



AN INTEGRATED APPROACH TO IMPROVE THE RANKING OF WEB PAGES

A.Salman Ayaz¹, A.Jaya²

¹Assistant Professor, ²Professor & Head
BS Abdur Rahman University

Abstract

In this study we propose a method to improve the ranking of the documents retrieved from the web. The user query is accepted and the search process is initiated by using an external search engine. Once more number of web pages are obtained from the web the ranking of those pages should be carried out to provide the most relevant pages. The ranking of those pages will be carried out using the novel ranking measure.

Introduction:

Search engine is an information retrieval system from WWW (World Wide Web). Generally Web-search engines hold title tag with URL and meta-tag information or first few words from the Web-page content in their main search result. Without visiting search result links Web-searcher cannot understand the Web-page content will fulfill their expectation or not [1]. The dearth of any specific structure and wide range of data published on the web makes it highly challenging for the user to find the data without any external assistance. It is a general credence [2,3] that a single general purpose search engine for all web data is improbable because its processing power cannot scale up to the fast increasing and unlimited amount of web data. A tool that swiftly gains approval among users is Meta search engines [4]. The Meta search engines can run user query across multiple component search engines concurrently, retrieve the generated outcomes and amassed them. The benefits of Meta search engines against the search engines are notable. [5]. The Meta search engine enhances the search coverage of the web providing higher recall. The

overlap among the primary search engines is generally small [6] and it can be small as three percentages of the total results retrieved. The Meta search engine solves the scalability issue of searching the web and facilitates the use of multiple search engines enabling consistency checking [7]. The Meta search engine enhances the retrieval effectiveness providing higher precision because of 'chorus effect' [8].

The ubiquity of the Internet and Web has led to the emergence of several Web search engines with varying capabilities. These search engines index Web sites, images, Usenet news groups, content-based directories, and news sources with the goal of producing search results that are most relevant to user queries. However, only a small number of web users actually know how to utilize the true power of Web search engines. In order to address this problem, search engines have started providing access to their services via various interfaces [11]. Search engine as a tool to investigate the Web must obtain the desired results for any given query. Success of a search engine is directly dependent on the satisfaction level of the user. Users desire the information to be presented to them within a short time interval. They also expect that the most relevant and recent information to be presented [9]. Most of the search engines cannot completely satisfy user's requirements and the search results are often very inaccurate and irrelevant [10].

There are already many researchers who have reported on about various aspects of search engines in [12-14]. Search tools for the web can be classified as Search Engines, Directory

Services, Meta-Search Engines, and Hybrid Search Services. Google, Inktomi, etc., are typical search engines. Each search engine has three key functional phases, namely, Web Data Acquisition (WDA), Web Data Indexing (WDI) and Web Data Rendering (WDR). They are divided into general purpose and special purpose search engines [9]. A meta-search engine is the kind of search engine to provide users with information services and it does not have its own database of web pages. It sends search terms to the databases maintained by other search engines and gives users the results that come from all the search engines queried [10]. The outcomes that a Meta search system gathers from its component engines are not similar to votes or any other single dimensional entities: Apart from the individual ranking it is assigned by a component engine, a Web outcome also incorporates a title, a small fragment of text which represents its significance to the submitted query [15-17] (textual snippet) and a uniform resource locator (URL). Ostensibly, the traditional rank aggregation techniques are insufficient for providing a robust ranking mechanism appropriate for Meta search engines, because they ignore the semantics accompanying each Web result.

Related Work:

Guang-ming and Wen-juan [18] have proposed the professional meta-search engine system, used the CC4 neural network algorithm to calculate web page relation degree, and obtained the high degree of professional web pages. They solved the problem of the breadth the people's access to information, proposed a good solution to search in the ocean of information. It fully analyzed the web page information, used CC4 neural network algorithm to judge the relevant web pages and optimal rank and then combined the professional dictionary to filter the sort results. Finally, the experimental results proved that the method improved the search quality to the specific specialty.

