Webology, Volume 16, Number 1, June, 2019

Ask Search Engine: Features and Performance identification

Mozaffar CheshmehSohrabi

Corresponding author, Associate Professor, Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran. ORCID: 0000-0003-1856-4210. E-mail: mo.sohrabi@edu.ui.ac.ir

Neda Abbasi Dashtaki

Master Degree, Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran. ORCID: 0000-0001-9935-4922. E-mail: nabbasi.d.69@gmail.com

Received January 12, 2019; Accepted April 30, 2019

Abstract

The structural and content features of the *Ask* search engine together with its assessment based on the three types of *keyword*, *phrase* and *question* queries run on information retrieval performance are identified in this article. This is an applied research run by adopting the survey and documentation methods. Two checklists were used to collect data: one for identifying the structural and content features and the other for recording the recall and precision ratios. The obtained data is then recorded to calculate the recall and precision. In total, 48 structural and 16 content features are identified. The findings indicate an average of 44.95 percent recall and of 31.54 percent precision in this search engine. This fact reveals that the *Ask* search engine performance is not appropriate. The obtained results emphasize the fact that the performance of information retrieval through *question search* method outperforms *keyword search* and *phrase search* methods.

Keywords

Ask search engine; Precision; Recall; Structural features; Content features

Introduction

The Web is a great information resource center where information is presented in the form of web pages interlinked with one another (Page et al., 1998). The growth Internet users count seeking information available in the web cause difficulties in finding or accessing relevant information or maintaining the information on any of search engine. The difficulty in extracting information is due to the fact that web content is presented primarily in a natural language, with human readers as its audience (Chahal, Singh & Kumar, 2013).

The users must apply the most practical and popular search tools to seek and retrieve information on the Web, which collect, index, classify and seek the information available in the Web environment. Internet users are provided with information through the simplest and most sophisticated search engines.

Search engines are applied on the web as a tool for information retrieval. Because web is a gigantic repository of heterogeneous and unstructured data, search engines are necessary to filter out relevant information from the irrelevant ones (Dutta & Bansal, 2016). Search engines are the Web power houses which response to queries every day, hence, meeting the user's requirements. Search Engines here have been and are the key component in Internet. There exist a great number of search engines in the world. Because users need to perform successful search through search engines, possibly with the least time and cost, they must be familiar with the effective search engines. Meanwhile, efficient search engines are identified and determined by assessing the function of *information retrieval* and verify their properties. Recall and precision are among the criteria applied in running a full assessment on the search tools' functionality (ies); consequently, the users must get acquainted with the structural and content characteristics of the search engines as well.

Structure of a page is linked to all other pages on the website; thus, allowing visitors to navigate through the available information upon need. Content refers to the information, features, or services offered and provided on the website, designed to be available and accessible to visitors (Huizingh, 2000). Content is what is inside a site and identifies the different types of information within (Robbins & Stylianou, 2003).

There exist many public and specialized search engines. *Ask* is one of the top 10 search engines in the world (Rashid & Wasif Nisar, 2016). *Ask.com* is initially known as *Ask Jeeves*, a web search engine operating based on the question/answer principle. *Ask.com* was innovated in 1996 by Garrett Gruener and David Warthen in Berkeley, California. *Ask.com* is criticized for its browser toolbar because its functions like a malware and when coupled with other software, become utterly difficult to dismantle. Initial objective of *Ask Jeeves* is to provide answers to the end users regarding the daily-life events in a natural language through traditional keyword searching. The advanced *Ask.com* is still supported through traditional search facilities like:

math, dictionary, and conversion questions. A unique feature provided by *Ask.com* is an encyclopedia, a type of unique reference work containing articles on different scientific topics (Dutta & Bansal, 2016).

Although, there exists many search engines on the web, their quality is not known in terms of interface features and retrieval performance (recall and precision). Given the fact that ordinary Internet users may have certain level of knowledge about search engines, the need to run studies like those already run by Spink et al (2006), Kumar & Prakash (2009), Deka and Lehkar (2010), Rathee et al (2013), Olajide and Matthew (2014), Sahu, Mahapatra and Balabantaray (2016), Parsania, Kalyani and Kamani (2016), Martinez-Sanahuja & Sanchez (2016) on *The evaluation of search engine performance* or on *Interface features of web search engine* Azizan et al., (2013), on *Keyword search & semantic search* (e.g., those performed by Rashid & Wasif Nisar (2016), and Mala and Lobiyal (2016) are outstanding in introducing the features and performance of search engines to the users, moreover it is contributive in choosing the most relevant search engines and save time.

