

An Overview Implementations of Automated Search Engine Based on Cloud Computing for Faster Searching

* Tanya Chouksey
** Pawan Patidar

ABSTRACT

There has been a seismic shift in the way we use search engines. The search engine is becoming a hub of a lot of information. It has become a market place in its own right. This is now become a challenging marketer to find new ways to connect with audiences. For search engines, this was a turning point in the services that they offer. It has become more and more important for the user to fetch prominent information that has been searched by the users. When before, "WEB SEARCH" brands like GOOGLE, YAHOO, BING, and BAIDU were consumers' first port of call, they are now rapidly losing share to new competition (particularly AMAZON, WE CHAT and FACEBOOK) and formats (primarily from VOICE and apps).

Personalization has become important due to an increase in the lot of work so the user wants to have a website that needless amount of memory and can get relevant data in a short time. Search engine which is newly creating or updating should be used so that it can bring a revolution in the web search engine history.

Keywords: - (WWW) World Wide Web, (SEO's) SEARCH ENGINE OPTIMISATION

* Tanya Chouksey, PG M.Tech student, Department of Computer Science and Engineering, Lakshmi Narain College of Technology, Indore, tanyachouksey.02@gmail.com

** Pawan Patidar, Assistant Professor, Department of Computer Science and Engineering, Lakshmi Narain College of Technology, Indore

I. INTRODUCTION

A search engine is an information retrieval system designed to help find information stored on a computer system for the user in the most relevant way. Some of the famous search engines are GOOGLE, YAHOO, SIRI, and NET EXPLORER. The search results are usually presented in a list and are commonly called hits. The search engine is evolving to counter this challenge. A web search engine or internet search engine is a software system that is designed to carry out a web search, which means to search the World Wide Web in a systematic way for particular information specified in a web search query. The search results are generally presented in a line of results, often referred to as search engine results pages (SERPs). The information may be a mix of webpages, images, videos, infographics, articles, research papers and other types of files.

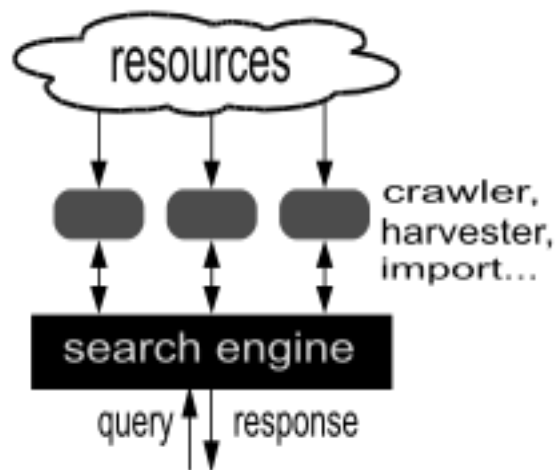


FIG 1: Index-based search engine

A Web search engine contains indexing, storage, query processing, spider (or crawler, robot), and user interface subsystems. The indexing subsystem aims to capture the information content of Web pages by using their words. During indexing, frequent words (that, the, this, etc.), known as stop words, may be eliminated since such words usually have no information value. Various statistics about words (e.g., number of occurrences in the individual pages or in all of the indexed Web pages) are usually stored in an inverted file structure. This organization is used during query processing to rank the pages according to their relevance scores for a given query. Hyperlink structure information about Web pages is also used for page ranking (Brin & Page, 1998; Kobayashi & Takeda, 2000).

For assessing the performance of search engines there are various measures such as **database coverage, query response time, user effort, and retrieval effectiveness.**

II. THE FEATURES OF WEB SEARCH ENGINE

Crawling:

The crawler, or web spider, is a vital software component that essentially sorts through the Internet to find website addresses and the contents of a website for storage in the search engine database. This allows the search engine to find current content on an hourly basis

Indexing:

Once the search engine has crawled the contents of the Internet, it indexes that content based on the occurrence of keyword phrases in each individual website.

Ranking:

The search engine after indexing gives the rank to the pages which have been searched the most as the top position.

Query understanding:

Maybe the query might have misspelled. Maybe the query way is too expensive(not precise) webpages before returning to the user, you want to clean the query and understand it.

Storage:

Storing web content within the database of the search engine is essential for fast and easy searching. The amount of content available to the user is dependent on the amount of storage space available.

