



Factors Affecting User Satisfaction with Search Engines: Survey Based Approach

Sunita Yadav*

Amity Institute of Information Technology
Amity University, Noida
syadav@ascs.amity.edu

Dr Ela Kumar

School of Information Technology
Gautam Budha University
Greater Noida
ela_kumar@rediffmail.com

Abstract: Today, trillion of pages are published on the web. Traditional search engines, based on exact keyword match, return too many documents in response to a user query, and most of the returned documents are irrelevant. So web searchers are facing problems like information overload phenomena. Therefore there is a need to filter the useful information from the excess of information published on the web. Information overload put users at stress and thereby causing difficulty in searching and filtering useful information. User satisfaction is one of the most important factors to measure the success of search engines. It is the key variable in the search process where the information that has been obtained is evaluated and utilized by the information seeker. In this paper we have analyze that there is a need for identifying the factors affecting user satisfaction with search engine and proposed a model comprises of various factor that affect user search satisfaction. Based upon the proposed model we have conducted a survey and have identified major factors namely up-to-date information, Search Result relevancy, Response Time and Reliability contributing to the user satisfaction which may further be used as guideline for accessing the information through search engine.

Keywords: Search Engine, User Interface, Response Time, Diversity and Freshness, long tail, geographic information.

I. INTRODUCTION

Success of search engine is measured by the satisfaction of its users. User satisfaction has played a key role in behavioral research in search engines and information system. The measure and analysis of end-user satisfaction is always motivated by the desire to improve the productivity of search engines and thus difficult to standardized. The standardized specification of factors affecting user satisfaction is a critical issue.

Several studies have been conducted on user satisfaction with information system (IS) and information retrieval (IR). Jillian et. al. [20] identified and categorized four major factors namely system, user, environmental and task which constitute user satisfaction and concluded user satisfaction as one of the factor to measure system performance. Yamin and Ramayah [22] reviewed user experience; domain knowledge and user search behaviour and found that these factors have significant effect on the search satisfaction. User with higher experience and good domain knowledge seems to be highly satisfied. Topi and Lucas [35] found that overall search satisfaction is subjective to the performance i.e. accuracy of the content, response time of the search system and user's attitude i.e. confidence and satisfaction. User satisfaction may be accessed by the user actions like picking up the results, spending time at these documents, printing, saving, bookmarking, mailing to someone or coping and pasting a portion of that document. In addition, more the time spent by users on a document, more importance is the information to that user [38, 39]. Search satisfaction is also defined based on the analysis of the relevance of the search result. Sihvonen and Vakkari defined search satisfaction by the number of relevant results in the top twenty retrieved results [7]. User search satisfaction is also defined by finding a website that contained the precise information defined in the search query[21].

Various models have been developed for measuring user satisfaction. According to Tessier et al. [19] satisfaction is

a state of mind experienced by the user. Belkin and Vickery [36] explained satisfaction as a material response to information seeking. The user makes a judgement based on the relevance and utility of the documents retrieved. According to Kuhthau [37], satisfaction represents an emotional response to the cognition and action associated with the final stage of the information seeking process.

II. PROPOSED FACTORS AFFECTING USER SATISFACTION MODEL

User satisfaction is one of the indicators of user's achievement of information need. User satisfaction may be calculated by the user's intention to repeatedly visiting the search engine. Besides this User friendly Interface, fast response time, good coverage of Web, diversity of content, discovery of new pages/content, interactivity and other factors have significant impact on user satisfaction. Short response time, easy navigation, appropriate use of graphics and high degree of interactivity are more likely to attract users.

