A Comparative Approach to Search Engine Ranking Strategies

Dharminder Singh[1], Ashwani Sethi[2]
Guru Gobind Singh Collage of Engineering & Technology
Guru Kashi University Talwandi Sabo, Bathinda, Punjab[1, 2]
dharminder.17@gmail.com[1],ashwani.gku@gmail.com[2]

Abstract-- Internet is a collection of vast information. In the modern internet era everyone search on internet for some information. To explore the internet Search Engines are primary targets. Everyone uses Search Engine from find a name of website to secure information. This vast need of Search Engines makes them an important tool for this challenging task. Generally Search Engine user hits top few results in SERP. Low ranking web pages got less intention of users, so ranking a web page according to their relevant information and popularity is a major task for Search Engine. Search Engines have primary responsibility to search query related information and rank them according to relevant information. This two main task makes a Search Engine popular or most preferable tool to explore internet. On the basis of information retrieval and query handling, Search Engines have two types, Keyword Based Search and Semantic Search. Keyword search directly matches keywords with web pages against the user query and a web page which has most matched keyword density, results high rank in Search Engine results. On the other hand Semantic Search uses knowledge base to extend the user query and adds more words by itself to query. Since the Semantic Search not only depends on user query keywords but also depends on Knowledge Base which gives high quality page results. This research paper presents an analysis of both techniques based on user query keywords.

Keywords: Search Engine Optimization (SEO), Search Engine Results Page (SERP), Website Ranking, Keyword Density, Semantic Search, Keyword Search, Inverted Index.

I. INTRODUCTION

Search Engine plays an important role to extract information from internet in the form of ranked web pages result. Searched results top few results got hits form users and low ranked pages. Web page ranking can be improved with Search Engine Optimization, which helps in analysis of Search Engine rules and develop a web page according to Search Engine ranking rules. SEO is a practice which helps to improve web page's popularity as well as content, which easily recognized by Search Engine. But judge a web page's quality and popularity is totally a responsibility of Search Engine. Every Search Engine works with some searching algorithms and strategies. On the behalf of searching techniques Search Engine can be divided into two major categories: Keyword Based Search and Semantic Search. Keyword Based Search focuses on the web pages keywords and directly matched with query keywords. Before starting search on internet Search Engines register every web page with their personal database. Search Engines spider are located over the web and make searching of web pages on web easier [8]. Every Search Engine firstly parse the web pages, ignore all the stop words like 'the' 'is' 'am' 'are' etc. Every article lies between <page> and </page> tags, where the page id, title, and text is separated by the corresponding tags.

Interrogative words are removed from the query and page ranking is performed on the basis of main keywords of the query [11]. Every Search Engine performs following three tasks:

A. Inverted Index: After parsing all web pages documents Search Engine Create an inverted index [2] for all documents which speed up information retrieval. E.g. in figure 1 'computer' is a term for inverted index and corresponding '[1, [0, 5, 15]]' is an index for 'computer' term in inverted index database. Here 1 point to web page index and [0, 5, 15] points to term's index within web page. Finally search starts against inverted indexed database.

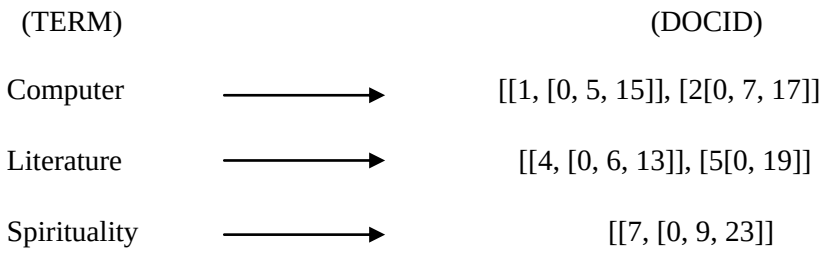

Figure 1: Inverted Index

Search Engines are mainly three types: Full Text, Meta and Directory Search Engine [4]. Search engine have two major searching techniques used to search web pages on the Internet and store them into Search Engine's personal database and ranking in SERP.

