I.J. Information Engineering and Electronic Business, **2012, 4, 9-18**
Published Online August 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2012.04.02

WebSEReleC – Optimized Web Implementation of SEReleC Using Google

Vishwas J Raval
M Tech CSE Scholar, Electronics & Computer Engineering Department, Indian Institute of Technology Roorkee, Uttarakhand, India
Email: vishwas.raval@gmail.com

Dr. Padam Kumar
Professor & Head, Electronics & Computer Engineering Department, Indian Institute of Technology Roorkee, Uttarakhand, India
Email: padamfec@iitr.ac.in

Abstract—The World Wide Web has immense resources for all kind of people for their specific needs. Searching on the Web using search engines such as Google, Bing, Ask have become an extremely common way of locating information. Searches are factorized by using either term or keyword sequentially or through short sentences. The challenge for the user is to come up with a set of search terms/keywords/sentence which is neither too large (making the search too specific and resulting in many false negatives) nor too small (making the search too general and resulting in many false positives) to get the desired result. No matter, how the user specifies the search query, the results retrieved, organized and presented by the search engines are in terms of millions of linked pages of which many of them might not be useful to the user fully. In fact, the end user never knows that which pages are exactly matching the query and which are not, till one check the pages individually. This task is quite tedious and a kind of drudgery. This is because of lack of refinement and any meaningful classification of search result. Providing the accurate and precise result to the end users has become Holy Grail for the search engines like Google, Bing, Ask etc. There are number of implementations arrived on web in order to provide better result to the users in the form of DuckDuckGo, Yippy, Dogpile etc. This research proposes development of a meta-search engine, called WebSEReleC (Web-based SEReleC) that provides an interface for refining and classifying the search engines' results so as to narrow down the search results in a sequentially linked manner resulting in drastic reduction of number of pages using power of Google.

Index Terms— Search Engines, Meta-Search Engine, HyperFilter, HyperUnique, HyperClass, WebSEReleC

I. INTRODUCTION

World Wide Web (WWW) has seen an unbelievable growth in the last decade. Both the number of the users and available data on Internet has grown rapidly. Following figure 1 gives facts about how Internet has grown in terms of populations across the globe. These figures indicate usage of Internet for each continent till December 31, 2011.

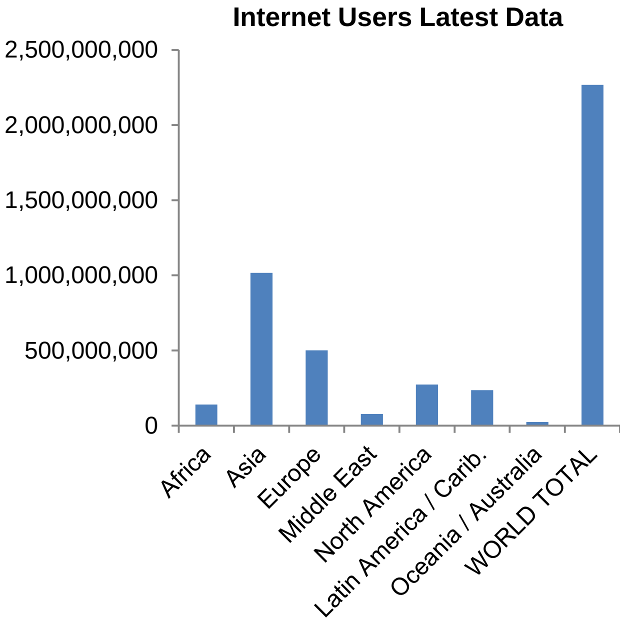


Figure 1. Internet Users in World by December 31, 2011

It is easily visible that overall Internet users in the world are now more than 20 billion and the growth in last decade is more than 500% which is quite huge. With this unbelievable growth, search engines have become dominant in day to day life for an Internet user may it be a computer professional or naïve user to locate piece of information. As per recent figures, Google's (www.google.com) worldwide market share peaked at 90.81% in April 2012 whereas Bing-powered Yahoo! (www.yahoo.com) and Bing (www.bing.com) are 3.8% and 3.46% respectively! Figure 2 shows the statistics of market share collected between July 1, 2008 and April 21, 2012. From the figure, it can be easily concluded that Google, Yahoo! and Bing are the only major giants in search engines.

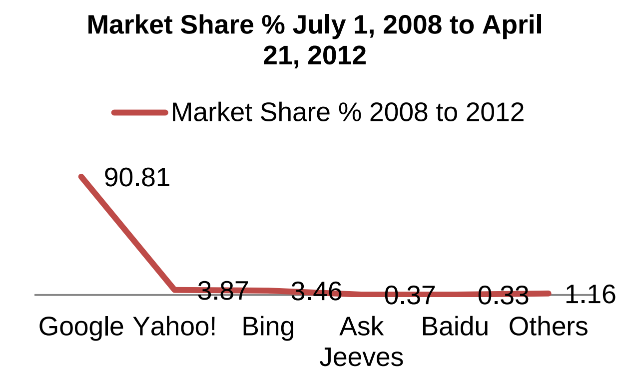


Figure 2. Top Search Engines between July 2008 to April 2012 (Source: gs.statecounter.com)