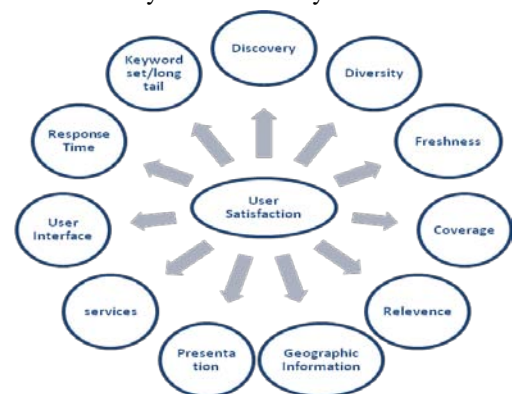


Figure 1:- Factors affecting User Satisfaction with Search Engine

There are number of factors that affects user satisfaction as user relies on search engine for information need as shown in figure 1 and are as follows:

A. User friendliness (User Interface)

The effectiveness of search engine is largely determined by the user friendliness of the interface and the interaction between the users and the information technology. Users generally reject complex interfaces for formulating advanced queries, and demand a fast response time. To overcome such problems, major search engines take a lot of tasks to provide the user with an intuitive interface to help formulate a query representing user's intention, browse long lists of documents, discover related topics. Ivan Berlocher *et al.* [40] proposed an efficient interface and fully automatic algorithm doing online clustering of words to return dynamically highly semantic related topics to a user query, helping the user to refine his search. Stability of graphics and logos build up an image in the mind of the user that in turn, enhances learning and the comfort level. A properly designed interface can stimulate system use and increase productivity while reducing learning time. The interface design should support user mistakes (like invalid input/data) as well as learning while doing tasks.

a) Response Time: User may get easily frustrated if the response time is large. John and Chris [28] showed in the research that slow system response time leads to dissatisfaction of users. The search response time may vary significantly for different kinds of queries with different query lengths and common or rare query terms in different databases [26]. Searching response time is also a very important component in source selection. Slow Response time may be caused by various factors like network congestion and large indexes. Various research has been done till date for finding methods to improve response time. Caching is an effective technique in search engines for improving response time, reducing the load on query processors, and improving network bandwidth utilization [27].

b) Presentation: Presentation of SERPs, the order as well as the number of results on a page, has significant effect on user satisfaction. Antti *et al.* [24] found that a six-item search result list was associated with higher satisfaction, confidence, and perceived carefulness than a 24-item list. Order effects do not appear with fewer than 15 documents, whereas with more items (15–75) there are [30, 31]. According to Schwartz [25] providing more number of highly relevant results will lead to inferior choice and degrade satisfaction. To improve presentation of Search Results, Shuo Wang *et al.* [41] proposed image search engine which presents the results in semantic clusters. Clustering and Open Search are the methods which can be used for improving presentation of search results. For satisfying user, the presentation of Search results is very important.

c) Coverage: Ali *et al* [3] defined Coverage as presence of content of interest in a catalog. According to Lawrence & Giles [23] there were approximately 800 million publicly index-able pages in February 1999. Whereas search engines cover only around 16% of the web. The reason may be many of the websites on the web are not user friendly or the website is not registered on the search engine. The search engines may be limited by the scalability of their indexing and retrieval technology, or by network bandwidth. The engines typically index a biased sample of the web [23]. The accessibility of information on the web is also biased by new search techniques and by the increasing use of features other than those directly describing the content of pages. Popularity based ranking methods are causing problem for new and unlinked pages to become visible in search engine result listings. This may delay or even stop the widespread visibility of new high quality information. The search engines are biased with national coverage differences along with evidence of probably unintentional domain bias in their page finding and page retention algorithms [32]. This biasing may lead to user dissatisfaction.

d) Keyword Sets/long tails: When the initial search is unsuccessful, 82% of the search engine users re-launches their search using the same search engine by adding more keywords to refine their subsequent search [1]. This shows a quite high level of confidence exists among users for search engines, as the majority seems to trust their search engine of choice to return to correct information more then they trust themselves to enter the appropriate keywords. So user satisfaction may be increased by Search Engines by expanding their keyword targeting by thoroughly examining their log file data and to find the long tail search terms. There are large numbers of queries that happen far less frequently than the leading terms at the head of the list. Into the tail, the traffic will be sizable as well as often proves to convert better than less general terms.