The **motivation** of this project is based on the fact that identifying the most effective Web search engines satisfying the current information-needs is important both at a personal and a business level. Among others, this information can be used for

- (a) Reduce the searching time of the web search engine for specific pages.
- (b) Highly customizable
- (c) Access to retrieve information even the internet is OFF.
- (d) lightweight application(required less space to Download, Upload and access the webpages searched.)
- (e) Learn individual behavior of search to find relevant information.

III. COMPARISON

Features	Faster search engine	Other search engine
Access time	Access time is less for e.g. (3sec 235)	Access time is more for e.g.(4 sec 23)
Customisable	Highly customisable	Less customisable
Access even internet is OFF	Yes	Not

Locally stored memory	yes	Not always
------------------------------	-----	------------

IV. ADVANTAGES OF FASTER SEARCH ENGINE

Sr.n	FEATURES	EXPLANATION
0		
1.	SECURITY	Data is secure because no one can externally delete the history. Until mobile had reset or format memory.
2.	APPEARANCE	Clear appearance. Not too much fancy background.
3.	ADVERTISEMENT	Does not show advertisement, so that user does not irritate when searching anything.
4.	SEARCH RESULT	More accurate and relevant search results.
5.	QUALITY	The quality of the search result is good.

V. LIMITATIONS

Monolingual search: the search results can only be retrieved if the information is present in the query language. However, the research to handle those issues is also being conducted and is usually known as cross-language information retrieval.

History update: since once searched it will save in the stored database. There is a problem of updating once a data is searched it will be stored in the same way it has been searched.

Administration problem: this app does not require administration permission that's why this is not able to access all the stored information properly.

Background process: it will support background process but sometimes create a problem to run behind process under different network so need child thread to access it.

No voice search capability: this application is not using the voice of ours to do searching it just accepting the keywords.

VI. EVALUATION APPROACHES

Evaluation involves accessing the performance of search engines to improve their effectiveness. Evaluation is a continuous process of investigating the new approach of study, appraisal, and improvement of the search engine:

It consists of some Evaluation parameters:

RESULT OF FOLLOWING EVALUATION:

Effectiveness: It measures the ability of the search engine to find the right information and the desired result. It will use the restored data of the user and then give the detailed back to the user so highly effective.

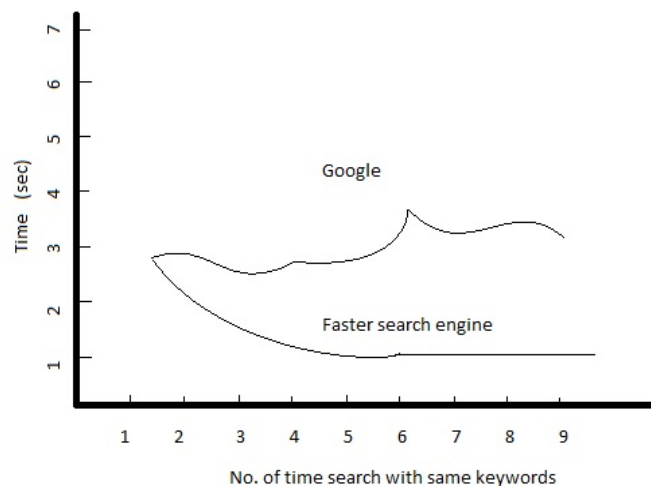


FIG: ACCESS TIME ACCORDING TO INTERNET SPEED

This graph shows the relationship between the total no of time required for the first search and for the same result of how much time is required.

Efficiency: It measures how quickly the right information is retrieved. It is defined in terms of time and space requirements.

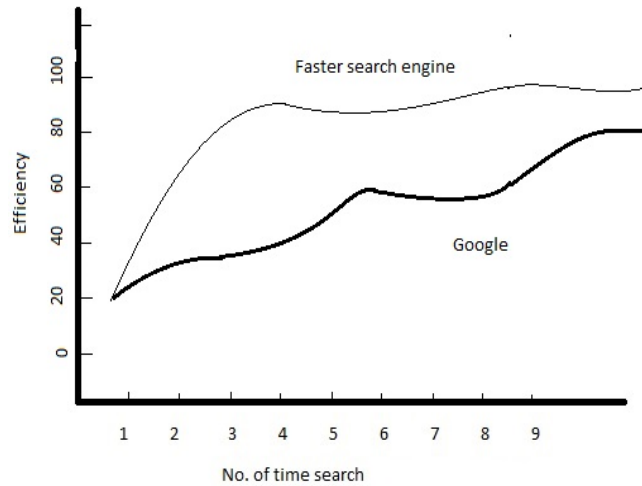


FIG: EFFICIENCY ACCORDING TO DIFFRENT RELEVANT KEYWORDS

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 it will show more précised data by using the restored data searches result.