All of these search engines are programmed to rank websites based on their popularity and relevance. Empirical studies indicate that there are various political, economic, and social biases in the information they provide [6]. These biases could be a direct result of economic and commercial processes (e.g., companies that advertise with a search engine can become also more popular in its organic search results) or political processes (e.g., the removal of search results in order to comply with local laws) [13]. Google Bombing is one example of an attempt to manipulate search results for political, social or commercial reasons. It is a kind of a mal-practicing by creating large numbers of links that cause a web page to have a high ranking for searches on unrelated or off topic keyword phrases, often for comical or satirical purposes. One of the Google bomb dates back in 1999, when a search for “more evil than Satan himself” resulted in the Microsoft homepage as the top result. Another example of the same kind, when a search for “Miserable Failure” resulted in Biography of George W Bush as top link! Fortunately, these bombs were killed later on by Google.

So the shocking fact is Google or any other search engine may not return the results accurately and hence “They are not perfect”! The robot in iRobot movie was good but not perfect, similarly all search engines are good but they are not perfect. They provide results in less than milliseconds but the results are more quantitative than qualitative. There are many meta-search engines (those who do not have their own crawlers or spiders and rely on other search engines like Google, Yahoo!, Bing etc.) in markets which try to provide better results by merging the results from several search engines or by re-ranking it or by doing some post retrieval processing, yet “They are better but they too are not perfect!”

II. SEARCH AND META-SEARCH ENGINES

Search Engines

Web search engines are the tools to search the contents stored across World Wide Web. The results generated may be pages, images, ppts or any other types of files. The results of search engines are displayed in the form of a list in which the numbers of pages might be in thousands or millions. The usual working of a search engines consists of following:

- They search the Internet or select pieces of the Internet based on important words, this is called crawling.
- They keep an index of the words they find, and where they find them which is referred as indexing.
- They allow users to look for words or combinations of words found in that index that is called searching.

Before moving ahead, let’s take a dip into the generations of search engines first. Around 1995-97, AltaVista, Excite, WebCrawler, etc. which are first generation used mostly on-page data (text and formatting) and was very close to classic Information Retrieval. They support mostly informational queries. In the beginning, search results were very basic and largely depended on what was on the Web page. Important factors included keyword density, title, and where in the document keywords appeared.

First generation added relevancy for META tags, keywords in the domain name, and a few bonus points for having keywords in the URL. <meta> tags allow the owner of a page to specify key words and concepts under which the page will be indexed. This can be helpful, especially in cases in which the words on the page might have double or triple meanings – the <meta> tags can guide the search engine in choosing which of the several possible meanings for these words is correct. There is, however, a danger in over-reliance on <meta> tags, because a careless or unscrupulous page owner might add <meta> tags that fit very popular topics but have nothing to do with the actual contents of the page. To protect against this, spiders will correlate <meta> tags with page content, rejecting the <meta> tags that don't match the words on the page. Due to these limitations of <meta> tag, search engines started using web crawlers or known as spiders too. Following figure 3 and text illustrates the basic architecture of a web crawler [1].

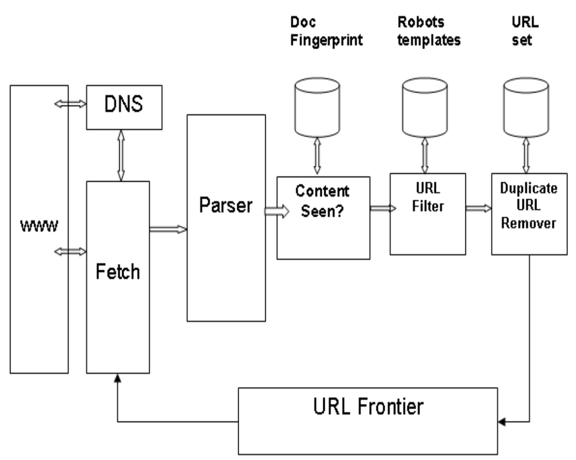


Figure 3. Architecture of Web Crawler

1. URL Frontier: It contains URLs yet to be fetched in the current crawl. At first, a seed set is stored in URL Frontier, and a crawler begins by taking a URL from the seed set.
2. DNS: Domain name service resolution which looks up IP address for domain names.
3. Fetch: It generally use the http protocol to fetch the URL.
4. Parse: The page is parsed. Texts (images, videos, and etc.) and Links are extracted.
5. Content Seen? This checks if a web page with the same content has already been seen at another URL. Need to develop a way to measure the fingerprint of a web page.
   a. URL Filter: Whether the extracted URL should be excluded from the frontier (robots.txt).
   b. URL should be normalized (relative encoding):
      i. en.wikipedia.org/wiki/Main_Page
      ii. Disclaimers
6. Duplicate URL Remover: The URL is checked for duplicate elimination so that spider does not fall into a recursive loop.

The Second generation search engines use off-page, web-specific data such as link analysis, anchor-text, and Click-through data. This generation supports both informational and navigational queries and started in 1998-1999. Google was the first engine to use link analysis as a primary ranking factor and DirectHit concentrated on click-through data. By now, all major engines use all these types of data. Link analysis and anchor text seems crucial for navigational queries.

