AN EXPLORATORY STUDY OF KEYWORD BASED SEARCH RESULTS

SHILPA S. LADDHA[a1] AND PRADIP M. JAWANDHIYA[b]

[a]Government College of Engineering, Aurangabad, India
[b]P. L. Institute of Technology & Management, Buldana, India

ABSTRACT

In this technical Era, Life without Internet is beyond imagination. Internet, which connects billions of people all around the world, is the fastest, easiest and most economical medium of communication. Internet provides many services like electronic-mail, E-Commerce, social networking, weather fore-casting etc. World Wide Web (WWW) is the biggest repository of information. Everyone is having web on their gadgets like laptops, mobile, tabs, desktops etc., in order to access the information on the web due to its friendly user interface and ample features. With the rapid development of WWW, search engines have become the main tool for people to search the web information. However, retrieval of relevant information is still a challenging task because the information in WWW is unstructured and human readable. Most of the keyword search engines get the answers syntactically correct but larger in amount. This paper presents analysis of the results of keyword based Search engine Google. This Paper analysis emphasizes need of a smart SEMANTIC information retrieval system enhancing the performance of keyword based search engine in terms of precision and recall.

KEYWORDS: E-Commerce, Social Networking, WWW, SEMANTIC.

This paper presents analysis of the results of keyword based Search engine-Google. Section A discussed the brief idea about the Search Engine using popular keyword based search engine Google. Section B presents the Literature survey and analyzes the popularity of Google Search Engine. Section C describes the working of Search Engine which is a three step process of Crawling, Indexing and Information Retrieval. Section D points out the advantages of Google Search Engine. Section E analyzes the performance of Google Search Engine using sample Queries in tourism domain where the objectives of the analysis is discussed. Section F presents the Google Search engine results of the sample queries in the tabular and the graphical form. And Finally the Paper concludes which emphasizes the need of a smart information retrieval system enhancing the performance of keyword based search engine in terms of precision and recall.

SEARCH ENGINE (SE)

A search engine is a software to search the information on the WWW. Keyword based SE searches documents and files with keywords given by the user and render list of the web sources with the keywords. It is a special type of program on the web that renders the requested information to the user from the huge repository WWW. The user inputs the keywords or queries into the SE interface, SE after processing displays the information related to the query on the user screen within no time called as SE results page which is a record of matching web pages with Meta information. Meta information comprises of title, Meta data and link for every matching web page. The user can use this Meta information to access information on the website by clicking on the link. Search Engine make use of especially designed ranking algorithms to rank the web pages based on the popularity or best content quality and accordingly the web page order in the result page will be decided. In this way SEs are very valuable to retrieve any information easily within no time. Using appropriate keywords the search results can be improved.

Google SE is the most widely used SE on the WWW owned by Google Inc., receives more than million queries per seconds all around the world. Google SE was originally designed in 1997 by Sergey Brin and Larry Page with many exceptional features. These include synonyms, sports scores, time zones, stock quotes, maps, weather forecasts etc.

LITERATURE REVIEW

Our study results indicate that the majority of people around the world prefer Google interface. It was found that the response time of SE is one of the most significant parameter for the popularity of any SE. According to Maxymuk, the web is a very important and productive layer where new paradigms frequently lead to different types of innovations. With the rapid growth of Google, others determinedly seem less popular. Lewandowski, did the Comparison of the performance of the popular search engines like Google, Yahoo!, and MSN. Large, said the WWW renders access to information on the web in multiple languages. Some designs make possible multilingual exploitation of the resources available on the web. Some SEs, for example, restrict the user to retrieve information in specific languages; some provides the end user with the

[1]Corresponding author

flexible interface in a choosing the language. Many popular web sites also provides their information in multiple languages, typically is English. According to Arnold, locating the required digital item on the web is very much challenging job. Google and other SE companies are becoming application platforms. Jamali, research result shows that Google is popular for problem-oriented information searching. The results also prove the increasing trust of scientists on general SEs, specifically Google, for finding academic articles.

