International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 1– No.4, February 2012 – www.ijais.org

An Overview of Search Engine Evaluation Strategies

Shikha Goel
Ajay Kumar Garg Engineering College
Ghaziabad,India

Sunita Yadav
Ajay Kumar Garg Engineering College,
Ghaziabad,India

ABSTRACT

Basically evaluation of search engine is the process of making judgment about the value, importance and quality of search engine, after considering search engines carefully. The evaluation of search engines has not been keeping up with the advancement of their development. Web search engines work differently based on different mode of interface, features, coverage of the web, ranking methods, delivery of advertising and many more such factors. It is not easy to evaluate them on a single basis. There are many strategies for evaluating search engines such as automatic evaluation, human relevance judgment based evaluation.

The purpose of this paper is to review the search engine evaluation strategies in order to propose an enhanced method for evaluating search engines.

Keywords:
Search Engine, Automatic Evaluation, User Feedback, Search query, Web search services, precision.

1. INTRODUCTION

Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day[1]. Search engines are tools designed to help user and to minimize the time required to find information over the vast Web of hyperlinked documents. It provides a medium for the user to get in touch with the huge information available over the internet though some locally maintained databases. The criteria for search are called a search query. The list of items that meet the criteria specified by the query is typically sorted or ranked in a group of pages to produce the desired information. There are three basic computer-aided techniques for searching traditional information retrieval collections: Boolean models, vector space models, and probabilistic models. These search models, developed in the 1960s, have taken decades to grow in form of new search models[2].

Web search engine consists of following main components[1]:

1.1 Crawler Module: As compared to traditional document collections which reside in physical warehouses such as the college's information available on WWW is distributed over the Internet. In fact, this huge repository is growing rapidly without any geographical constraints.

1.2 Page Repository: The downloaded Web pages are temporarily stored in a local storage of search engine, library, called page repository.

1.3 Indexing Module: The indexing module takes each new uncompressed page from the page repository extracting suitable descriptors, creating a compressed description of the page.

1.4 Indexes: The indexes hold the valuable compressed information for each web page.

Evaluation involves assessing the performance of search engines to improve their effectiveness. Evaluation is a continuous process of investigating the new approaches of study, appraisal, and improvement of search engine. Evaluation is the key for making progress in building here better search engines and to understand the working of search engines. Evaluation makes us up to come up with the new faces of problems associated with the search and searching results. The significance of evaluation is twofold: to help Web users in their choice of search engines and to inform the development of search algorithms and search engines. This paper is arranged to present the evaluation parameters first. Secondly, it gives a brief literature survey on evaluation strategies. Then the paper concludes with the study of various evaluation strategies under the automatic evaluation, survey based evaluation and other models with the idea of a new approach for evaluating the search engines.

2. EVALUATION PARAMETERS

The various criterions for the evaluation of search engines are:

2.1 Effectiveness: It measures the ability of the search engine to find the right information [3].It is the capability of producing the desired result. When a search engine is deemed effective, it means it has an intended or expected outcome, or produces a deep, vivid impression.

2.2 Efficiency: It measures how quickly right information is retrieved. It is defined in terms of the time and space requirements for the algorithm that produces the ranking. Both effectiveness and efficiency will be affected by many factors such as the interface used to display search results and techniques such as query suggestion and relevance feedback[3].

2.3 Coverage Evaluation: Whether the result list covers most of the correct URL list[4].

2.4 Sequence Evaluation: Whether the result list puts the most important URL's on the front and secondary URL's on the back[4].

2.5 Precision: It is the fraction of a search output that is relevant for a particular query. Precision is scored by dividing the total number of pages found by the number of relevant pages found.

2.6 Relevance Scores: It is the measure the accuracy of the search results-in other words it's a measure of how close the documents listed in the search results are to what the user are looking for.

2.7 Quality: It is a very subjective term, but includes things like result ranking ,timeliness, one click access to information, volume of content and lack of spam.

2.8 Ability to retrieve top ranked pages: It is the engine's ability to retrieve top ranked pages[10].

2.9 Stability: Three measurements for stability are (i) the stability of the number of pages retrieved; (ii) the number of pages among the top 20 retrieved pages that remain the same in two consecutive tests over a shorter time period ; and (iii) the number of pages among the top 20 retrieved that remain in the same ranking order in two consecutive tests over a shorter time period[10].