The Third generation which is now emerging is attempting to blend data from multiple sources in order to try to answer "the need behind the query". For instance, when user searches for New York, the engine might present direct links to a hotel reservation page for New York, a map server, a weather server etc. Thus third generation engines go beyond the limitation of a fixed corpus, via semantic analysis, context determination, natural language processing techniques etc. The aim is to support informational, navigational, and transactional queries. The famous and most widely used search engines are Google, Yahoo!, Bing shown in figures 4, 5 and 6.

Figure 4. Google Search Engine

Figure 5. Bing Search Engine

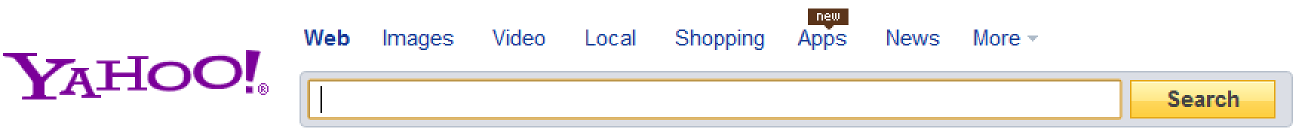


Figure 6. Yahoo! Search Engine

Early search engines used to hold an index of a few hundred thousand pages and documents, and received maybe one or two thousand inquiries each day. Today, a top search engine indexes hundreds of millions of pages, and respond to tens of millions of queries per day. This is done using proprietary algorithms, which work based on the assumption that if a page is useful, other pages covering the similar topic are likely to provide a link to it [1]. This is known as Page Relevance.

Meta-Search Engines

Unlike search engines, meta-search engine doesn't have a database of indexed pages of its own. They base their services on several individual search engines. They borrow services provided by their member search engines and return the integrated results. They neither own an index database or a classification directory, which is the biggest difference with individual search engines [14]. They are also termed as a web-based service that aggregates data from a number of search engines. Instead it "sends a user's query to multiple search engines and blends the top results from each into one overall list." Figure 7 resembles this functionality.

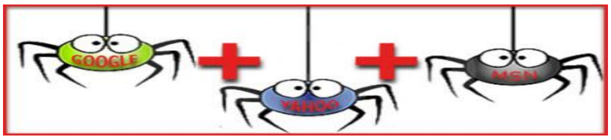


Figure 7. A typical meta-search engine functionality

There are many meta-search engines like Dogpile (www.dogpile.com), Yippy (www.yippy.com), DuckDuckGo (www.duckduckgo.com) etc. running on World Wide Web (WWW) which take user input, pass it to other search engines, process the result and return it to the user in better way. A typical architecture of a meta-search engine is given in figure 8 [12]. The term "meta-search" is frequently used to classify a set of commercial search engines but is also used to describe the paradigm of searching multiple data sources in real time.

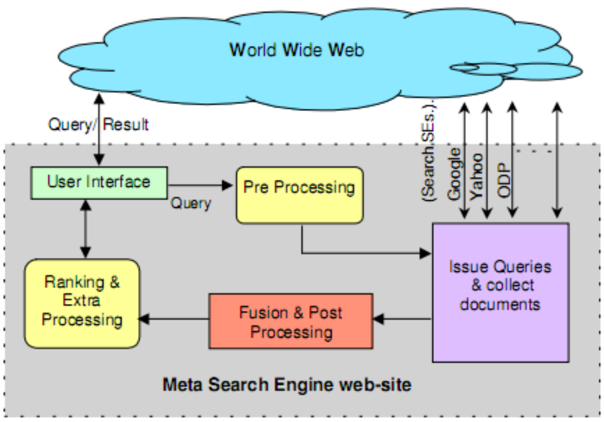


Figure 8. Architecture of a typical meta-search engine

Figures 9, 10 and 11 show three famous meta-search engines.

Figure 9. Yippy Meta-Search Engine

Figure 10. DuckDuckGo Meta-Search Engine

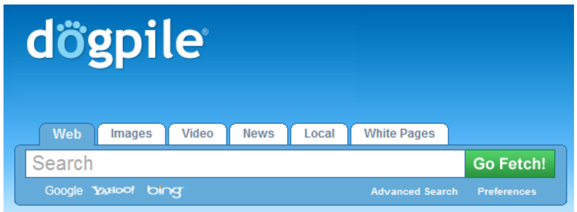


Figure 11. Dogpile Meta-Search Engine

Though the search and meta-search engines provide an ocean of pages as a response to the users' query, there are few very primitive limitations and problems that have remained unaddressed till date in spite of many attempts. Due to this, common and especially naïve users have to waste their precious time to dig through the returned result-set to land onto their desired pages by manually refining it. The next section explains problems of existing search and meta-search engines.