The status of ASK performance in terms of related documents' retrieval in some of the available studies is tabulated in Table 1.

Table 1. The details of Ask performance in some of the previous studies

Row	Author	Pub. Year	ASK ranked in terms of related documents' retrieval				
			1st	2nd	3rd	4th	5th
1	Demirci, Kişmir & Bitirim	2007	Google	Yahoo	AlltheWeb	MSN	**Ask**
2	Lewandowski	2008	Google	Yahoo	**Ask**	MSN	Seekport
3	Deka & Lahkar	2010	Google	Yahoo	Live	**Ask**	AOL
4	Wang et al	2012	Bing	Yahoo	**Google**	**Ask**	-
5	Balabantaray, Swain & Sahoo	2013	Google	Yahoo	AOL	Bing	**Ask**
6	Sunitha, Meena Preethi & Akshay	2013	Google	HotBot	AltaVista	**Ask**	Yahoo
7	Sahu, Mahapatra & Balabantaray	2016	Google	Yahoo	Bing	**Ask**	-
8	Bitirim & Görür	2017	Yahoo	Google	Bing	**Ask**	-

The objective of this research is to: 1) identify the interface features of the *Ask* search engine, and 2) assess its performance through recall and precision ratios.

Methods and Materials

The objective here is to analyze the structure and the content of the *Ask* search engine and to determine its information retrieval performance status. For this purpose, two checklists are applied: one for identifying the structural and content features and the other for recording the recall and precision ratios. The obtained data is then recorded and applied in calculating the recall and precision. Moreover, to examine the search engine performance, three sets of experiments are run where 20 queries are applied as follows: 1) Question Search (QS), 2) Phrase Search (PS), and 3) Keyword Search (KS). In both the *phrase and keyword* search, for each query, a synonym word is searched. To assess this search engine's performance, the first 20 results are examined. The data are assessed through Excel. The following two equations are applied to determine the recall and precision ratio of the results:

$$\textbf{Precision}\ (\text{percent}) = \frac{\text{Total percentage of relevant retrieved contents}}{T = (NR + RR)}$$

$$\textbf{Recall}\ (\text{percent}) = \frac{\text{Total percentage of relevant retrieved contents}}{(\text{KS} + \text{PS} + \text{QS}) - \Sigma \text{R.R}}$$

PS = Phrase Search
KS= Keyword Search
QS=Question Search
NR=Non-repetitive Record
RR = Repetitive Records
T = Total count of the relevant retrieved records based on KS

These equations are based on fuzzy logic and the relevant degree of retrieved information. The relevance of retrieved information with the user's request is determined within 0 and 100 degrees. To answer each question, the relevant degree of retrieved information to the users' requirements is determined (e.g., in question 2, *Who has won the Nobel prize twice?*, Table 2, the response is *Marie Curie, Frederick Sanger, Linus Pauling and John Bardeen*, thus, 25 percent for each correct answer). Therefore, unlike the available studies run on search engines' performance assessment with precision and recall criteria, based on relevance or non-relevance(0 or 1), the assessment criteria in determining performance in this study is not of classical logic (0 or 1), but is of correct answer percentage.

Results

The results are presented in two sections: 1) ones concerning the structural and content features of *Ask*, and 2) ones concerning the *Ask* performance assessment.

a) The Structural and Content Features of Ask

The 48 structural and 16 content features of the *Ask* search engine properties obtained through the *Ask* search engine site observation, both the engine's guide section and the relevant scientific literature review are tabulated in Table 2.

