

Customised Search Engine for Users' Needs

Iman Alansari

Department of Computer Science
Taibah University
Madinah, Saudi Arabia
iansari@taibahu.edu.sa

Ibtehal Nafea

Department of Computer Science
Taibah University
Madinah, Saudi Arabia
inafea@taibahu.edu.sa

Abstract— The Web has become an important information store. Searching for information on the Web is part of a human's everyday needs. One generic way to search the Web is by using search engines which are accessible programs for searching information on the Internet using keywords, cover titles of documents, URL's, fields, content or full texts. There are usually thousands of pages that contain these keywords, but the user is interested in a much smaller subset contained within the huge list of information. In addition, the desired pages may not achieve the user's needs. This paper deals with a survey on the search classification needs and habits of the web and explores how Internet users prefer to find information in web search engines.

As a result of this study, a methodology for easing a search by using search engine classification is proposed, to minimize time and give accurate related results.

Keywords—search engines; adapting Web Search; information retrieval; search classification

I. INTRODUCTION

Users use the Internet in different ways, such as accessing online journals, downloading software or texts, chatting, discussion, E-mail services and for finding related references. The Web plays an important part as a source to access information resources that are steadily increasing. This increase is due to education and research use [1]. With the growth of information on the Internet and the development of searching tools, it becomes possible to find information and answers to real questions. Also, finding information and answers to real questions is possible with the growth of information on the Internet and the development of searching tools. The main use of the Internet is information retrieval from search engines. Generally, there are two modes of retrieval based on keyword (ad-hoc retrieval) and categories (ontology) in information retrieval terms [2]. The purpose of this paper is to develop and examine users views towards search engines as a tool for retrieving information. In addition it should also provide users with adaptable searches for their needs and identify trusted websites.

The underlying question of the research presented in this paper is how Internet users prefer to find trusted information in web search engines. Survey forms are presented in Arabic and English languages, which address many questions that aim to develop a search engine that is appropriate to users' needs in order to determine their search. We consider a search engine's customization by providing classified tabs along with the search

page, to help users find desired information that is related to our proposed classification, such as a search within education and health sectors.

The remainder of the paper is structured as follows: Section 2 describes related works. Section 3 presents the methodology used in this study. Section 4 describes the survey results and their analysis. Section 5 concludes the paper.

II. RELATED WORKS

Search engines are used extensively every day, but what is the reason behind this web search? Taxonomy of web searches can be classified into three classes namely, navigational, informational and transactional. In navigational style, the user's intent is to reach a particular website, in informational style the user's intent is to get information from one or more web pages and finally in transactional style, the user's intent is to perform some web-mediated activity like shopping or looking for data in the Yellow Pages [3].

The authors identify three stages in the evolution of web search engines: The first generation uses classical information retrieval that covers informational queries. The second generation uses off-page web-specific data such as link analysis, anchor-text and click-through data, which supports both informational and navigational queries. Navigational queries were first used by Google. Finally, the third generation tries to accomplish the search that answers "the need behind the query", which supports transactional queries via semantic analyses that blends various external data bases. The rapidly changing landscape covers semantic analysis, context determination, dynamic data base selection, and much more [3].

Web search engines are different, in various aspects. Search engines can be compared in terms of their search capabilities and retrieval performances such as Boolean logic, truncation, field search, word and phrase search, retrieval precision and response time. These facilities determine search engines' performance. A study was conducted comparing three search engines in terms of their search capabilities and retrieval performances. Ten search queries extracted from real reference questions posed by the librarians at Long Island University showed that different search engines may outperformed others in both search facilities and retrieval performance, although some of these search engines had the largest coverage of web resources. The authors concluded that they would apply their proposed methodology to a wider range of search engines, to enable Web users to select a search engine that meets their specific search needs. It would

also help Web search engine developers design even better ones for the Internet community [4].

Several studies deal with the Multimodal information retrieval (MMIR) field, which meets users needs by allowing them to express their query in any form suitable to them and to retrieve content in several forms with a complete view of the retrieved information.

Content Objects (COs) is a complete solution for search and retrieval of rich multimedia content over modern databases [5]. It combines the advantages of multimodal search with those of annotation propagation into a unified system. The performance of the proposed method was evaluated by using two properly constructed multimodal datasets. The method achieves quite promising results, both in terms of automatic annotation propagation and multimodal search and retrieval.

Another framework called Multimedia Document (MMD), defines a set of multimedia objects (images, audio and text) that carry the same semantics [6]. A ranking algorithm was applied after creating a Multimedia Correlation Space (MMCS) and each MMD is represented as a data point. A local linear regression model is used for each data point and then it globally aligns all of them through a unified objective function.