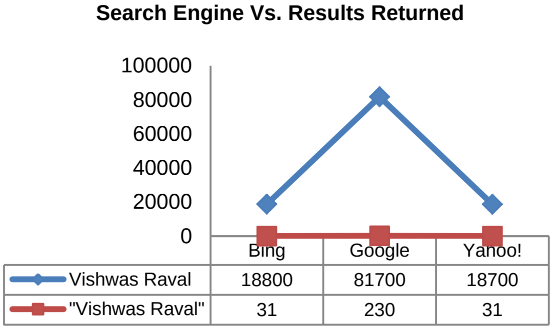
III. PROBLEMS OF EXISTING SEARCH AND META-SEARCH ENGINES

This section will take you dip into the problems of existing search and meta-search engines. In order to discuss the problems, several experiments have been carried out using three basic and most famous search engines Google, Yahoo, Bing and three meta-search engines Yippy, DuckDuckGo and Dogpile. Though most of these search engines have advanced options for search, by an in depth study, it is identified that all of these search engines and hence meta-search engines have either or all of the below mentioned basic problems in the results returned. The experiments are done mainly for a normal search and an exact search. A Normal search is the usual search which is performed by all the users most of the time. An Exact search is the search which is performed by putting keywords in double-quotation mark which finds exact sequence of the words in the search string. Following sections explain those experiments.

Plenty of Irrelevant Results

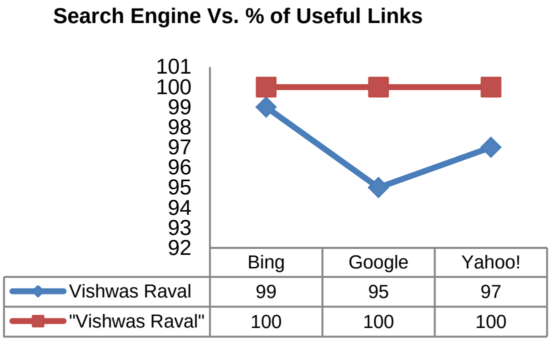
A link is useful if that link references the page that matches the search keywords exactly. Search engines, mainly Google, returns millions of links in response to the query of which only few links are useful to the user which user is actually interested in. Other fact which is true is that no one is going to check these many millions of links one by one and that is time consuming too!

Here, in the experiment, when the search keywords were Vishwas Raval, search engines returned around 81700 links of the pages which contained even Vishwasan or Ravalia, Ravalgaon etc. words which are not correct and relevant to the given keywords. This happened since these results are based on Page Relevance, PageRank and many other factors that Google might have considered and are not based on exact match as discussed earlier. However, when search query was "Vishwas Raval", search engines returned those links that matched exact word in the same sequence. Noticeable thing is that in case of exact match only 100% accurate results are obtained (Refer figure 12 and figure 13) which normally a user does not perform. So some automated functionality is required for exact search.



	Bing	Google	Yahoo!
Vishwas Raval	18800	81700	18700
"Vishwas Raval"	31	230	31

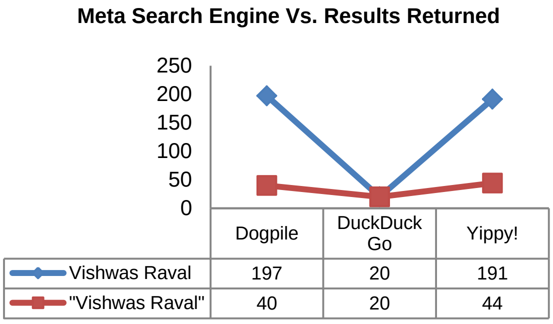
Figure 12. Number of Results returned by search engines



	Bing	Google	Yahoo!
Vishwas Raval	99	95	97
"Vishwas Raval"	100	100	100

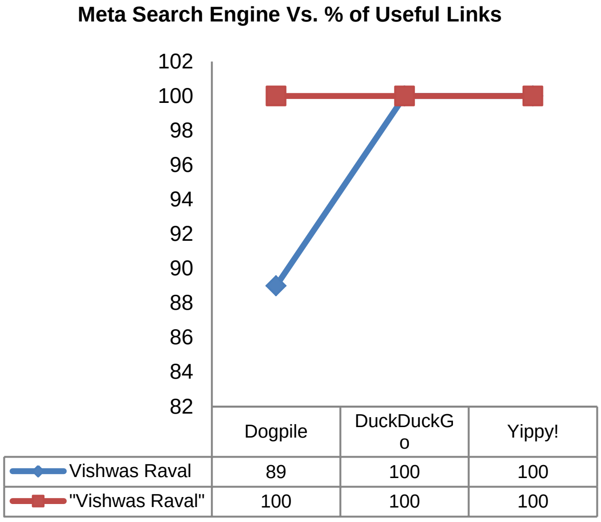
Figure 13. Useful links in first 100 or less links returned by search engines

In case of meta-search engines, when Vishwas Raval was the search string, the percentages of useful links were not up to the mark for Dogpile with reference to the results returned by it. See figure 14 and figure 15.



	Dogpile	DuckDuck Go	Yippy!
Vishwas Raval	197	20	191
"Vishwas Raval"	40	20	44

Figure 14. Number of results returned by meta-search engines



	Dogpile	DuckDuckG o	Yippy!
Vishwas Raval	89	100	100
"Vishwas Raval"	100	100	100

Figure 15. Useful links in first 100 or less links returned by meta-search engines

Lack of Combinatorial Search

Another important thing is that all the search engines including Google don't perform combinatorial search. Combination of keywords plays a big role in accuracy. A query, for instance, "Vishwas Raval" should also return the results containing "Vishwas", "Raval", "Vishwas Raval" and "Raval Vishwas" too for a search since these all could be reference to the same person. Searching with all possible combinations of search keywords is called Combinatorial Search. Omitting combinatorial search could miss some important relevant links which user might be interested in. A naïve Internet user usually does not know the combination of keywords to give to search engines or rather it is quite time-consuming and a drudgery kind of thing to give such combinations every time during search. Hence many a times user misses the important results which might not have been returned by search engine due to the factors that a search engine looks in while searching. So a method that provides combinatorial search with 100% accuracy is required.