WORKING OF SEARCH ENGINE

WWW is the biggest repository of information. To locate the required information on the web, the SE plays the key role. Without search engines, locating anything on the Web is challenging job, unless we know the exact URL on which the information is present.The three step search engine information retrieval process is as shown in Figure 1 and discussed below:

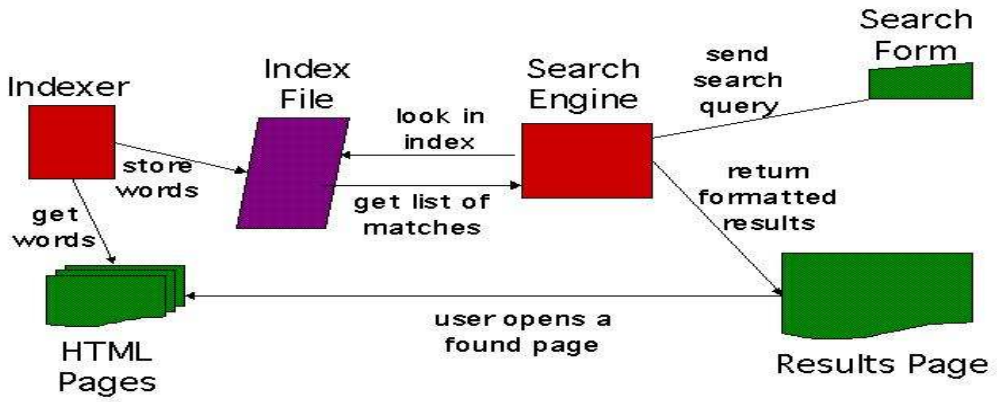

Figure 1: Information Retrieval Process of a Search Engine

Crawling

Crawling, is the most delicate application used for interacting with hundreds of thousands of web servers and various name servers which are all beyond the scope of the system and retrieving a list of everything over there the page title, keywords , images, other pages it links etc. The crawler then provides all the collected data for indexing to a central repository. In order to cover the hundreds of millions of web pages, the crawler will revisit these sites at regular interval to check for any new information. This interval is decided at the administrator level. To locate required information on the hundreds of millions of Web resources, a SE makes use of special software robots, called spiders, for constructing lists of the words found on different URLs. The process of constructing the lists by the spider is called Web crawling. In order to construct and retain useful list of words, a SEs spiders have to visit lots of pages. The question may arise that about the spiders starting point over the Web. The normal starting points are lists of very popular pages and heavily used servers. The spider usually starts its traversal from a popular site, index the words on its pages and repeat this process for every link found within the website. Google has a fast distributed crawling system.

Indexing

Indexing is the process of analyzing the crawled contents and keep in huge repository. In this, SEs uses the information that spiders can view and understand and maintain it in huge repository in a suitable manner helpful for information retrieval. The simplest SE just store the keyword and the URL with its source where it was found. The popular SEs maintain more information along with the word and URL to render more relevant results by assigning weight to each entry by using some mathematical formula. This is one of the key reasons that a search for the same keyword will render different search results on different search engines, with the search pages presented in different order to the end user. Indexing makes the information retrieval faster. There are various methods to create the index out of which the most efficiently used method to create index is hashing in which a hash table is created using a formula to numerical value which is attached with each word.

Retrieve Information

User input the query in the SE interface which in turn matches the indexed content with users search query or keywords and retrieve the most relevant and reliable list of records from their repository.

The goal of searching is to provide quality search results efficiently. Many of the large commercial search engines seemed to have made great progress in terms of efficiency. Therefore, we have focused more on quality of search in our research. Google maintains much more information about web documents than typical search engines.

Advantages of Google

It helps the user to search proficiently by providing appropriate guidelines. The most significant guideline is to use particular words to illustrate what we are searching. When typing the keywords, Google's suggestion for the full key words is very much useful. Because if the query comprises of vague words, the success rate of getting the relevant result will be reduced. It facilitates many types for the search results, like web, images, videos, shopping, news etc. With the selection of specific type, a user can get specific results on the webs. The search results show the title in the first line, the Meta information of webpage in the next several lines, and the URL. This provides sufficient information to the user to rapidly locate and find information from other web pages in the ranked results. This proves the efficient working of search engines. Google on its URL, provides the very simple search interface, the Google trademark picture and the search types understandable by any common man. Google provides relevant results quickly and precisely than other popular SEs.

