



## Website Link Analysis by Keyword

### KEYWORDS

Website, Search Engine Optimization(SEO), Crawling, Reverse Engineering approach, Analysis.

**Kansagra Deep Mukeshbhai**

School of Information Technology & Engineering, VIT ,University Vellore, India.

**Baldaniya Paresh Mavjibhai**

School of Information Technology & Engineering, VIT ,University Vellore, India.

**Kamalakkanan J.**

School of Information Technology & Engineering, VIT ,University Vellore, India.

**ABSTRACT** In the world of the Internet Website to have a good rank when any of the query/keyword is been search in any search engines[17] but for having weak rank they should have a good analysis for different types of search queries. The Analysis on the particular search query(i.e search terms , people ,things etc.) this all are based on many applications , tools and different methods such as analysis of data(data mining), search engine optimization[17] and different tools available on the internet like analytic tool by Google . there is a high dependency on above products we need some extra research on that products for SEO(Search Engine optimization) field In this paper we are doing same things in different way which is above tools are doing. To improve the rank of website in this paper we have have applied reverse engineering approach i.e based on analysis trying to improve the rank of the your website in the search engine. In this paper we are focusing on crawling the search engine for particular input parameter(search query, search engine ,website etc.) and then applying some rules and doing analysis on that particular given input parameter.

### 1.Introduction

In world of Internet your website have good rank in the any search engine , if you search any query on search engine your website should come on first page ,that is all about SEO [7], hows your website is googlized[21] ,so we need good SEO work on website.

In this paper we are covering this aspects by reverse engineering of SEO, in this paper we are analyzing the links rank for the particular query and gives the report to the SOE teams that team works on this report and by different method they are trying to improve the rank of the that website for that query , in this paper we are searching for particular search query on search engine by considering some parameters like search engine , number of pages to be crawled , website on that we are doing research and IP address of the search engine and for that particular search query we are searching on the search engine and counting the rank of the that website which is coming in search engine results for particular search query and gives that report for SEO .This counted results storing in the database for further analysis. by this data we are generating the different kind of comparison graph and generating business reports of different kind.

Here also we fetching the result from the database and comparing different dates data. Same things we are applying for different search engine.

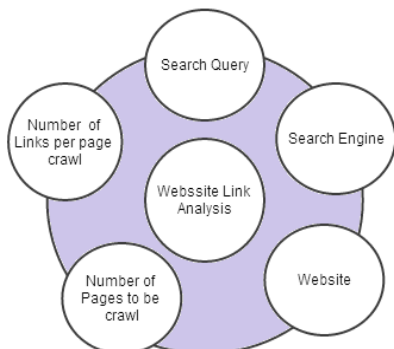


Fig.1 Model For Website link Analysis by Keyword

Above Model for the Website Link Analysis are explained in graphical way . Here we are first fix some input parameters and then applying on this tools and generate the results and store in database . and generating different kind of reports.

Here considering all those inputs parameters and we are counting results by ignoring those data/content which is irrelevant to topic, means we are ignoring those links like image links , back links , video links [17] etc.

### 2.RELATED WORKS

Here we are going to explain some of the related tasks to my paper ,Google Analytics[21] , Google Webmaster Tool[18],Google Trends and Google Insights for Search[19], Alexa Internet [20] , above motioned tools are help to SOE(search engine optimization) for the analysis of your website , improve your marketing strategy ,business etc. this paper (WEBSITE LINK ANALYSIS BY KEYWORD) that is one of the part of the Analysis tool that helps to the SEO for analysis of websites link rank[17].

#### A .Google Analytics.

It is the free analytical service it is helps to SEO for the analysis of the website it is gives the different kind of reports that how the end user touch with any portal , in this google analytics lots of features are available , some of the feature of this tools are given bellow It is monitoring the website and gives the result like number of users , number of pages viewed , traffics of your website even traffic of particular pages, module etc. it is also gives the reports like list of visitor that often visiting your site.

#### B. Google Webmaster Tool

This tools gives the reports that behavior of website that is seen by the google search engine, it is also free of cost service by the google some of the features of the webmaster tools are given bellow .It inform to the end user that which pages or module are not able to crawl.Website owner can submit the sitemap and also check the site map .Display the how many pages , module or sites that is blocked by the google.Liked page to other site.Error in any HTML pages.It also gives the like for particular search query how many times your site visited, number of listing clicked of your website on google etc.

