

An Exploratory Study to Evaluate the Working of Different Search Engines

^[1] Jyothi S, ^[2] Shanthi S

^[1] PG Student ^[2] Assistant Professor (SLG)

Kongu Engineering College, Perundurai, TamilNadu, India

^[1] jyothishivakumar.mca@gmail.com

Abstract: Today search engines are becoming necessity of most of the people in day to day life for navigation on internet or for finding anything. Search engine answer millions of queries every day. With the initiation of the Web and search engines, online searching has become a common method for obtaining information. People search for information on a daily basis, for various purposes: for professional tasks, or for non-professional tasks[4]. Today, there so many search engines available on the internet globe like Google, Yahoo, Bing, MSN, etc., each competing to provide the best recall results.

Here, a comparative study is made on the different search engines like Google and Yahoo to analyse their precision rates, to predict the quality of the retrieved results[1]. The information provided after the search query can be categorized in different ways: relevant, irrelevant, sites that cannot be accessed, and links. The results are provided based on the type of queries posted, for popular queries the engine works even faster by providing the suggestions to the users. Different search engines use different techniques to provide the results and these techniques are studied and analysed. Here, an attempt is made to provide the suggestion regarding the effective search.

Key words: Search Engines, Google, Yahoo, Search, Information Retrieval, Suggestion.

I. INTRODUCTION

Search engine is a software program that is designed to search for the information on the World Wide Web. It can also be defined as a type of website that helps in locating the information on the web. Search engines support multiple languages, in some of the search engines these language features remain at the lexical level.

The main aspect the user looks in a search engine is the speed. The other aspect is the accuracy of the information retrieved. The information can be categorised as follows:

- Relevant
- Irrelevant
- Links

Relevant information is the actual result expected by the user. Irrelevant information refers to the unrelated data to the query given by the user. Some retrieval provides only the link access to the related information and they are referred to as Links.

The web contains a huge amount of data which has to be filtered and then presented to the user according to the query they pass. This quality of the search engine in filtering the data and searching the right information correctly is referred to as Precision Value[1]. The Precision rates vary based on search engines, the algorithms they

use, based on the database size, etc.,. It also affects the type of information retrieved.

The search engines can be classified into three different categories [2]:

- Crawler Supported – Ex: Google
- Human Driven – Ex: Look Smart
- Hybrid – Ex: Yahoo

Crawler supported search engines implement crawling technique to search and retrieve the information. Human driven search engines facilitate the manual input of the web pages. Hybrid search engines are the combination of both the above types.

II. WORKING OF SEARCH ENGINES

Search engine is a very complex program which is a key to find the specific information on WWW. Basically, a search engine represents an answer machine which finds the answers to all the queries posted by the users.

The working of a search engine involves three main stages [4]:

- Crawling
- Indexing
- Information Retrieval

STAGE I - CRAWLING

Crawlers are automated software agents that scans the entire website, gets the complete list of everything like page title, images, keywords, links, etc.,. It returns all the information to the Central Pool where the data is indexed. Crawlers periodically check the site for the updates in information.

Special software robots called Spiders are employed by the search engines to find the information on the web pages. It builds the list of words found on the websites. When the spider is building the list, the process is referred to as Crawling. The crawling always starts at the popular site.

The crawling is the stage where the content is actually discovered from the website. It involves the searching of database, spinning the records based on the queries, etc.,.The spider begins the search from a popular site and move on with indexing the site and the words, phrases found on it, following the every next link available on the site.

STAGE II - INDEXING

Indexing is a process where the content obtained from crawling is analysed and placed into the huge repository of databases. The main purpose of indexing is to make the information available to the users as quick as possible. The search engine after taking a screen shot of the data for the spiders to analyse, places the same into the huge databases. Indexing helps in knowing how many times a particular word was searched, how many time a particular link was visited, etc.,. It stores a single word or couple of words into the database and provides a link to the site.

The search engine assigns a unique value to the words contained in the list. These values are stored into the index of the search engines which helps in information retrieval. The most effective way to create these unique values is known as Hashing. The hashing function helps in generating different key values and is thus allocated to the words.

The hashing technique involves the creation of a hash table where the hash keys are stored. This also contains the word and the unique value attached to it. This helps in the next stage of working of search engine i.e., the information retrieval.

STAGE III - INFORMATION RETRIEVAL

After the website has been crawled and the content is retrieved, the indexing of the pages or words is done to classify the information. The next stage is the retrieval of information where the user enters the query and the respective result is displayed on the search screen.