V Raval and Padam [19] have proposed a meta-search engine called EGG that was intended to use power of the Google for more accurate and combinatorial search. They achieved through simple manipulation and automation of Google functions that are accessible from EGG through the Google. The proposed technique achieved through simple manipulation and automation of the existing Google functions. The proposed

meta-search engine supported search based on "Combinatorial Keywords" and "Normal Search". A detailed evaluation demonstrated how one can harness the capability of Google cluster architecture through its programmable Web services by creating advanced search features at a third party user application level.

Chaurasia *et al.* [20] have presented the priority assisted and user profile based meta search engine. This meta-search engine was able to improve search performance by querying multiple search engines at once. The work had also achieved parallelization of the crawling system, overlapped downloading of related document along with reduction of database redundancy. The results and analysis proved that the method improved the search quality of the database and specific search quality was also improved.

V Raval and Padam [21] have proposed a meta-search engine called EGG that was intended to use power of the Google for more accurate and combinatorial search. They achieved through simple manipulation and automation of Google functions that are accessible from EGG through the Google. The proposed technique achieved through simple manipulation and automation of the existing Google functions. The proposed meta-search engine supported search based on "Combinatorial Keywords" and "Normal Search". A detailed evaluation demonstrated how one can harness the capability of Google cluster architecture through its programmable Web services by creating advanced search features at a third party user application level.

Leonidas Akritidis *et al.* [22] have presented a QuadRank technique which considered the additional information regarding the query terms, collected results and data correlation. They have implemented and tested the QuadRank in real world Meta search engine. They comprehensively tested QuadRank for both effectiveness and efficiency in the real world search environment and also used the task from the TREC-2009 conference. They demonstrated that in most cases their technique outperformed all component engines.

Hideaki Ishii *et al.* [23] have proposed a technique to reduce the computation and communication loads for the Page Rank algorithm. They developed a method to systematically aggregate the web page into groups by using the sparsity inherent in the web.

For each group, they computed an aggregated page rank value that can be distributed among the group members. They provided a distributed update scheme for the aggregated Page Rank along with an analysis on its convergence properties. They provided a numerical example to illustrate the level of reduction in computation while keeping the error in rankings small.

D.A. Adeniyi *et al.* [24] have presented a study of automatic web usage data mining and recommendation system based on current user behavior through his/her click stream data on the newly developed Really Simple Syndication (RSS) reader website, in order to provide relevant information to the individual without explicitly asking for it. The K-Nearest-Neighbor (KNN) classification method has been trained to be used on-line and in Real-Time to identify clients/visitors click stream data, matching it to a particular user group and recommend a tailored browsing option that meet the need of the specific user at a particular time. To achieve this, web users RSS address file was extracted, cleansed, formatted and grouped into meaningful session and data mart was developed. Their result showed that the K-Nearest Neighbor classifier was transparent, consistent, and straightforward.

Rajhans Mishra *et al.* [25] have developed a system that considers sequential information present in web navigation patterns, along with content information. They also considered soft clusters during clustering, which helps in capturing the multiple interests of users. The proposed system had utilized similarity upper approximation and singular value decomposition (SVD) for the generation of recommendations for users. They tested their approach on three datasets, the MSNBC benchmark dataset, simulated dataset and CTI dataset. They compared their approach with the first order Markov model as well as random prediction model.

Kibeom Lee *et al.* [26] have proposed a graph-based recommender system that uses only positively rated items in users' profiles to

construct a highly-connected, undirected graph, with items as nodes and positive correlations as edges. Using the concept of entropy and the linked items in the graph, the proposed system can find recommendations that are both novel and relevant. They tested the system on Last.fm data to recommend music to users and show that the proposed recommender system was indeed able to provide novel recommendations while keeping them relevant to the user profile, consistently outperforming a state-of-the-art matrix factorization-based recommender.

Proposed Architecture:

The primary intention of this research is to design and develop a proficient technique for meta search engine using efficient ranking measure. Collection and searching of web has become an essential part of most people's lives because of the simple access to the internet. Searching on the web using Search engines such as Google, Bing, yahoo has become an extremely common way of locating information. Web-page recommendation plays a significant role in intelligent web systems. Useful knowledge discovery from web usage data and satisfactory knowledge representation for effective Web-page recommendations are crucial and challenging. The process of the proposed meta search engine is divided into two important steps, (1) searching through various search engine, and (2) Ranking the results through the designed ranking algorithm. Initially, the query given by the user is inputted to the various search engine such as, google, bing and yahoo considered in the technique. Once more number of pages obtained from the web, the ranking of those pages should be carried out to provide the most relevant pages. The ranking of those pages will be carried out using the novel ranking measure designed in the proposed technique. Finally, the experimentation will be done to prove the efficiency of the proposed ranking measure in meta search engine with the existing techniques. The proposed method will be implemented on JAVA. The proposed approach is shown in figure 1.

