International Journal of Web Technology
Volume: 03 Issue: 02 December 2014 Page No.31-37
ISSN: 2278-2389

Refining Serp – Search engine result page for Enhanced Information Retrieval

C.Pabitha [1], G.Sangeetha[2]
[1 2] Assistant Professor, Department of Computer Science and Enginnering, Valliammai Engineering College, Kattankulathur , Kancheepuram District.
Email: contactpabi@gmail.com, sangeethagsam@yahoo.co.in.

Abstract-With the exponential growth of the Internet, it has become more and more difficult to find information. Web search services such as AltaVista, InfoSeek, and MSN Web Search were introduced to help people find information on the web. A web search engine is designed to search for information on the World Wide Web. It works by storing information about many web pages, which they retrieve from the html itself. The search results are generally presented in a list of results and are often called *hits*. A search engine results page (SERP), is the listing of web pages returned by a search engine in response to a keyword query. The results normally include a list of web pages with titles, a link to the page, and a short description showing where the keywords have matched content within the page. A SERP may refer to a single page of links returned, or to the set of all links returned for a search query. Such a result page may not always satisfy the user's needs as there are wide list of result ranked by relevance to the user entered query. Moreover user tends to restrict their sightings to minimum number of result pages as they satisfy their needs thus prohibiting the usage of entire search result. Our aim is to let the user, refine the result page in such a way that all the result pages are covered thus providing various information and views relevant to the query. We design an algorithm that allows the user in refining the search result as per their needs. This is done by grouping the result page in such a way that helps the user in choosing the document according to the categorization.

Key words-SERP, SEO, Search Engine, Query Processing, IR-Information

I. INTRODUCTION

Retrieval Engines have become an integral part of the Internet. For beginners, Search Engines are web based programs that index pages from across the web allowing people to find what they require. Billions of searches are made on different topics using search engines. So when you type in a phrase in a search engine trying to get what you want, the search engines pop up with millions of results. Organic Search Results are those that rise up in results naturally. They are the primary results of the search you made. These results are placed in the **left side** of the search engine result page(**SERP**). Basically they are called Natural Search Results, simply because they are not influenced by money – paid listings. Organic Search Results at the top of the search are mostly informative. But it depends upon the algorithms that the search engine uses to index the pages. Google's success as a search engine is highly due to the more relevant results for the Search term. So a search engine's success depends on the algorithm.

II. LITERATURE REVIEW

It In merely fifteen years the Web has grown to be one of the major information sources. Searching is a major activity on the Web [1,2], and the major search engines are the most frequently used tools for accessing information [3]. Because of the vast amounts of information, the number of results for most queries is usually in the thousands, sometimes even in the millions. On the other hand, user studies have shown [4-7] that users browse through the first few results only. Thus results ranking is crucial to the success of a search engine. In classical IR systems, results ranking was based mainly on term frequency and inverse document frequency. Web search results ranking algorithms take into account additional parameters such as the number of links pointing to the given page [9,10], the anchor text of the links pointing to the page, the placement of the search terms in the document (terms occurring in title or header may get a higher weight), the distance between the search terms, popularity of the page (in terms of the number of times it is visited), the text appearing in metatags [11], subject-specific authority of the page [12,13], recency in search index, and exactness of match [14]. Search engines compete with each other for users, and Web page authors compete for higher rankings with the engines.

This is the main reason that search engine companies keep their ranking algorithms secret, as Google states [10]: "Due to the nature of our business and our interest in protecting the integrity of our search results, this is the only information we make available to the public about our ranking system …". In addition, search engines continuously fine-tune their algorithms in order to improve the ranking of the results. Moreover, there is a flourishing search engine optimization industry, founded solely in order to design and redesign Web pages so that they obtain high rankings for specific search terms within specific search engines (see for example Search Engine Optimization, Inc., www.seoinc.com/). It is therefore clear from the above discussion that the top ten results retrieved for a given query have the best chance of being visited by Web users. This was the main motivation for the research we present herein, in addition to examining the changes over time in the top ten results for a set of queries of the largest search engines, which at the time of the first data collection were Google, Yahoo and Teoma (MSN search came out of beta on February 1, 2005 in the midst of the second round of data collection [15]). We also examined results of image searches on Google image search, Yahoo image search, and on Picsearch (www.picsearch.com/). The searches were carried out daily for about three weeks in October-November, 2004 and again in January-February, 2005. Five queries (3 text queries and 2 image queries) were monitored. Our aim was to study changes in the rankings over time in the results of the individual engines, and in parallel to study the similarity (or rather non-similarity) between the top ten results of these tools.

