International Journal of Engineering & Technology, 7 (4.33) (2018) 1-4

International Journal of Engineering & Technology

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

Research paper

A Comparative Evaluation of Search Engines on Finding Specific Domain Information on the Web

Azilawati Azizan[1]*, Zainab Abu Bakar[2], Nurazzah Abd Rahman[3], Suraya Masrom[1], Nurkhairizan Khairuddin[1]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Tapah Road, 35400 Perak, Malaysia
[2]Al-Madinah International University, Shah Alam, 40100 Selangor, Malaysia
[3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, 40450 Selangor, Malaysia
**Corresponding author E-mail: azila899@perak.uitm.edu.my*

Abstract

Recently search engines have provided a truly amazing search service, especially in finding general information on the Web. However, the question arises, does search engine perform the same when seeking domain specific information such as medical, geographical or agriculture information? Along with that issue, an experiment has been conducted to test the effectiveness of today's search engines from the aspect of information searching in a specific domain. There were four search engines have been selected namely Google, Bing, Yahoo and DuckDuckGo for the experiment. While for the domain specific, we chose to test information about the popular fruit in Southeast Asia that is durian. Precision metric has been used to evaluate the retrieval effectiveness. The findings show that Google has outperformed the other three search engines. Nevertheless, the mean average precision value 0.51 given by Google is still low to be satisfied neither by the researcher nor the information seekers.

Keywords: *Search engine evaluation; Precision; Specific domain; Durian.*

1. Introduction

The Web has become the largest unorganized repository of data and information [1]. In fact, in the present it has turn out to be an information deluge which causing information search to be more challenging. Hence, it is not easy to find a piece of information without assistance of the search engine. Search engine also has become a primary need since searching activity has been a daily routine nowadays. Therefore we need to have a very good search engine so that it can fulfill the user's needs.

The purpose of this experiment is to evaluate the effectiveness of the commercial search engine on searching domain specific information. So, the research is done to confirm the needs of improvements in this searching technology. The findings from this experiment also proved that the general problem statement in Information Retrieval field (to retrieve all relevant documents to a user query while retrieving as few non-relevant documents as possible) is still relevant until today [2].

This paper is organized as follows: Section 2 shares several previous works related to search engines evaluation. Then, section 3 describes the methodology employed to evaluate the relevance of the search results in terms of precision value. While section 4 exhibits and discusses the result and the last section concludes the paper including the issues and challenges on searching the Web.

2. Related Works

Many studies about search engine effectiveness have been done by various researchers worldwide, and mostly is a comparative type study. Most of them tested the effectiveness by using general topic query. Among the comparisons ever done, were against the keyword-based search engines and the semantic-based search engines [3-4]; commercial search engines against dedicated search engines [5] and English search engines against other language search engines [6-7]. Some researcher also did compare the effectiveness using short queries and long queries [8]; natural language queries [9], reformulated queries [10] and many more.

Even so, the comparative study involving specific domain search is still lacking. Among the available publications is the research done by [11]. They evaluated the search engines effectiveness on finding health information domain. They chose to compare between general search engines (Google, Bing, Yahoo, Sapo) and health-specific search engines (MedlinePlus, SapoSaude, WebMD). They found that general search engines have surpassed all the health-specific search engines and Google has the highest precision value in the top ten results.

In [12] has evaluated three search engines' application programming interfaces (API) on finding geographic web services. They chose Google, Bing and Yahoo and they reported that discovering geographic web services using search engine does not require the use of advanced search operator. They also reported Yahoo has outperformed the other search engines in discovering the geographic web services domain.

In [13] compared the performance of 4 international search engines (Google, Yahoo, Altavista, Exalead) and 4 Greek search engines (Google.gr, In.gr, Robby.gr,Find.gr) in the point of view of Greek librarians. He concluded that most librarians were satisfied and preferred to use international search engines.

Due to this limited analysis addressed by the researches in comparing search effectiveness involving specific domain search, we decided to conduct an experiment on comparing the current popular

search engines in finding information on fruit domain. The inspiration for this study is to motivate researchers and search engine providers towards producing better search technology in the future.