e) Diversity: Diversity is related to the breadth of the content and also to the quantification of query intent. Search engines' ability to handle the ambiguity and subjectivity of the query will also affect user satisfaction. People have diverse goals when they issue the same query to a search engine. The ability of search engine to address such diversity significantly affects user's satisfaction with search engine. As the number of possible users and the size of the search results increases, the best approach for ordering the results such that maximum users are provided with relevant pages in the top few results is diversification [18]. Queries having more than one interpretation are called ambiguous. Song *et al* [15] showed in their experiments that 16% of the queries are ambiguous queries. Therefore while ranking documents relevant to an ambiguous query; a search engine should consider document relevancy as well as the approach to satisfy different interpretations of the query. Yin *et al* developed a model based on knowledge of subtopics of a query, which returns a set of

documents covering all the subtopics of a given query and re-rank the retrieved documents using statistical information of user's intention on different subtopics [6].

Chen and Karger [16] considered that users are satisfied with few numbers of relevant results, rather than needing all relevant documents. They worked at the goal of maximizing the chances that the user will get an answer to their interpretation of the query. In their subtopic retrieval work, Zhai et al [17] reordered the results so that the top includes some results from each subtopic. Agarwal et al [18] formulated the problem of diversifying search results theoretically by considering that usually users are interested in top few results. They worked for maximizing the possibility of finding at least one useful result within the top few results for average users. One of the goals of search engines should be to minimize the chance of returning no relevant results.

f) Discovery & Latency: When new pages are posted on the web, discovery and latency are no. of pages listed in the search engine results and time it took to get the new pages in the search engine results respectively [3]. Search engines have strong freshness requirements at various timescales. For breaking event, users may want to access the related information within minutes. The less immediate content may be accessed by the user in hours or days. The crawler's ability of to procure new content in a timely manner may be limited by bandwidth, effectiveness of graph generators and even the effectiveness of the crawling model. Various studies have highlighted the high arrival rate of new content on the web. Dasgupta et al [13] studied the extent to which this new content can be efficiently discovered by a crawler.

The inconsistency between the content of the search engine repository and the current content of the live Web leads to user dissatisfaction. Pandey & Olston [14] derived an equation showing the quality of search engine repository is very much related to the likelihood of viewing the particular page by the user. Study described how to schedule web pages for selective re-downloading into a search engine repository to maximize the quality of the user experience

g) Relevance: When user issues a query to a search engine and receives an ordered list of results, User Satisfaction depends on how much the retrieved results met user's information need [3]. According to Huffman and Hochster [4] the relationship between relevance and session satisfaction is very strong. Figure 2 referred from [4] show a scatter-plot of session satisfaction versus first query relevance. The plot is scaled so that 0 and 1 correspond to the maximum and minimum observed value for each dimension.

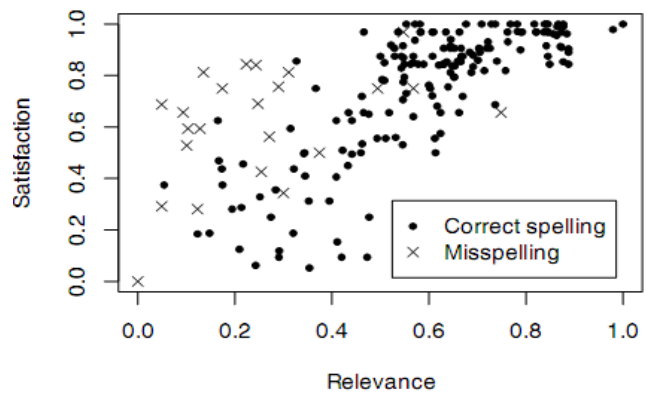


Figure 2: Session Satisfaction vs. First-query Relevance

As the plot shows, there is a reasonably strong relationship between satisfaction and relevance.

h) Freshness: Freshness of content in search results is also an important aspect for user satisfaction. In measuring search engine quality, quality of search results is one of the measures to be considered. Index freshness is very important part of quality measurement. Search engines should provide up-to-date information. Web contents are ever-changing. New pages are built, old pages are deleted, and links are changed, all at a very high rate. Because of the growth of the web, the number of old pages that no longer change has also increased drastically. Search engines need to find ways to show to the user pages that meet user's up-to-datedness criteria. In some cases, older pages may be helpful, but in the majority of cases, user prefers current ones. Search engine need to design efficient refresh policies to adapt the changes on the web. Freshness can be a critical factor when a user wants to find only current information.