International Journal of Web Technology
Volume: 03 Issue: 02 December 2014 Page No.31-37
ISSN: 2278-2389

In addition, we examined the changes in the results between the two search periods. Ranking of search results is based on the problematic notion of relevance (for extended discussions see [16,17]. We have no clear notion of what is a "relevant document" for a given query, and the notion becomes even fuzzier when looking for "relevant documents" relating to the user's information seeking objectives. There are several transformations between the user's "visceral need" (a fuzzy view of the information problem in the user's mind) and the "compromised need" (the way the query is phrased taking into account the limitations of the search tool at hand) [18]. Some researchers (see for example [19]) claim that only the user with the information problem can judge the relevance of the results, while others claim that this approach is impractical (the user cannot judge the relevance of large numbers of documents) and suggest the use of judges or a panel of judges (e.g. in the TREC Conferences, the instructions for the judges appear in [20]).

On the Web the question of relevance becomes even more complicated; users usually submit very short queries [4-7.Most previous studies examining ranking of search results base their findings on human judgment. In an early study in 1998 by Su et al. [21], users were asked to choose and rank the five most relevant items from the first twenty results retrieved for their queries. In their study, Lycos performed better on this criteria than the other three search engines examined at the time. In a recent study in 2004, Vaughan [22] compared human rankings of 24 participants with those of three large commercial search engines, Google, AltaVista and Teoma, on four search topics. The highest average correlation between the human-based rankings and the rankings of the search engines was for Google, where the average correlation was 0.72. The average correlation for AltaVista was 0.49. Beg [23] compared the rankings of seven seven search engines on fifteen queries with a weighted measure of the users' behavior based on the order the documents were visited, the time spent viewing them and whether they printed out the document or not. For this study the results of Yahoo, followed by Google had the best correlation with this measure based on the user's behavior.

Here, our aim is to let the user, refine the result page in such a way that all the result pages are covered thus providing various information and views relevant to the query. We design an algorithm that allows the user in refining the search result as per their needs. This is done by grouping the result page in such a way that helps the user in choosing the document according to the categorization.Search engine optimization (SEO) is the process of improving the visibility of a web site or a web page in search engines via the "natural" or un-paid ("organic" or "algorithmic") search results. SEO techniques are classified by some into two broad categories: techniques that search engines recommend as part of good design, and those techniques that search engines do not approve of and attempt to minimize the effect of, referred to as spamdexing. Some industry commentators classify these methods, and the practitioners who employ them, as either white hat SEO, or black hat SEO. White hats tend to produce results that last a long time, whereas black hats anticipate that their sites will eventually be banned once the search engines discover what they are doing.

A. SEO

Fig .1 SERP Categorization

A SEO tactic, technique or method is considered white hat if it conforms to the search engines' guidelines and involves no deception. As the search engine guidelines are not written as a series of rules or commandments, this is an important distinction to note. White hat SEO is not just about following guidelines, but is about ensuring that the content a search engine indexes and subsequently ranks is the same content a user will see. White hat advice is generally summed up as creating content for users, not for search engines, and then making that content easily accessible to the spiders, rather than attempting to game the algorithm. White hat SEO is in many ways similar to web development that promotes accessibility,[40] although the two are not identical. White Hat SEO is merely effective marketing, making efforts to deliver quality content to an audience that has requested the quality content.Traditional marketing means have allowed this through transparency and exposure. A search engine's algorithm takes this into account, such as Google's PageRank. Black hat SEO attempts to improve rankings in ways that are disapproved of by the search engines, or involve deception. One black hat technique uses text that is hidden, either as text colored similar to the background, in an invisible div, or positioned off screen. Another method gives a different page depending on whether the page is being requested by a human visitor or a search engine, a technique known as cloaking. The goal is to increase your site's visibility by achieving favorable listings within the search engine results page(s) ('SERPs'). The pages are divided into 'natural' / 'organic' listings, achieved through effective SEO, and sponsored 'pay-per-click' (PPC).

