

Study and Comparison of Various Search Engines' Browsing Capabilities

^ISridhar Neralla, ^{II}Renuka Devi J, ^{III}Nirmala A, ^{IV}Swarna M

^IAssociate Professor, Dept. of CSE, GMRIT, Rajam, Andhra Pradesh, India

^{II,III,IV}4th Year B.Tech IT, GMRIT, Rajam, Andhra Pradesh, India

Abstract

The primary objective of this paper is to understand the browsing capabilities of search engines. There are many search engines were available. Each offers a unique set of features and indexes. Moreover no two engines are same. In this paper we are focusing on the browse capabilities of the various search engines, their advantages and limitations. So that a researcher can work further to implement a new search engine which overcomes the drawbacks of the previous ones and also with more optimization techniques. This paper provided a comparative study of various existing browsers made.

Keywords

Search Engines, Browsing Capabilities, Ranking, Zoning, Web Technologies, Information Retrieval

I. Introduction

Now-a-days finding the required information quickly and easily on the web remains a major challenge, if the searcher has little prior knowledge of search strategies and search techniques of search engines.

Once the search is complete, Browse capabilities provide the user with the capability to determine which items are of interest and select those to be displayed. So browse capabilities plays a major role in information retrieval.

This paper divided into 7 sections, section I contains introduction of web browsers, section II and III focuses on literature review and basics browser capabilities that are required by browser. Section IV reviews general architecture of a browser. Section V discusses about various existing browsers and their capabilities. In Section VI elaborates various differences between browsers along with their respective pros and cons. Final section concludes the paper.

II. Literature Review

The importance of browse capabilities would be lessened, if searcher resulted in high precision [2]. The main intention of browse capabilities is to support the user in focusing on items that have the highest probability in meeting his need. There are several researchers done comparison works related to search engine performance issues. Gayakwad and Phulpagar focused on comparative analysis of re-ranking models [3]. Working process of meta crawler was elaborated in the works of Erik Selberg et al. [1].

Lawrence and Giles simplified the process of control meta search engine process [4] for improving search engine process. Further resource aggregation on the Web proposed by Selberg and Etzioni for meta crawler architecture. Dreilinger and Howe proposed architecture for processing queries based on multi-agent model [5].

III. Minimum Browsing Capabilities

The browse capability mainly focuses on the users' interest in displaying the items. There are mainly two ways in displaying summary of items according to the query: line item status and data visualization. Based on these summary displays the user selects some of the specific items.

A. Zoning

Once the query is entered the user wants to see the minimum information needed to determine if the item is relevant. Once the user identifies that the item is possibly relevant then the item is displayed for the detailed review. As the limited display screen sizes depends on the select ability of portions of an item a user needs to see. For example topic title, author is sufficient information for a user to predict the potential relevance of an item.

Zoning makes the multiple items to be displayed on a single display screen. Zoning is the idea of locality and passage based search and retrieval. In this case the basic search unit is not the complete item, but an algorithmic defined subdivision of the item. Zoning issues in popular search engines can be observed by incorporating a query related to a place. The query contains word "Hyderabad", it verifies locations that are in India rather than other country, because query is being submitted from an Indian location. Table 1 shows the search results from some of the popular search engines. For the query: "Pilgrim places in Hyderabad".

Table 1: Search Engine Results for a Query

SNO	SEARCH ENGINE	RESULTS NUMBER	TIME(in sec)
1.	GOOGLE	31,60,000	0.23
2.	YAHOO	591,000	0.29
3.	BING	5,91,000	0.22
4.	DIRECT HIT	10	-
5.	ALL THE WEB	395,000	0.19



Fig. 1 Query typed in Google as “how to make a flying robot with paper” and some of the highlighted words shows paper, flying, robot, makes.

B. Highlighting

Highlighting is an indication of why an item was selected. It lets the user quickly focus on the potentially relevant parts of the text to check for item relevance. The strengths of highlighting indicates how strongly the highlighted word participated in selection of item and allow subsequent jumping to the next highlight. An additional capability is that identifying the passage in the document which is most relevant to the query. Highlighting has always been helpful in Boolean systems to specify the cause of retrieval. This is because of the straight mapping between the term in search and the terms in the item.

