Comparative Study of Web Search Engines and User-Centric Search Engine

Mubashshirahbanu Shekh, Vikramsinh Sisodiya, Ms. Jikitsha Sheth, Dr. Kalpesh Lad

ABSTRACT

Searches of the entire World Wide Web using search engines such as Google, Yahoo!, Bing, and Ask have become an extremely common way of locating information. Search engines are providing great facilities to the Internet users to search intended information from hundreds of millions of Web pages within a part of second. Today search engines are become more powerful and efficient by using various algorithms and technologies to provide a best result which demanded by user. Search engines now provide various added services too. One more area where search engine can also improve by keeping the track of user activity and history of visited sites which help user to carry their previous visited sites among different Web browsers, now a day it is done by Web browser only. This paper presents critical comparison of various popular search engines based on added features. A detailed analysis is presented and results are provided.

Keywords: Search Engine Evaluation, Search Engine Statistics, Feature Comparison, Search Engine Evaluation, World Wide Web.

1. INTRODUCTION

World Wide Web is a complex and unique source of multimedia information 3. It plays an important role in business, education, health, banking, communication, agriculture etc. Every day a bulk of information which includes thousands of documents and images are added to the web 28. Search engines are the resources that help the people to find online information available on the web. People can search the web by using search engines that use different algorithms and techniques 6. Search engines play a major role for searching images, sounds, graphics and other kind of files. A number of search engines are available on the web with their own abilities, characteristics, searching behavior and features 4.

In 2011 around 27% of the global population has access to the Internet at home 16.With more than 12 billon searches being performed each month as of January 2009, approximately 400 million web searches are performed every day. This means that on average more than 4500 searches are performed every single second of every day. 8And Google reported in July 2008 that it knew of about a trillion pages on the web 12. In December 2008, the worldwide Internet audience of people ages 15 and older surpassed one billion visitors for the first time 18. In January 2007, there were 106,875,138 on the Internet 13. As of January 2011, there were 273,301,445 14. This represents a 156% increase in 4 years. So, today people use web search engine to select item to eat from menu.

2. HISTORY OF SEARCH ENGINE

The Internet starts from the 1950s and 1960s with the development of computers. This began with point-to-point communication between mainframe computers and terminals, expanded to point-to-point connections between computers and then early research into packet switching. In 1982 the Internet Protocol Suite TCP/IP was standardized and the concept of a world-wide network of fully interconnected TCP/IP networks called the Internet was introduced. The Internet was commercialized in 1995 when National Science Foundation Network (NSFNET) was decommissioned, removing the last restrictions on the use of the Internet to carry commercial traffic. A search engine thus behaves like a reference librarian who helps user to find information sources **Error! Reference source not found.**.

Internet was new to the Information Technology as it become more popular many website added to network and it was expanding with light speed. The first search engine Archie was developed as a school project by Alan Emtage, a student at McGill University in Montreal in 1990.**Error! Reference source not found.** In 1993, Matthew Gray produced the first web robot, the Perl-based World Wide Web Wanderer, and used it to generate an index called 'Wandex', that crawl the web indexing and searching the catalog of indexed pages on the web in the form that is known as search engines today. Netscape was looking to give a single search engine an exclusive deal to be the featured search engine on Netscape's web browser. There was so much interest that instead a deal was struck with Netscape by five of the major search engines, Yahoo!, Magellan, Lycos, Infoseek, and Excite.28 The major search engines found in use today originated in development between 1993 and 1998.

3. HOW SEARCH ENGINE WORKS

There are differences in the ways various search engines work, but they all perform three basic tasks: 11.

- They search the Internet or select pieces of the Internet based on important words.
- They keep an index of the words they find, and where they find them.
- They allow users to look for words or combinations of words found in that index.

The most important measure for a search engine is the search performance, quality of the results and ability to crawl, and index the web efficiently. The

primary goal is to provide high quality search results over a rapidly growing World Wide Web.