III. QUERY CACHING

Some search engines cache SERPs for frequent searches and display the cached SERP instead of a live SERP to increase the performance of the search engine. The search engine updates the SERPs periodically to account for new pages, and possibly to modify the rankings of pages in the SERP. SERP refreshing can take several days or weeks which can occasionally cause results to be inaccurate or out of date, and new sites and pages to be completely absent. Major search engines like Google, Yahoo! and Bing primarily use content contained within the Metadata tags of a web page to generate the content that makes up a search snippet. The title tag will be used as the titles of the snippet while the most relevant or useful contents of the web page (description tag or page copy) will be used for the description. If the web page is not available, information about the page from dmoz may be used instead.

A. Serp Tracking

International Journal of Web Technology
Volume: 03 Issue: 02 December 2014 Page No.31-37
ISSN: 2278-2389

Webmasters use Search engine optimization to increase their website's ranking on a specific keyword's SERP. As a result, webmasters often check SERP's to track their Search engine optimization progress. To speed up the tracking process, programmers created automated software to track multiple keywords for multiple websites. . The rise and fall of keyword rankings can sometimes mean the difference between a website making millions or nothing at all. For that reason, search engine optimizers use applications that help track search engine result pages – commonly known as SERP Trackers.

While most SERP Trackers are great at monitoring keyword ranking trends, all of them suffer from the same problems, which include:

- To Many Variables
- With the onset of universal, personalized and local results, actual keyword ranking can be different for different users
- Difficult to Get Results – Google previously provided an API that would return search results (and therefore ranking data), but they abandoned it a couple years ago and introduced a crippled AJAX version. They still allow access to their old search API, but it's not supported, isn't available to new users and provides notoriously bad data.
- Uses Server Resources – With the lack of a commercially available API, most (if not all) SERP Trackers rely on some method of scraping to get their data. Most legitimate businesses do not like doing this, but have no other choice

Because ranking your site in the proper SERP position will be the difference between success and failure in your search marketing campaigns**.**
Consider the following:

- 68% of searchers select a result on the first page of search results
- There is a disproportionate number of clicks (around 40 percent) on the number one listing
- Reaching the first page in paid search is equally important
- Most people click on the first page, and most of *those* people click on earlier results.

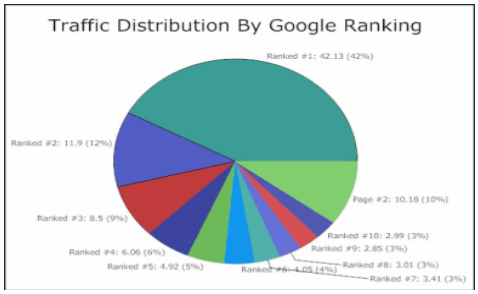

Fig 2. Traffic distribution

While better keyword rankings in SERPs are great, the fact remains that:

- Improving SEO SERPS is Hard - Generating a better search engine optimization ranking for competitive keywords is often as complicated as it is profitable.
- Buying Better Paid SERP Position Can Be Expensive - Meanwhile, consider the fact that in paid search you are literally being charged more money for higher SERP positions.

Improving your SERP position can be done in a few (often not so easy) steps:

- Research Keywords - You first need to discover search engine keywords that will drive traffic AND actual business.
- Organize (Group) the Keywords - You then have to create a logical hierarchy for these keywords; you need an effective information architecture for SEO, and you need to create tight Ad Groups and well-organized ad campaigns for PPC.
- Manage Your SEO Workflow - You need to be able to determine which areas will give you the best return on your time and resource investments; ensure you're not spending *months* building out an SEO campaign to a set of keywords that won't convert for *you* and your business.
- Manage Your PPC Workflow - This works in much the same way that SEO workflow does: you need to prioritize which Ad Groups and campaigns will be worth the effort of writing ad creative and developing landing pages for.
- Act On the Analytics - You then need to do something with all this keyword data and prioritization: you need to write the Web copy, create the ad campaigns, actually achieve the SERP positions you were after.
- Observe the Results, & Repeat - Finally, you need to develop processes for iteratively improving upon the results you've generated. You need to automate certain aspects of your campaigns, and ensure that you are continually having new keywords researched, workflow dynamically suggested, and that you are equipped with the tools to perpetually build on this initial "SERP improvement to-do list".