Redundant Links

Among the useful links from the result returned, many of the links were found redundant. For e.g. if Vishwas Raval is to be searched and it is found on various pages of some link www.abc.com then, just one link, stating existence of word Vishwas Raval on www.abc.com, is enough. Rests of the same links are not required as user would never go to same link again. This happens mainly due to the fake sites which have been made for the purpose of earning money and steals information from other sites and create links on their own website of such information. This is how redundant links are defined. Figure 16 and 17 shows the percentage of redundant links found with reference to the useful links found in figure 13 and figure 15.

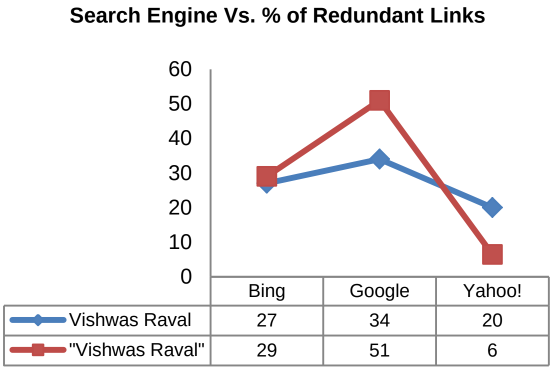


Figure 16. Redundant links in first 100 or less links returned by search engines

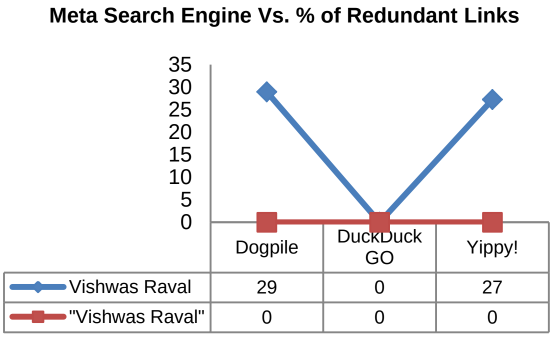


Figure 17. Redundant links in first 100 or less links returned by meta-search engines

Unclassified Results

Though many of the above mentioned search and meta-search engines classify the results but they are not based on search keywords. For instance, Google classifies results based on chronology, images, shopping, blogs etc., whereas Yahoo! and Yippy classify based on several famous and most frequently words related to the search keywords. Not any of the search or meta-search engines classifies the results based on the keywords in the search query hence a Search Keyword based Link Classification scheme using Combinatorial Exact Search is required.

IV. RELATED WORK

Many attempts have been made to resolve the issues discussed in Section 3. Following texts discusses work done so far.

The best example of such work in this direction is Google API Search Tool by Softnik Technologies [2]. It is a simple but powerful Windows software tool for searching Google. It is completely free and is not meant to be a commercial product. All that the users need to do is register with Google for a license key and they will be entitled to pose 1000 queries a day. It is also an efficient research tool because it allows the users to record the search results and create reports of their research easily and automatically. But Google Search API has been deprecated now from market.

Another work is GuidedGoogle [4] carried out by Dr. Buyya et al. which is implemented using Google Search API to guide google search engine for accurate search. This paper proposed a guided meta-search engine which provides meta-search capability developed using the Google Web Services. It guides and allows the user to view the search results with different perspectives. This is achieved through simple manipulation and automation of the existing Google functions. Its meta-search engine supports search based on "combinatorial keywords" and "search by hosts". This work gives an important concept of combinatorial search for finding correct results.

Work of Dou et al. [5] is about finding whether a web query is a personal name and showing profile related information once query is determined as personal name.

Amrish Singh et al. in [7] proposed an approach to presenting web search results that supports personalization, taking into consideration users' perspectives. It is basically a post-retrieval algorithm which uses document classification techniques to organize search results into a meaningful hierarchy of topics, based on the perspective of the user performing the search, represented as a taxonomic ontology.

David Vogel et al. [8] demonstrates that most web search queries contain only two or three terms and therefore provide very limited information about the user's information need to the search engine. Utilizing this information is a key factor to constructing effective web search engines. One way of approaching this problem computationally is to approximate the intended meaning of a query by a node, or a set of nodes, in a given subject taxonomy. For instance, a query "the raven" can indicate that a user searches for information on entertainment/movies or on zoology. Thus, the intuitive problem of capturing the intended meaning of a query is reduced to the computational problem of mapping the query string to a set of nodes in a given – fixed, but arbitrary – subject taxonomy.