Apart of their main task of providing link to user by their request, today search engine have to provide many features and security to the user to stand in the competitive market. For providing new feature they always are in process of development and add new features to recent product in regular time of periods.

4. COMPARATIVE STUDY OF SEARCH ENGINES

Generally most of search engines having own different techniques and methodologies. Here the authors have described the details of the most popular search engines and it's some Statistics like Rank and Market Share. 212010

Search Engine	Compete Rank	Alexa Rank	Market Share in May 2011	Market Share in December 2010	Unique Visitors(08/2011)	Estimated Unique Monthly Visitors
Google	1	1	82.80%	84.65%	165,067,996	900,000,000
Yahoo!	8	4	6.42%	6.69%	69,471,098	160,000,000
Bing	13	22	3.91%	3.29%	93,690,606	165,000,000
Ask	11	52	0.52%	0.56%	60,378,295	125,000,000

Table 1: Search Engines Statistics

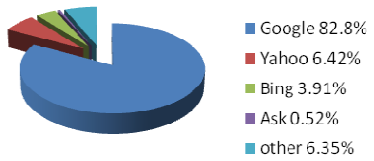

Fig. 1: Search engine market share of October 2011

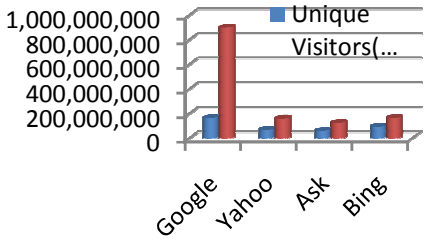

Fig. 2: Search Engine Usage Data

4.1 FEATURES OF SEARCH ENGINES

Twenty first century is so much dependent on the web and web searching is one of the first choices for number of people. This paper presents the comparison of

four search engines i.e. Google, Yahoo, Bing, Ask on the basis of features, facilities, speed and many more things. A number of studies have developed approaches to evaluating web search engines 5.Authors have compared these search engines on the basis of the features and results provided by Search Engines.

Features	Google	Yahoo	Ask	Bing
Website	www.google.com	www.search.yahoo.com	www.ask.com	www.bing.com
Search Operator	AND, OR, NOT	AND, OR	AND, OR	AND, OR, NOT
Search Web	√	√	√	√
Search Images	√	√	√	√
Search Videos	√	√	√	√
Search News	√	√	√	√
Search Maps	√	X	X	√
Search Books	√	X	X	X
Advance Search	√	√	√	√
Change Background	√	X	√	√
Change Search Settings	√	√	√	√
Shopping	X	√	X	X
Translation Services	√	X	X	√
Multi-Language Support	√	X	X	X
Questions/Answers	X	√	√	X

Directory	√	√	X	X
Advertising Programs	√	√	X	X
Business Solution/Services	√	X	X	X
Themes	X	X	X	√
Case Sensitive	X	X	X	X
Finance	√	√	X	X
Safe Search	√	√	√	√
Search Pad	X	√	X	X
Careers	X	√	√	X
Preferences	√	√	√	√

Table 2: Features available in Four Search Engines

Table 1 shows the options available on the home page of these four search engines. From the above table, it is clear that Google provides more options as compared to other search engines. It has better search capabilities than others. There are many others features which are unique to a particular search engine. Following is the description of the unique features provided by the each search engine.

Features	**Explanation**
Search Operator	The operators used internally by the search engine for retrieving the results.
Search Web	Search the information from the web.
Search Images	Search for images on the web.
Search Videos	Search online videos, T.V. shows from the web.
Search News	Search for news, top stories from the various search engines.
Search Maps	Search engine enable users to search for directions from one location to another and more.
Search Books	Search and preview millions of books from libraries and publishers worldwide.