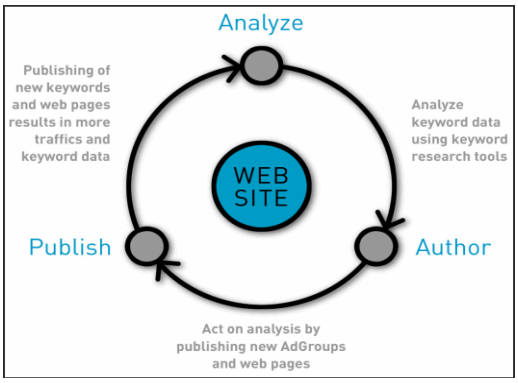

Fig. 3: Analyzing the Process.
A simplified version of the process looks like this.

IV. SERP PROCESS

When returning results to a user, search engines use various resources based upon the particular search or query. The basic information published on a Web page may not be visible to a user, even when the page is displayed in an Internet browser, such as Internet Explorer or FireFox. Some of this information is used by the search engines to determine the topic (content) of a particular Web page. Search engines use many specific

International Journal of Web Technology
Volume: 03 Issue: 02 December 2014 Page No.31-37
ISSN: 2278-2389

characteristics to evaluate a Web page for a particular query string or keyword search. Many of these characteristics differ from one search engine to another and may even vary based upon the query string itself. Because of this, this user help focuses only on the "snippets" or information regarding a particular page, published in organic listings. These are the natural or non-sponsored listings, on search engine results pages, referred to as SERPs.

Traditionally, search engines provide three specific bits of information that make up the snippet in the SERPs. These bits of information are:

- Page Title
- Page Description
- Page URL

Each part of the snippet provides the user with important information related to their query and, as such, gives them partial information for their review. This allows them to select the best result for their specific request. Each search engine controls what is and what is not included in its index. The index contains all of the pages that it feels are relevant and would be important to users for specific queries. Just like the index, the search engine also controls what information is published as the snippet in the SERPs.The search engines look for a few different things when returning results. They want to provide the most relevant results, so that users will return again and again. Therefore, each search engine applies a ranking algorithm to a user's query and then orders the results quickly. Obviously, human interaction with results would be difficult, based upon the massive amount of information available and the almost infinite amount of possible queries. I am not going to discuss conspiracy theories that the search engines do or do not manually rank individual results for specific queries. Publishers of Web pages have the ability to influence the proper ranking of a Web page in the SERPs by publishing relevant content, titles, descriptions, keywords and page names. Some of these can even influence the actual snippet in the SERPs. Typically, the Title tag, META Description tag and the URL with the Page Name, which are part of the web page's code, are used in the SERP snippet.

For consistency purposes, the search engine may truncate, or cut off, long titles, descriptions or URLs if they are too long to fit in within its SERP template. These lengths do not change frequently; however, each search engine has its own standards.

- Title – In the event that a publisher fails to let the search engine know what the title of a page should be, or if the page title is not related to the content on that page, the search engine may use other "trusted sources" to publish the title of the snippet. These "trusted sources" are typically DMOZ, also known as the Open Directory, and the Yahoo Directory. Both of these directories have titles, descriptions and URLs for domains that are in these directories.
- Description – The description displayed in the snippet may not be from the META tag. Sometimes, especially for longer queries, a search engine may return a short section of the content of the individual page that is related to the specific query.
- URL – The URL that is shown in the snippet may be too long for the taste of the search engine, so they may truncate the "http://," "www," or even cut out directories after the domain name to return keywords related to the specific query.

The search engines have given publishers the ability to control whether they want the snippet to be influenced by DMOZ or the Yahoo Directory. There are additional META tags that are supported by the search engines that can be included in the page code that will tell the search engines to stop using this information in the SERP snippets.

DMOZ
All Engines – <META NAME="ROBOTS" CONTENT="NOODP">
Yahoo – <META NAME="SLURP" CONTENT="NOODP">
Google – <META NAME="GOOGLEBOT" CONTENT="NOODP">
MSN/Live – <META NAME="MSNBOT" CONTENT="NOODP">
Yahoo Directory
All Engines – <META NAME="ROBOTS" CONTENT="NOYDIR">
Yahoo – <META NAME="SLURP" CONTENT="NOYDIR">
Google – <META NAME="GOOGLEBOT" CONTENT="NOYDIR">
MSN/Live – <META NAME="MSNBOT" CONTENT="NOYDIR">