Table 2. The structural and content features of the Ask search engine

	Components	Count of features	Features
The structural features of Ask	Search	**18**	Advanced Search, Search Web, Search Images, Search Videos, Search News, Search operator, Change Search Settings, Safe Search, Careers, Preferences, Service Ask.com, Questions/Answers, Ask the forum members, Traditional search possibility, Product search, Search for celebrities, Map search and Ability to ask questions instead of words search
	Display	**7**	Ability to view the URL of the retrieved page, Ability to display by relevance, Insert title, Brief description and site URL in retrieved results, Highlight display of search keywords in retrieved records, Proper layout, & Comfortable reading of available text in terms of font size
	Retrieval	**7**	Search shortcuts, Offer a search suggestion, offering multimedia results, Possibility to open the results page in a new browser, Ability to browse and to examine the results, Hide the time spent to retrieve the research and Hide the total number of retrieved results
	User friendly	**16**	Service themes, Service terms, Maintaining the technical information of the user's computer, Maintenance of user IP, Maintenance of log file of the user, Maintaining the URL of the last web page before visiting the site, Using the pixel tag, Possibility to configure the cookies in other browsers, Ability to delete or change the user's personal information, help service, the possibility of contact with the supporter, Help search, Ask toolbars Help, Links to other pages on each page, Ask Eraser service and Possibility to provide feedback
The content features of Ask		**16**	Frequently Asked Questions, Encyclopedia, Dictionary, Currency exchange, Mobile content, Help charity by answering questions, Matches, Survey, Message forums, Chat groups, IAC software, Social Networks, Games, Careers service, Smart Answer, Related Search and Ads related to.

b) Performance assessment of Ask

To assess the performance of this search engine, a total of 100 queries of: main keywords, synonym keywords, main phrases, synonym phrases and questions, 20 each are searched and the recall and precision ratios of the *Ask* search engine are tabulated, Table 3.

Table 3. Precision and Recall of results

No.	Query	Recall (%)	Precision of KS (%)	Precision of QS (%)	Precision of PS (%)
1	What are the symptoms of heart attack?	62.35	36	45.50	36.00
2	Who has won the Nobel prize twice?	37.45	31.25	47.50	47.05
3	How many websites are there on the Internet?	20	3	20.00	21.43
4	Where is the Apple company located?	30.61	30.00	35.00	15.00
5	Who is Sigmund Freud?	61.4	58.75	70.00	33.75
6	Who discovered penicillin?	78.2	75	95.00	53.85
7	What constitute the elements of water?	4.34	2.5	10.00	2.5
8	What is semantic retrieval?	45	35	35.50	1.27
9	When was Jesus born, the year?	77.77	55	35.00	00.00
10	Who is the current president of united states?	34.24	50	65.00	3.45
11	How many people were killed in Vietnam war?	72.81	38.25	75.00	00.00
12	What is the population of Switzerland?	40.35	23.25	30.50	1.11
13	Which is the most famous museum in world?	26.22	23.5	37.00	33.75
14	What is the hypertext markup language?	47.53	53.5	65.50	22.08
15	What does crawling robot mean in search engines?	32.77	24	18.00	00.00
16	What is the oldest British political party?	23.4	20	30.00	00.00
17	Which are the best procedures in weight loss?	41.92	29.37	42.50	13.14
18	What are the advantages of pear?	45.83	24.37	45.00	25.66
19	What is the Japanese New Year customs?	88.48	50	55.00	00.00
20	What are the reasons of violence among athletes?	28.24	17.5	45.00	00.00
Total average		**44.95**	**34.01**	**45.1**	**15.50**
			31.54		

According to data displayed in Table 3, the precision average of the *Ask* search engine is 34 percent for *KS*, 15.5 percent for *PS*, and 45 percent for *QS* methods, indicating that *QS* precision ratio is higher than that of the *KS* and *PS*. The recall and precision averages of the *Ask* search engine are 44.95 and 31.54, respectively, Table 3. The spectrum of the precision ratio of query samples based on three PS, KS and QS types of searches in *Ask* search engine are diagrammed in Figure 1.

