generating hypotheses by simultaneously exploring data on a number of dimensions.

Relatedly, faceted navigation is another important feature of the ESSs. Basically, the facets are a representation of metadata that relates to the objects being searched for and which could plausibly be used to represent the data in a profound manner. Additionally, faceted navigation has the capacity to sieve through desired results and explore the collection to identify what is relevant and what is not in a particular search. On another note, visual representation of data is also important because it arouses searchers' insight and helps prepare them on their next line of exploration[31]. Accordingly, visual representation is considered by demonstrating the extent of resources obtainable per facet resource count. Hence, this research examines the impact of these features.

## 4. Faceted Search Systems (FSS)

Faceted search system applies classic facet theory in the online digital environment. It is the grouping of free, unstructured text search, with faceted navigation. In the position of White and Roth, they noted that faceted search interfaces are interfaces with an impeccable combination of keyword searches and browsing. By this, people are allowed to quickly and flexibly access information based on what they remember about the reason for their search. Faceted interfaces help searchers from 'being lost' to system explorers. Many SEs offers faceted browsing or navigation tech-

nique. Faceted browsing allows application of dynamic filters in exploring data in several phases. That is, whenever a filter is applied, the results are shown to the user, who has the prerogative to apply additional filters or modify existing ones[32, 33]. The power of FSS rests in the ability of users to create their own custom navigation by combining various perceptions rather than constraining them through a definite path. FSS allows a multi-dimensional exploration of a subject. Interestingly, notwithstanding that keyword searches typically pop up a ranked result list, in FSS, users have the liberty to filter the result set by stating one or more preferred characteristics of the dimensions[34, 35]. Traditional faceted navigation styles allow searchers to drill down into a subject matter in search of specific documents. One limitation to this, however, is the likelihood that searchers will get a very "narrow" view of the issue(s) in search [10].

# 5. Literature Review

Search engine has become the primary mechanism for IR on the World Wide Web. Particularly, foremost web SE recently reported that they respond to a total of 100 billion queries per month, and track over 30 trillion unique URLs[36, 37]. Nonetheless, the massive content of data on the Web is not the only challenge faced by current web SE. basically, discovering information is not the primary essence of a search. Search processes are just a means to an expected end. Hence, in order to better serve their huge customer base, it is pertinent for SE managers to conduct exploratory conversations with customers' to really know what the customers are really in need of, and the reasons behind it. For example, a customer who looks for a local map may be on the way to a train station, which may be closed. In this case, just finding the information does not help the final goal. Characteristically, searching is challenging. However, the challenges are dues to a number of reasons:

## 5.1. Query Problems

On the web, a user typically expresses his/her information need with only three terms or less. By this, a knowledge-gap is thus created which makes the searcher not to be able to generate the correct query. Therefore, several SE have allowed the user to specify his/her domain of interest or at best describe his/her interest in a profile[38] so as to assist the search process. Such approaches create some limitations in the domain for which the search is being conducted. It is also not a universally accepted process, and not every user will be able or willing to provide such information[39]. Sometimes, users also search with some form of ambiguity, which is capable of not guiding the SE properly. This in turn causes the user to abandon the originally query. In sensibly addressing this challenge, it is sensible to expand the scope of the searcher and related documents retrieved for a query. There is thus the likelihood that at least one of these documents will meet the actual information need of the user.

Furthermore, though the usefulness of the present generation SE is noted, they tend to perform optimally only for easy queries. For example, queries to find homepages or find popular/known topics. Generally, when the information need of a user is well-defined in addition to a very good knowledge of target web page(s), the user can articulate an effective query process which prompts the SE in returning relevant results on the top. In such circumstances, a query with a few keywords should be adequate. Regrettably, when a user is bereft of any particular target pages or does not have a good grasp of the topic to be searched, as is often the case in ES and informational search[27], such short keyword queries will turn out ineffective. Technically, a number of reasons are attributable to why a query is ineffective[40]. Three of the plausible reasons are discussed below:

### 5.1.1. Ambiguity

Queries should not be ambiguous. In cases where a query contains some vague words, but the user is only aware of one specific sense, search results may give the best of results since they are controlled by an undesired sense or mixed by multiple senses. For example, the results in the first page returned from Google for the ambiguous query "jaguar" comprise at least four different senses of "jaguar" (i.e., car, animal, and a sports team). For a more refined query such as "jaguar team", the results will still be quite ambiguous, including at least four different jaguar teams: a wrestling team, a jaguar car team, Southwestern College Jaguar softball team, and the Jacksonville Jaguar football team.