However, the last methods show significant retrieval accuracy only when the query objects belong to the dataset. However, for other query objects that do not belong to the dataset, the methods provide low support. These methods can produce acceptable results by applying relevance feedback.

One other trend is the popularity of social networking sites, such as Facebook, MySpace and LinkedIn, which have introduced a new option for finding information online. In a study exploring the pros and cons of using a social networking tool to fill an information need, as compared with a search engine, the authors describe a study in which 12 participants searched the Web while simultaneously posing a question on the same topic to their social network [7]. They compare the results they found using each method. Although participants preferred using search engines to look up information, using a social network gives personalized answers and instills more confidence in search results, suggesting querying search engines and social networks in parallel. Moreover, Google provides Internet data analysis at the enterprise level, which makes it easy for developers to extract live data for analysis.

III. METHODOLOGY OF THIS STUDY

This study is meant for researchers in the field of an adequate search engine. In order to adequate search engines, we need to develop our understanding of user behavior and their feedback. Search engines reflect a set of user needs and we assume that understanding these needs will help to improve search engines, adequate those to help users find information as they prefer and give more specified accurate results. Our research goals include:

Investigating how users find information using search engines to support their research and determining whether search behavior varies according to education levels.

Understanding aspects of the preference of searchers, such as searching by field, or by content. They have also been asked

if they trust search results from official websites, or their selection of service types as seen in their search.

Based on the data gathered from usage patterns in the survey, a new suggestion for search structure is presented to improve search engines and this will lead to providing users with their needs during search sessions.

To simulate the suggested design, Customized Search Engine (CSE) which is powered by Google, has been used to enable dynamic selection within a search engine's feature, so that users can find what they need on their website and narrow their search results. CSE can be used to refine and categorize queries based on Google search. To create such a search engine, the user has to have a Google or Gmail account. Customization is different from normal search as it emphasizes the results above anything else on the web and some features are removed, such as personalized results [8]. We have created two search engines namely, *EduSearch* which focuses its search on Education in general and *HealthSearch* which focuses on the health sector, as seen in Figure 1.



FIG 1. SCREEN SHOT FOR BOTH SEARCH ENGINES USING CSE CONTROL PANEL.

Figures 2 and 3 show a screen shot for each developed search engine.

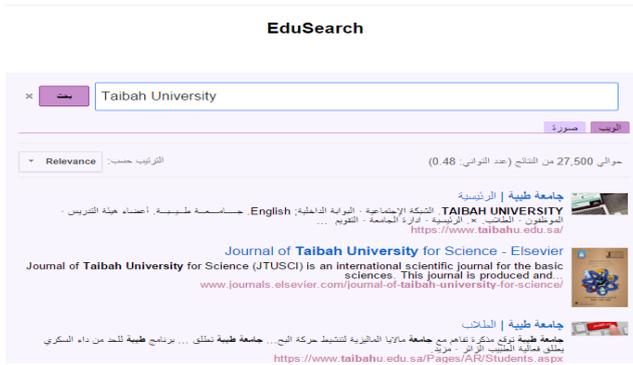


FIGURE 2: EDUSearch CUSTOMIZED SEARCH ENGINE.¹

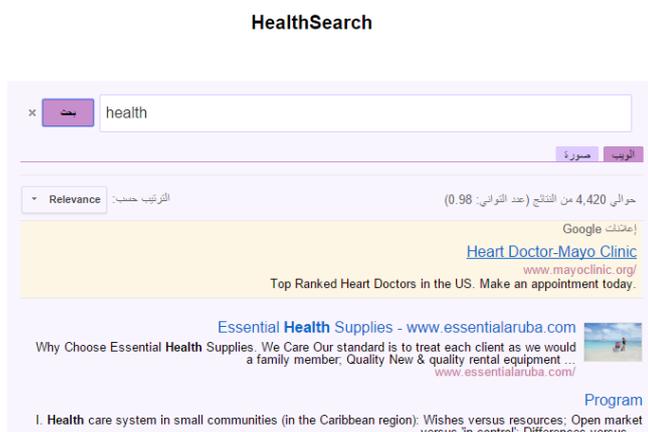


FIGURE 3: HEALTHSEARCH CUSTOMIZED SEARCH ENGINE.²

Data was collected online using the survey tool Google Docs. The survey was available in English and Arabic, and distributed and shared through social media channels on WhatsApp and distributed via e-mail. The survey was available from 15/4/ 2015 to 28/4/2015. A total of one hundred and ninety (190) individuals responded.

63.7% of our participants are in the age group 25-40 which is reasonable for the users who are developing educational skills, because they have selected the Education option in their search preference.

As a case study for our proposed design we choose Education tab as it is the most frequent selected choice in our survey. Analysis will cover time, number of related websites and relevance for selective search.