3. Methodology

The methodology being employed in this experiment is adopting a common approach being used in many search engine evaluation research works. Generally the first step starts by selecting the search engine, and then a list of search queries will be identified [14]. The setting of the queries might be chosen from a variety of features such as simple, complex, natural language or multi language queries. Next step is to submit or run the queries to the chosen search engine and subsequently record all the search results. Before the analysis is made, the researcher will first identify the eligible person and resources to do the relevance judgment process. Lastly, the analysis is made based on the standard evaluation measure that are precision and recall metric. There are also many other evaluation measures can be used such as Mean Average Precision (MAP), Average Precision at *n* (P@*n*), R-Precision, Precision Histogram, Mean Reciprocal Rank (MRR), E-Measure and F-Measure [2]. Though, the most widely used measurement metric is the standard precision measure.

In order to find out how those search engines performed when searching for domain specific information, we decided to do a comparative experiment among the popular search engines. Therefore, the selection of the search engines to be tested in this experiment is based on the most popular and most successful search engine rated by several search engine optimization websites such as Search Engine Watch [15], Search Engine Journal [16] and Alexa.com [17]. They have listed more than 10 popular search engines based on their traffic statistics, market shares and user responses. In the list, Google has been always on the first ranking compared to other search engines. For that reason, we chose Google and another 3 top ranking search engines that are Bing, Yahoo and DuckDuckGo. Table 1 shows the list and URL of the selected search engines.

Table 1: Selected Search Engine for the Experiment

No.	Search Engine	URL
1	Google	https://www.google.com/
2	Bing	https://www.bing.com/
3	Yahoo	https://yahoo.com/
4	DuckDuckGo	https://duckduckgo.com/

3.1. Search queries and domain

Prior to the selection of the queries, a survey has been done through online forums / groups, social media and also from the Google suggestion system features (Google Instant). The purpose of doing the survey was to get a general idea of the questions that users often ask about durian. Apart from getting the collection of queries, the survey also indirectly revealed the information that user commonly want to find about durian. Initially 290 queries about durian were collected from the survey. In order to validate the queries and facts about durian domain, we do collaborate with the domain experts that are the durian experts from MARDI (Malaysian Agricultural Research and Development Institute) and the durian farmers. Finally we decided to run the pre-test by using only 8 queries that has been identified and validated by MARDI as a very commonly question being asked by the user about durian. List of the queries is shown in Table 2.

Each query listed in Table 2 was submitted to each of the selected search engine and the results were captured. It retrieved tons of results, but only the first top 20 results (links) being analyzed. This is because many studies in search behavior field reported that most Web users will only inspect the top 10 search results [18] and it is relatively uncommon for a user to inspects beyond the top 20 results [19].

Table 2: Test Queries

Query Number	Queries
Q1	List of insect pests that attack the durian tree
Q2	When is the durian season in Malaysia
Q3	What are the varieties of durian in Malaysia
Q4	What are the characteristics of good quality durian
Q5	How to plant durian
Q6	How to control durian tree disease
Q7	What are the products of durian
Q8	What are the side effects of eating durian to health

3.2. Evaluation criteria

In order to maintain the evaluation quality of the web search engines, it typically uses human judgements to indicate which results are relevant for a given query [2]. Therefore, in this experiment, all the links (search results) were evaluated using human relevance judgment. The judgment is made based on the facts provided by MARDI. Each link has been classified as 'relevant' or 'not relevant'. All those steps were repeated until all the queries being run on all 4 selected search engines. In total, 640 links have been evaluated by the same author so that the judgments made are more consistent. Besides that, all the searches and evaluations were performed in minimal time space to ensure a stable performance measurement of the search engines.