Similar work has been carried out by Milos and Mirjana in CatS [9]. CatS operates by forwarding the user query to a major Web search engine, and displaying the returned results together with a tree of topics which can be browsed to sort and refine the results. It is actually a meta-search engine that utilizes text classification techniques to improve the presentation of search results. After posting a query, the user is offered an opportunity to refine the results by browsing through a category tree derived from the dmoz Open Directory topic hierarchy. This paper describes some key aspects of the system (including HTML parsing, classification and displaying of results), outlines the text categorization experiments performed in order to choose the right parameters for classification, and puts the system into the context of related work on meta-search engines. The approach of using a separate category tree represents an extension of the standard relevance list, and provides a way to refine the search on need, offering the user a non-imposing, but potentially powerful tool for locating needed information quickly and efficiently.

Debjyoti et al. shows in [10] that the existing search engines sometimes give unsatisfactory search result for lack of any categorization of search result. If there is some means to know the preference of user about the search result and rank pages according to that preference, the result will be more useful and accurate to the user. In their work a web page ranking algorithm has been proposed based on syntactic classification of web pages. Syntactic Classification does not bother about the meaning of the content of a web page. The proposed approach mainly consists of three steps: select some properties of web pages based on user's demand, measure them, and give different weightage to each property during ranking for different types of pages. The existence of syntactic classification is supported by running fuzzy c-means algorithm and neural network classification on a set of web pages.

Classification of search queries is a complex and computationally challenging task [11]. Normally, search queries are short, reveal very few features per single query and are therefore a weak source for traditional machine learning. In [11], Isak et al. presents a method that combines limited manual labeling, computational linguistics and information retrieval to classify a large collection of web search queries. A short set of manually chosen terms that are known a priori to be of interest to a particular class is used to cull a small number of actual queries from a commercial search engine log. These queries are then submitted to a commercial search engine and the returned search results are used to find more class related terms.

Rose et al. [15] analyzes a classification system that uses web directory search results as an extended feature of the query. They showed that capturing the user intent behind a query statement is crucial for any search engine and is equivalent to figuring out the category to which the query belongs to.

Instead of using the entire web page for classifying Web documents, in [16], Alamelu et al. emphasizes the need for automatic web page classification using minimum number of features in it. A method for generating such optimum number of features for web pages is also proposed.

In summary, existing search and meta-search engines, as well as, the related research works failed to adequately address all the afore-mentioned problems. So this research attempts to address these open issues. In this research, the technique of Combinatorial Search [4] has been enhanced to provide accurate results and named as Combinatorial Exact Search [17]. One more technique of Link Classification has also been proposed. Both of these techniques are based on search keywords. The next section discusses the *WebSEReleC* in details with results.

V. WEBSERELEC —A META-SEARCH ENGINE

Looking towards the unaddressed problems of search and meta-search engines discussed in Section 3, a web-based meta-search engine, *WebSEReleC*, has been developed based on concept of SEReleC [18] which is actually optimized version of SEReleC. It is categorized as a meta-search engine as it works one layer above on existing search engines and dig through the search engines' results and hence it resembles functionality of a meta-search engine. Similar to other meta-search engines, it too does not have its own search database or own crawlers like actual search engines possess. Following figure 18 shows the interface of *WebSEReleC*.

Figure 18. WebSEReleC Interface

Results

With reference to the legends and results discussed in Section 3, figures 19 to 24 presents the results of SEReleC. The results were derived considering both kind of options of Normal and Exact Search. Later, in the final version, the normal search option has been removed.

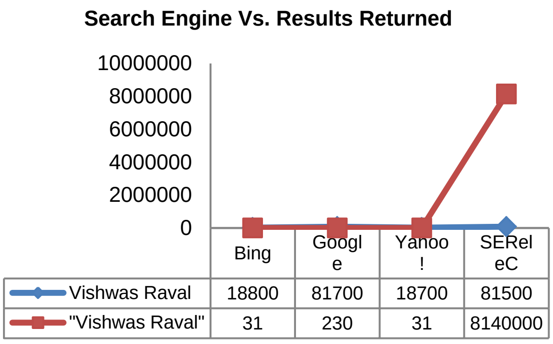
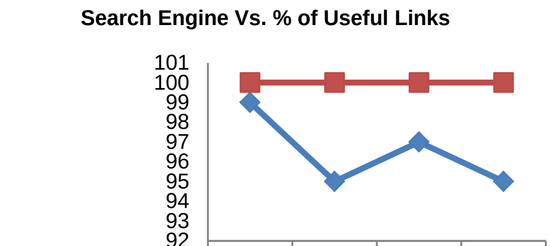
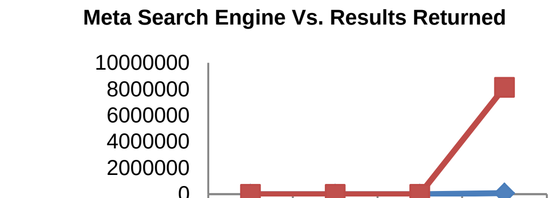


Figure 19. Number of results returned by search engines



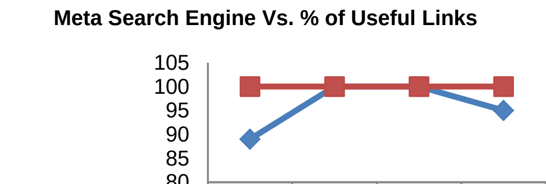
	Bing	Google	Yahoo!	SEReleC
Vishwas Raval	99	95	97	95
"Vishwas Raval"	100	100	100	100