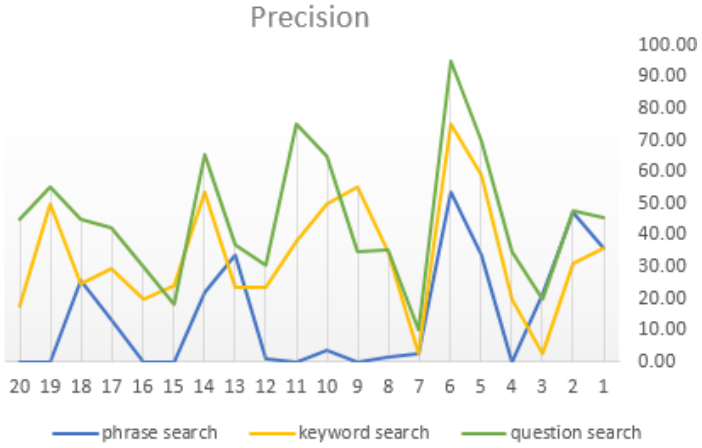

Figure 1. The precision of query samples based on the type of search

Discussion

According to the findings here, *Ask* consists of 48 structural and 16 content features. Among the structural features, components of *Search* and *User-friendly*, with 18 and 16 properties are in a much better state than that of the *Display* and *Retrieval* components, respectively. The unique content feature of the *Ask* search engine is its being evaluated with an encyclopedia. A precision of 31.54 and a recall of 44.95 percent indicate that the performance of the *Ask* search engine is not up to par. However, the information retrieval performance here, as to QS method outperforms that of the KS and PS, and the results obtained indicate that the volume of structural and content features influence the performance of this search engine. Multiple structural features of this search engine have a direct effect on the recall ratio, while few content features have an indirect effect on the precision ratio and efficiency.

Based on the *Phrase Search* (PS) for 6 of the queries (counts: 9, 11, 15, 16, 19 and 20), no results are retrieved, Table 3. In *Question Search* (QS) and *Keyword Search* (KS), the lowest precision percentages are 10 and 2.5 percent for *elements of water* query, respectively. The recall ratio of this query is 4.34, the lowest among the 20 queries. Here, it is assumed that the low level of precision of this query is due to the low coverage of documents and contents related to *elements of water* subject.

In *Phrase Search (PS) method*, in 19 out of the 20 searched phrases, the precision ratio is less than 50 percent. In query number 6, the precision ratio for *QS*, *Who discovered penicillin?*, is 95 percent, and the same is 75 percent for *KS* (with keywords: penicillin + discoverer or penicillin + inventor), while, the precision ratio of *penicillin discoverer* or *penicillin inventor* phrases is 53.85 percent in PS, Figure 1 and Table 3. Now, the question is: Why the precision ratio of the found information by the *KS method* is more than that of the *PS method*? In brief, the percent of the relevance of the found information by the *PS method* is lower than that of the two other methods.

In *KS method*, in 6 cases from 20 searches, the precision ratio of the retrieved information is more than 50 percent (counts: 5, 6, 9, 10, 14 and 19), while in the *PS method*, for items 5 and 6, no relevant information is retrieved. The fact that in most cases, the percentage of the relevance of the retrieved information is equal or close to the search results of the *QS method* is underlined through a diagram in Figure 1. In *QS method*, the results indicate that the relevance rate of the retrieved information of 17 out of 20 (80%) queries is more than that of the two KS and PS methods.

Conclusion

The 48 structural and 16 content features are identified for the *Ask search* engine. As to the finding here, in this search engine, the QS is more powerful than the KS and PS. Despite *ASK.Com* technology for responding to question, phrase, or single-word searches of users and the existence of features like Advanced Search, Search Web, Search operator, Questions/Answers, Product search, Search for celebrities, Map search, Smart Answer, Related Search etc., the precision ratio of phrase and keyword searches is not appropriate. In the past, the evaluation of the performance of search engines was based on a classical equation of precision and recall, (i.e., classical logic 0 and 1, relevance and non-relevance of the record), with many drawbacks, that is the precision and recall rates were not being calculated in real terms, while here, the performance of the search engine is based on the precision and recall fuzzy equation. Here, 45 percent precision of the search engine, in the QS, is significant. It is suggested that this newly proposed equation be applied in evaluating other search engines and compare their results

Acknowledgements

This work is partially supported by the Semantic Retrieval Project (SRP) at the University of Isfahan, Department of Knowledge and Information Science.

References

Azizan, A., Bakar, Z. A., Ismail, N. K., & Amran, M. F. (2013). Interface features of semantic web search engine. In *2013 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, 142-147. doi:10.1109/ic3e.2013.6735981.

Balabantaray, R. C., Swain, M., & Sahoo, B. (2013). Evaluation of web search engines based on ranking of results and features. *International Journal of Human Computer Interaction (IJHCI)*, 4 (3), 117-127.

Bitirim, Y., & Görür, A. K. (2017). A comparative evaluation of popular search engines on finding Turkish documents for a specific time period. *Technical Gazette*, *24*(2), 565-569.