Advance Search	Allow User with advanced options to write specific query and return more precise results.
Change Background	User can customize page i.e. User can change the background settings according to their own choice.
Change Search Settings	User can change the search settings.
Display No. of Results	Search engines display the number of results fetched.
Shopping	Facility of buying online products.
Translation Services	Search engine translate text and web pages to another language.
Multi-Language Support	Search engine support multiple languages?
Answers	Good facility where people ask and answer questions on any topic and can share facts, opinions and personal experiences community.
Directory	Facility to search the web, organized by topic or category.
Advertising Programs	Search engines provide the facility to the users to advertise their business and products.
Business Solution/Services	Search engine provide business solution facility to promote and help user's business.
Themes	User can change theme according to his own choice.
Case Sensitive	Search engine is case sensitive or not.
Finance	Information regarding the stock market.
Safe Search	Allow the user to filter out explicit, adult-oriented content from results.
Search Pad	To keep track of the websites you choose from the search results and to make notes on them.
Careers	User can browse various jobs according to his choice.
Preferences	User can search information exactly what they want.

Table 3: Explanation of various features mentioned in Table 2

Generally all search engines frequently updated their feature and provide some easy and powerful tool for user's comfort. Many Research work is running for adding a new feature. Following Table 3 Shows the Country wise Search Engine Usage Statistics 1719.

Country	Google	Yahoo!	Microsoft	Ask	Baidu
Argentina	89.40%	-	-	-	-
Australia	91.81%	2.38%	4.20%	0.30%	-
Brazil	89.00%	-	-	-	-
Canada	80.96%	4.20%	8.60%	3.70%	-
China	19.60%	-	-	-	75.50%
Colombia	92.20%	-	-	-	-
Hong Kong	58.02%	23.51%	-	-	6.08%
India	88.40%	-	-	-	-
Japan	38.00%	51.00%	-	-	-
Latin America	90.50%	-	-	-	-
Russia	-	-	-	-	-
Singapore	86.51%	10.17%	1.74%	-	-
South Korea	7.30%	4.10%	-	-	-
UK	91.65%	2.51%	2.88%	1.36%	-
US	65.50%	15.90%	14.10%	2.90%	-

Table 4: Country wise Search Engine Usage

From the above Table 3 Statistics, The Google is the Leader of the Search Engines.

4.2 SEARCH RESULT ON THE BASIS OF SOME QUERY

When a search is carried out in response to a search query, many times the user is enabling to retrieve the relevant information. The quality of searching the right information accurately is said to be the precision value of the search engine Table 4 shows the search results of Google and Yahoo! on 9 different Queries of APPENDIX I.

Search Query	**Total Num. of Site**	**Total Num. of Site**
Q#1	3,110,000,000	328,000,000
Q#2	1,390,000,000	309,000,000
Q#3	6,320,000,000	2,200,000,000
Q#4	120,000,000	665,000,000
Q#5	35,300,000	42,400,000
Q#6	127,000,000	124,000,000
Q#7	58,100,000	62,600,000

Q#8	5,960,000	20,100,000
Q#9	101,000,000	59,500,000

Table 5: Search Result Based on Queries

APPENDIX 1

Simple 1 Word Query	Simple Multi word query	Complex Multiword Query
1. Computer	4. Internet Resources	7. How Internet Search Engine Works
2. Database	5. Search Engine	8. Evaluation of website
3. Internet	6. Cyber crime	9. Designing of Library Building

5. USER-CENTRIC SEARCH ENGINE

Recently search engines like Google, Yahoo, Bind, Ask etc. are providing various features. Authors have proposed other features which will provide user based more relevant results for a search done by a user.

A. General search

General search is also known as blinded search. This is an ordinary search like currently in trend; General search include all type of web pages from fresh pages, past result pages and other pages which are not belong from these categories. The algorithms for general search are simple as today's blinded search. It displays the link from crawled content by general ranking algorithms.

B. Visited pages search

In this search the results for a given search query is found only in those pages which are already visited by that user. Thus the user is not required to remember the links or other information which he/she has visited earlier.

