A Literature Survey on Keyword Query Processing in Web Search Engines

Kalyani Wagaj[#1], Prof.Subhash Pingale [*2]
#Computer Science And Engineering Department, Solapur,
SKN Sinhgad College of Engineering,Korti,Pandharpur,India

Abstract—**This survey paper is for web search engines to get a good organic rank over Internet. Is a kind of SEO analysis for web sites where search engines uses keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. A multilevel scoring mechanism is proposed for computing the relevance of routing plans based on scores at the level of keywords, data elements, element sets, and sub graphs that connect these elements., This processing of the keywords from the user query helps to get the URLs with all relevant sources of web data. Further, result of query and analysis for the list of search engine result page (SERP) provides similar functionality as search engine optimization method.**

Keywords—**Search Engine Optimization, Search Engine Result Pages,Page Rank.**

I. INTRODUCTION

Search engine algorithms take the key elements of a web page, including the page title, content and keyword density, and come up with a ranking for where to place the results on the pages. Each search engine's algorithm is unique, so a top ranking on Yahoo! does not guarantee a prominent ranking on Google, and vice versa. The number of potential results may increase exponentially with the number of sources and links between them. Yet, most of the results may be not necessary especially when they are not relevant to the user. The routing problem, we need to compute results capturing specific elements at the data level. Routing keywords return all the source which may or may not be the relevant sources.

A search engine is a web-based tool that enables users to locate information on the World Wide Web. Popular examples of search engines are Google, Yahoo!, and MSN Search. Search engines utilize automated software applications (referred to as robots, bots, or spiders) that travel along the Web, following links from page to page, site to site. The information gathered by the spiders is used to create a searchable index of the Web.

Every search engine uses different complex mathematical formulas to generate search results. The results for a specific query are then displayed on the SERP. [8]
The face of SEO is changing, and it's mainly due to incredible research and development into artificial intelligence (AI) baked into search algorithms. Early times, search engine marketers could create a site with some keywords in specific places, drop some links around the web, and start ranking within a short amount of time.

Future of SEO, is in hand of Google's market. Google wants to give a variety of websites to search engine users. It's important to understand that the search engine user is Google's customer. Once an SEO is able to identify the quality of a website from the searcher's point of view, he's able to build a better brand that attracts both human customers and bots.

Google continues to improve its search engine algorithms through AI and spokespeople for the search engine have continually confirmed that its goal is to "understand" content in relation to user experience, back links, and user behaviour patterns. Instead of ranking for one common factor, the future of SEO should take all factors into consideration. This is, after all, the goal for AI and search engine optimization[1].In this paper, section II describes the SEO analysis factors.

II. WORKING PRINCIPLE

Search engine optimization plays a critical role in web commerce. Without this set of techniques, most websites would be unable to acquire high rankings in searching results.

1. Building web crawler to provide the list of web pages or URL's.
2. Prepare soft code for the routing plan for all the entered keywords in query.
3. Find the inter relationship at multi levels of the keywords with the use of available routing plans.
4. Applying the relevance ranking order to the sorted SERP's(Search Engine Result Pages).

Finally integrating all the above modules to build whole system as per the implementation requirements some added functionality modules can be included also.

A.SEO for Website
The Internet is made up of content and most of that content is in written words. Most engines function by reviewing this written content to determine which web pages are the most pertinent to any given search. The purpose of these searches is to find answers to questions. Any search engine or directory exists to help users answer these questions as quickly as possible. Ideally, the first result of a search is the perfect answer to such questions. Statistically, the first result of any search gets more clicks than any other. These clicks turn into money either through users clicking on

advertisements on highly ranked websites or through the direct purchases of products sold on those web pages.

1 .Beat Your Competition
Optimized websites are the fastest, easiest way to dissolve your competition. But the fact is Web users rarely search through pages and pages of search engine listings to find your business. It's a race to the top of the listings, so who will get your potential clients' attention first.

2. Cost-Effective Marketing
SEO is one of the most affordable and effective marketing strategies when you work with the right professionals. At Web Marketing Pros, we make sure you have the right SEO and design package for your company so you are spending money on services you don't need.[7]
A set-level data graph can be derived from a given schema or a generated pseudo schema. We consider the search space as a set of Linked Data sources, forming a web of data.

The search space of keyword query routing using a multilevel inter-relationship graph. The inter-relationships between elements at different levels are above A keyword is mentioned in some entity descriptions at the element level.