Differentiate the descriptions for different pages. Using identical or similar descriptions on every page of a site isn't very helpful when individual pages appear in the web results. In these cases we're less likely to display the boilerplate text. Wherever possible, create descriptions that accurately describe the specific page. Use site-level descriptions on the main home page or other aggregation pages, and use page-level descriptions everywhere else. If you don't have time to create a description for every single page, try to prioritize your content: At the very least, create a description for the critical URLs like your home page and popular pages.Include clearly tagged facts in the description. The meta description doesn't just have to be in sentence format; it's also a great place to include structured data about the page. For example, news or blog postings can list the author, date of publication, or byline information. This can give potential visitors very relevant information that might not be displayed in the snippet otherwise. Programmatically generate descriptions. For some sites, like news media sources, generating an accurate and unique description for each page is easy: since each article is hand-written, it takes minimal effort to also add a one-sentence description. For larger database-driven sites, like product aggregators, hand-written descriptions can be impossible.

In the latter case, however, programmatic generation of the descriptions can be appropriate and are encouraged. Good descriptions are human-readable and diverse, as we talked about in the first point above. The page-specific data we mentioned in the second point is a good candidate for programmatic generation. Keep in mind that meta descriptions comprised of long strings of keywords don't give users a clear idea of the page's content, and are less likely to be displayed in place of a regular snippet. Any one can manage the SERP of the future by creating and embracing holistic strategies to fully

International Journal of Web Technology
Volume: 03 Issue: 02 December 2014 Page No.31-37
ISSN: 2278-2389

manage owned, earned and paid content that lives on the SERP. When developing a comprehensive search strategy for SERP domination, also be sure to take into account all three types of content – paid, owned and earned. Specific tactics to capitalize on this content include:

- Consumer Research to understand the why, what, when and where of consumer queries to more effectively and efficiently target consumers on an increasingly personalized SERP.
- Social Listening to understand how consumers are talking about you brand to facilitate the creation of positive UGC and prevent the creation of negative UGC.
- Data Mining and Compilation Technology to leverage the mass volumes of search and consumer data to make more efficient dynamic optimization and buying decisions on the increasingly personalized SERP.
- Hypertext-Matching Analysis: Our search engine analyzes page content....and our technology analyzes the full content of a page and factors in fonts, subdivisions and the precise location of each word. We also analyze the content of neighboring web pages to ensure the results returned are the most relevant to a user's query.

This is the second way on how a search engine determines which websites will be presented at the top search results. This method does not require the use of PR but rather the relevance of the content based on the keywords being searched by Internet users. This is the reason why some websites even with zero or low PR or which are new in the Internet lands at the top ten search results. This is also the reason why some webmasters might say to go for SERP and not PR, not knowing that PR also plays an important role in SERP rank. They don't realize that the relevance of content matched against the keywords being searched is the method being used on who lands at the top search results. So, if two competing Web pages having similar content and targeting the same set of keywords are indexed, the one with the higher PR will be rightfully chosen. These two are just an overview on how a search engine presents its SERP, it may be the use of PR, the use and relevance of keywords or both. Some key differences in 2010 SERP of Google

- More paid search advertisers - particularly in the space above the organic listings.
- Increase in the variety of results such as "news results" and YouTube videos (versus regular Web pages).
- Multiple page listings for a given site (e.g., Wikipedia and IRS both show two listings each).
- The "Google" logo is smaller.
- The "books" link in the blue bar directly below the search box has been replaced with a "show options" link.

By far the biggest difference is the sheer volume of results. When you look at the first SERP on any given query, you will see a blue bar below the search box, on which the right-hand side will read, "Results 1-10 of about X." This is telling you that the SERP is displaying the top 10 organic results of the total number of results (X). The 2008 SERP for the query "tax return" showed 13.6 million results. That same query in 2009 now shows 67.7 million. That's a five-time increase in two years.

V. EFFECTIVE SEARCH ALGORITHMS

Every search engine has its own relevancy algorithm (rules for ranking), which is continuously developing as search engines are learning from their experience, but let's start from the defining search engine algorithm. It's a complex mathematical formula used by a search engine to rank the webpages that it finds by crawling the Web. The search engines' algorithms are very top secret; there are many people who work outside of the search engine industry who attempt to reverse engineer the code so they can understand how they work in an attempt to help people to rank higher in the results. An algorithm is a set of rules that a search engine uses to rank the listings contained within its index in response to a particular query. Describing search engine algorithms will not be complete without differentiating for types of engines, which, of course, have distinguishing principles of ranking. The first type is crawler-based search engines such as Google and MSN – top search engines.They are also called "traditional." Another type is directories that use human resource for indexing. Most directories do not have their own ranking mechanism; they use some obvious factor to sort the URLs such as alphabetic sequence or Google Page Rank. The next type is hybrid engines, which include META engines and those using other engines' results. The last type is PPC and paid inclusion engines that provide paid listings. Crawler-based search engines apply a sophisticated technique to determine how relevant your pages are to search words and phrases. Though each search engine has its own distinguishing features, the main idea lies in regarding many on-the-page and off-the-page factors, and only after this is determined is your page given a certain position or rank