B. Query Analysis: User puts a query to search some type of information on the internet. Query works as a feed to Search Engine, and express users view for search. On the basis of query keywords query can be divided into three categories: Navigational, Transactional and Informational. Query analysis is processes which refine query with shorten keywords. Interrogative words are removed from the query and page ranking is performed on the basis of main keywords of the query [11].

C. Ranking Web Pages: Finally Search Engine rank a web page with some predefined policy, strategies. Ranking a web page may depend on several factors of web pages like Popularity, contents, structure of web pages etc. After search on personal database Search Engine finally rank web pages according to their content frequency matched with query keywords.

On the basis of query all Search Engines can be categorize as Keyword Based Search Engine or Semantic Search Engine. Both types of Search Engines follow all the upper steps but difference is based on the searching techniques.

Keyword Based Search: This search mainly focus on the query keywords and match theses query keywords directly with the web page keywords without any proper query analysis or knowledge. Search mainly points out the title tag, anchor text, page heading tag, Meta tag's keywords, anchor tag, ALT tag, URL. To optimize a web page in search result keywords density, distribution and keywords selection is more important. Every value in a field considered as a small text document that can be used for the Keyword Based Search [5]. Keyword matching process finds a set of tuple in database [7], which has matched keywords in user query. With the help of inverted index it is easy to determine matched keywords and their index in a webpage. A web page which has exact query matched keywords and has more strength of query related keywords gets high rank in search result. Because of Search Engine gives priority to keywords so Black Hat techniques [9]. Keyword Based Search Engines are Google, Yahoo etc.

Semantic Search: This is also called a search with meanings. The Semantic Search determines the content and contextual meaning of query keywords. The Semantic Search Engine interpreted extracting the relevant concept from the sentence [6]. A query also itself determines the user's view. Interrogative words are used to change the meaning of query like what, why, when etc. The Semantic Search Engines are an alternative to the Keyword Based Search Engines. The difference of Semantic Search Engines from conventional search engines is that the semantic search engines are meaning-based [12]. Semantic Search is combined with conventional keyword-based retrieval to achieve tolerance to knowledge base incompleteness [10]. The Semantic Search uses a knowledge base to extend the user query. A user uses shortcut keywords but want more results, so this is the Search Engine's responsibility to extend the user's query to find relevant results to user query. Knowledge base is a group of related terms which has similar type of related terms [1]. Any keyword which exists in knowledge base can be a part of user query indirectly. Knowledge base mostly may be in form of database with different related terms collection. Each term has pre assigned weight in a database. Term's weight further used to calculate the web page's weight for ranking in search results. The Semantic Search also retains the exact meaning of query and relate with results.

II. LITERATURE SURVEY

Zhou Hui *et al*. highlights in "Study on Website Search Engine Optimization", [4] the factor effecting searches ranking of search engine optimization. This paper presents the three factor Webpage correlation, Link Weight and Time Based Factors of SEO with three types of Search Engines Full Text Search Engine, Meta Search Engine and Directory Search Engine. According to this paper Keyword optimization can be done with Keyword Selection, Keyword Density and Keyword Distribution. A web page optimization can be described with both internal and external links which purely describe web page popularity. An effective way to optimize a web page is to submit it to Search Engine or submit to open directory library for future indexing.

Ping-Tsai Chung *et al."*A web server design using search engine optimization techniques for web intelligence for small organizations" [13] this research paper highlights the ideal model for developing web servers for small organizations by including search engine optimization techniques. As we now today is a globalization era, a new way of business and marketing is developed by the use of Internet web services. Search engine optimization is the process of improving the visibility of a website I the search engine result via "natural" search result. Small organization such as schools, banks, government agencies, libraries, retailers, restaurants, post offices could build their web services in higher quality for improving their business in surviving in today's competitive world.