### C. Google Trends and Google Insights for Search

Google Insights for Search[19] is launched and after that it is merged with Google trend[19]. Some of the features of this tools are given below. It is used to analyze the traffic for particular search term i.e popularity of the any person or any things. For comparison of the between term that is search on google search engine , for the time period graph like which particular keyword search frequently , for particular time period and for particular region .

### D. Alexa Internet

This tools is same like other google tools that gives the different kind of analytical reports for the business purpose.it is from Amazon it gives the report like overall traffic . it provides most of functionality that provides google tools in addition is also provides lots of developer tools ,APIs etc. It is provide alexa rank that is globe to all website[20].

## 3.WLAK(Website Link Analysis By Keyword) Relationship MODEL

### I.Website Link Analysis By Keyword

On internet many tools and software available for examination of the website.Here we are presenting one of the website link analysis tool that help to the SEO , sketch of this tools given below on the based on considering some parameters like search engine , search queries , websites etc . i.e here in this paper I firing particular search query (ex. "sony vio"), on different search engine like google ,yahoo or bing and considering specific website counting the hyperlink rank [17].

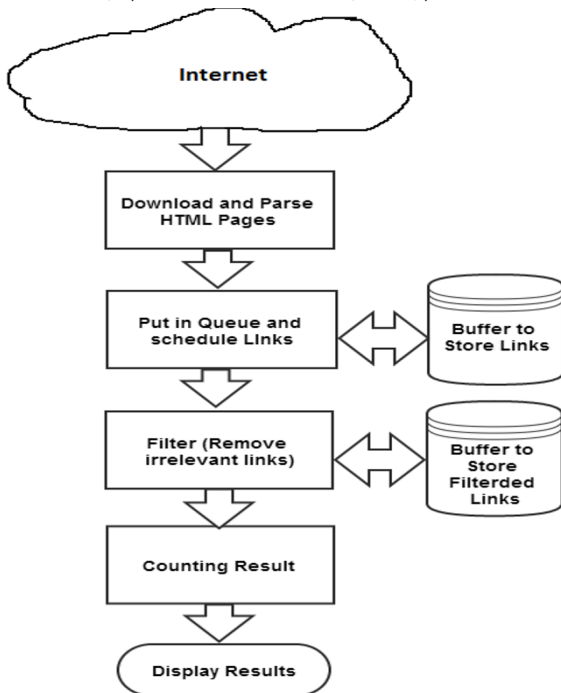


Fig2 .Graphical Represents of tool

### II .Download and Parse HTML Pages

Considering the search query , search engine, IP address of the search engine and some other parameters like number of pages to be parse , number of links per page [17] etc. for each search query this tool analyze the result ,here crawling the internet on bases of given above parameters , there is lots of difficulties during the crawling process because On the internet wide verities of content hyperlinks are available like violation and lots of tags and similar names[17] etc . how we are crawling the internet that is totally depend on algorithm logic and search engine pages pattern . on changes of time period we have to change the crawling algorithm because pattern of search engine pages changes frequently.

**III. Put in Queue and Schedule Links.** Put this crawled links in to the queue and schedule this links ,this links temporary stored in the buffer . this links combinations of different kind of links including image links , adds links , news links video links [17]etc. scheduler schedule the links and put it in to the buffer temporary for the further use . further scheduler fetches these links from the buffer and go ahead for the next step for filtration of the links.

### IV. Filter (Remove Irrelevant links)

From previous step we have scheduled links from the buffer and putted in to the filter for removing the irrelevant links from the crawled links. This is the main part of the our tasks and it is very challenging process to remove unrelated content, here in this process we are dividing the two partition one for relevant links and another irrelevant links that are not part of our paper , these relevant links process for further study and we are putting in to temporary buffer for use of next operation . In next step we are counting the result.

### V. Counting Result.

After filtration process , we are fetching these links from the buffer and counting the result of the given input website that and search for that website links from the buffer and counting the number of the links[17].

### VI. Display Results

On the bases of the previous results we are displaying the results. For displaying result we are using different technique i.e result displaying by graph , by excel sheet etc.