For the purpose of retrieval evaluation, a standard precision and recall metrics were used to evaluate the retrieval quality. Precision is defined as the fraction of relevant documents retrieved to the number of total documents retrieved. It is formulated as in (1). While, recall is the fraction of relevant documents retrieved to the number of relevant documents in the collection as shown in (2).

$$\text{Precision} = \frac{\text{No. of relevant document retrieved}}{\text{No. of document retrieved}} \quad (1)$$

$$\text{Recall} = \frac{\text{No. of relevant document retrieved}}{\text{No. of relevant document in the collection}} \quad (2)$$

In the case of evaluating commercial search engine, recall value is quite impossible to be calculated since we do not know the total number of relevant document in the entire search engine collection. Therefore, we only do the precision measure, which we considered the 'links retrieved' (search results) as the 'document retrieved'.

Comparing retrieval evaluation for different algorithm or methods over a set of queries is commonly use the average precision values as in (3), where $\bar{P}(r_j)$ is the average precision at recall level r_j and $P_i(r_j)$ is the precision at recall level r_j for the *i*-th query over N_q total number of queries.

$$\bar{P}(r_j) = \sum_{i=1}^{Nq} \frac{P_i(r_j)}{N_q} \quad (3)$$

Single value summary of the evaluation can be presented using Mean Average Precision (MAP). The mean value precision over a set of queries is defined as in (4), where $\sum_{q=1}^{Q} AveP(q)$ is the summation of average precision obtained for all relevant documents and N_q is the total number of queries.

$$MAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Nq} \quad (4)$$

It is a common practice when evaluating search engine, precision of the search results always being measured at the top positions in the ranking. Typically, precision is measured at cut-off point 5, 10 and 20 [2]. It means the precision value is being calculated when 5, 10 and 20 documents / links have been seen. In practice, it is being

written as precision at 5 (P@5), precision at 10 (P@10) and precision at 20 (P@20). In our evaluation, we did consider P@15 as an additional value to be analysed.

4. Results and discussion

The number of relevant links retrieved for all search engines according to query is shown in Table 3. The data shows that Google has retrieved the most relevant link followed by Yahoo, DuckDuckGo and Bing. Google has retrieved 61 relevant links out of 160 retrieved links overall which gives 38.13%. Google surpasses Yahoo by 3.75%; Yahoo surpasses DuckDuckGo by 3.13% while DuckDuckGo surpasses Bing very thinly by 0.62%. Data in Table 3 also showed that query number 8 (Q8) has the highest total relevant retrieved, while Q4 has the lowest relevant retrieved by all search engines.

Table 3: Relevant Links Retrieved for each Query and Search Engines

Query	Google	Bing	Yahoo	DuckDuckGo
Q1	4	4	4	4
Q2	5	3	4	4
Q3	14	9	11	10
Q4	3	1	2	1
Q5	10	8	11	11
Q6	5	7	5	7
Q7	8	2	6	3
Q8	12	15	12	10
Total	61	49	55	50
Relevant Retrieved %	38.13	30.63	34.38	31.25
Mean of relevant	7.63	6.13	6.88	6.25

Average precision for all search engines was compared and portrayed in a line graph in Figure 1. We can see clearly Bing has the lowest line in the graph, while to identify the highest line is quite difficult because the graph lines for Google, Yahoo and DuckDuckGo do not show much difference.

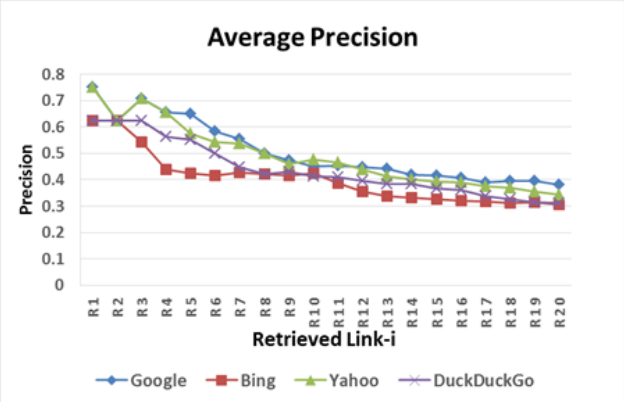

Fig. 1: Average Precision Over Retrieved Link-i

Since the average precision graph in Figure 1 does not help much in seeing the difference, so we analyzed the results at several cut-off points. Table 4 shows the average precision at cut-off point 5, 10, 15 and 20 for all search engines.