### 5.1.2. Vocabulary Mismatch

In an instance when a user searches for very precise information in a domain that is unfamiliar to him/her, the user may perhaps be unaware of the proper terminology with which to label the information he/she needs. Consequently, the possibility is there that keywords in the query may not fit the terms used in the appropriate documents. This could cause a challenge of vocabulary mismatch. For instance, if a user desires to obtain information relating to retrieval functions used by SE but is unaware of the right IR terminology to search with, queries such as "SE formulas", "SE query execution methods", and SE scoring methods" may be used by the searcher Interestingly, none of these will give the desired result. A better and direct query should be "SE retrieval functions".

### 5.1.3. Lack of Discrimination

In a domain that is unfamiliar, users always find it difficult to think of more specific terms and the consequential queries that is likely to arise therefrom. This may lead to a citation of not being able to get the desired documents as sought for by the searcher. For example, a query such as "auto quotes" can produce diverse results. Some results will relate to automobile insurance quotes and some about automobile sale prices. In such an instance, it would be useful to add "insurance" or "sale" to make the query more distinctive. This would ensure coherence and usefulness in the results for a refined query for the end users.

## 5.2. Information Overload

The Web is a repository of everything. Anything that cannot be found on the Web is definitely not worth finding. To the user, information on the web seems more credible than information gotten from other sources due to the ease in accessing such information. However, the vast information lies inside the backend databases. In a usual search, the desired result lies in first 10 results which are ranked by ranking algorithms. Usually, this ranking is plausibly a function of the frequency of keywords in the web page or on more sophisticated methods of evaluating linking, as is the case for Google. Nonetheless, it is only the user that can determine the usability of the first ten results generated by the search, as there is no assurance that the first ten results are the best [41].

"Searching is easy; finding is difficult". Specifically, today's SE is made for searching information and searchers can effortlessly find results based on their query. However, the ease in easily finding results of a search is a function of the appropriateness of the query or expertise of user. Successful search requires time to ensure better findings.

The WWW is loaded with a huge quantum of information that is decentralized, redundant and inaccurate which makes the use of whatever is generated cumbersome. This is referred to as "information overload". However, this challenge has been addressed by advanced technologies in information retrieval, which powers the web SE and make finding of resources easy. However, these solu-

tions are not enough as they fundamentally ignore the actual meaning of the processed data that would enable more sophisticated analysis and filtering of the results.

## 6. Research Evaluations and Validations of SE

Some success has been gained by the introduction of certain systems in various information overload and ambiguous query. However, a short-coming of the functionality of those systems are noted when presenting information and in term of the speed of the system.

IR systems have been evaluated and compared for many years. Dunaiski, et al. [42], and Cleverdon, et al. [43]listed six criteria that could be used to evaluate an information retrieval system: (1) recall (2) precision, (3) coverage, (4) time lag, (5) presentation and (6) user effort. Of these criteria, recall and precision are the most applied in measuring IR. Presentation and user effort are two other criteria that assesses IR in user terms. Unlike recall and precision, there is a paucity of research examining presentation and user effort, and there is no consensus on how they should be measured.

## 7. Search Engine Evaluation

One of the critical factors considered in building a viable SE is evaluation. The essence of evaluation is to access the effectiveness of the SE in a specific application. Characteristically, the purpose of evaluation of the SE is to ascertain its level of effectiveness vis-à-vis it levels of efficiency. While, the ability of the SE to locate the right information is a measure of its effectiveness, efficiency measures the speed at which the locations are done. However, the effectiveness and efficiency of the SE is a function of a number of factors. The factors are, the interface used to demonstrate search results and techniques such as query suggestion and relevance feedback. Hence, it is crucial to note that the focus of IR research is on improving the effectiveness of search, and when a technique has been recognized as being theoretically useful, the focus shifts to finding efficient implementation mechanisms[44].

Accordingly, recall and precision are the two most customary effectiveness measures used in the Cranfield studies for summarizing and comparing search results. Basically, recall measures the functionality of the SE at finding all the relevant documents for a query, while precision measures how well the SE is doing at throwing-out non-relevant documents. Further illustration is presented in Table 4.

**Table 4:** Sets of Criteria for Documents Precision and Recall

| | Retrieved & Irrelevant | Not Retrieved & Irrelevant |
|---|---|---|
| Relevant | Retrieved & Relevant | Not Retrieved but Relevant |
| | Retrieved | Not Retrieved |

A number of effectiveness measures can be defined using this table. This study is mainly interested in the following two equations:

$$\text{Recall} = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of relevant documents}} \quad (1)$$

$$\text{Precision} = \frac{\text{Number of relevant documents retrieved}}{\text{Total of number of documents retrieved}} \quad (2)$$

In view of the fact that users will only glance through a limited number of the results pages of a search, the measurement of recall would not be very useful for the Web as such. Basically, recall on the Web is related to SE coverage. In essence, if the coverage of the SE is higher, that is, it indexes a larger proportion of the Web, and supposing that the standard of its search algorithms are satisfactory, a high level of precision can be achieved.