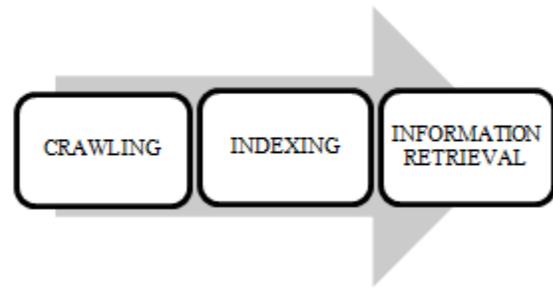


Fig 1: Working of Search Engine

III. PURPOSE AND SCOPE OF STUDY

Searching is one of the most common tool for seeking information over the network. It helps us in knowing the information according to our needs within minimum amount of time. They are the resources that help the users to search any kind of information on the network. It is a simple and easy way to get to know the information and to learn the things.

The purpose of this study is to know how the search engines work, how one search engine differs from the other in retrieving the information, what is the precision rate of different search engines. This is achieved by providing the search engines like Google, Yahoo, and Look Smart. Here the comparison is made between the search engines based on its classification types.

A comparative study is made on the different search engines to analyse their precision rates, to predict the quality of the retrieved results. Different search engines use different techniques to provide the results and these techniques are studied and analysed. Here, an attempt is made to provide the suggestion regarding the effective search.

IV. CASE STUDY

Google, Yahoo and Look Smart are the three different search engines considered for this study. Few queries were selected for the study of these search engines. The queries used were Single-word queries. The queries included simple technical computer related terms.

In this study, the information retrieved by these three engines is classified into three categories:

- Relevant (R)
- Irrelevant (I)
- Links (L)

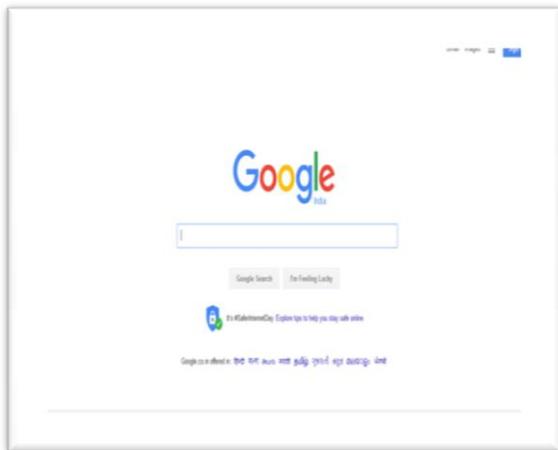


Fig 2: Google home page

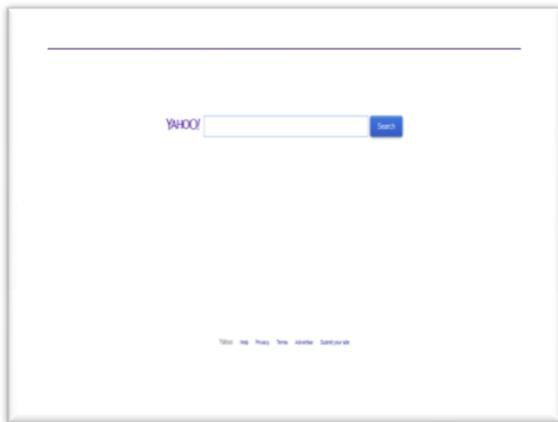


Fig 3: Yahoo home page

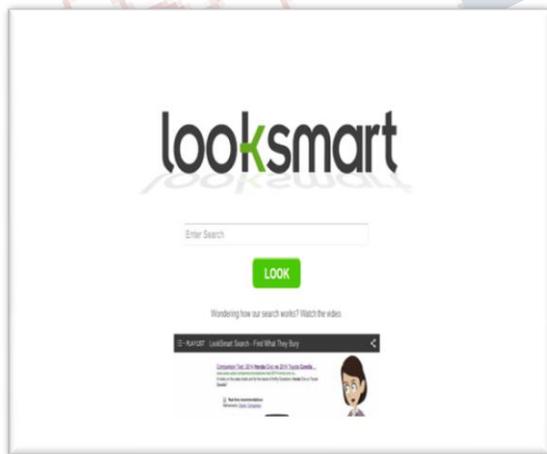


Fig 4: LookSmart home page

V. RESULT

The different queries were put across the three different search engines which provided different results. The results were classified based upon the timings, the first 4-5 results displayed on the search screen.

Crawler supported Google search engine provided the results in a combination of Relevant and Links for the queries given where few results had only relevant and the others had both. The Hybrid Yahoo search engine had a mixture of all the three categories of information namely, Relevant, Irrelevant and Links. The human driven Search Engine LookSmart provided a combination of Relevant and Links for all the queries placed.