The screenshot shown in figure 1 shows how highlighting can be seen in Google: Top part of the figure shows the query entered into Google as “how to make a flying robot with paper”. The results are shown in bottom part of the figure 1, as one can easily observe the use of highlighting feature. Here first result showing flying, robots and paper as highlighted ones, similarly for other results. Highlighting features catches the eyes of users so that they can easily identify the required relevant results.

III. Ranking

Ranking is mainly based upon the prediction of the relevant values, the status summary displays the relevant score associated with the item along with the brief descriptor of the item. Most IR systems figure a numeric score on how well each object in the database matches the query, and rank the objects according to the value. The popular search engines when they begin their journey they used ranking algorithms as given below:

Yahoo: Search Engine Optimization Algorithm

Google: Google search is based on priority rank called “Page Rank”.

Direct Hit: popularity ranking algorithm

IV. Search engine Architecture

1. User Expectations from Browsers

Web users have so many expectations from the browsers some of them are

- The browser should be fast enough that it should not waste

the user’s time.

- It should bring all the information with people seeking answer.
- Provide the best user experience. I.e. look and feel of the browser should be good.
- Should solve the search problems.
- Provide democracy on the work of the user and it should provide the open source implementations.
- It should update with the new features.
- It can be deployed in any computer device.

2. How Browsers Works

Browser is a software program that searches for the information needed by the user and that we designated as the search terms. Every browser will maintain its own database to retrieve the information. Browsers automatically creates website listings by using spiders that “crawl” web pages, index their information and follows that site links to other pages. Spiders return to already crawled sites to check for updates and everything that the spiders finds goes to its own database. All the browsers have same basic search process i.e. the searcher type the query into a browser .Then the browser quickly sorts literally millions of its pages in the database to find matches to this query then the browser results are ranked in order of the relevancy.

3. General Browser Architecture

It is very useful to understand how a search engine put together. The figure 2 shows schematically way of describing search engine architecture which was proposed by Aravind [7]. Every search engine depends up on the “crawler”. The crawlers are small programs that search web on the search engine behalf and to how a human user will follow link to reach different pages. They are given an initial set of URLs whose pages they retrieve. They extort URLs that come out on the crawled pages and give this information to the crawler control module.

The crawler control modules decide which pages to visit next and give their URLs back to the crawlers. The crawlers also put the retrieved pages into a page repository. Crawlers continue visiting the web, until local resources, such as storage, are completed. The crawl control module is responsible designed for directing the crawling operation. After completion of at least one crawling cycle the crawl control module informed by several indexes that were created during the earlier crawl. In sometimes the crawler may get the data from the private companies. The page repository will provide the index searchable by users.

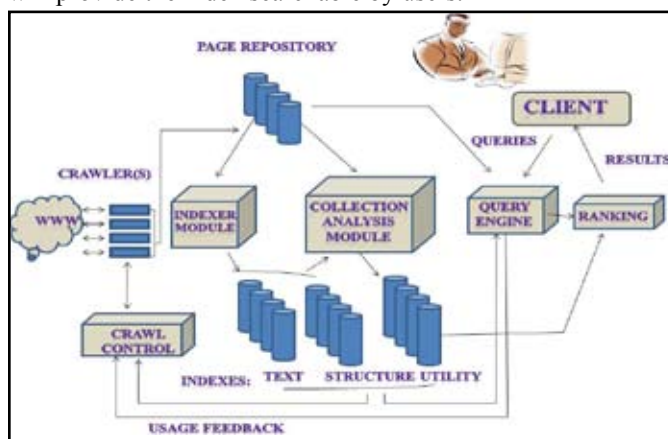


Fig. 2 : General Architecture of a typical search engine, which consists of Indexer Module, Crawl Control, Collection Analysis

Module, Query Engine and Ranking Module

General model [7] is shown in figure 2. The indexer module collects words from each page it visits and records its URLs. It became into a large lookup table that gives a list of URLs pointing to pages where each word occurs. The table lists those pages, which were covered during the crawling process. The collection analysis module creates a variety of other indexes like text, structure and utility. The utility indexer provides an access to pages of given length or a page contain a certain number of pictures in it. During the crawling and indexing, a search engine stores the pages it retrieves. They are stored for the short term in a page repository.