## 8. Findings

Current search engines use the query and response (lookup) process. Exploratory Search (ES) allows for open-ended search. Faceted search (FS) is a one feature in ES that can be used to improve the overall search. Many studies have shown that replacing traditional search engines with ES by using the features of ES can reduce the data overload. ES handles uncertainty in the search and the possible lack of knowledge about the domain and FS aid users in exploring items of interest within a large data repository. The overall results illustrate that faceted search combine's faceted navigation with full text search. This provides the user with the opportunity to work successfully with content that is semi-structured. Full text search is used to attain obtain those results which do not have structural characteristics. While on the other hand, faceted navigation provides a means of browsing and refining by metadata-structured information. This greatly reduces the chance of generating no results, while still providing refinements when too many results are returned.

## 9. Conclusions

A critical review of IR, the current classification of the SE, types of SE with specific focus on current exploratory search initiatives is examined in this paper. Also examined in this study are a number of approaches adopted by researchers in effectively improving the SE needed by the user. Both the ES and FS techniques are discussed. Characteristically, the main focus of all the approaches is directed towards maximizing the level of satisfaction a user gets in the use of a particular process, and to amplify the visibility of the relevant documents in the SE. This paper also compares various SE and discusses the methods needed to measure the quality of the results. It is noted that these methods should effectively increase the IR process of documents.

## Reference


[1] A. Chandurkar and A. Bansal, "Information Retrieval from a Structured KnowledgeBase," in Semantic Computing (ICSC), 2017 IEEE 11th International Conference on, 2017, pp. 407-412.

[2] R. Campos, G. l. Dias, A. M. Jorge, and A. Jatowt, "Survey of temporal information retrieval and related applications," ACM Computing Surveys (CSUR), vol. 47, p. 15, 2014.

[3] D.-I. D. Backhausen, "Adaptive User Support in Interactive Information Retrieval Processes," 2017.

[4] J. Xu and W. B. Croft, "Quary Expansion Using Local and Global Document Analysis," SIGIR Forum, vol. 51, pp. 168-175, 2017.

[5] A. N. Langville and C. D. Meyer, Google's PageRank and beyond: The science of search engine rankings: Princeton University Press, 2011.

[6] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, et al., "Systematic literature reviews in software engineering–a tertiary study," Information and Software Technology, vol. 52, pp. 792-805, 2010.

[7] A. Figueroa and G. Neumann, "Context-aware semantic classification of search queries for browsing community questionâ€“answering archives," Knowledge-Based Systems, vol. 96, pp. 1-13, 2016.

[8] R. Qumsiyeh and Y.-K. Ng, "Searching web documents using a summarization approach," International Journal of Web Information Systems, vol. 12, pp. 83-101, 2016.

[9] G. Simonini and S. Zhu, "Big data exploration with faceted browsing," in High Performance Computing & Simulation (HPCS), 2015 International Conference on, 2015, pp. 541-544.

[10] B. Kules and B. Shneiderman, "Users can change their web search tactics: Design guidelines for categorized overviews," Information Processing & Management, vol. 44, pp. 463-484, 2008.

[11] M. Kathuria, C. Nagpal, and N. Duhan, "Journey of Web Search Engines: Milestones, Challenges & Innovations," 2016.

[12] Z. Huang and N. Mamoulis, "Location-Aware Query Recommendation for Search Engines at Scale," in International Symposium on Spatial and Temporal Databases, 2017, pp. 203-220.