S.G. Choudhary *et al. "*Semantic search algorithm based on page rank and ontology: A review" [10] **t**his research paper introduces some Semantic techniques and provides different algorithms based on page ranking and ontology. The algorithms based on Semantic Search are the following ways: PSSE: Personalized Semantic Search Engine use the user's profile with the ranking score and ontology for calculation of personalized factor which helps to get more personalized result. Google search engine use user profile to personalized search result for their location and on the basis of their previous search interest result. The architecture of PSSE has two parts: Offline and Online. The Offline part consists of crawling and preprocessing processes.

Shikha Goel *et al. "*Search Engine Evaluation Based on Page Level Keywords" [3] **t**his Paper highlights the approach of search engine evaluation which is based on page level keywords. Page level keywords are the keywords found in individual pages of website. Page level keyword is an impotent factor to measure the relevance of search engine results. A user create a query and search engine designer design the database for this query and later the queries are run by the users to calculate the page level keywords.

Phyo Thu Thu Khine *et al.* "Keyword Searching and Browsing System over Relational Database" [14] **t**his paper highlights the searching of keywords in relational database for increase the searching speed of desired keywords. A user doesn't need the knowledge of database schema or SQL. A user submit a list of keywords them system search for the relevant records and ranking them on their occurrence basis. Three systems DISCOVER, BANKS and DBXplorer user keyword based search over relational database. A query gives a set of keywords and finds rows in relation database for keywords. Indexing Relational Database: Indexing is used to speed up the retrieval of records. Indexing is useful when database has large number of Text fields. Each value in such a column considered as a small text documents that can be used for keyword based search. Query Cleaning: System takes a query as input and produces a 'clean' query output. This is achieved by filtering the stop words from query. These words are meaningless. So the result occur with them may not satisfy the user. Keyword matching: Once the cleaned query is produced, this system can match the keywords. System matches the query keywords with database tuples. A keyword matching algorithm may different for multiple keywords queries. Record Scoring: After query results, calculation of the score for each result is need. The record is determines which record is relevant to user query. This process is also called ranking process of documents in result.

Duygu Tumer *et al.* "An Empirical Study on Semantic Search Performance of Keyword-Based and Semantic Search Engine: Google, Yahoo, Msn and Hakia" [12] this paper analyzes the semantic search performance of search engines. This paper took three keyword based search engines Google, Yahoo, Msn and a Semantic search engine Hakia. Different queries of different topics analyze the performance of these search engines. Web search engines are computer programs which allow users to search their desired information from websites. The most popular search engines are Google, Yahoo, and Msn with 71.9, 71.7% and 4.2 volume of search ratio respectively. Hakia is the publicly available semantic search engine. This paper has a table with ten different types of queries. These queries were run on the both keyword-based search engine as well as semantic search engine. Keywords were used to replace phrase. Yahoo and Msn retrieved approximately 75.5%, 63% and 78% non-relevant documents respectively. Hakia retrieved 62.5% non-relevant documents

III. ANALYSIS

Both Keyword Based Search and Semantic Search deal with query keywords and web pages keywords. A Keyword is a key point in both searching techniques. A Comparative analysis of both techniques will show the merits and demerits of both techniques against each other. Comparison can be done on the basis of Time taken by a particular search and Accuracy of produced results against query. For analysis we have taken 20 keywords as shown I table 1 for query to both types of searches against hundred different web pages used for experimental results. A combination of more than one keyword can be used to find relevant results.

TABLE 1: Selected Query Keywords

Sr. No	Query Keywords	Sr. No	Query Keywords
1	Compute	11	ROM
2	Computer science	12	Search
3	Literature	13	Density
4	Spirituality	14	Poem
5	Spiritual science	15	Essay
6	Literature is science	16	GOD

7	RAM	17	English Literature
8	math	18	SEO
9	Research	19	Punjabi mp3
10	Cricket	20	IPL

All above mention keywords we put one by one on both searching techniques and store experimental results separately for time and accuracy. We apply all upper keywords in the following Keyword Based Search algorithm one by one for experimental results. Every query keywords directly matched with web page keywords in Search engine inverted indexed database.

Algorithm for Keyword Based Search:[3]

The algorithm describing the sequence of computation steps involved in present work is given below:-
Search keyevalu()
1. Read a key from the user input.
2. Pick up the page from the page repository.
3. Match each key with the entire existing keyword list with each web page database.
4. Calculate the frequency of keyword in the page.
5. Show keyword frequency in keyword table.
6. Repeat the step from 2 to 5 for each key.