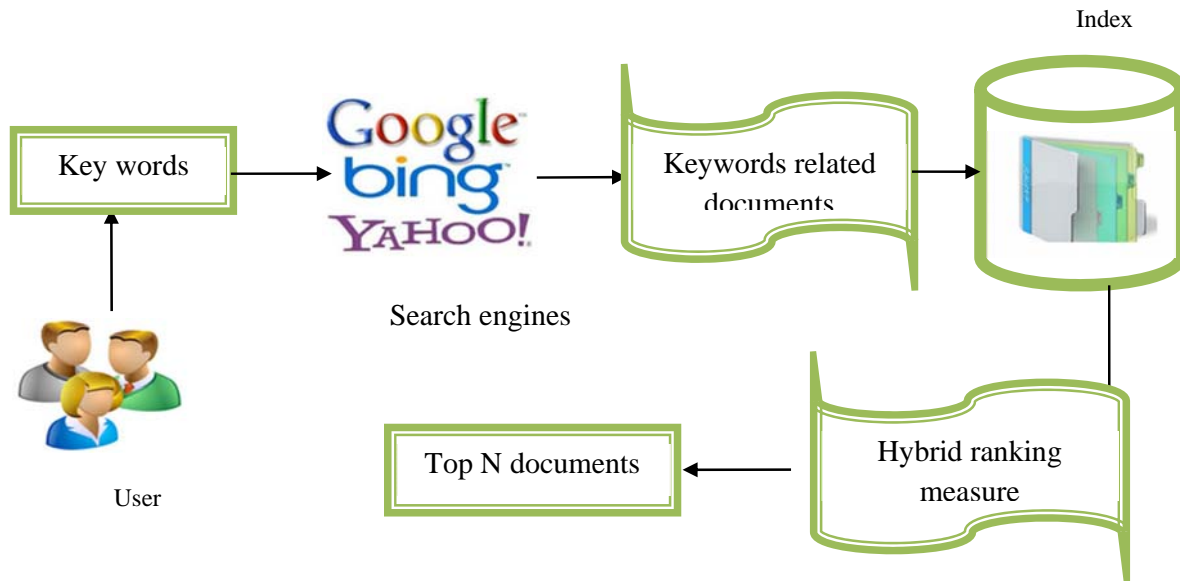


Figure 1: Flow diagram of proposed system

Conclusion:

Search engines are enhancing their search algorithms so that it is possible to provide the answer to the user queries effectively. The major problem of many search engines is the presentation of many choices to the client at a time; this usually results to strenuous and time consuming task in finding the right product or information on the site. Our algorithm employs the novel ranking measure. This will help in delivering the most relevant pages to the user. The efficiency of our ranking measure will be proved by comparing it with the existing techniques.

References:-

- [1] Sukanta Sinha ,Rana Dattagupta and Debajyoti Mukhopadhyay , " Designing an Ontology based Domain Specific Web Search Engine for Commonly used Products using RDF", 2012.
- [2] Sugiura, A., Etzioni, "Query routing for Web search engines: architecture and experiments", *Computer Networks*, vol. 33, pp.417–429, 2000.
- [3] Manning, C.D., Raghavan, P., Schutze, H., 2008. *Introduction to Information Retrieval*. Cambridge University Press
- [4] Meng, W., Yu, C., Liu, "Building efficient and effective meta search engines" *ACM Computing Surveys*, vol. 34, no. 1, pp.48–89, 2002.
- [5] Spink, A., Jansen, B.J., Blakely, C., Koshman, S., "Overlap among major Web search engines", In: *Proceedings of the IEEE*

International Conference on Information Technology: New Generations (ITNG), pp. 370–374, 2006.