Figure 20. Useful links in first 100 or less links returned by search engines



	Dogpile	Duck Duck Go	Yippy!	SEReleC
Vishwas Raval	197	20	191	81500
"Vishwas Raval"	40	20	44	8140000

Figure 21. Number of results returned by meta-search engines

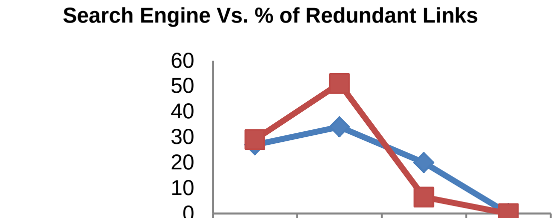


	Dogpile	DuckDuckGo	Yippy!	SEReleC
Vishwas Raval	89	100	100	95
"Vishwas Raval"	100	100	100	100

Figure 22. Useful links in first 100 or less links returned by meta-search engines

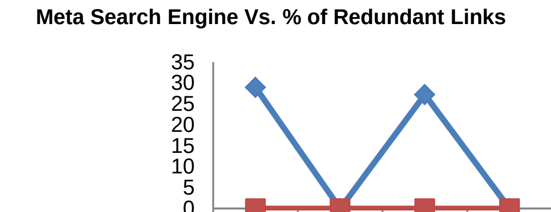
For the SEReleC, figure 19 and 21 shows large number of results. This total number of results was derived based on Combinatorial Exact Search. The reason behind this huge number is that it counted the results for all possible combinations of the entered keywords and searched for exact match of each. In final version of SEReleC, this number can be limited by the end-users. Figure 20 and 22 shows that SEReleC gives 95% accuracy for normal search and 100% accuracy for exact search which is desired and hence final version of SEReleC contains only exact search functionality. In this way SEReleC addresses and resolves the first two issues discussed in Section 3.

One can notice from figures 23 and 24 that the redundancy of the results in case of SEReleC is zero which was one of the objectives in this research and addresses the issue discussed in Section 3. The final issue discussed in Section 3 has been resolved by classifying the links in the results based on search keywords and generated combinations which are shown figure 18.



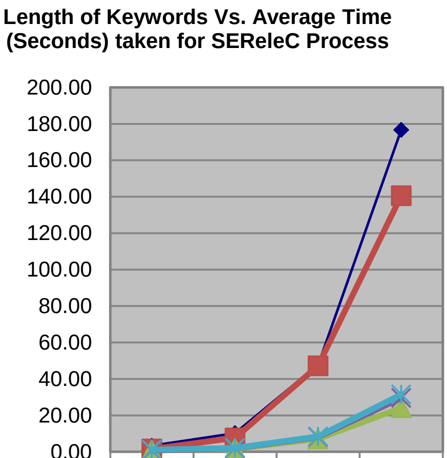
	Bing	Google	Yahoo!	SEReleC
Vishwas Raval	27	34	20	0
"Vishwas Raval"	29	51	6	0

Figure 23. Redundant links in first 100 or less links returned by search engines



	Dogpile	DuckDuckGO	Yippy!	SEReleC
Vishwas Raval	29	0	27	0
"Vishwas Raval"	0	0	0	0

Figure 24. Redundant links in first 100 or less links returned by meta-search engines



	1	2	3	4
SEReleC# (Yahoo!)	3.35	10.24	47.76	176.71
SEReleC# (Bing)	1.59	7.60	47.22	140.59
Optimized SEReleC# (Google)	0.79	1.68	7.05	24.19
Optimized JSEReleC (Google)	0.78	1.92	8.18	29.61
WebSEReleC (Google)	1.13	2.06	8.37	31.47

Figure 25. Average time taken by SEReleC process for search keywords of length one, two, three and four

The most important achievement and result among all the results derived is shown in figure 25. It shows the average time required for the whole SEReleC process for 10 different keywords of length one, two, three and four respectively with specification of maximum number of links to be retrieved 10 for all versions of SEReleC on a broadband connection. Here one can observe that the maximum time it takes is less than a minute which is much lesser than the time spent by an average user, could be few minutes or hours or greater, for landing onto page of his/her interest by going through each and every links manually or extracting the links manually for some other purposes.

VI. CONCLUSION

In this research, SEReleC - a meta-search engine has been developed in order to address and remove limitations of existing search and meta-search engines. In which two innovative techniques of search-keyword based Combinatorial Exact Search and Link Classification have used. WebSEReleC provides meta-search capability developed using power of Google. It guides and allows the users to view the search results with different perspectives and that too is achieved through applying the proposed techniques with simple manipulation and automation of the existing search engine functions. Users can save quite good amount of time by reducing his/her efforts in digging through the set of links returned by search engines to land onto their desired page and can have those results to on his computers drive permanently for his own purpose.

ACKNOWLEDGMENT

We are thankful to The Omnipotent GOD for making us able to do something. We express our gratitude to our department of Electronics & Computer Engineering and the management of Indian Institute of Technology Roorkee for providing us research opportunities and motivating environment. Finally, our acknowledgement cannot end without thanking to the authors whose research work helped us in this research.