7. Find a page with high frequency of keywords in a keyword table.

8. Show results from high to low frequency keyword pages.

The Semantic Search uses a knowledge base to extend the user query keywords. Knowledge base is a collection of related terms, so thousands of related terms stored in knowledge base previously. A web page's rank in the Semantic Search can be calculated with following algorithm.

Algorithm for the Semantic Search:[3]

The algorithm describing the sequence of computation steps involved in present work is given below:-
Search Semevalu()
1. Read a key from the user input.
2. Match it from the entire existing keyword list in the knowledge base.
3. Pick up the page from the page repository.
4. Match each key with the entire existing keyword list in each web page database.
5. Calculate the frequency of keyword in a web page.
6. Show keyword frequency in keyword table.
7. Repeat the step from 2 to 5 for each key.

8. Find a page with high frequency of keywords in a keyword table.

9. Show results from high weighted pages to low weighted pages.

IV. RESULTS

After compare both techniques on the basis of Time an Accuracy parameters the figure 3 shows the final results. Where Time Efficiency describes the Time taken by a Search technique to find query related web

pages and Accuracy describes the relevant results to user query produced by a particular search technique. Both techniques have different time and accuracy results for same query keywords as shown in Table 2 and figure 3. Keyword Based Search is comparative fast to Semantic Search but produced less relevant results to user query. And Semantic Search is comparative slow to Keyword Based Search but produced more relevant results to user query. The Keyword Based Search is 68 % Time Efficient and the Semantic Search is 40% Time Efficient which is very less than the Keyword Based Search as shown in Table 2.

When we search with following query of two keywords for example "English literature" as shown in figure 2, the Keyword Based Search goes with only these two keywords "computer is god" and searches web pages which have both these keywords. As we know these two keywords are not relate and have no sense which to be searched for. But meaning of query in not important, only keywords are important.

English literature

Figure 2: Query keywords

On the other hand with the Keyword Search there are more possibilities for spam web page got high rank in SERP. The Black Hat SEO mostly uses keywords density to illume the Search Engine and user doesn't find any relevant results. The Keyword Based Search depends only user's query keywords, if user puts a wrong query then this technique produce.

TABLE 2: Analysis Result of Keyword Based Search and Semantic Search

Parameters\Technique	Keyword Based Search	Semantic Search
Time Efficiency	68%	40%
Accuracy	41%	80%

When we go with all the keywords as shown in table1and search, the Semantic Search found more relevant results as compare to the Keyword Based Search. The Keyword Based Search is comparatively fast to the Semantic Search. But a user wants accurate results to his query, instead of time limits. If user doesn't find any relevant results, then user will put a query again with different keywords, so time is not a big issue for The Semantic Search if it produces relevant results to a user.

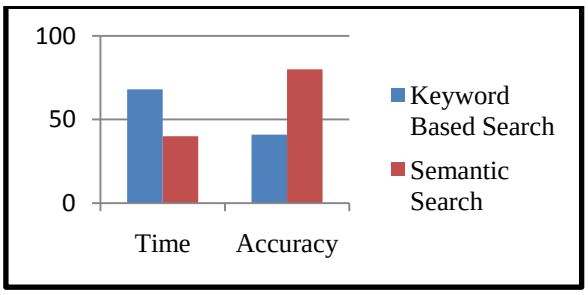

Figure 3: Analysis of Keyword and Semantic Search

The Semantic Search produces relevant information but there is a disadvantage as compared to the Keyword Based Search to takes more time to extract information from web. In the Semantic Search not only query keywords searched, but also all knowledge base keywords searched against web pages. So as query keywords increased, knowledge base keywords will also be extended. Complexity of the Semantic Search extends as query keywords increases. The Semantic Search has more capability to extract the relevant

information from web as compare to the Keyword Based Search. The Semantic Search is most effective with multiple keywords query. The Semantic Search has only disadvantage of time complexity. In advantage the Semantic Search has fewer possibilities to use as spamming for Search Engine because of knowledge base keywords.