Another situation is with PPC engines. Paid inclusion engines require certain fees to list your page as well as some difference for working systems such as re-spidering or top-ranking for keywords you choose. The fundamental principle that lies at the heart of PPC process is the that higher you bid, the higher your position will be for the particular search terms. In this system, keywords and phrases are associated with a cost-per-click (CPC) fee. This well-made auction system is popular, and most major Internet search engines utilize such schemes as a part of their indexing and ranking system. Yahoo!, for example, owns Overture's pay engine and provides search results to AltaVista, AllTheWeb, MSN, Overture's supplemental results, and other Inktomi-powered engines such as HotBot.Sponsored results from Overture are found at the top, side, and bottom of the search results pages fed by Yahoo!, but, of course, this search engine ranks pages looking on their on-the-page and off-the-page factors such as MSN and Google top search engines. Google search engine is one of the most widely know for it demands towards submitting pages. Though many consider Page Rank system Google's main search engine algorithm, it has other ways to rank pages, too. Page Rank is an absolute value that is regularly calculated by Google for each page it has in its index. Here, you should to know that the number of links you've got from other sites outside your domain matter greatly as well as the link quality. There are two

International Journal of Web Technology
Volume: 03 Issue: 02 December 2014 Page No.31-37
ISSN: 2278-2389

other terms related to Google's way to rank pages: Hilltop and Sandbox

Hilltop is an algorithm that was created in 1999. Basically, it looks at the relationship between the "Expert" and "Authority" pages. An "Expert" is a page that links to lots of other relevant documents. An "Authority" is a page that has links pointing to it from the "Expert" pages.
In theory, Google would find "Expert" pages and then the pages that they link to would rank well. Pages on sites like Yahoo, DMOZ, college sites, and library sites can be considered experts. Google also checks age of your page by help of Sandbox algorithm, which detects how old your page is and how long ago it was updated. Usually, pages with stale content tend to gradually slip down the result list while the new pages just crawled initially have higher positions than they would if based on Page Rank only. In other words, Google considers new pages to have more relevant and up-to-date content and gives them a certain advantage over the stale pages. Therefore, constantly updating your pages can help keep them up the list.

VI. CONCLUSION

The newest long term search engine optimization strategy is to learn how to present a web site search results in what is called SERP vision. SERP stands for (Search Engine Results Pages). Through the use of SERP vision search engine optimization, the tiny plan10 little spots available on Google now will expand into over 200 exciting results. This will be done with the use of SERP vision search engine optimization expansions, hovers, and drop boxes. This is where a search results will get very interesting in the fact that it will express text, picture, video, and audio combined with the additional expansion features. With the many combinations and connections with other websites that each website has, this will prove to be a very informative and enjoyable search. That is expanded further with far more content connecting each search results with related web sites and information. The search engine will also have further capabilities of changing format presentation so as to be seen in a more desirable individualized format preference that will be set by the reader. With each search engine page one result, the readers as will as the business ad owners will love the resulting benefits.

