



An approach for Search Engine Optimization Using Classification - A data Mining Technique

Rajesh Singh¹, S.K. Gupta²

¹Scholar in Bhagvant University, Ajmer (Rajasthan)

² A.P.,CSED, BIET, Jhansi (U.P.)

Abstract

Search engines play an important role in information retrieval on the web. Given a query, search engines, such as Google, Yahoo! and Bing, return a ranked list of results.. For different queries, the result set includes documents on a variety of topics, rather than a single topic. This variation is probably due to the matter that a typical query contains about only two to three terms, which is not sufficient to locate the desired information unambiguously. For example, the query "God" will often return documents referring to both the God and the good. While the user is only interested in one topic, it is not possible for a search engine to know which topic is relevant based on the query alone. Moreover, the standard ranking of the documents in the result set is independent of the topic. So, the rank-ordered result should have an arbitrary topic ordering. Referring to the "GOD" example, this means that a user should scroll through a ranked list in which many documents may be wrong.

Keywords: Web Crawler, Structure, Policies, Cluster.

Introduction

In this Paper, we have searched the issue of improving search engines by employing classification. It is managed by the following issues .1) Information growth on the web is increasing at an exponential rate, 2) discovering structure within an information domain has been proved to be an effective way in practice of organizing large data sets. For example, the time-consuming manual classification of information has successfully been applied by librarians for information organization with a long history. Classification deals with such type of problem that the retrieved results from traditional search engines are topic-independent

World Wide Web provides us with heavy amount of necessary data digitally available as hypertext Data may be WebPages, images, information and other type. [12] This hypertext pool is dynamically changing due to this reason it is more difficult to find useful information. So Web Crawler for automatic Data and Web Mining is Useful to Us. The economic importance of web will enhance the academic interest. The Database Administrator, Management persons or others wishing to perform data mining on large number of web pages will require the services of web crawler or its based tools. For these reasons crawlers are normally multi threaded by which millions of WebPages may be extracted parallel by only one process..

This paper shows distributed approach for web Crawlers including data mining. It works as a distributed system with a central Control Unit fixing or providing the jobs to different computer. Which are connected with a Network? This proposal is not exactly new as large search engine can also distribute processing power by using different number of computers Systems linked with each other.

Search Engine

Search Engine provides the gateway for most of the users trying to explore the huge information base of web pages.[14] Search engines are programs that search documents for specified keywords on search for information on the World Wide Web and returns a list of the documents where the keywords were found. A Search Engine is really a class of programs; however, the term is often used to specifically describe systems like Google, Bing and Yahoo! Search that enable users to search for documents on the World Wide Web.

Types of Search Engine

According to functioning three types of search engine [7].

1) Crawler Based Search Engine:

They create their listings automatically. Spider builds them. Computer algorithm ranks all pages. These types of search engines are heavy and often retrieve a lot of information. For complex search it allows to search within the results of previous search and enable you to refine search results.

2) Human Power Directories:



These are designed by human selection means they depend on professional to create listings. These never contain full text or webpage they link to.

3) Hybrid Search Engine:

These are different from traditional text oriented search engine such as Google or directly based searched engine such as Yahoo in which each program operates by comparing a sets of metadata

Search Engine Optimization

Search Engine Optimization is the procedure of improving the visibility of a website or webpage in search engine via the natural or unpaid searched results [11]. Optimization may target different types of search like image search, local search, video search, academic search, new search, industry specific vertical search .It can also be define as the process of affecting the visibility of a website or webpage in search engine.

In search engine optimization updation or modification of all variables to get a better location in the search engine take place. We start with Search Engine Optimization and how it can be used to formulate internet marketing. Strategy as well as Technical aspects of SEO. [4]

- A) Using SEO as a marketing strategy it can be described as a method of getting our website to rank higher in search engine as Google; Yahoo,
Means that if user likes to search for a list of optimized keywords the chances are that the visitors see your site on first few places may be good.
- B) Parameters for evaluating SEO of websites- Page Rank- Page rank of each page depends on the page rank of pages pointing to it.
- C) To enhance our site page rank few key ideas are inbound links, outbound links, Dangling links, domain and File names and Broken links [3]