Search engines maintain a cache of pages they visit so that retrieval of already visited pages saved. The query module is responsible for taking the requests from users and full filling that requests. The ranking module will sorts the results such that the results near the top are most relevant to the user. This search engine architecture is modified in a distributed architecture of search engine. This search engine architecture has gatherers and brokers. Gatherers collect indexing information from web servers and brokers give the indexing mechanism and the query interface.

V. Search Engines and Their Capabilities

A search engine is a software structure that is intended to exploration of info on the Internet. Today there are various search engines for retrieving the information. But the interesting thing about the search engines is that they are not limited for desktops itself. In fact they are extended to smart phones, tablets etc. That means we have search engines for both desktops and handheld devices. Some of the search engines and the year of invention is shown in table 1.

Table 1: Search Engines and Inventions

Year	Search Engine Name	Invented By
1994	Yahoo	Jerry Yang and David Filo
1996	Google	Larry Page and Sergey Brin
1998	Direct Hit	Gary Culliss and Mike Cassidy
1999	All the Web	
2009	Bing	Steve Balmer

A. Yahoo

Informally, Yahoo stands for “Yet Another Hierarchical Official Oracle”. But initially it was called as “Jerry and David’s guide to the World Wide Web”. It maintains the database in hierarchical manner. In the early stage itself it has a rapid growth and by 1998 it became the starting point of web users. Initially it used Google for search but later it developed its own technology for searching. From 2014, it was the second leading search encyclopedia on the net by query size. It uses Search Engine Optimization Algorithm for ranking the websites. Currently Yahoo is going to have a deal with Microsoft.

Yahoo Search Features

Yahoo has some interesting search features which we do not find in any other search engines. These features need to be typical and beneficial and are ample to draw the attention and make users switch to Yahoo or at least practice it as a secondary search engine. The features are:

1. Search Assistant: This feature is a combination of autocomplete and related searches. Yahoo was inspired by

this feature from Ask.com search engine.

2. Site Advisor’s warnings: This is another interesting feature of Yahoo which adds Site Advisor’s warning following to search results. These are the security ratings from McAfee Site Advisor. They are created on automatic security tests of Web sites. Also they are improved by response from volunteer critics. Although other search engines serves the same purpose, this feature of Yahoo provides more information and safety regarding the threats.
3. Search Monkey: It is an impressive feature of Yahoo which is a way for site vendors to enhance the snippets with organized info. With the help of this feature the users were directly provided the info through ratings, reviews etc.
4. Semantic Web standards: This feature is used to obtain organized info based on the users’ opinion regarding the search results.
5. Yahoo has its own style of “Universal search” which includes image previews and videos from Flickr.

If Yahoo promotes all these features then it becomes more popular and may be become replacement of various popular search engines.

B. Google

Google also has some informal expansion such as “Global Organization of Oriented Group Language of Earth “. Google had a rapid growth since its evolution because of its simple design. It generally handles about three billion queries each day. Google search is based on priority rank called “PageRank”. It provides a customized search. It is not only meant for textual search, but it also provides retrieval of quotes, meanings etc. It has special features for calculations, conversions etc. It analyzes the frequency of search terms. It handles both indexable and non-indexable data. It is associated with universal search even. It maintains a series of localized websites. It allows the user to type simple search queries but goes for a query expansion. Google has a fascinating feature called “I’m feeling lucky” button which directly pushes the user to first search result.

Google Advanced Features:

Today Google is one of the top search engines. Google has about 40 options for searching. Apart from normal features it includes nearly 22 special features like patents, language translation, package tracking etc.