Performance Analysis using Sample Queries

In this technical era, life is very much dependent on the internet to get all type of information very quickly all over the world. The Web has significantly changed the way of locating the required information from the large amount of information available on the web. User friendly interface to access the SE plays the key role for analyzing the best SEs. So, there is a need to analyses the relevancy of the result provided by the SE. For analyzing the performance of Google SEs, we conducted the result analysis by providing around 57 queries related to tourism domain to the Google SE.

Objective of the Analysis: This analysis is conducted to know the: Time taken by search engine to execute the query, total number of results rendered by the search engine to execute the query, number of links displayed on first page, number of irrelevant links displayed on first page, percentage of irrelevant results, number of relevant links displayed on first page, percentage of relevant results. Analysis results will be used to find out the performance of Google search engine with respect to sample queries in the tourism domain. The analysis is as shown in Table 1, Fig.2, Fig. 3 and Fig. 4.

RESULTS

Table 1: Result analysis

Sr. No	Domain Area	Query		Google						
				Ex. Time (second)	Total Results	Page 1 links	Irrelevant links	% of Irrelevant Results	Direct Results	% of Direct Results
1	Bus Service	Q1	bus from delhi to Shimla	0.76	4,89,000	17	5	29.4117647	0	100
		Q2	bus for pune to mumbai	0.8	9,96,000	17	3	17.6470588	0	100
		Q3	bus between aurangabad to nagpur	0.62	5,86,000	17	3	17.6470588	0	100
		Q4	bus between bhopal and indore	0.53	5,84,000	17	3	17.6470588	0	100
2	How to reach	Q5	how to reach from aurangabad to mumbai	0.42	5,20,000	18	3	16.6666667	0.5	50
		Q6	how to reach jaipur	0.54	7,43,000	12	0	0	0.5	50
		Q7	how to reach aurangabad maharashtra	0.59	4,81,000	16	2	12.5	0.5	50
		Q8	How to travel manali	0.62	5,42,000	17	4	23.5294118	0	100
		Q9	tourist places in	NA	1170000	12	0	0	1	100

			kerala		0					
3	Tourist Spots	Q 10	visiting spots in agra	NA	440000	13	2	15.3846154	1	100
		Q 11	find visiting places in kolkata region	0.67	16,20,000	14	3	21.4285714	0	100
		Q 12	famous places in mysore karnataka	NA	553000	12	3	25	1	100
4	Train	Q 13	train from mumbai to pune	0.7	7,54,000	10	0	0	0	100
		Q 14	train bhopal to jabalpur	0.6	4,14,000	10	0	0	0	100
		Q 15	train kolkata to darjeeling	1.08	4,15,000	11	2	18.1818182	0	100
		Q 16	train between aurangabad to nagpur	0.65	4,55,000	10	0	0	0	100
5	Weather	Q 17	weather pune	0.59	7,18,000	11	1	9.09090909	1	100
		Q 18	today weather report in mumbai	0.46	6,55,000	17	4	23.5294118	1	100
		Q 19	today palakkad weather in kerala	0.47	1,86,000	15	4	26.6666667	1	100
		Q 20	surat weather	0.44	7,01,000	14	3	21.4285714	1	100
6	Distance	Q 21	indore to jabalpur km	1.04	6,18,000	10	5	50	1	100
		Q 22	km from agra to shimla	0.74	5,00,000	10	2	20	1	100
		Q 23	distance from nagpur to aurangabad	0.8	4,81,000	11	1	9.09090909	1	100
		Q 24	show distance between kolkata and darjeeling	0.97	3,25,000	13	4	30.7692308	1	100
7	CityState	Q 25	display cities of goa	0.69	4,23,000	16	6	37.5	0	0
		Q 26	get cities of manipur	0.6	2,00,00,000	17	6	35.2941176	0	0
		Q 27	state of kullu	0.48	4,47,000	13	3	23.0769231	0	0
		Q 28	agra in which state	0.61	1,03,00,000	16	4	25	1	100
8	State List	Q 29	indian state	0.45	98,90,00,000	17	5	29.4117647	0.5	50
		Q 30	get states of bharat	NA	768000	11	3	27.2727273	1	100
		Q 31	find states	0.56	1,93,00,00,000	11	11	100	0	0
		Q 32	states	0.48	4,16,00,00,000	9	7	77.7777778	0.5	50
9	About City	Q 33	about pune	0.47	42200000	10	0	0	1	100
		Q 34	abt shimla	0.67	1030000	10	1	10	1	100
		Q 35	abut nashik	0.59	1320000	10	0	0	1	100
		Q 36	ABOUT ooty	0.6	872000	12	3	25	1	100
10	Best Time Visit	Q 37	best time to visit shimla	0.43	422000	10	1	10	0	0
		Q 38	Best Season To Visit srinagar Place	0.65	523000	14	4	28.5714286	0	0
		Q	Best Time To	0.49	663000	10	0	0	0	0