Although most of the users intend to search for different reasons, our results show that most of them search to acquire information on a particular subject, to perform an activity and finally, to find a particular website. This emphasizes the most desirable search mechanism, which is to search using words or phrases other than website links or by doing an activity.

The users responses showed that it important to give more attention to developing search engines due to extensive daily usage, at least several times per day.

Google search engine was ranked first on the list in our listed search engine. It shows that almost 99.5% of users have used Google in their search due to their retrieval information being more accurate and related to users needs, as 90.5% of the participants always/mostly find what they need. This is not surprising, as current statistics show that Google has become the search interface of choice for many faculty and students, to address their information needs and to use library catalogs [8].

60% of the participants agreed that the information they usually find is accurate and trustworthy, but 35.3% of them agreed that not all of the search results are confidential. This raises the need to increase the percentage of trusted information. This guides our attention to finding certifying a mechanism for searching or certified search.

Each search engine displays the results depending on their importance, by using different algorithms. In conducting a study [9], they investigated the relationship between website quality and website importance. They considered that if quality is positively related to website importance, a majority of first-page search results will contain high quality information, as seen in Figure 4.

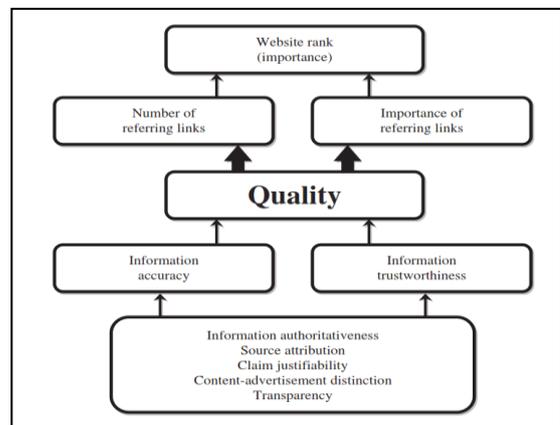


Figure 4: CONCEPTUAL MODEL OF THE RELATIONSHIP BETWEEN WEBSITE SEARCH RANK AND QUALITY [9].

Note, that government, universities, organizations and education websites can be considered trustworthy due to their high ranking and visibility, which is supplied by Search Engine Optimization (SEO). In another example, The Health On the Net (HON) which assumes the Code of Conduct in order to provide reliability of health information that is found on the search engine. HON certifies websites against this code upon request, which allows certified websites to display quality of health information [10].

IV. RESULTS DISCUSSION

¹ <https://cse.google.com/cse/publicurl?cx=005139713032654573167:eobao-iej4a>

² https://cse.google.com/cse/publicurl?cx=005139713032654573167:i6gqlmmn_js

The users reported that they found a great amount of information (77.9%) and most of it was clear and related as reported (1.1% found obscure search results, and 8.9% said they found unrelated search results). Only 17.9% participants said that they found conflicting search results.

The results showed an almost equal percentage in user search preference. 45.8% preferred to search by field and 52.6% preferred to search by content. Equal attention has to be given to developing advanced web searches. Also, users have been asked to select what options they like to see in web search engines in order to personalize and adapt their search.

TABLE I. PERCENTAGE FOR USERS SELECTION IN WEB SEARCH OPTION

Percentage	Search source results
96.80%	official websites
6.30%	Personal Home Pages
5.80%	Forums
5.80%	high indexed websites
3.20%	I do not care about the source of information

The results showed the variety in user selection, which indicates that they adapt their searches for their needs. As seen in table 1 the highest percentage was related to education, reasoning that most of the participants are educated and using search engine mostly in education and research. Secondly, users selected to search videos and images which mostly related to their informational needs. Others needed to see the options of services, such as the language option, email, maps and calendar.

This concludes that users' needs focused on three options which can be summarized as, search by field, content and services. Our proposed search engine has to give users an adaptable search for their needs and from trusted websites, as 96.8% asked for trusted websites, see table 2. Forums, personal homepages and high indexed websites are not likely to be a search source. Note, that most people do not have enough knowledge about the benefits of high indexed websites and it may be because Google has become completely automated.

TABLE II. PREFERRED USER SEARCH SOURCE PERCENTAGE

Percentage	Web Search options
61.60%	Education
56.80%	Video, Images
55.80%	Health
44.20%	Language translator
42.60%	Technology
42.60%	Mail
36.80%	News
34.70%	Science
34.20%	Entertainment
33.20%	Government
31.10%	Travel
29.50%	Social network: Facebook, Twitter, Instagram
29.50%	Language option
26.30%	Arts
24.20%	Personal Webpages
24.20%	Maps
22.10%	History
20.50%	Calendar
20%	Religion
18.40%	Network
16.80%	Forums
12.60%	Games
2.60%	Other

In regards to customize web searches such as choosing a preference or restricting the source of information, 62.1% participates preferred to increase child search safety. 52.6% considered customization for giving more specified accurate results, and it is worth mentioning that most participants in this group are educated. Another important factor to consider is to increase the search speed, which is selected by 26.3% participants. A small percentage of the participants (7.9%) cared about powerful web search, without customization.