1. General features: Google’s general features are functional in nature. Its features comprise a definition link for maximum explorations, including dictionary arguments. It provides as many results as possible. It even offers results for the misspelt queries and many more.
2. Google implements “Term ordering” means that more definite terms have a better outcome on the search outcomes.
3. I’m Feeling Lucky: Generally home page of Google is associated with a button “I’m Feeling Lucky”. This feature bypasses the query results. This feature directly navigates the user to the first search outcomes.
4. Rich snippets: This feature of Google is an enhanced schematic markup. This feature is provided for location pages.
5. All these words: In this feature you need to specify what are the words you want in your search results. That means you need to type the important words of your needed information. This is an “Advanced search feature” which results the search results that containing the specified words.

6. This exact word or phrase: In this feature we need to specify the required words in quotes. Then it will give the results with these specified words.
7. Any of these words: In this feature we specify the needed words with a separator "OR". Then it will give results with contains either of the information.
8. None of these words: Sometimes we want exclude some results to get more relevant results. This can be achieved by putting "-before the results.
9. Numbers ranging from-to: With the help of this feature we would be able to range of numbers. This feature is helpful for searching information related to time period, prices of various products etc.
10. Language: Google allows to find the pages in our needed language. This provides more flexibility to understand the required information.
11. Region: Using Google we can find the pages published in a particular region. This reduces our effort of finding the information to some extent.
12. Last update: Using this feature we can find the pages which are updated within the specified time.
13. Site or Domain: In this feature of Google we can specify the required site or domain like oracle.com, stackoverflow.com etc.
14. Terms appearing: With this feature we can hunt for terms in the whole page, web address, or links to the page you're viewing.
15. Reading level: With this we can get the search results at one reading level or observe the level information.
16. Safe search: Google allows to block incorrect or unambiguous images from the search outcomes. This feature helps to avoid most of the mature info.
17. File type: This feature allows to specify the type of files we want to get as a search results. For example we can specify file types as pdf, ppt etc.
18. Usage rights: This feature allows to find search pages which are free to use on our own.

C. Direct Hit

Direct Hit was a search engine that involved consideration for its innovative use of ranking sites. Because of the "Automated search technology" it is different from other search engines. It ranks various sites by tracking information that consumers were searching for, the amount of time they spent on different websites that were visited from the search outcome, and how often they are back to the outcomes. It gained impression apart from various search engines as it determines the sites popularity based on users pattern of search. It used to work for various other search engines like MSN, HotBot etc. Because of its attracting technology it was bought by Ask Jeeves. It uses a popularity ranking algorithm.

Unique Features of Direct Hit:

1. Direct Hit is popular for searching "Popular Topics".
2. It searches multiple terms by matching with previous searches.
3. Direct Hit finds the information based on earlier queries but not against the records. Because of this many of the features are not used directly.
4. Direct Hit gives only 10 search results. For every search results, it provides a title attached to the URL, which is an extract, and the URL.
5. In Direct Hit Boolean search, Proximity search, Truncation,

Case sensitivity, Field search and Limits are not used directly.

6. It makes use of "visit through" and "visit duration" to know the popular sites for a specific topic.
7. It uses "Click popularity" for organizing the search results. And also this Click popularity gives the popularity to the sites based on the clicks done by various users for that site.
8. It uses a special feature called "Stickiness" which measures the time spent by the user in a particular site.

D. All the Web

At a point of time AlltheWeb was a rivalry to Google in terms technology and size. But it could never overcome Google and is even declines when it was bought by Overture in 2003. It has various advantageous search features than popular search engines like Google. It is associated with more advanced search features, customizable look, updated database etc. It is even provided with direct image search. From 2011 it was redirected to Yahoo search.