Requirement of data mining in Search Engines

There have been several studies of web crawling in its relatively short history, but most of them have had a focus rather different from ours. Some have concentrated on aspects relating to caching, e.g., [13] and [9]. Others have been principally interested in the most accurate and effective way to update axed size database extracted from the web, often for some specific function, such as data mining, see e.g. the work of Cho et al. [5, 6, 7]. These studies were performed over time periods ranging from a few days to seven months. However, for differing practical reasons, these crawlers were restricted to subsets of web pages. Several authors, e.g., Co man et al. [8], approach crawling from a theoretical point of view, comparing it to the polling systems of queuing theory, i.e., multiple queue-single server systems. However, the problem equivalent to the obsolescence time of a page is unexplored in the queuing literature. A common assumption has been that page changes are a Poisson or memory less process, with parameter λ as the rate of change for the pages. Brewington and Cybenko [1] and Cho and Garcia-Molina [6] con_rm this within the limits of their data gathering. This is somewhat undermined by another study, based on an extensive subset of the web by Brewington and Cybenko [2] showing that most web pages are modified during US working hours, i.e. 5am to 5pm (Silicon Valley Standard Time), Monday to Friday. Nevertheless, the widely accepted Poisson model forms the basis for a series of studies on crawler strategies. These lead to a variety of analytical models designed to minimize the age or maximize the freshness of a collection by investigating:

how often a page should be crawled in what order pages should be crawled should a crawling strategy be based on the importance of pages or their rates of change?

Proposed Architecture

To represent our models for classification based search, we require a search engine to integrate a classification module. We will describe architecture of such a classification-based search engine which is optimized

We have use meta search engine, it combines results from one or more search engines and presents one ranked list to the user through a given common interface. The main profit of meta search engine over each individual search engine is improved performance. All search engines cover only a small portion of the size of the web. Searching multiple search engines simultaneously via meta search will improve the coverage significantly. If combined with vertical search engines, meta search can improve the precision of individual search engine to some extent, this relates to the query classification problem.[14].

Another advantage is the meta search does not require any crawler and indexing because meta search provides a free independent platform to access the commercial search engines. One issue with meta search engines is that they are at the mercy [13]

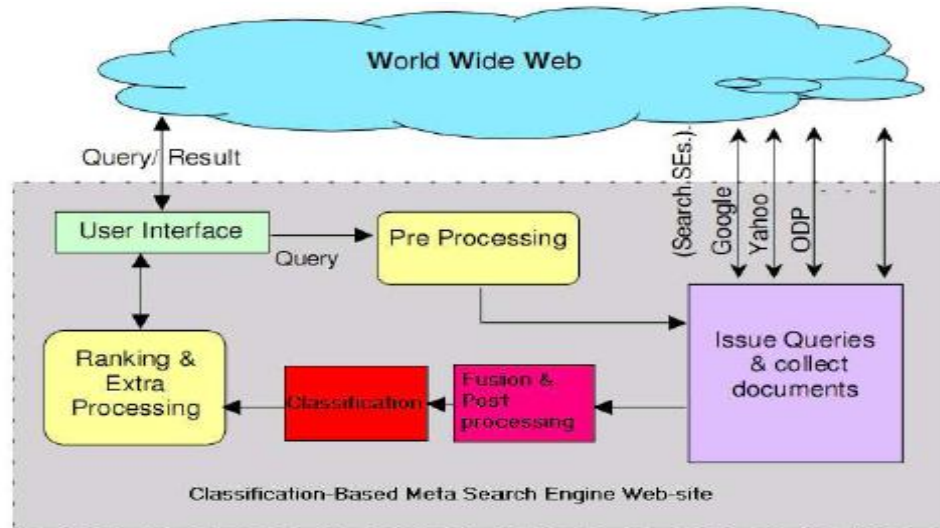


Figure : An architecture of a classification-based meta search engine.

of the search engines they wish to query, and some have banned meta search engines from free riding on top of them. This leads some meta search engines to negotiate a commercial arrangement with the major search engines [16].

The classification-based meta search engine is similar to a typical meta search engine except that it consists of one extra module the classifier, [10] hence the ranking module involves the additional computation of ordering classes and documents within each class.

The architecture of a classification-based meta search is shown in Figure The classification module involves two parts: query classification and results classification.

Future Work

This Paper made use of the class distribution of query classification with newly derived class features to provide a ranking for classes. Most previous work had used class size or the alphabet of the class label to rank the class. Two types of queries were described in the experiment: one-click queries and multiple-click queries. The required information of one-click queries is the result clicked on, while for multiple-click queries, we assumed the last click through was the target document.

It can have a better performance for documents with low rank and the same performance as traditional search engines for documents with high ranks, which form the bulk of click through.

Acknowledgement

It gives me immense pleasure in Research on topic An approach for Search Engine Optimization using Classification - A data mining technique. I would like to express my special appreciation and thanks to my advisor Professor Dr. S.K.Gupta. Department of Computer Sci. & Engg. , BIET,Jhansi, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scholar. Your advice on both research as well as on my career have been priceless. I would also like to take this opportunity to thank Dr. Navneet Singh Sikarwar, Editor In Chief, for their very helpful comments and suggestions..