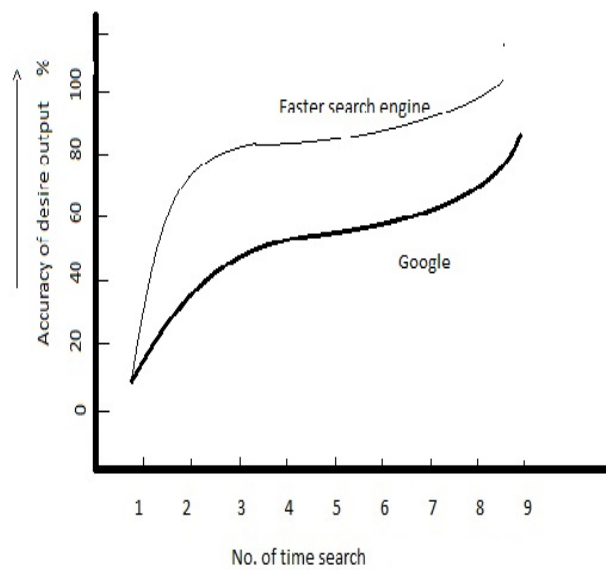
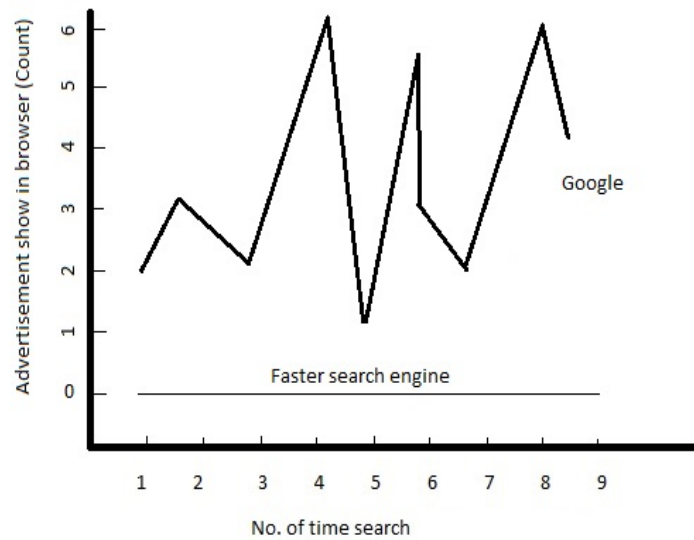


FIG: PRECISION OF DESIRED SEARCH RESULT

ADVERTISEMENT: The faster search engine does not show advertisements in between the accessing of information whereas Google shows a lot of advertisements in between.



VII. RESULT

We have tried to create a web search engine which will not only overcome these problems but also have some special features as:

Scheduled re-indexing:

Each time your websites change the free find spider will visit your site and build an index of your pages.

You can schedule re-indexing daily, weekly, monthly, or ask for immediate reindexing at any time.

Predictive: SIRI and APPLE spotlight is easy for people to quickly find applications, emails, and contact.

Personal space: So that the user can store the previously searched data. They did not need to search for the previous search data. It will automatically store in the user internal space storage.

Auto filing of information: From the personal stored memory data it can auto-filled the information.

Alert: In this project, it will automatically generate an alert regarding there completion time and important dates to an individual account.

Process bar: We will get to know how much project is completed in how much time.

On-demand services: When needed it will automatically be accessed from anywhere if you are not having the internet it will then also search the previously searched data.

Location-based: The project will provide us location-based services. It will track our initial location as well as the current location.

No fixed page limit:

This website indexes sites of all sizes. Many of our sites have in excess of fifty thousand pages. If you have a big, high traffic site you've come to the right place.

User's intent: The ability of the searched engine to more correctly determine a user search intent will serve to increase advertising revenue further.

Super-fast setup: free find is the fastest way to add search capabilities to your website. All you do is enter your email and website address in the form below and press the sign-up button.

Internet connectivity: Due to some technical problem or if you do not have the internet connection then no need to worry you can even access the previously searched data without the internet connection.

Low access time: since we have searched the previously visited page once again it will need less time as you can in the app that first time searched data need more than **5 sec(GOOGLE)** to do the searching and when again visited **take 2 sec (FASTERSEARCHING)** time to access data.