V. CONCLUSION AND FUTURE SCOPE

Both techniques are directly depends on user query keywords. A user's query keyword(s) works as feed for these both techniques, only text is a way to be searched for, no image or graph search support. As figure 3 shows the Semantic Search is more effective then the keyword search. But a time complexity of the Semantic Search is a big issue. The Keyword based search is also effective if user has some knowledge about searching fields. Otherwise the keyword search produced keyword bounded results which may not appropriate for a user. With same keywords both techniques produces different results. The Semantic Search uses knowledge base and produces more relevant results. Future scope for this research is to refine the Semantic Search time complexity. A mixture of both techniques can be defined to take advantages of both techniques. With the mixture of both techniques there are possibilities to overcome the Black Hat techniques which bypass the Search Engine rules.

REFERENCES

[1] Ping-Tsai, Sarah, H. Chung and H. Chun-Keung, "A web server design using Search Engine optimization techniques for web intelligence for small organizations, *"Systems, Applications and Technology Conference IEEE,* ISBN: 978-1-4577-1342-2, 2012.

[2] Christopherd Manning, R. Prabhakar and S. Hinrich, "*An Introduction to Information Retrieval"*, Cambridge University Press, 2009.

[3] G. Shikha, Y. Sunit and G. Raj Kumar, "Search Engine Evaluation Based on Page Level Keywords", *IEEE*, ISBN: 978-1-4673-4529-3, 2012.

[4] H. Zhou, S. Qin., J. Liu,**"**Study on Website Search Engine Optimization," *International Conference on Computer Science and Service System IEEE,* 2012.

[5] J. *Yi, L*. Zhuying and L. Hongwei, "The Research of Search Engine Based on Semantic Web, *"International Symposium on Intelligent Information Technology Application Workshops IEEE,* 978-0-7695-3505-0, 2008.

[6] K. Junaidah Mohamed and R. Mahathir," Introduction to Semantic Search Engine, "*International Conference on Electrical Engineering and informatics IEEE,* ISBN: 978-1-4244-4913-2, 2009.

[7] K. Phyo Thu Thu, W. Htwe Pa Pa, T. KhinNwe Ni, **"**Keyword Searching and Browsing System over Relational Databases," *IEEE*, ISSN : 978-1-4577-1539-6/11, 2011.

[8] M. Debajyoti , S. Manoj , J. Gajanan , P. Trupti, P. Adarsha, "Experience of Developing a Meta-Semantic Search Engine," *IEEE*, ISSN : 978-0-4799-2235-2/13, 2013.

[9] P. Swati and P. Ajay.," Search Engine Optimization: A Study," *Research Journal of Computer and Information Technology Sciences*, Vol. 1(1), P.P 10-13, February 2013.

[10] S.G. Choudhary, S.R. Kalmegh and Dr. S. N. Deshmukh, "Semantic Search Algorithms based on Page Rank and Ontology: A Review",*3rd International Conference on Intelligent Computational Systems IEEE,* 2013.

[11] S. Robin , K. Ankita and B. Priyanka, "Web Page Indexing through Page Ranking for Effective Semantic Search, *"Proceedings of 7th International Conference on Intelligent Systems and Control IEEE,* ISBN: 978-1-4673-4603-0, 2012.

[12] T. Duygu, S. Mohammad Ahmed and B. Yıltan, "An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia," *Fourth International Conference on Internet Monitoring and Protection IEEE,* ISBN: 978-0-7695-3612-5, 2009.

[13] T. Ping, Sarah H. Chung and Chun-Keung Hui, "A web server design using search engine optimization techniques for web intelligence for small organizations," *Systems, Applications and Technology Conference IEEE,* ISBN: 978-1-4577-1342-2, 2012.

[14] T. PhyoKhine, Htwe Pa Win and KhinNwe Ni Tun, "Keyword Searching and Browsing System over Relational Databases," *IEEE,* ISBN: 978-1-4577-1539-6, 2011.