References

- [1] A web crawler design for data mining' Mike Thelwall Journal of information Science, 27 (5) 2001 PP. 321.
- [2] A Heydon and M. Najork Mercator- A scalable extensible web crawler, World Wide Web 2 (1999) 219-229.
- [3] M. Thelwall, web impact factors and search engine coverage, Journal of documentation 56(2) (2000) 185-189.



- [4] Koyoro Shadeo, Trends in web Based Search Engine 'Journal of emerging trends in computing and information Sciences' Vol 3, No-6, June 2012, ISSN – 2079-8407.
- [5] Dr. S. Sarvankumar, A New methodology for search engine optimization without geeting sandboxed 'International journal of Advanced research in computer and communication Engineering Vol 1, issues, Sept 2012, PP- 472-475.
- [6] Reinhard O,otto and Johnson, Search Engine Optimization for the website of a South African Organisation as an Internet Marketing strategy, Proceeding of International Multiconference of Engineers and Computer Scientists 2012, Vol1, IMECS-2012, March, 14-16. HongKong, ISSN. 2078-0966.
- [7] I.va Gregurec, Petna Grd' Search engine optimization website analysis of selected faculties in croatia. Central European conference on information and intelligent systems, sep 19-21, 2012 PP-213.
- [8] M.P.S. Bhatia, 'Discussion on web crawlers of search engine Proceeding of 2nd National Conference on Challenges and Opportunity in Information Technology, March 29-2008. PP. 227-229.
- [9] Muhammad Akram, Search Engine Optimization Techniques Practiced in Organisations, "A Study of Four Organisation", Journal Of Computing, Vol-2, Issue-6, June-2010, ISSN- 2151-9617.
- [10] Meng Cui, 'Search Engine Optimization Research for Website Promotion' , International Conference of Information Technology computer Engineering and Management Science, Dalian Chine, 2011, PP-100-103.
- [11] Alex Berson, Stephen J. Smith, Data Warehousing, Data mining and OLAP, TATA McGRAW Hill, ISBN-0-07-058741-8 PP-418.,
- [12] R Singh, S K Gupta "Search Engine Optimization - Using Data Mining Approach" International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 9, September 2013, PP 28-32
- [13] Zheng Zhu "Improving Search Engine Via Classification" Dissertation Submitted to Birkbeck College, University of London
- [14] R Singh, S K Gupta "Search Engine Optimization Using Crawler Optimization- A Technique" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 4, July – August 2013, 393-395
- [15] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41:288,297, 2011
- [16] M. Levene. An Introduction to Search Engines and Web Navigation. Addison Wesley, Essex, England, 2005

AUTHORS:-



Rajesh Singh Received a B.E. in Computer Sci Engineering from Faculty of Engg. & Technology, RBS College, Bichpuri, Agra, in 2000 and M.Tech in Information Technology from Karnataka State University, Karnataka He is currently a Ph.D. candidate in Computer Science at Bhagwant University, Ajmer (Raj.). Some of his research interests include the Data Mining, Computer Architecture, human computer interaction, Search Engines, scalability of information access interfaces.



Dr. S.K. Gupta is presently working as A.P. in Comp. Sc. & Engg. Dept. of BIET, Jhansi, U.P., India.. Initially graduating in Computer Science & Engineering from H.B.T.I., Kanpur, U.P., INDIA in 1993 and then M.E. from M.N.R.E.C., Allahabad, U.P., India in 2001 and after that completed Ph.D. in Computer Science & Engineering from Bundelkhand University, Jhansi, U.P., India in 2010. 05 Ph.D.'s are going on. 04 M.Tech. thesis and 40 B.Tech. /M.C.A. projects supervised. Published 10 Papers in International Journals, 01 in International Conference, 06 in National Conferences, attended 0International Conference, 01 Winter School, 06 STTP, 09 International/National Workshops / Conferences. Other responsibilities i.e. Joint Controller of Examination, Mahamaya Technical University,Noida, Head, Deptt. of I.T. at BIET, Jhansi, Head Examiner, for different subjects of C.S. and I.T. in Central Evaluation of U.P.T.U., Lucknow, U.P. Member of Inspection Committees, constituted by UPTU, Lucknow for approval of New Institutions / Courses / Extension of Approval / Seats time to time Member of Board of Studies / Paper Setter / Practical Examiner in different Institutes / Universities time to time i.e. UPTU, Lucknow, RGPV, Bhopal, M.P., Jivaji University, Gwalior, M.P., MGCGV, Chitrakoot, M.P., B.U., Jhansi, Jaypee Institute, Guna, M.P., MITS, Gwalior, M.P. etc. Approved Counsellor / Project Evaluator of I.G.N.O.U. for M.C.A. and B.C.A. courses