C. Fresh Search

This feature allow user to visit new web pages only. Generally user want to find a specific contain from web and search engine provides the same list of links; hence he get frosted by thinking that the information he/she is searching may not available on web. This feature allows him/her to have Search Engine Result Page show the result from those pages which are not visited by him earlier, or cleaned up from his history. For this feature implementation, search engine is required to keep the track of all web page history which is visited by the user.

Fig. 3: Fresh Search Retrieval

As shown above the UCSE collect the entire link from general search algorithm and minus result with history links and find links which are new to the users.

D. Past activity

Using this feature, a user is able to know, which URL has been visited how many times by him/her till now. Also, keyword inputted, date and time are also associated with such URL. Details regarding what have been uploaded or downloaded on that website by the user, if any, till now is also described in this feature. The user is facilitated to view and remove browsed page history. For the visited website URLs, the user can add tags to them. For example, user is allowed to give a tag like favorite or a customizing tag i.e. office data, research data, etc. This will help user to form a logical grouping of URL.

5.1 TAGS MANAGEMENT

Search engines like Google keep history of user search keywords and web site visited. User is not required to depend on web browser to keep the history of browsed sites. Visited history is available to the user wherever user goes. Authors have proposed different mechanism for tagging the visited links which are as follows:

A. Favourite

This feature allow search engine to keep some web site link as favorite like bookmark. It help user to access favorite site fast. As author mentioned above UCSE keeps the user visit links stored in database; user can tag this links for easy categories. It is the favorite link tag which always shown to user in search engine result page.

B. Custom Tags

User is able to create or modify his/her own tag which provides grouping of web sites. Later on, the user can do exclusive search on such tagged links. For example, the tags created by a user is Personal, Research and Entertainment. If he/she requires to search a research related article already tagged with Research tag, user can now search that article in the visited URLs having tag “Research”.

C. Remember me

This is a special tag given to web page link when user is in hurry or is not able to decide the category for that URL. Giving such tags to any web page URL helps the user to retrieve data immediately from the same page in later use. It will give all links list which are tagged as Remember me.

6. CONCLUSION

Searching is one of the common tool for seeking online information. Search engines are the resources that help the people to search any kind of information on the web in a simple and easy way. The purpose of this study is to present the extensive review and analysis of the four search engines. Authors conclude that Google is best till date. People like to search information on Google as it provides better interface, features and ease of use to the users. As user expectations are not static they demand more and more facility for comfort, easiness and fast retrieval from search engines, authors have proposed some features that can make the Search Engine as User-Centric Search Engine.

7. REFERENCES:

1. "Internet History - Search Engines" (from Search Engine Watch), University Leiden, Netherlands, September 2001, web: LeidenU-Archie.
2. "Yahoo! And Netscape Ink International Distribution Deal"
3. Goodrum, A. Spink, “Image Searching on the Exite Web search engine”, Information Processing and Management 37, 2001, pp. 295-311.
4. Howe and D. Dreilinger, “SavvySearch: A Metasearch Engine That Learns Which Search Engines to Query”, AI Magazine, Vol. 18 No. 2, 1997.
5. Spink, “A User Centered Approach to Evaluating Human Interaction with Web Search Engines: An Exploratory Study”, Information Processing and Management, 38 (2002) 401-426.
6. Spink, D. Wolfram, B. Jansen and T. Saracevic, “Searching the Web: The Public and Their Queries”. Journal of the American Society for Information Science and Technology, Vol. 52 Issue 3, Feb. 1, 2001.
7. Archive of NCSA what's new in December 1993 page
8. Art of SEO Mastering Search Engine Optimization [Book- First Edition] (Chapter 1, Page 1)
9. Browser Deals Push Netscape Stock Up 7.8%. Los Angeles Times. 1 April 1996

10. Middleton and R. Baeza-yates, "A comparison of open source search engines," Net Marketshare – World

11. http://computer.howstuffworks.com/internet/basics/search-engine.htm [Access on Nov 2011]