## 4.Algorithm

### I.Initialize the variables

- i.Parsed URL
- ii.Count of URL
- iii.Website Domain to be searched
- iv.Search Engine (Google.com/Google.co.in/Yahoo/Bing)
- v.Number of page to be parsed
- vi.Number of link to be parsed per page (i.e per page 10 links or 20 links etc. )
- vii.IP address of search Engine.
- viii.Search query /keyword (Samsung mobile ..etc)

### II. Repeat

- III. Based on IP address of the search engine and considering search engine ,number of link to be parsed per page(i.e 1 to 10 link per page etc.) download HTML page for particular Search query.
- IV. Store this(Unfiltered) HTML page in buffer for further use.
- V. Repeat this process till given rang of pages to be parse
- VI. Start filtering process of links from stored HTML page

Using particular algorithm filters all irrelevant links from HTML page. Iterate page URL one by one and filter other links like Adds link, duplicate links , Image links etc. and store this filtered link in other buffer.

7. Repeat this step for other page 20 to 30 , 30 to 40 ...till given range etc.

8. Iterate this Stored filtered URL and match this URL domain with initialized website Domain, if found then count the place of that link

9.Display that count (Result).

### 5.Experimental Results

A1. This Graph generated from the fetching the results from database for particular keyword and different websites for different search engine , here we are also adding one more parameter "date " , on different date we are displaying results by fetching the results from the Database.

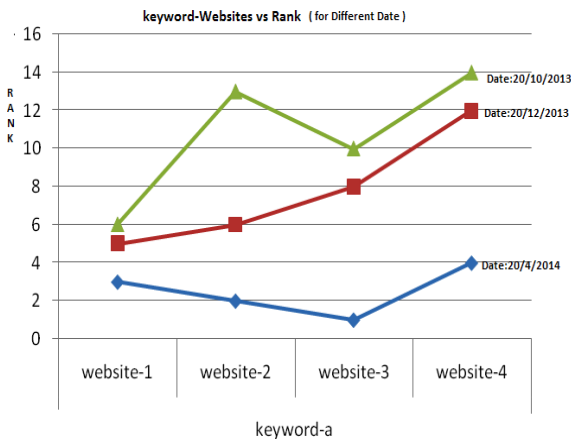


Fig.3 A1 Keyword –website vs ranks Graph for Different Dates.

A2. Here keyword-website vs rank graph are given for the different keywords and websites i.e we are considering some bunch of keywords and some websites on which we want to experiment , and represent in graphical way by fetching the data from database.

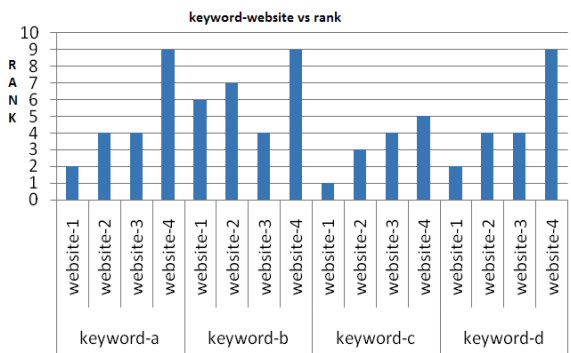


Fig.4 A2 Keyword –website vs ranks Graph for different search queries(keyword) and websites

A3. Comparison graph of Keyword –website vs ranks Graph for different search engine

Here we are doing same experiment that explained above doing for different search engine i.e yahoo, google and bing

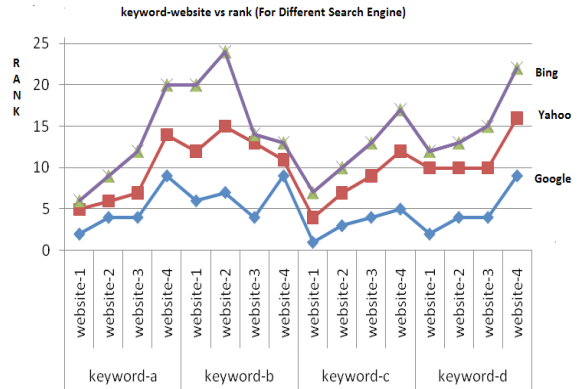


Fig.5 A3 Keyword –website vs ranks Graph for different search engine

Note: Here all results are dummy and taken on the assumption for understating the paper work.