REFERENCES

[1] Sergey, B. and Lawrence, P.; “The Anatomy of a Large-Scale Hypertextual Web Search Engine”; Proceedings of the 7th World Wide Web Conference (WWW7), pp.107-117, http://www-db.stanford.edu/~backrub/google.html, April 1998. (Conference Proceedings)

[2] Softnik Technologies; “Google API Search Tool”; http://www.searchenginelab.com/common/products/gapis/docs/, 2003. (Internet Draft)

[3] Alex, D.; “Meta Search Engine Web services with .NET & Java”; EPFL, Lausanne, 2003. (Thesis)

[4] Choon, H. and Rajkumar, B.; “Guided Google: A Meta Search Engine and its Implementation using the Google Distributed Web Services”; International Journal of Computers and Applications, Vol. 26(3) pp.181-187, March 2004. (Journal Publication)

[5] Dou, S., Zheng, C., Qiang, Y., Hua-Jun, Z., Benyu, Z., Yuchang, L. and Wei-Ying, M.; “Web-page Classification through Summarization”; Proceedings of ACM SIGIR-04, pp.242-249, 2004. (Conference Proceedings)

[6] Vaughan, L. and Thelwall, M.; “Search Engine Coverage Bias: Evidence and Possible Causes”; Journal of Information Processing & Management, Vol. 40(4) pp.693-707, May 2004. (Journal Publication)

[7] Amrish, S. and Keiichi, N.; “Hierarchical Classification of Web Search Results Using Personalized Ontologies”; Proceedings of HCI International, 2005. (Conference Proceedings)

[8] Vogel, D., Bickel, S., Haider, P., Schimpfk, R., Siemen, P., Bridges, S. and Scheffer, T.; “Classifying Search Engine Queries Using the Web as Background Knowledge”; ACM SIGKDD Explorations, Vol. 7(2) pp.117-122, 2005. (Journal Publication)

[9] Milos, R. and Mirjana, I.; “CatS: A Classification Powered Meta-Search Engine”; Proceedings of International Conference on Advances in Web Intelligence and Data Mining, pp.191-200; 2006. (Conference Proceedings)

[10] Debjyoti, M., Pradipta, B. and Young-Chon, K.; “A Syntactic Classification based Web Page Ranking Algorithm”; Proceedings of IEEE International Conference on Information Technology, pp.83-92, 2006. (Conference Proceedings)

[11] Isak, T., Sarah, Z. and Amanda, S.; “Using Web Search Logs to Identify Query Classification Terms”; Proceedings of IEEE International Conference on Information Technology, pp.469-474, 2007. (Conference Proceedings)

[12] Manoj, M. and Elizabeth, J.; “Information Retrieval on Internet using Meta-Search Engines: A Review”; Journal of Scientific & Industrial Research, Vol. 67 pp.739-746, October 2008. (Journal Publication)

[13] Segev, E. and Ahituv, N.; “Popular Searches in Google and Yahoo! - A Digital Divide in Information Uses”; The Information Society Journal, Vol. 26(1) pp.17-37, January 2010. (Journal Publication)

[14] Lin, G., Tang, J. and Wang, C.; “Studies and Evaluation on Meta Search Engines”; Proceedings of 3rd IEEE International Conference on Computer Research and Development (ICCRD), pp.191-193, 2010. (Conference Proceedings)

[15] Lovelyn, R. and Chandran, K.; “Web knowledge and WordNet based Automatic Web Query

Classification”; International Journal of Computer Applications, Vol. 17(7) pp.23-38, March 2011. (Journal Publication)

[16] Alamelu, M. and Santhosh, K.; “A Novel Approach for Web Page Classification using Optimum Features”; International Journal of Computer Science and Network Security, Vol. 11(5) pp.252-257, May 2011. (Journal Publication)

[17] Vishwas, R. and Padam, K.; “EGG (Enhanced Guided Google) – A Meta Search Engine based on Combinatorial Keyword Search”; Proceedings of 2nd IEEE International Conference on Current Trends in Technology, December 2011. (Conference Proceedings)

[18] Vishwas, R. and Padam, K.; “SEReleC (Search Engine Result Refinement & Classification) – A Meta Search Engine based on based on Combinatorial Search and Search Keyword based Link Classification”; Proceedings of IEEE International Conference on Advances in Engineering, Sciences and Management, March 2012. (Conference Proceeding)

Vishwas Raval: He is Assistant Professor at Charotar University of Science & Techology, Changa Gujarat (INDIA). He is MTech in Computer Science & Engineering from Indian Institute of Technology Roorkee, India. He has many international publications at conferences and journals of repute. He received many awards for his outstanding research work at conferences. His research areas are Web Technologies and IR.

Padam Kumar: Dr. Kumar did his B Tech (Electrical Engg.) and M Tech (Radar Systems) from IIT Delhi in 1970 and 1972 respectively. He diversified himself into Computer Science area and did his Ph.D. in the use of Functional Programming Languages in Multiprocessing Systems from University of Roorkee in 1990. Currently, he is Professor and Head of E&C Dept. at Indian Institute of Technology Roorkee. His research areas are Parallel Processing, Multiprocessor Networks, Load Balancing and Scheduling, Grid Computing, Cloud Computing, Information Retrieval and Real Time Systems.