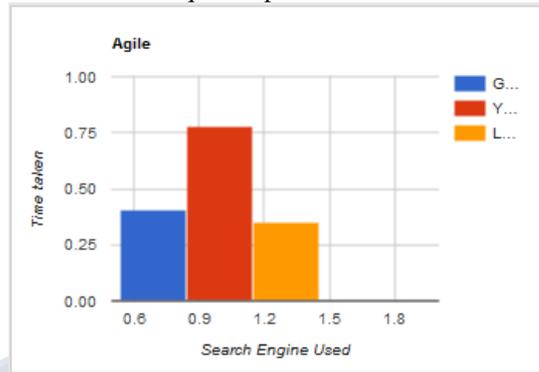


Fig 5: Result for "Agile" term



Fig 6: Result for "Linux" term

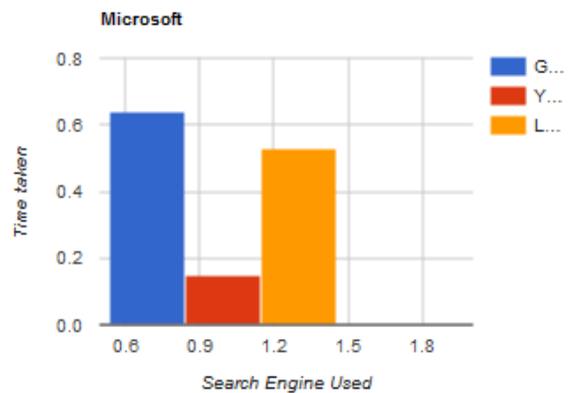


Fig 7: Result for "Microsoft" term

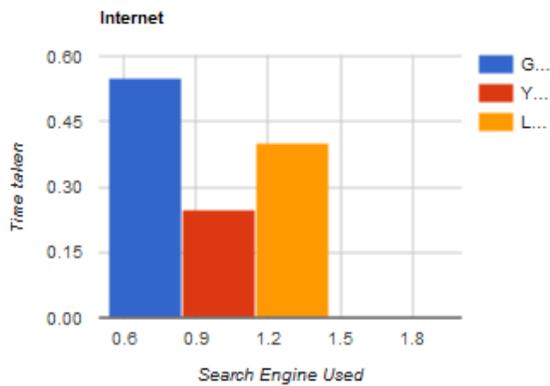


Fig 8: Result for "Internet" term

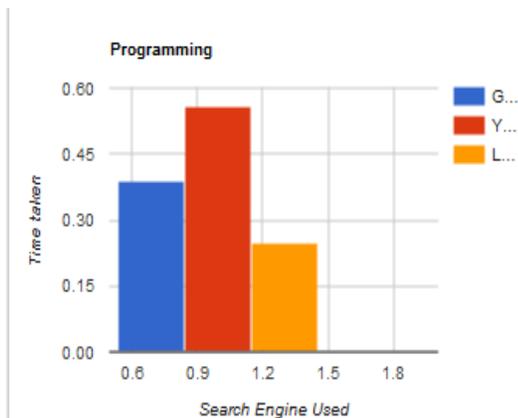


Fig 9: Result for "Programming" term

VI. CONCLUSION

Nowadays there is a great demand for the development of the Search Engines that provide Effective and Accurate results. The gigantic size of the web, the vast variety of users' on the internet, the needs and interests of the people has made the growth of internet and search engines remarkable.

Every search engine follows their own algorithms and methodologies to provide the results to the queries. Google can be said as the most effective and time saving search engine because of the time it takes to provide the results, the quality of the links provided is most of the time Relevant.

For people who do not have an idea of the exact information to be searched can use the Human driven search engines like LookSmart. LookSmart provides with different links that is relevant to the query which will be helpful to the user to traverse the web. Yahoo search engine is a moderate type of engine which provides the mixture of results to the users.

REFERENCES

- [1] Kailash Kumar, VinodBhadu, A COMPARATIVE STUDY OF BYG SEARCH ENGINES, 2013, American Journal of Engineering Research (AJER)
- [2] NagarajuMamillapally, TrivikramMulukutla, Impact of Meta search Engines on Information Retrieval Effectiveness:A Case Study, 2013, International Journal of Advanced Trends in Computer Science and Engineering
- [3] Krishan Kant Lavania, Sapna Jain, Madhur Kumar Gupta, and Nicy Sharma, Google: A Case Study (Web Searching and Crawling), 2013, International Journal of Computer Theory and Engineering.
- [4] Sushma Malik,A Comparative Study of two major Search Engines – Google Yahoo .<http://computerscijournal.org/?p=706>
- [5] Dirk Lewandowski, New perspectives on Web Search Engine Research, 2012, Bingley: Emerald Group Publishing.
- [6] Amanda Spink, A User centered approach to evaluating human interaction with web search engines: An Exploratory Study, <http://eprints.qut.edu.au>
- [7] Jan Brophy and David Bawden, Is Google Enough? Comparison of an internet search engine search with academic library resources, 2005, Emerald Group Publishing Limited.