Entities at the element level are associated with a set-level element via type. A set-level element is contained in a source. There is an edge between two keywords if two elements at the element level mentioning these keywords are connected via a path. We propose a ranking scheme that deals with relevance at many levels.[2]

III. .LITERATURE SURVEY
Linked data sources are efficient way to store web data with use of RDF, rather RDBMS and Keyword relationship graph at set and element level are to be searched. Finally multi-interrelationship between different elements is used to find out ranking order of the retrieved documents. Routing keywords only to relevant sources can reduce the high cost of searching for structured results that span multiple sources. The routing plans, produced can be used to compute results from multiple sources.[1]

Space Search Ordering algorithm is taken to perform the aggregation on the data set and finds out relevance score for the each keyword in entered search query. This is similar to OLAP analysis on data cube, searches the relevant k cells on cube by the space search ordering approach.[2]

Here, RDF is modeled as undirected data graph and Minimum keyword tree generation method is used to reduce the searching complexity. Searching and retrieval of data item is done from generated minimum keyword tree and at the end to put ranking on result score function is used[3].

Ranking keyword search results with Collective Importance Ranking called as CI-Rank. Random walk with message passing(RWMP) approach is used to find out the cohesiveness in the result tree of data retrieved. CI –Rank will help to find out the better ranking, as it can be later useful for finding precision of the pages to calculate mean reciprocal rank,as it helps for relevance of web pages[4].
For the RDBMS search area, CI rank is used to check the performance of search engine basically these databases are normalized to remove redundancy and compare normalization used for wed data.[5]

To find general approach to study learning objectives via search engines like SCORM and CORDRA which are different from the general purpose search engine as Google. And also to compare ranking order this add use and generalized structure for SERPs Search Engine Retrieved Page Links.[6]

Referred Paper	Purpose	Conclusion
Keyword Query Routing IEEE Transactions On Knowledge And Data Engineering, 2014.	Routing plans preparation methods are provided at element and set level graphs of web data.	If no any semantic between keywords is found methods will not be applied.
Efficient Keyword Based Search for Top K-cells in Text cube IEEE Transaction vol 23, no 12, DEC 2011.	It gives idea about keyword search on DB for top k relevant result similar working as RDBMS	This is not applied for all the database schemas except RDBMS structure.
Paper2:Keyword Proximity Search Over Large and Complex RDF Database 2012 IEEE Transaction.	To find search procedure on RDF database reduces searching complexity.	Dynamic search is adapted for RDF databases.
CI-Rank –Ranking Keyword search results based on collective Importance 2012 IEEE Conference.	Used for ranking keyword search results and cohesiveness of the answer is kept limited .	Random walkthroughs message passing scheme is complex to find CI rank.

Table1.. Summary of Literature Review

IV. SCOPE
1. Web information systems such as search engines have to keep up with the growth and changes of the Web. For marketing, research and in business: To get a better handle on search engine optimization, it's important to understand why people use search engines, at all. Generally, people use search engines for one of three things: research, shopping, or entertainment.

2. In Search Engine Optimization Companies When people search for your products and services, you obviously want to appear as high in the search engine rankings as possible, but the reasons for this are more than just because you want them to click through to your website. In fact, there is a certain amount of value in simply appearing in search results for terms directly related to business.

3. For instance, most searchers don't just simply search once, click on some websites, and be done with it. Instead, they search, click on some websites, edit their search terms, search again, click on some websites, further repeats their search terms, search again, and so on. To avoid such repeating of search procedure the system will provide the optimized search for the query[7].

4. This system can also be used as efficient method for big data analysis and come up with better solution for SEO techniques used over web analytics.[9]

V. CONCLUSION

Now a day's large volume of data on web is generated, to keep the relevant data available as result to user keyword queries, this system will be framework of web analytics and e commerce applications to get the marketplace value on basis of page popularity metric.

REFERENCES

[1] Thanh Tran And Lei Zhang, "Keyword Query Routing",IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014.

[2] Lin,han,zhai,"efficient On Knowledge And Data Engineering Keyword Based Search For Top K-cells In Text Cube" ,IEEE Transaction Vol 23, No 12, Dec 2011.

[3] Niu,zheng,xia And Li,"Keyword Proximity Search Over Large And Complex Rdf Database",IEEE/WIC/ACM International Conference On Web Intelligence And Intelligent Agent Technology, 2012.

[4] Yu And Huxia,"CI-RANK –Ranking Keyword Search Results Based On Collective Importance", IEEE International Conference On Data Engineering 2012..

[5]J oel Coffman, Alfred C. Weaver,"An Empirical Performance Evaluation of Relational Keyword Search Systems", IEEE Transactions on Knowledge and Data Engineering, (Volume: 26 , Issue: 1) Year:2014.

[6] Yen, Shih and Chao,"Ranking Metrics and Search Guidance for Learning Object Repository", IEEE Transactions On Learning Technologies, Vol. 3, No. 3, 2010.

[7] http://firstviewonline.com/why-seo-is-important-for-businesses/

[8] http://www.webseoanalytics.com/free/seo-tools/serp-analysis.php

[9] http://searchengineland.com/guide/what-is-seo

[10] http://www.pimall.com/nais/n.engine.html for Search Engine Strategy, Tips And Techniques.