In our design we added the mainly preferred options in our survey which includes: Education, Health, Technology, Science and governments.

As a simulation to our design we have used CSE to test our proposed design in Education tab, as shown in Figure 5.



Figure 5: SIMULATION DESIGN FOR CUSTOM SEARCH

The following Table provides a comparison in results between our customized Educational proposed search tab in Google and regular Google search results.

TABLE III. COMPARISON RESULTS BETWEEN CUSTOMIZED AND REGULAR SEARCH

keywords	Google Site Search			Customized Search Engine (Education tab)		
	Time/second	Result count	Relevance	Time/sec	Result count	Relevance
Taibah University	.54	197.000	80%	0.32	25.300	95%
Saudi education	.44	142.000	95%	.40	37,100,000	95%

As seen in Table 3 we consider the three main features of search: time of getting results, results account and percentage of relevance of results in first twenty appearances. We submit keywords (Taibah University, Saudi education) in both search boxes.

Comparative results from Google Site Search to the results of the Customized Search Engine show the efficiency of the latter, as it is faster and more relevant. However, the result total of the Google site is greater than the Customized Search Engine, but it does not mean that most of these results are gathered from social websites such as Facebook and Twitter.

V. CONCLUSION

Our paper was designed as a survey to investigate how users utilize search engines and how their behavior is effective to adequate the search engine. Our purpose is to hold an exploratory study in order to expand our understanding of aspects of the preference of searchers, such as search by field or by content. They may also trust search results from official websites, with their selections of service types visible on their

search engines. We observed that search behavior and perceptions depend on the education level. Data was collected by using a questionnaire that was distributed on line and used social applications. Results showed that all the respondents are frequently using the Internet search engines to look for information that they need. It was observed that the web is normally used in research for educational purposes. Also, it is observed that the Google and Yahoo search engines are more widely used compared to other search engines, due to their performance quality, which much better than the other engines. The analysis shows that 90.5 percent of the participants always find what they need and 60% of the participants find accurate information on the Internet. 52.6 % preferred to search by content. 96.8 percent of the participants believed that trusted information is available on the Internet, while 96.80% of the participants preferred their search results to be official websites.

The proposed paper provided classified tabs to help users find desired information that is related to our proposed classification such as, search within education and health sectors.

In the future, we can allow our user to add his/her own tabs, as presented in Figure 5 in (custom search) at the bottom.

REFERENCES

- [1] Edwards, S. L., & Bruce, C. S. (2002). Reflective Internet searching: an action research model. *The Learning Organization: An International Journal*, 9(4), 180-188.
- [2] G. Hubert and J. Mothe, "An adaptable search engine for multimodal information retrieval". *J. Am. Soc. Inf. Sci.*,60: 1625-1634. doi: 10.1002/asi.21091, 2009.
- [3] Broder, Andrei. "A taxonomy of web search." *ACM Sigir forum*. Vol. 36. No. 2. ACM, 2002.
- [4] Chu, Heting, and Marilyn Rosenthal. "Search engines for the World Wide Web: A comparative study and evaluation methodology." *PROCEEDINGS OF THE ANNUAL MEETING-AMERICAN SOCIETY FOR INFORMATION SCIENCE*. Vol. 33. 1996.
- [5] Lazaridis, Michalis, et al. "Multimedia search and retrieval using multimodal annotation propagation and indexing techniques." *Signal Processing: Image Communication* 28.4 (2013): 351-367.
- [6] Y. Yang, D. Xu, F. Nie, J. Luo, Y. Zhuang, Ranking with Local Regression and Global Alignment for Cross Media Retrieval, *ACM MM*, Beijing China, 2009.
- [7] Morris, Meredith Ringel, Jaime Teevan, and Katrina Panovich. "A Comparison of Information Seeking Using Search Engines and Social Networks." *ICWSM 10* (2010): 23-26.
- [8] <https://cse.google.com/cse/>, accessed on 11/11/2015.
- [9] J.R. Griffiths and P. Brophy, 2005. "Student searching behavior and the Web: Use of academic resources and Google," *Library Trends*, volume 53, number 4, pp. 539-554.
- [10] Kitchens, Brent, Christopher A. Harle, and Shengli Li. "Quality of health-related online search results." *Decision Support Systems* 57 (2014): 454-462.
- [11] Health on the Net Foundation. retrieved on May 26, 2011 at <http://www.hon.ch>