2.10 Recall rate: It is the proportion of relevant documents that are retrieved.

2.11 Tendency Degree: It is a measure of whether the search engine results has a good presentation[15].

2.12 Coverage Degree: It is a measure in terms of search engine results retrieval effectiveness[15].

3. EVALUATION STRATEGIES

3.1 Automatic evaluation strategies

Jinbiao et al.(2009) suggested a simple, accurate, effective, automatic and safe system which is used to automatically evaluate search engines. It has four modules-sampler, crawler, refinery, evaluator. System design is given below in Figure 1.C# is used as a development tool in this work. At the end, Author gives score to search engine on the basis of coverage evaluation(result list covers most of the correct URL's) and sequence evaluation(result list puts most important URL's on front and secondary URL's on back).

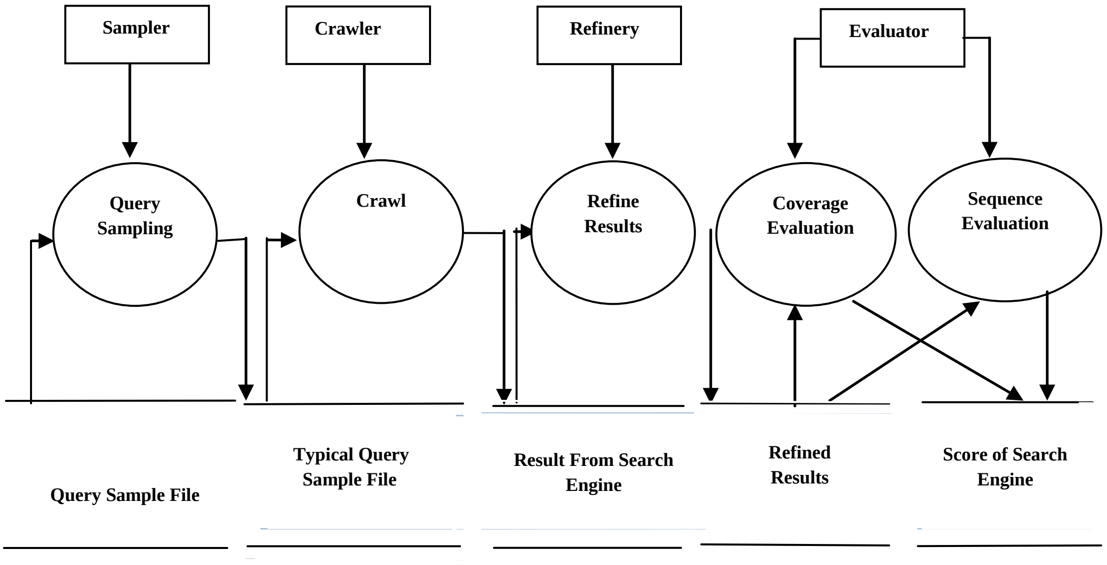

Figure 1.**System Design**

Abdur et al.(2002) had shown comparison on five (Lycos, Netscape, Fast, Google, HotBot) search services based on known item searching; comparing the relative ranks of the items in the search engines' rankings using three query sets 500, 1000,and 2000 issued to each search engine. The overall score of search engine is mean reciprocal rank(MRR). Experimental results showed that while most engines are roughly the same in terms of effectiveness, there is a considerable gap between the best and worse in terms of MRR.

Himanshu et al.(2005) organised an evaluation in an organization setting with 12 manufacturing and quality engineers from a major manufacturing organization based on user search actions as implicit judgments of document relevance.Actions are implicit feedback interactions such as scrolling, saving, printing ,bookmarking , adding to favorites and copying.