12. http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html

13. http://news.netcraft.com/archives/2007/01/05/january_2007_web_server_survey.html

14. http://news.netcraft.com/archives/2011/01/12/january-2011-web-server-survey-4.html

15. http://wrg.upf.edu/WRG/dctos/Middleton- Baeza.pdf. Dec. 03 2008. [Online]. Available

16. http://www.budde.com.au/Research/Global-Internet-Search-Engines-Websites-Internet-User-Statistics.html

17. http://www.chandlernguyen.com/2011/03/search-engine-market-share-by-country-mar-2011.html

18. http://www.comscore.com/Press_Events/Press_Releases/2009/1/Global_Internet_Audience_1_Billion

19. http://www.comscore.com/Press_Events/Press_Releases/2011/5/Google_Sites_Accounts_for_9_of_10_Searches_Conducted_in_Latin_America -- ChandlerNguyen Sun Jul 10 2011 05:07:44 GMT+0700 (SE Asia Standard Time)

20. http://www.ebizmba.com/articles/search-engines(Last Updated: September 12, 2011)

21. http://www.karmasnack.com/about/search-engine-market-share/

22. Internet Search Engine Statistics, pwebs.net/news/wp-content /plugins /...pdf.../ kalins_pdf_create.php?

23. J. Bernard and A. Spink, "A Comparison of Nine Search Engine Transaction Logs", Information Processing and Management 42(1), pp. 248-263.

24. M. Kaur, N. Bhatia and S. Singh "Web Search Engines Evaluation based on Features and end-user Experience", International Journal of Enterprise Computing and Business Systems ISSN (Online) : 2230-8849

25. P. Panigrahi "A language to create web page", Vidyasagar University Journal of Library and Information Science; 1997; 2; p 61-70.

26. S. Lawrence and C. Lee Giles, "Searching the World Wide Web", Vol. 280 no. 5360 pp. 98-100, 3 April 1998.

27. The Anatomy of a Large-Scale Hypertextual Web Search Engine, http://infolab.stanford.edu/~backrub/google.html

28. Y. Rui, S. Thomas, Huang and Shih-Fu Chang, "Image Retrieval – Past, Present and Future", Journal of Visual Communication and Image Representation, 1997, pp: 1-23.

AUTHOR'S PROFILE

Ms. Mubashshirahbanu Shekh, MCA (Pursuing) is Student of Shrimad Rajchandra Institute of Management and Computer Application, affiliated to Gujarat Technological University, Ahmedabad. She is "IBM Tivoli Directory Server V6.1" and "IBM DB2 9 Fundamental" certified. She also got the 1st Rank during Graduation. She has written one Article. She is Active Member of Computer Society of India.

Mr. Vikramsinh Sisodiya, MCA (Pursuing) is Student of Shrimad Rajchandra Institute of Management and Computer Application, affiliated to Gujarat Technological University, Ahmedabad. He is "IBM Tivoli Directory Server V6.1", "IBM DB2 9 Fundamental" and GIL [Gujarat Informatics Limited] core Java Training certified. He has written one Article. He is Member of Computer Society of India.

Ms. Jikitsha Sheth, GSET qualified, is working as Asst. Professor in Shrimad Rajchandra Institute of Management and Computer Application, Gopal Vidyanagar and has 7 years of teaching experience. Her research interest areas are sata mining and natural language processing. She has presented and published research papers in national journals and conferences. She can be contacted at jikitsha.sheth@srimca.edu.in

Dr. Kalpesh B. Lad, B. Sc. (Computer Science), MCA, Ph. D is presently working as Assistant Professor at Shrimad Rajchandra Institute of Management and Computer Application, Uka Tarsadia University, Maliba Campus, Tarsadi. He has more than 9 years of teaching as well as research experience. He has presented/published more than 15 papers in international and national conferences/journals. His area of interest includes programming languages, compiler construction and open source technology. He can be contacted at kalpesh.lad@srimca.edu.in.