		39	Visit coimbatore							
		Q 40	best season to visit kolkata place	0.67	16400000	10	0	0	0	0
11	Flight	Q 41	Flight from delhi to mumbai	0.54	12700000	15	0	0	0.5	50
		Q 42	flight from delhi to srinagar	0.52	446000	14	0	0	0.5	50
		Q 43	Flight from lucknow to guwahati	0.5	448000	14	0	0	0.5	50
		Q 44	flight from mumbai to bhopal	0.3	486000	15	0	0	0.5	50
12	Hotel	Q 45	hotels in coimbatore	0.66	1570000	17	0	0	1	100
		Q 46	find list of hotels at trivandrum	0.77	540000	17	2	11.7647059	1	100
		Q 47	accommodation in srinagar	0.59	718000	17	3	17.6470588	1	100
		Q 48	resort in bhopal	0.56	553000	17	4	23.5294118	1	100
		Q 49	guest house in ooty	0.45	516000	17	0	0	1	100
13	Hotel Type	Q 50	cheap hotel at Nagpur	0.66	489000	17	2	11.7647059	1	100
		Q 51	list of cheap hotels in Ahmedabad	0.83	543000	17	2	11.7647059	1	100
		Q 52	find budget hotels of pune	0.76	25800000	17	1	5.88235294	1	100
		Q 53	budget hotel in Chennai	0.55	674000	17	1	5.88235294	1	100
14	Hotel Ratings	Q 54	four star hotels in Amritsar	0.68	526000	17	0	0	1	100
		Q 55	find 3 star hotel in mangalore	0.61	508000	17	1	5.88235294	1	100
		Q 56	2 star hotels in maharashtra Mumbai	1.01	19300000	17	0	0	0	0
		Q 57	five star accommodation in indore	0.9	292000	17	2	11.7647059	0	0

Figure 2: Time Required for the Execution of the Query

Figure 3: Graphical Analysis of Irrelevant Results

Figure 4: Graphical Analysis of Relevant Results

CONCLUSION

Search Engine is one of the commonly used popular tools for searching the information on WWW. SE's provides the re-sources that help the end user to search any sort of information on the web in a simple and convenient way. The purpose of this study is to presents the extensive review and analysis of most popular search engine Google .From our analysis of the results provided by Google SE, we conclude that Google is providing the best results till date. People like to search information on Google as it provides user friendly interfaces. It performs better than the other search engines but still the scope is there to design smart and intelligent SEMANTIC SE to improve the performance by providing more direct, relevant and precise results in less time. In the future, our work will focus on the deeper and broader research in the field of intelligent semantic search, with the purpose of concluding the current situation of the field and promote the further development of intelligent semantic search engine technologies.

REFERENCES

Maxymuk J., 2008. "No search limits", Bottom Line: Managing Library Finances, the, **21**(4):132–134.

Lewandowski D., 2011. "The retrieval effectiveness of search engines on navigational queries", Aslib Proceedings, **63**(4):354–363.

Large A. and Moukdad H., 2000. "Multilingual access to web resources: an overview", Program: electronic library and information systems, **34**(1):43–58.

Arnold S., 2006. "Search: the new application platform", Electronic Library, **24**(2):121–125.

Jamali H. and Asadi S., 2010. "Google and the scholar: the role of Google in scientists' information-seeking behavior", Online Information Review, **34**(2):282–294.

Reprint of: The anatomy of a large-scale hypertextual web search engine Sergey Brin, Lawrence Page Computer Science Department, Stanford University, Stanford, CA 94305, USA.

http://en.wikipedia.org/wiki/Web_search_engine.

http://searchenginewatch.com/article/2065173/How-Search-Engines-Work.