The three strategies described above are automatic evaluation strategies but each of them work on a different criteria. Jinbao used a recall and accuracy rate in evaluating the web queries. It uses a simple structure developed in c# which has limited the scope of advancement. The total number of referenced urls is not sufficient for a record result. In an another reference work by abdur, suggest the use of query document pairs but concludes to a considerable gap of Mean Reciprocal Ranking. However himanshu et al uses implicit user feedback which is a considerable approach to capture user response. But it has a limitation of calculating metrics due to errors in the capture of user actions such as clicks on a link .Popup windows opening, slow processor and heavy web pages slow down the whole process. This comparison states the differences in approaches of various studies.

3.2 Survey based evaluation strategies

Maninder et al.(2011) compared and evaluated five search engines (Google, yahoo, bing, ask, AltaVista) on the basis of their search capabilities into two sections[9]. In the first section, features of five search engines are compared which are available to the user while searching the information. In second section, performance and capability is analyzed from the user's point of view. For this, they had taken a survey in which 263 participants participated and examined their interests in search engines. From this survey, they find out which search engine provides best utility and services to the user and most likely used by the people and they find out that users give highest rank to Google.

Ya-Lan et al.(2007) proposed two major factors hygiene factor and motivation factor. Hygiene factors are those more fundamental requirements for a search engine and make users willing to use a search engine, and motivation factors are those more additional services of a search engine and make users willing to keep using the same search engine.

The author had surveyed 758 people in Taiwan. The survey had three main components:

1) Demographic questions, the results showed that the age of 95% of the respondent's centers on the range from 18 to 30, and most of the participants are students

2) Experiences of using computer, Internet, and search engine, the results showed that more than 75% of the participants have experiences of using computer and Internet for more than five years. More than 95% of them use computer and surf on the Internet everyday for at least one hour.

3) Perceptions of search engines, test the hygiene-motivation hypothesis of search engine proposed in this research paper.

Maninder et al evaluated five search engines but based on limited user review.Whereas Ya-Lan et al have used different factors for user liking and behavior, the results are dependent of various previous studies and the factors ought to take a unidirectional approach.

3.3 Relevance scoring method based evaluation

Longzhuang et al. evaluated six search engines Alta vista, Fast, Google, Go, iWon and Northern light based on four popular relevance scoring methods. The methods include vector space model, okapi similarity measurement, cover density ranking and a three level scoring method. They had calculated relevance scores of query results of using these methods and google shows outstanding performance.

3.4 Continuous relevance ranking by human subjects based evaluation

Vaughan et al.(2004) compared three search engines (Google, AltaVista, Teoma).The author measured quality of result ranking, ability to retrieve top ranked web pages and stability using four queries and first ten links from each search engine and 24 graduate students to rate the links. The performance of Google is best in quality, ability to retrieve top ranked pages and stability. Vaughan et al has a higher point of view for the evaluation strategy and discussed the argues over the new evaluation methods. It concludes with the suggestion of choosing the query topics effectively.

3.5 Four level relevance judgement based evaluation

Gordon et al.(1999) compared eight search engines (AltaVista, Excite, Info seek, Open Text, Hot Bot, Lycos, Magellan, and Yahoo!) using four level relevance judgment (highly relevant, somewhat relevant, somewhat irrelevant, and highly irrelevant) and one human subject. The author used 33 queries and top 200 links to find out that alta vista, Lycos and open text had high precision and recall rate.

3.6 Evaluation using clickthrough data and a user data model

George et al.(2007) suggested a model to evaluate search engines on the click through data of past users. The model used two variables i.e. A(attractivity) and C(consideration) to determine the probability of choosing a snippet out of the list of relevant pages through which he successes to a distance d ; after considering upto distance d-1 portions. The conclusion of evaluation shows that the distance model represents the data better than popularity model. The complete evaluation illustrates that the positional biasing of relevancy can be resolved by click through data. Here it may seem counter-intuitive to use this model to measure performance. This toy model is unable to represent clearly the user behavior but it can be further improved to implement click through data methods.

3.7 Evaluation using Rough Set Based Rank Aggregation

Rashid et al.(2009) devised an automatic web search evaluation system based on rough set based rank aggregation technique.Basically , different ranking results obtained from different techniques are combined.Two phases are used, ranking rules learning phase and rank aggregation phase.Author used 15 queries in rank learning phase.The output of this phase is a set of ranking rules.