Dirk [10] showed in his study that search engines have large shortcomings in updating their databases. Search engines are not offering the ideal solution for the user in which a comprehensive database of the web is updated according to the actual updates of the pages themselves. They found that search engines are not providing up-to-date copies even for the daily updated pages.

Ntoulas et al [11] found in his study that a huge number of web pages are changing on a regular basis. The study results found that there are about 320 million new pages every week. Approximately 20 percent of the Web pages disappear within a year and about 50 percent of all contents will be changed within the same period. The speed of change of link structure is faster. About 80 percent of all links are changing or will be new within a year. Although the absolute values may be out of date now, the results show that it is very important for the search engines to keep their databases refresh and up to date.

Freshness in Web-based research can be seen as a factor in information quality [12]. It is important for the user to get

information that is current. The out-of-date information, may lead the user to wrong conclusions for the desired information.

i) *Geographic Information*: User's information need can be satisfied to a great level by discovering user's explicit and implicit geographic intention in web search. Many times a user's information need has some kind of geographic boundary associated with it. For example, when the user wants information about shops, malls etc. Previous research has shown that more than 13% of web queries contain Geographic information [5, 33]. By identifying geographic information of user queries for various retrieval tasks, search engines can personalize retrieval results based on the geo information in the query to improve a user's search satisfaction. Also search engines can provide better advertisement matching and deliver more information about local goods and services that in which the users may be interested. Earlier researches have demonstrated that by incorporating related geo information the retrieval performance for a query can be improved when this information explicitly appears in the query or is known in advance [34]. Also it has been found that only about 50% of queries where the users expected the results to be contained within some geographic area had explicitly mentioned location names [5]. Identifying implicit geographic intent and precisely discovering missing location information is very necessary for satisfying the search engine user. Xing & Chris [8] developed a system for automatically discovering user's specific geographic intent in the web search at the city/location geo level, even when the explicit geo information is missing.

k) *Extra Services*: The modern web search engines provide extra services like spelling suggestions, current news, images, local results, stock charts, etc. to users. For the evaluation of the effectiveness of search engines to satisfy users' needs, these extra elements are very important to consider. It is found that the queries for which search engines provide spelling suggestion along with search results the user session satisfaction is much higher [4]. Advance facilities like assisted tool in commercial search engine have been found to have a considerable effect on the search engine's performance, satisfaction and confidence [35].

III. SURVEY TECHNEQUE

A survey was conducted from sample size of 100 diverse search engine users to identify the importance of the above mentioned factors with additional three factors namely availability, 'Help & Support' features and Search Engine's future evolution updates for user as they query search engine. In a written questionnaire, people were asked to rank the factors according to the importance they give to these factors as they select a search engine for searching information. Survey sample consisted of heterogeneous group of participants consisting of research scholars, corporate employees, faculties, tech-savvy people and some naive users.

IV. SURVEY RESULT/BEHAVIOUR BASED EVALUATION

After carefully studying and scrutinizing the data collected from the questionnaire the factors affecting user satisfaction are categorized as primary(P) and secondary(S).

Table 1: Weight Value for Search Engine Selection Factors

P-Primary, S- Secondary	Factors	Normalized Score
P1	Response Time	0.037
P2	Up- to- date Information	0.027
P3	Reliability	0.037
P4	Search Result relevancy	0.034
S1	Presentation of Search Results	0.088
S2	'Look and feel' , friendliness & easiness of User Interface	0.076
S3	Ability to handle different interpretation of a query(Diversity)	0.073
S4	Ability to provide results based on geographic requirements	0.086
S5	Extra services like news, stock chart etc.	0.113
S6	Ability of providing useful information for the rare queries too	0.073
S7	Ability to provide all the information which is available on the web	0.076
S8	Availability	0.067
S9	Help & Support' features for users	0.104
S10	Search Engine's future evolution updates	0.110

Table 1 show that up-to-date information is the most preferred criteria for search engine selection while users gave least preference to future evolution updates. Thus up-to-date information, Search Result relevancy, Response Time and Reliability are the most important factors for satisfying the maximum users.