[13] Y. Chen, G. Y. Jeon, and Y.-M. Kim, "A day without a search engine: an experimental study of online and offline searches," Experimental Economics, vol. 17, pp. 512-536, 2014.
[14] M. Levene, An introduction to search engines and web navigation: John Wiley & Sons, 2011.
[15] V. Ramaswamy, R. Konow, A. Trotman, J. Degenhardt, and N. Whyte, "Document Reordering is Good, Especially for e-Commerce," 2017.
[16] B. J. Jansen and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," Information Processing & Management, vol. 42, pp. 248-263, 2006.
[17] S. Dutta, "Access to the Web Resources through Google: a Brief Study in Popularity and Service Features," International Journal of Innovative Knowledge Concepts, vol. 2, 2016.
[18] M. Peshave and K. Dezhgosha, "How Search Engines Work: And a Web Crawler Application," Citeseer, 2005.
[19] U. Directory, "Umdum Helps You Find the Perfect Websites While Avoiding all the Junk," 2018.
[20] A. Schuth, "Search engines that learn from their users," 2016.
[21] S. Goel and S. Yadav, "Search engine evaluation based on page level keywords," in Advance Computing Conference (IACC), 2013 IEEE 3rd International, 2013, pp. 870-876.
[22] Z. Dou, Z. Jiang, S. Hu, J.-R. Wen, and R. Song, "Automatically mining facets for queries from their search results," IEEE Transactions on Knowledge and Data Engineering, vol. 28, pp. 385-397, 2016.
[23] D. Sullivan, "How search engines work," SEARCH ENGINE WATCH, at http://www. searchenginewatch. com/webmasters/work. html (last updated June 26, 2001)(on file with the New York University Journal of Legislation and Public Policy), 2002.
[24] H. Agrawal and S. Yadav, "Search Engine Results Improvement--A Review," in Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on, 2015, pp. 180-185.
[25] B. Sarrafzadeh and E. Lank, "Improving Exploratory Search Experience through Hierarchical Knowledge Graphs," in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 145-154.
[26] G. Singer, "Web search engines and complex information needs, Doctoral dissertation," Doctoral dissertation, 2012.
[27] G. Marchionini, "Exploratory search: from finding to understanding," Communications of the ACM, vol. 49, pp. 41-46, 2006.
[28] R. W. White and R. A. Roth, "Exploratory search: beyond the query-response paradigm (Synthesis lectures on information concepts, retrieval & services)," Morgan and Claypool Publishers, vol. 3, 2009.
[29] L. Gao, Y. Lu, Q. Zhang, H. Yang, and Y. Hu, "Query expansion for exploratory search with subtopic discovery in Community Question Answering," in Neural Networks (IJCNN), 2016 International Joint Conference on, 2016, pp. 4715-4720.
[30] R. W. White, G. Marchionini, and G. Muresan, "Evaluating exploratory search systems: Introduction to special topic issue of information processing and management," ed: Pergamon, 2008.
[31] D. Tunkelang, "Faceted search," Synthesis lectures on information concepts, retrieval, and services, vol. 1, pp. 1-80, 2009.
[32] W. Kong and J. Allan, "Precision-Oriented Query Facet Extraction," in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, 2016, pp. 1433-1442.
[33] M. N. Mahdi, A. R. Ahmad, and R. Ismail, "A Real Time Visual Exploratory Search Engine for Information Retrieval in a Cloud," International Journal of Future Computer and Communication, vol. 4, p. 216, 2015.
[34] D. R. Harris, "Modeling Faceted Browsing with Category Theory for Reuse and Interoperability," 2017.
[35] M. N. Mahdi, A. R. Ahmad, and R. Ismail, "Paradigm Extension of Faceted Search Techniques A Review," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 9, pp. 149-153, 2017.
[36] C. Li, N. Yan, S. B. Roy, L. Lisham, and G. Das, "Facetedpedia: dynamic generation of query-dependent faceted interfaces for wikipedia," in Proceedings of the 19th international conference on World wide web, 2010, pp. 651-660.
[37] S. Quinton and L. Simkin, "The digital journey: Reflected learnings and emerging challenges," International Journal of Management Reviews, vol. 19, pp. 455-472, 2017.
[38] R. Kelly and S. J. Payne, "Collaborative web search in context: a study of tool use in everyday tasks," in Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, 2014, pp. 807-819.
[39] H. Wachsmuth, M. Potthast, K. Al Khatib, Y. Ajjour, J. Puschmann, J. Qu, et al., "Building an argument search engine for the web," in Proceedings of the 4th Workshop on Argument Mining, 2017, pp. 49-59.
[40] J. Savoy, "Why do successful search systems fail for some topics," in Proceedings of the 2007 ACM symposium on Applied computing, 2007, pp. 872-877.
[41] J. Azimi, A. Alam, and R. Zhang, "Ads keyword rewriting using search engine results," in Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 3-4.
[42] M. Dunaiski, J. Geldenhuys, and W. Visser, "How to evaluate rankings of academic entities using test data," Journal of Informetrics, vol. 12, pp. 631-655, 2018.
[43] C. Cleverdon, J. Mills, and E. Keen, "An inquiry in testing of information retrieval systems.(2 vols.)," Cranfileld, UK: Aslib Cranfield Research Project, College of Aeronautics, 1966.
[44] P. Sirotkin, "On Search Engine Evaluation Metrics," arXiv preprint arXiv:1302.2318, 2013.