Chahal, P., Singh, M. & Kumar, S. (2013). Ranking of web documents using semantic similarity. In *2013 International Conference on Information Systems and Computer Networks*. Retrieved October 23, 2018, from http://www.leonsoftsolutions.com/ieeepapers/datamining/datamining4.pdf

Deka, S. K. & Lahkar, N. (2010). Performance evaluation and comparison of the five most used search engines in retrieving web resources. *Online Information Review*, 34 (5), 757-771.

Demirci, R. G., Kismir, V., & Bitirim, Y. (2007, May). An evaluation of popular search engines on finding turkish documents. In *Second International Conference on Internet and Web Applications and Services (ICIW'07)*, May 2007, pp. 61-61.

Dutta, M., & Bansal, K. L. (2016). A review paper on various search engines (Google, Yahoo, AltaVista, Ask and Bing). *International Journal on Recent and Innovation Trends in Computing and Communication*, 4 (8), 190-195.

Huizingh, E.K.R.E. (2000). The content and design of web sites: An empirical study. *Information & Management*, 37, 123-134.

Kumar, B. S., & Prakash, J.N. (2009). Precision and relative recall of search engines: A comparative study of Google and Yahoo. *Singapore Journal of Library & Information Management*, 38, 124-137.

Lewandowski, D. (2008). The retrieval effectiveness of web search engines: Considering results descriptions. *Journal of Documentation*, 64(6), 915-937.

Mala, V., & Lobiyal, D.K. (2016). Semantic and keyword based web techniques in information retrieval. *In International Conference on Computing, Communication and Automation (ICCCA2016)*, 23-26. doi:10.1109/ccaa.2016.7813724.

Martinez-Sanahuja, L., & Sanchez, D. (2016). Evaluating the suitability of web search engines as proxies for knowledge discovery from the Web. *Procedia Computer Science*, 96, 169-178.

Olajide, A.O., & Matthew, E. D. (2014). Performance evaluation of selected search engines. *Computer Engineering and Intelligent Systems*, 5 (1), 1-13.

Page, L., Bri, S., Motwani, R., & Winograd, T. (1999). The page rank citation ranking: bringing order to the Web. *Stanford InfoLab.* Retrieved October 12, 2018, from http://ilpubs.stanford.edu:8090/422/

Parsania, V. S., Kalyani, F., & Kamani, K. (2016). A comparative analysis: DuckDuckGo vs. Google search engine. *GRD Journals- Global Research and Development Journal for Engineering*, 2 (1), 12-17.

Rashid, J., & Wasif Nisar, M. (2016). A study on semantic searching, semantic search engines and technologies used for semantic search engines. *International Journal of Information Technology and Computer Science (IJITCS)*, 10, 82- 89.

Rathee, D., Atri, J., Rathee, J., & Nath, R. (2013). A framework to evaluate search engines. *International Journal of computer Science & Communication (IJCSC)*, 4 (2) 2, 184- 190.

Robbins, S. S., & Stylianou, A. C. (2003). Global corporate web sites: An empirical investigation of content and design. *Information & Management*, 40 (3). 205–212.

Sahu, S. K., Mahapatra, D. P., & Balabantaray, R. C. (2016). Comparative study of search engines in context of features and semantics. *Journal of Theoretical and Applied Information Technology*, *88*(2), 210-218.

Spink, A., Jansen, B. J., Kathuria, V., & Koshman, S. (2006). Overlap among major web search engines. *Internet Research*, 16 (4), 419-426.

Sunitha, C., Meena Preethi, B., & Akshay, M. (2013). A comparative study over search engine optimization on precision and recall ratio. In *Proceedings of National Conference on New Horizons in IT-NCNHIT*, 35-39. Retrieved November 9, 2018, from http://www.conference.bonfring.org/papers/met_ncnhit2013/ncnhit12.pdf

Wang, L., Wang, J., Wang, M., Li, Y., Liang, Y., & Xu, D. (2012). Using Internet search engines to obtain medical information: A comparative study. *Journal of medical Internet Research*, 14 (3), 1-19. Retrieved October 26, 2018, from https://www.jmir.org/2012/3/e74/pdf

Bibliographic information of this paper for citing:

CheshmehSohrabi, Mozaffar, & Abbasi Dashtaki, Neda (2019). "Ask Search Engine: Features and performance identification." *Webology*, 16(1), Article 180. Available at: http://www.webology.org/2019/v16n1/a180.pdf