- [6] Aslam, J.A., Montague, "Meta search consistency", In: *Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 386–387, 2001.
- [7] Vogt, C.C., 1999, "Adaptive combination of evidence for information retrieval", Ph.D. Thesis. University of California at San Diego
- [8] Dwork, C., Kumar, R., Naor, M., Sivakumar, "Rank aggregation methods for the Web", In: *Proceedings of the ACM International Conference on World Wide Web (WWW)*, pp. 613–622, 2001.
- [9] K.Satya Sai Prakash and S. V. Raghavan, "DLAPANGSE: Distributed Intelligent Agent based Parallel Architecture for Next Generation Search Engines", Department of Computer Science & Engineering, Indian Institute of Technology Madras, India, 2001.
- [10] Z. Li, Y. Wang and V. Oria, "A New Architecture for Web Meta-Search Engines", *Seventh Americas Conference on Information Systems*, CIS Department, New Jersey Institute of Technology, 2001.
- [11] Abawajy, J.H., Hu, "A new Internet meta-search engine and implementation", *The 3rd ACS/IEEE International Conference on Computer Systems and Applications*, 2005.

- [12] A. Araus, "Searching the Web", ACM Transactions on Internet Technology, Vol. 1, pp. 243, 2001
- [13] J. W. Green, "Hyper Dog: Up to Date Web Monitoring through Meta Computers", MS Report, Baltimore, Maryland, October 2000.
- [14] S. Lawrence, L. Giles, "Searching the World Wide Web", Science, VOL. 280, No. 5360, pp: 98-100, 1998.
- [15] Fabrizio Lamberti, Andrea Sanna and Claudio Demartini, "A Relation-Based Page Rank Algorithm for Semantic Web Search Engines", IEEE transactions on knowledge and data engineering, vol. 21, no. 1, 2009.
- [16] Rashid Ali and M. M. Sufyan Beg, "Aggregating Subjective and Objective Measures of Web Search Quality using Modified Shimura Technique", Springer-Verlag Berlin Heidelberg, 2008.
- [17] Stefan Endrullis, Andreas Thor and Erhard Rahm, "WETSUIT: An Efficient Mashup Tool for Searching and Fusing Web Entities", Proceedings of the VLDB Endowment, Vol. 5, No. 12, pp. 1970-1973, 2012.
- [18] LI Guang-ming , GENG Wen-juan, "Research and Design of Meta-search Engine Oriented Specialty," International Symposium on Intelligence Information Processing and Trusted Computing, 2010.
- [19] Vishwas Raval and Padam Kumar, "EGG (Enhanced Guided Google) - A Meta Search Engine for Combinatorial Keyword Search," Institute of Technology, Nirma University, Ahmedabad -382 481, 08-10 December, 2011.
- [20] Brijesh kumar Chaurasia, Shashi Kant Gupta and Rishi Soni, "Meta search engine based on prioritizer", International conference on computational intelligence and communication systems, 2011.
- [21] Vishwas Raval and Padam Kumar, "EGG (Enhanced Guided Google) - A Meta Search Engine for Combinatorial Keyword Search", Institute of Technology, Nirma University, pp.382 481, 2011.
- [22] Leonidas Akritidis, Dimitrios Katsaros and Panayiotis Bozanis, "Effective rank aggregation for metasearching", The Journal of Systems and Software, vol. 84, pp. 130-143, 2011
- [23] Hideaki Ishii, Roberto Tempo and Er-Wei Bai, "A Web Aggregation Approach for Distributed Randomized PageRank Algorithms", IEEE transactions on automatic control, Vol. 57, No. 11, pp. 2703-2717, 2012.
- [24] D.A. Adeniyi, Z. Wei, Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method, "Applied Computing and Informatics, Vol.36,2015.
- [25] Rajhans Mishra, Pradeep Kumar, Bharat Bhasker, "A web recommendation system considering sequential information," Decision Support Systems, Volume 75, pp.1-10, July 2015.
- [26] Kibeom Lee, Kyogu Lee, "Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items, "Expert Systems with Applications, Volume 42, Issue 10, 15 June 2015, Pages 4851-4858.