6.Conclusion and Future Enhancement

It is hard to represent the accurate result of the SEO(search engine optimization) for the particular search queries , website and search engine . By using different search engine here we are comparing results of different search queries and represent in graphical way and counting the rank of the links.

As time goes we have to update the algorithm of our paper as for more accurate result then previous result and algorithm of search engine is changing very frequently and also we will add some extra features and graphical represent of data will be improve.

REFERENCE

[1] T. Alan Keahey, Stephen G. Eick, "Visual Path Analysis," IEEE Computer Society Proceedings of the IEEE Symposium on Information Visualization 2002 (InfoVis'02). | [2] D. Hati, B. Sahoo, A. Kumar, "Adaptive Focused Crawling Based on Link Analysis," 2nd International Conference on Education Technology and Computer (ICETC) v-4 pp.455-460. | [3] A. Pal, D. S. Tomar and S.C. Shrivastava, "Effective Focused Crawling Based on Content and Link Structure Analysis", (JCSIS) International Journal of Computer Science and Information Security, Vol. 2, No.1, June 2009. | [4] Y. Sun, P. Jin, L. Yue, "A Framework of a Hybrid Focused Web Crawler" Second International Conference on Future Generation Communication and Networking Symposia 2008 pp.50-53. | [5] Y. Jin, J. Zhang, P. Ma, W. Hao, S. Luo, Z. Li, "Applying PageRank Algorithm in Requirement Concern Impact Analysis," | [6] 33rd Annual IEEE International Computer Software and Applications Conference 2009 pp.361-366 | [7] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Digital Libraries SIDL-WP-1999-0120, 1999. | [8] Ka-I Pun, Yain-Whar Si, "Audit Trail Analysis for Traffic Intensive Web Application," IEEE International Conference on e-Business Engineering 2009 pp.577-582. | [9] M.Cui, S.Hu, "Search Engine Optimization Research for Website Promotion," International Conference of Information Technology, Computer Engineering and Management Sciences 2011 pp.100-103. | [10] A. Grzywaczewski, R. Iqbal, N. Shah, A. James, "E-Marketing Strategy for Businesses," IEEE International Conference on E-Business Engineering 2010 pp.428-434. | [11] I. S. Altıngöçde and Ö. Ulusoy, "Exploiting Interclass Rules for Focused Crawling," IEEE INTELLIGENT SYSTEMS and Published by the IEEE Computer Society 2004. | [12] Jyh-Jong Tsay, Chen-Yang Shih, Bo-Liang Wu, "AuToCrawler: An Integrated System for Automatic Topical Crawler," Proceedings of the Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05) 2005. | [13] Shwu-Min Horng, "Analysis of Users' Behavior on Web 2.0 Social Network Sites: An Empirical Study," Seventh International Conference on Information Technology 2010 pp.454-459. | [14] Dilip Singh Sisodia, S. Verma, "Web Usage Pattern Analysis Through Web Logs: A Review," Ninth International Joint Conference on Computer Science and Software Engineering (JCSSE) 2012 pp.49-53. | [15] B. Vrdoljak, G. Gledac, Z. SkoCir, "An Application for Multidimensional Analysis of the Web Site Traffic," at 10th Mediterranean Electrotechnical Conference, MELECon 2000, vol.1 0-7803-6290-X Published by the IEEE, pp. 383-386. | [16] CHEN Li, Li Zhi-shu, YU Zhong-hua, HAN Guo-hui, "Classifier-guided topical crawler: a novel method of automatically labeling the positive URLs," 2009 Fifth International Conference on Semantics, knowledge and Grid Published by the IEEE Computer Society 2009 pp.270-273. | [17] C. Zhou, X. Sun, H. Guo, "The URL Search Strategy Based on the Content and Link Analysis," 978-1-4244-4507-3/09 IEEE 2009. | [18] Google,Google Webmaster Tools, 2014, <https://www.google.com/webmasters/tools>, accessed on 5/3/2014. | [19] Google,Google Trends and Google Insights for Search, 2014, <http://www.google.com/trends/>, accessed on 2/2/2014. | [20] Amazon,Alexa Internet, 2014, <http://www.alexa.com/>, accessed on 1/2/2014. | [21] Google, Google Analytics, 2014, <http://www.google.com/analytics>, accessed on 10/3/2014.