Table 4: Average Precision at *n*

Search Engine	P@5	P@10	P@15	P@20
Google	0.650	0.450	0.417	0.381
Bing	0.425	0.425	0.325	0.306
Yahoo	0.575	0.475	0.392	0.344
DuckDuckGo	0.550	0.413	0.367	0.313

The values show that Google has outperformed at three cut-off point that are P@5, P@15 and P@20. On the other hand, Yahoo has the highest average precision at cut-off point 10. The comparisons among all the search engines can be seen clearly in Figure 2. There is quite a significant difference of precision value at cut-off point 5, while at cut-off point 10, all search engines achieved almost similar value.

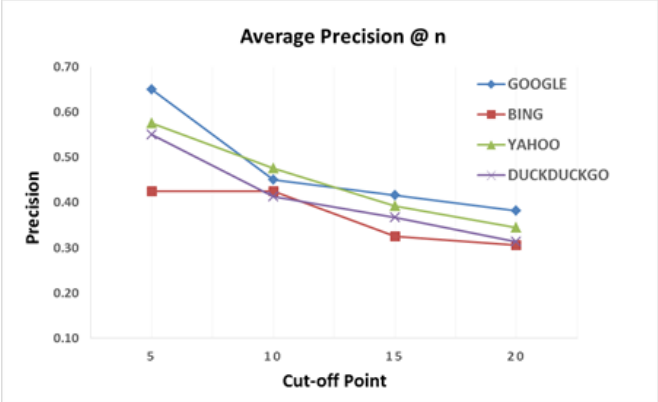

Fig. 2: A Graph for Average Precision at *n*

To summarize the results, mean average precision (MAP) for each search engine over queries were calculated and illustrated in Figure 3. Google achieved 0.505 mean value, followed by Yahoo (0.488), DuckDuckGo (0.440) and Bing (0.403). Mean value between the highest (Google) and the lowest (Bing) is 0.102 which gives 20.2% difference.

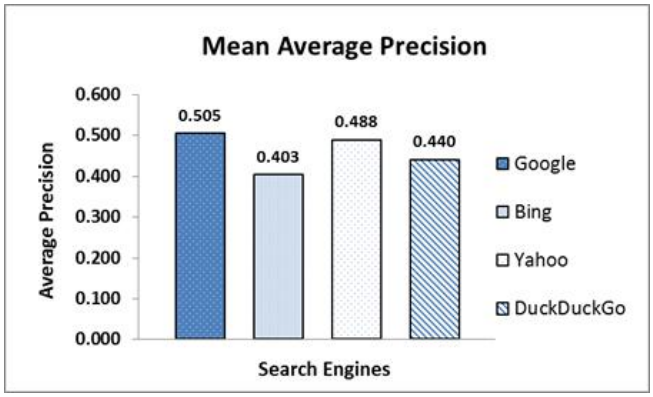

Fig. 3: A Graph for Average Precision at *n*

Many comparative studies reported that Google always outperform other search engines [7, 11]. As an example result from Deka's evaluation [20] reported that Google has the highest rate of performance, followed by Yahoo and Live, Ask and AOL search engines. Findings in our experiment also reported almost similar results which Google is at the top rank and followed by Yahoo and other search engines. It proved that Google always surpass all his competitors.

5. Conclusion

The result shows that Google surpass the precision of other search engines at three cut-off points (P@5, P@15, P@20), while Yahoo has the highest precision at cut-off 10. Many other researchers also reported Google always outperform in their experiments, for example in [19] claimed that Google has outperformed Hakia in his experiment as Google had mean precision at 0.64 as compared to Hakia at 0.54 for general topic search. Whereas in our experiment, Google achieved lower mean average precision that is 0.51 for specific domain search (durian fruit information). So we concluded that even though Google always outperformed other search engines, but mean precision value 0.51 given by Google for finding specific domain information particularly in durian fruit is still unsatisfactory. This means Google only achieve half from the perfect mean value that is 1.0. This analysis also reveals how search engines differ in their responses when seeking for specific domain information such as fruit information (e.g.: durian) on the Web.