V. CONCLUSION AND FUTURE WORK

Our study has shown that user satisfaction is affected by several factors and few of them are presented in our work. User satisfaction is proved to have a strong relationship with relevance, freshness, presentation of search results. It is believed that higher satisfaction can be achieved by providing user friendly interface, fast response time, good coverage of Web. Further there is evidence that diversity of content and discovery of new pages/content will play an important role

towards user satisfactions. There are several human factors which may contribute to the overall level of user satisfaction. Based upon a survey conducted it is concluded that up-to-date information, search result relevancy, response time and reliability are most important factors for satisfying the maximum users. This work has not shed light on factors related to user expectation & behavior and psychology. These factors can also affect the users' satisfaction. Moreover we also are working on how to validate these factors with the help of empirical study. In our future work we try to propose a model using Neural Network to automatically predict the user satisfaction with above proposed factors.

VI. REFERENCES

- [1] iProspect Search Engine User Behavior Study", http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf
- [2] Keynote Customer Experience Rankings / Search Engine Industry Research Abstract / June 2004
- [3] Ali Dasdan, Kostas Tsioutsoulis, Emre Velipasaoglu, "Web Search Engine Metrics for Measuring User Satisfaction" Tutorial @ 18th International World Wide Web Conference, April 20-24, 2009
- [4] S.B. Huffman and M. Hochster (2007), "How Well does Result Relevance Predict Session Satisfaction?" SIGIR'07.
- [5] M. J. Welch and J. Cho. Automatically identifying localizable queries. In ACM SIGIR, pages 507–514, 2008.
- [6] Dawei Yin, Zhenzhen Xue, Xiaoguang Qi, Brian D. Davison, "Diversifying search results with popular subtopics"
- [7] Sihvonen, A., & Vakkari, P. "Subject Knowledge, Thesaurus Assisted Query Expansion and Search Success". Journal of the American Society for Information Science, 55(11), 2004, pp: 963-969.
- [8] Xing Yi Hema Raghavan and Chris Leggetter, "Discovering Users' Specific Geo Intention in Web Search" WWW 2009, April 20–24, 2009, Madrid, Spain.
- [9] A. Dasdan and X. Huynh, User-centric content freshness metrics for search engines, WWW'09.
- [10] Dirk Lewandowski, "A three-year study on the freshness of Web search engine databases" Journal of Information Science, XX (X) 2008, pp. 1–17
- [11] A. Ntoulas, J. Cho and C. Olston, What's New on the Web? The Evolution of the Web from a Search Engine Perspective. In: Proceedings of the Thirteenth WWW Conference, New York, USA (2004).
- [12] S. Adams, Information Quality, Liability, and Corrections, Online 27(5) (2003) 16-23.
- [13] A. Dasgupta, A. Ghosh R. Kumar, C. Olston, S. Pandey, and A. Tomkins (2007). The discoverability of the Web, WWW'07.
- [14] S. Pandey and C. Olston. User-centric web crawling. In Proc. 14th WWW, pages 401–411, 2005.
- [15] R. Song, Z. Luo, J.-R. Wen, Y. Yu, and H.-W. Hon. Identifying ambiguous queries in web search. In WWW '07: Proceedings of the 16th international conference on World Wide Web, pages 1169–1170, New York, NY, USA, 2007. ACM
- [16] H. Chen and D. R. Karger. Less is more: probabilistic models for retrieving fewer relevant documents. In SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pages 429–436, New York, NY, USA, 2006. ACM
- [17] C. Zhai, W.W. Cohen, and J. Lafferty. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. In Proceedings of ACM SIGIR 2003, pages 10–17, 2003