The same set of ranking rules is used in rank aggregation phase.The output of this phase is aggregated ranking.A coorelation coefficient is computed between search engine ranking and aggregated ranking for 543 queries.The correlation coefficients obtained for 543 queries are averaged and the search engines are rated on the basis of this coefficient.The results showed that Google is best out of the five search engines.

3.8 Evaluation using judgements of Meta search engine

Hamid et al.(2011) proposed tendency degree and coverage degree for evaluation of three search engines(Google,Ask,Bing).For each search engine the weighted average of similarity degree between its ranked result lists and those of its meta search engines is measured. To compute the similarity degree, these two new measures were proposed. The results showed that Google gave outstanding performance.The effectiveness of methods were also compared with human-based ones.

4. CONCLUSION

This paper provides an overview of the various strategies to evaluate the search engines based on different methods and models proposed in the field of search engine evaluation. The diverse Automatic evaluation strategies based on recall rate, accuracy rate, and implicit user feedback and know item searching had been discussed here. Moreover, survey based evaluation, relevance scoring method; continuous relevance ranking by human subjects, four level relevance judgements based evaluation strategies had been investigated. Survey based evaluation have a efficient approach towards the user understandings but it depends on different user categories and hence tends to give different results. On this study based on different aspects; it can be seen that there is need to improve the search engine query retrieval system and new methods are yet to be found out to do a more effective search.The supplementary enhancements in the technology have led to the changes in the user perspectives; there are various upcoming approaches for the user based ranking model for evaluation of search engines. The more involvement of user in search methodologies has lead to the drastic changes in the search engine indexing methods; the more new search evaluation strategies subjected on the human ranking based evaluation are further to be explored.

5. REFERENCES

[1] Sergey Brin and Lawrence Page., (1998). “The Anatomy of a Large Scale Hyper textual Web Search Engine”, Computer Networks and ISDN Systems, pp.107-117.

[2] D.C. Free Net – ServInt Internet Services- “Working With Search Engine”.

[3] Addison Wesley., (2008).” Evaluating Search Engines”, pp.1-40.

[4] Jinbiao Hou., (2009). “Research on Design of an Evaluation System of Search Engine”, ETP International Conference on Future Computer and Communication,pp.12-18.

[5] Abdur Chowdhury, Ian Soboroff., (2002). “Automatic Evaluation of World Wide Web Search Services”,ACM, pp.421-422.

[6] Himanshu Sharma, Bernard J. Jansen., (2005).”Automated Evaluation of Search Engine Performance via Implicit User Feedback” The Pennsylvania State University,ACM,pp.649-650.

[7] Maninder Kaur, Nitin Bhatia, Sawtantar Singh., (2011).” Web Search Engines Evaluation Based on Features And End-User Experience”, International Journal of Enterprise Computing and Business Systems,Vol. 1 issue 2.

[8] Ya-Lan Chuang, Ling-Ling Wu., (2007).”User-Based Evaluations of Search Engines: Hygiene Factors and Motivation Factors, National Taiwan University, Proceedings of the 40th Hawaii International Conference on System Sciences, pp. 1-10.

[9] Longzhuang Li, Yi Shang, and Wei Zhang, “Relevance Evaluation of Search Engines Query Results” University of Missouri-Columbia.

[10] Liwen Vaughan., (2004)” New measurements for search engine evaluation proposed and tested” Information Processing and Management, Vol. 40, No. 4, pp. 677-691

[11] Gordon & Pathak., (1999).“Finding information on the World Wide Web: the retrieval effectiveness of search engines”. Information Processing and Management , pp. 141-180.

[12] Georges Dupret ,Vanessa Murdock, Benjamin Piwowarski., (2007).”Web search evaluation using click throughdata and a user model.

[13] Rashid Ali,M.M. Sufyan Beg., (2009).”Automated Performance Evaluation of Web Search System using rough set based rank aggregation”Proceedings of the first international conference on Intelligent Human Computer Interaction, Springer, pp.344-358.

[14] Hamid Sadeghi.,(2011). “Automatic Performance Evaluation of Web search Engines using judgements of Meta search Engines”, Online Information Review,ISSN:1468-4527,Emerald Publishing Limited, pp.957-971.