Popularity of All the Web in 2000s

1. AlltheWeb search is a showcase search engine for FAST search and Transfer.
2. It has a smart interface that unabashedly appeals on the "best practices" initiated by Google, AltaVista, and various other major search engines.
3. Its features are structural in nature.
4. AlltheWeb presents the results in a single column format.
5. This search engine doesn't have Banner ads.
6. It automatically determines the default language for search outcomes using browser's IP address. It provides 8 default languages.
7. AlltheWeb automatically rewrites confusing queries for good results.
8. It supports Boolean searching.
9. The RANK operator increases the occurrence of a search term, rearranges the outcome set with the specified terms itemized higher up.
10. It maintains a link to "Rerun Advance Search" which proceeds us to the advanced search form itself. And it is place next to the query.
11. It has a "URL Investigator" which gives us specific information about the site bearing the URL
12. It is provided with a search tool option which is used to create search button which allows us to highlight the info on web pages and this can be used as a query.

E. Bing

It was initially called Live search, MSN search. It is provided by Microsoft. It makes use of semantic technology. It includes the cataloging of search proposals when the queries are entered and a list of linked searches. It has a format of three column display. Unlike other search engines it maintains a back ground image and updates it daily. Third party site info can be used using enhanced view. It provides video and image search with adjustable setting. It uses advanced filters and advanced operators. It is integrated with Hotmail, Facebook etc. It is provided in various languages based on locations. It has a tie-up with Yahoo and in 2013 new-look Bing was released.

Recent Enhancements in Bing

1. Bing supports the ads based on the device preference.
2. New dynamic text variables: If dynamic text is added to URL

- then it specifies the destination URL what to use depending on whether the using device is mobile or not.
3. New bid adjustment options: This feature now includes negative bid adjustments, change in bid adjustment ranges etc.
 4. Ad group level site link extensions: This feature can be useful for overriding the site link extensions.
 5. Keywords Distribution Graph: This feature allows you to display the keyword info graphically. With this we can find outliers and refine in on problem regions in a faster way.
 6. Change History Graph: Using this feature we can have a relationship between actions and performance.
 7. Improved Opportunities Tab: Now Bing's opportunities Tab has been improved in such a way that it is now providing filtering and it now gives more suggestions for keywords.
 8. Bing Maps Preview: This feature is now provided for Windows8.1. This feature gives a more realistic user interface with 3-D imagery.
 9. Page zero: This feature resolves a query when the user is entering it, by showing links, data and actions on the fly before the outcomes displays.
 10. License: This feature is associated with searching of images in Bing. This results the images which are provided in public domain, which are restricted etc.
 11. Visual Search: This new feature of Bing allows us to get the information about various things with the help of their images.

VI. Detailed Comparison with Features

The table 2 shows the completed comparison of various search engines and their browsing capabilities.

Table 2: Browsers and Search Engine Capabilities

Sl. No.	Features	Google	Yahoo	Bing	Direct Hit	Allthe Web
1.	Term ordering	Y	Y	N	Y	Y
2.	I'm Feeling Lucky	Y	Y	N	N	N
3.	Rich Snippets	Y	N	N	N	N
4.	Terms Appearing	Y	N	N	N	Y
5.	All The words	Y	N	N	N	N
6.	Language	Y	Y	Y	Y	Y
7.	Safe search	Y	Y	N	N	N
8.	Site or Domain	Y	N	Y	N	Y
9.	Visual search	Y	Y	Y	N	Y
10.	Keywords Distribution Graph	N	N	Y	N	N
11.	Opportunities Tab	Y	Y	Y	N	Y
12.	Live Search	Y	Y	Y	Y	Y