- [18] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In WSDM '09: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pages 5–14, New York, NY, USA, 2009. ACM.
- [19] Tessier, J.A., Crouch, W.W. and Atherton, P. (1977) 'New Measures of User Satisfaction with Computer-based Literature Searches', Special Libraries 68: 383–9.
- [20] Jillian R. Griffiths, Frances Johnson and Richard J. Hartley, 'User satisfaction as a measure of system performance', Journal of Librarianship and Information Science 2007; 39; 142
- [21] Nachmias, R., & Gilad, A. "Needle in a Hyper stack: Searching Information on the World Wide Web". Journal of Research on Technology in Education, 34(4), 2002, pp: 45-486.
- [22] Fadhliah Mat Yamin, T.Ramayah "An Overview of Web Search Satisfaction," Communications of the IBIMA, Volume 3, 2008.
- [23] Lawrence, S. and Giles, C. L. (1999), "Accessibility of information on the web, Nature, Vol. 400, pp. 107-109.
- [24] Antti Oulasvirta, Janne Hukkinen, Barry Schwartz, "When More Is Less: The Paradox of Choice in Search Engine Use", SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.
- [25] Schwartz, B. The Paradox of Choice: Why More Is Less. Harper Perennial, 2005.
- [26] Dan Zhang, Luo Si, "Modeling Search Response Time " SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.
- [27] Ricardo Baeza-Yates¹, Aristides Gionis¹, Flavio Junqueira¹, Vanessa Murdock¹, Vassilis Plachouras¹, Fabrizio Silvestri, "The Impact of Caching on Search Engines ", SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands.
- [28] John A. Hoxmeier, Chris DiCesare, System Response Time and User Satisfaction: An Experimental Study of Browser-based Applications, in Proceedings of the Association of Information Systems Americas Conference (2000).
- [29] Rushinek, A. and S. Rushinek, "What Makes Users Happy?," Communications of the ACM, 29, 7 (July 1986), 594-598.
- [30] Parker, L., and Johnson, R. Does order of presentation affect users' judgment of documents? JASIS 41, 7 (1990), 493-494.
- [31] Huang, M., and Wang, H. The influence of document presentation order and number of documents judged on users' judgments of relevance. JASIST 55, 11 (2004), 970-979.
- [32] Mike Thelwall, Commercial Web Sites: Lost in Cyberspace? Internet Research Volume 10 Number 2, 2000, pp 150-159.
- [33] R. Jones, W. V. Zhang, B. Rey, P. Jhala, and E. Stipp. Geographic intention and ~~motion~~ in web search. International Journal of Geographical Information Science (IJGIS), March 2008.
- [34] B. Yu and G. Cai. A query-aware document ranking method for geographic information retrieval. In ACM GIR, pages 49–54, 2007.
- [35] Topi, H., & Lucas, W. "Searching the Web: Operator Assistance Required". Information Processing & Management, 41, 2005, pp: 383-403.
- [36] Belkin, N.J. and Vickery, A.(1985), Interaction in information systems. West Yorkshire: The British Library Broad.
- [37] Kuhlthau, C.C (1991), "Inside the search process: Information seeking from the user's perspective." Journal of the American Society for Information Science 42(5), 361-371.
- [38] Beg, M.M.S. "A Subjective Measure of Web Search Quality". An International Journal of Information Science, 169, 2005, pp: 365-381.
- [39] Beg, M.M.S., & Ahmad, N. "Web Search Enhancement by Mining User Actions". Journal of Information Science, 177, 2007, pp: 5203-5218.
- [40] Ivan Berlocher, Kyung-il Lee, Kono Kim "TopicRank: Bringing Insight to Users" SIGIR'08, July 20–24, 2008, Singapore. ACM 978-1-60558-164-4/08/07.
- [41] Shuo Wang, Feng Jing, Jibo He, Qixing Du, Lei Zhang "IGroup: Presenting Web Image Search Results in Semantic Clusters" CHI 2007, April 28–May 3, 2007, San Jose, California, USA.