Highly customizable:

We do not need to have any extra application to install on the computer for any of the above features. It will automatically work according to the available information. It will integrate free find seamlessly with your website by customizing the search results page to match your website easy to use free find wizards give you complete control over the look of your search results including your own titles, text, logo, background and more.

Database: When we spend some time in this search engine it automatically learn about searched information and stored it in the local user database.

HIGHLY SECURE:

This website is highly secure because without administration permission nobody can delete the stored website pages. Since it is locally stored the user has to go to the ram to delete the data.

Once you delete the application then also the saved webpages will not be deleted.

Another possibility is having classrooms connected through satellite communication so that classes for similar courses could be streamed live to be shown to students in distant places with access to question and answers just like video conferencing sessions.

The rural infrastructure needs to be made ICT ready at the Panchayat level and the government needs to bring in a national policy through Public Private Partnership or some other route. With the most of the rural workforce being school drop-out, ICT enabled vocational education at the village level can make them employment ready.

VIII. CONCLUSION

TAKING THE STEP FROM TEN BLUE LINKS TO “POSITION ZERO”:

Customer attention spans are dropping to seconds. This is a behavioral change that will underpin all future developments in the search space. As the customer's appetite for knowledge grows, we provide more information and the technology to allow for a faster, more informal way to consume this content.

As users seek to click less, the number of featured snippets showing for queries is rapidly increasing and there is an ever growing number of rich result formats being launched into the wild.

These not only provide answers directly to searchers but in many cases, offer a similar experience to the apps that search engines are losing out to. And yet, for some, “build a responsive site”, pre-mobilegeddon.

Unlike mobile, optimizing for answer queries is difficult. To start with, the latest stats put the number of new, unique queries seen every day at around 15%. The optimist will say this is an opportunity and should be a key area of focus for growth. However, the realist will rightly ask how you can create a response to a query that does not yet exist.

REFERENCES

1. References Abbate, J. (1999). *Inventing the Internet*. Massachusetts: The MIT Press.
2. An Overview of search engine evaluation strategies by Shikha Goel Ghaziabad, India.
3. Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: ACM Press.

4. Bar-Ilan, J. (2002). Methods for measuring search engine performance over time. *Journal of the American Society for Information Science and Technology*, 53(4), 308–319.
5. Brin, S., & Page, L. (1998). The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117.
6. Chowdhury, A., & Soboroff, I. (2002). Automatic evaluation of World Wide Web search services. In *Proceedings of the ACM SIGIR conference*, vol. 25 (pp. 421–422).
7. Gordon, M., & Pathak, P. (1999). Finding information on the World Wide Web: the retrieval effectiveness of search engines. *Information Processing and Management*, 35(2), 141–180.
8. Hawking, D., Craswel, N., Bailey, P., & Griffiths, K. (2001). Measuring search engine quality. *Information Retrieval*, 4(1), 33–59.
9. Hull, D. (1993). Using statistical testing in the evaluation of retrieval experiments. In *Proceedings of the ACM SIGIR conference*, vol. 16 (pp. 329–338).
10. Jansen, B. J., & Pooch, U. (2001). A review of Web searching and a framework for future research. *Journal of the American Society for Information Science and Technology*, 52(3), 235–246.
11. Kobayashi, M., & Takeda, K. (2000). Information retrieval on the Web. *ACM Computing Surveys*, 32(2), 144–173. Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the Web. *Nature*, 400, 107–109.
12. Meng, W., Yu, C., & Liu, K.-L. (2002). Building efficient and effective metasearch engines. *ACM Computing Surveys*, 34(1), 48–89.
13. Mowshowitz, A., & Kawaguchi, A. (2002). Assessing bias in search engines. *Information Processing and Management*, 35(2), 141–156.
14. Salton, G., & Buckley, C. (1988). Term weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), 513–523.
15. Salton, G., & McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill. Soboroff, I., Nicholas, C., & Cahan, P. (2001). Ranking retrieval systems without relevance judgments. In *Proceedings of the ACM SIGIR conference*, vol. 24 (pp. 66–73).
16. Spink, A., Dietmar, W., Jansen, B. J., & Saracevic, T. (2001). Searching the Web: the public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226–234.
17. Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From E-sex to E-commerce: Web search changes. *Computer*, 35(3), 107–109.
18. Voorhees, E. M. (2000). Variations in relevance judgments and the measurement of retrieval effectiveness. *Information Processing and Management*, 36(5), 697–716.