Acknowledgement

We would like to thank Mr. Bahari Mohd Nasaruddin (Director of MARDI Perak Malaysia), Mr. Muhamad Afiq Tajol Ariffin (Senior Scientist-Senior Research Officer, Horticultural Center, MARDI Sintok Kedah Malaysia) and durian farmers in Kedah and Perak for collaborating and also to Universiti Teknologi MARA for the financial support of this project.

References

[1] R. Baeza-Yates (2003), Information retrieval in the Web: Beyond current search engines. *Int. J. Approx. Reason.* 34 (2–3), 97–104.
[2] R. Baeza-Yates & B. Ribeiro-Neto (1999), *Modern Information Retrieval*. ACM Press. Addison Wesley.
[3] J. Singh (2013), A comparative study between keyword and semantic based search engines. *Proceedings of the International Conference on Cloud, Big Data and Trust*, pp. 130–134.
[4] D. Tümer, M. A. Shah & Y. Bitirim (2009), An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, Yahoo, Msn and Hakia. *Proceedings of the Fourth Int. Conf. Internet Monit. Prot.*, pp. 51–55.
[5] Y. Peng & D. He (2006), Direct comparison of commercial and academic retrieval system: An initial study. *Proceedings of the International Conference on Information and Knowledge Management*, pp. 1–2.
[6] Y. Bitirim & A. K. Görür (2017), A comparative evaluation of popular search engines on finding Turkish documents for a specific time period. *Teh. Vjesn. - Tech. Gaz.* 24, 565–569.
[7] J. Zhang, W. Fei & T. Le (2013), A comparative analysis of the search feature effectiveness of the major English and Chinese search engines. *Online Inf. Rev.* 37, 217–230.
[8] A. K. Mariappan & V. S. Bharathi (2012), A comparative study on the effectiveness of semantic search engine over keyword search engine using TSAP measure. *Proceedings of the International Conference on E-Governance and Cloud Computing Services*, pp. 4–6.
[9] N. Hariri (2013), Do natural language search engines really understand what users want? A comparative study on three natural language search engines and Google. *Online Inf. Rev.* 37, 287–303.
[10] A. Azizan, Z. A. Bakar & S. A. Noah (2014), Analysis of retrieval result on ontology-based query reformulation. *Proceedings of the IEEE International Conference on Computer, Communication, and Control Technology*, pp. 244–248.
[11] C. T. Lopes & C. Ribeiro (2011), Comparative evaluation of web search engines in health information retrieval. *Online Inf. Rev.* 35, 869–892.
[12] F. J. Lopez-Pellicer, A. J. Florczyk, R. Béjar, P. R. Muro-Medrano, & F. Javier Zarazaga-Soria (2011), Discovering geographic web services in search engines. *Online Inf. Rev.* 35, 909–927.
[13] E. Garoufallou (2012), Evaluating search engines: A comparative study between international and Greek SE by Greek librarians. *Program* 46, 182–198.
[14] D. Hawking, N. Craswell, P. Bailey & K. Griffiths (2001), Measuring search engine quality. *Inf. Retr. Boston.* 4, 33–59.
[15] Search Engine Watch (2018). https://searchenginewatch.com/.
[16] Search Engine Journal (2018). https://www.searchenginejournal.com/.
[17] Alexa.Com-TopSites (2018). https://www.alexa.com/topsites/category/Computers/Internet/Searching/Search_Engines.
[18] B. J. Jansen, D. L. Booth & A. Spink (2009), Patterns of query reformulation during web searching. *Journal of the American Society for Information Science and Technology*, 60(7), 1358–1371.
[19] M. Andago, T. P. L. Phoebe & B. A. M. Thanoun (2010), Evaluation of a semantic search engine against a keyword search engine using first 20 precision. *Int. J. Adv. Sci. Arts* 1(2), 55–63.
[20] S. K. Deka & N. Lahkar (2010), Performance evaluation and comparison of the five most used search engines in retrieving web resources. *Online Inf. Rev.* 34, 757–771.