REFERENCES

[1] M. Madden, America's online pursuits: The changing picture of who's online and what they do, PEW Internet & American Life Project, 2003 Availablefromhttp://www.pewinternet.org/pdfs/PIP_Online_Pursuits_Final.PDF
[2] D. Fallows, The Internet and daily life, PEW Internet & American Life Project, 2004 Available from http://www.pewinternet.org/pdfs/PIP_Internet_and_Daily_Life.pdf
[3] D. Sullivan, Nielsen Netratings search engine ratings. In Search Engine Watch Reports, 2005. Available from http://searchenginewatch.com/reports/article.php/2156451
[4] C. Silverstein, M. Henzinger, H. Marais, M. Moricz, Analysis of a very large Web search engine query log, ACM SIGIR Forum, 33(1) (1999). Available from http://www.acm.org/sigir/forum/F99/Silverstein.pdf
[5] Spink, S. Ozmutlu, H. C. Ozmutlu, B. J. Jansen, U.S. versus European Web searching trends, SIGIR Forum, Fall 2002 (2002). Available from http://www.acm.org/sigir/forum/F2002/spink.pdf
[6] B.J. Jansen, A. Spink, An analysis of Web searching by European Alltheweb.Com Users. Information Processing and Management , 41(6) (2004), 361-381.
[7] B. J. Jansen, A. Spink, J. Pedersen, A temporal comparison of AltaVista Web searching. Journal of the American Society for Information Science and Technology, 56(6) (2005), 559-570.
[8] R. A. Baeza-Yates, B. A. Ribeiro-Neto, B. A., Modern information retrieval. ACM Press, Addison Wesley: Harlow, England, 1999.
[9] S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine. In Proceedings of the 7th International World Wide Web Conference, April 1998, Computer Networks and ISDN Systems, 30 (1998), 107 - 117.Available from http://www-db.stanford.edu/pub/papers/google.pdf
[10] Google. Google information for Webmasters, 2004. Available from http://www.google.com/webmasters/4.html
[11] Yahoo, Yahoo! Help: How do I improve the ranking of my website in the search results, 2005. Available from: http://help.yahoo.com/help/us/ysearch/ranking/ranking-02.html
[12] J. M. Kleinberg, Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), (1999) 604-632. [13] Teoma (2005). Adding a new dimension to search: The Teoma difference is authority. Retrieved March 26, 2005, from http://sp.teoma.com/docs/teoma/about/searchwithauthority.html
[13] MSN Search, Web search help: Change your search results by using Results Ranking, 2005. Available from
[14] http://search.msn.com/docs/help.aspx?t=SEARCH_PROC_BuildCustomizedSearch.htm
[15] C. Payne, MSN Search launches, 2005. Available from
[16] http://blogs.msdn.com/msnsearch/archive/2005/01/31/364278.aspx
[17] T. Saracevic, RELEVANCE: A review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 1975, 321-343.
[18] S. Mizzaro, Relevance: The whole history. Journal of the American Society for Information Science, 48(9) (1997), 810--832.
[19] R. S. Taylor, Question-negotiation and information seeking in libraries. College and Research Libraries, 29 (May) (1968),178-194.
[20] M. Gordon, P. Pathak, Finding information of the World Wide Web: The retrieval effectiveness of search engines. Information Processing and Management,35 (1999),141–180.
[21] TREC, Data – English relevance judgements, 2004. Available from http://trec.nist.gov/data/reljudge_eng.html
[22] L. T. Su, H. L. Chen, X. Y. Dong, Evaluation of Web-based search engines from the end-user's perspective: A pilot study, In Proceedings of the ASIS Annual Meeting, vol. 35, 1998, pp. 348-361.
[23] L. Vaughan, New measurements for search engine evaluation proposed and tested, Information Processing & Management, 40 (2004), 677-691.
[24] M. M. S. Beg, A subjective measure of web search quality, Information Sciences, 169 (2005), 365-381.

Authors:

The Author C.PABITHA was born in Chennai , Kancheepuram District in the year 1986. She completed her M.E Degree in Computer Science & Engineering at Jerusalem College of Engineering, Anna University, Chennai in the year 2009. She got her B.E in Computer Science and Engineering from Valliammai Engineering Colllege, Anna University near Kattankulathur,Kancheepuram District in the yaer 2007. She is an active member in IEEE Madras Section , ISTE since 2010. Her interest in Data Mining has made her publish four papers in National/International Conference and a Journal in Information Retrieval at CiiT International Journal. She has 4 years of teaching experience in academic institutions and she is currently working as a Assistant Professor in Computer Science & Engineering in Valliammai Engineering College, Kattankulathur.

G.Sangeetha was born in Pondicherry in the year 1977. She received her B.E degree in Computer Science and Engineering from University of Madras Chennai in the year 1999, and M.E Degree in Computer Science and

International Journal of Web Technology
Volume: 03 Issue: 02 December 2014 Page No.31-37
ISSN: 2278-2389

Engineering from Sathyabama University, Chennai in 2005. She has thirteen years of teaching experience in various academic institutions. Currently she is working as Assistant Professor in the Department of Computer Science and Engineering at Valliammai Engineering College, Chennai. She has authored a book “Computer Practice I” in the year 2004. She is a life time member of CSI (Computer Society of India) since 2005.