13.	Search Assistant	Y	Y	Y	Y	Y
14.	Site Advisor's warnings:	Y	Y	Y	Y	Y
15.	Search Monkey	Y	Y	N	N	N
16.	Semantic Web standards	Y	Y	Y	Y	Y
17.	Universal search	Y	Y	Y	N	N
18.	Searching Popular Topics	Y	Y	Y	Y	Y
19.	Searching multiple terms	Y	Y	Y	Y	Y
20.	Limited search results	Y	N	N	Y	Y
21.	Various Types of searches	Y	Y	Y	Y	Y
22.	Visit through & visit Duration	Y	Y	N	Y	N
23.	Click popularity	Y	Y	N	Y	N
24.	Stickiness	Y	Y	Y	Y	Y
25.	FAST search and Transfer.	Y	Y	Y	N	Y
26.	Smart interface	Y	Y	Y	N	Y
27.	Structural in nature	Y	Y	Y	Y	Y
28.	Banner ads.	Y	Y	Y	N	Y
29.	Default language for search	Y	N	N	N	Y
30.	Automatic rewrite	Y	Y	Y	Y	Y
31.	Boolean searching	Y	Y	Y	Y	Y
32.	RANK Operator	Y	Y	N	N	Y
33.	Return Advanced Search	Y	Y	Y	N	Y
34.	URL Investigator	Y	Y	Y	N	Y
35.	Highlighting	Y	Y	N	N	Y

VII. Conclusion

Search Engines are day-to-day improving their browsing potentials. This paper majorly focuses on various browsing capabilities and review current capabilities of different search engines. After understanding the architectural issues and browsing features, one can create novel architecture by adding unique features to it.

References

- [1] E. Solberg and O. Etzioni, "Multi-Service Search and Comparison Using the MetaCrawler," *Proc. World Wide Web Conference*, Elsevier, North Holland, 1995
- [2] Gerald J. Kowalski Mark T. Maybury, "Information Retrieval System Capabilities", *Information Storage and Retrieval Systems*, pp.38-40, Springer Publisher, 2000
- [3] Gayakwad M. D. and Phulpagar B.D. "Research Article Review on Various Searching Methodologies and Comparative Analysis for Re-Ranking the Searched Results", *International Journal of Recent Scientific Research*, Vol. 4, pp.1817-1820, November, 2013;
- [4] Lawrence and Giles, "Context and Page Analysis for Improved Web Search", *IEEE Internet Computing*, August, 1998;
- [5] D. Dreilinger and A. Howe, "An Information Gatherin Agent for Querying Web Search Engines," *Tech. Report CS-96-111*, Computer Science Dept., Colorado State Univ., Fort Collins, Colo., 1996.
- [6] E. Selberg and O. Etzioni, "The MetaCrawler Architecture for Resource Aggregation on the Web," pp. 11-14, *IEEE Expert*, Jan.-Feb. 1997,
- [7] Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke and Sriram Raghavan. "Searching the Web", *ACM, ACM Transactions on Internet Technology*, vol. 1, no. 1, August 2001, pp 2-43, U.S.A. General Architecture Model
- [8] www.google.com
- [9] www.yahoo.com
- [10] www.alltheweb.com
- [11] www.bing.com
- [12] www.directhit.com



Ms. Nirmala A pursuing her B.Tech in G.M.R.I.T, Rajam. Currently she is 4th Year of IT Department. She is an active member in CSI (Computer Society Of India) and ISTE (Indian Society Of Technical Education) Member and participates in all activities. She communicated paper in "DNA May soon used for storage" for CSI journal. She is a hard worker about for her academics



Ms. Swarna M pursuing her B.Tech in G.M.R.I.T, Rajam. Currently she is 4th Year of IT Department. She is an active member in CSI (Computer Society Of India) and ISTE (Indian Society Of Technical Education) Member and participates in all activities. She is a hard worker about her academics



Sridhar Neralla is presently working as Associate Professor in CSE Department, GMRIT, Rajam and submitted thesis in Andhra University under the supervision of Prof.P.S.Avadhani and Dr.D.Lalitha Bhaskari. He authored five text books including *Information Technology*, *Advanced Data Structures*, and *Web Technologies*. He received his M.Tech (IT) from Andhra University. He is a Life

Member of CSI and ISTE. His research areas include Network Security, Cryptography, Cyber Forensics and Web Security.



Ms. Renuka Devi J pursuing her B.Tech in G.M.R.I.T, Rajam. Currently she is 4th Year of IT Department. She is an active member in CSI (Computer Society Of India) and ISTE (Indian Society Of Technical Education) Member and participates in all activities. She communicated paper on "3D Printing" for CSI journal. She is a hard worker about for his academics.