

Website Quality Evaluation Based on Search Engine Queries using Web Rank Position Algorithm (WRPA)

Chandran M*, Ramani A.V

Department of Computer Science, Sri Ramakrishna Mission Vidyalaya,

College of Arts and Science, Bharathiar University

SRKV (Post), Coimbatore, Tamil Nadu, India, +91989432150.

Corresponding author, e-mail: onchandran@gmail.com*,avvramani@yahoo.com

Abstract

The research work is about to test the quality of the website and to improve the quality by analyzing the hit counts, impressions, clicks, count through rates and average positions. This is accomplished using WRPA and SEO technique. The quality of the website mainly lies on the keywords which are present in it. The keywords can be of a search query which is typed by the users in the search engines and based on these keywords, the websites are displayed in the search results. This research work concentrates on bringing the particular websites to the first of the search result in the search engine. The website chosen for research is SRKV. The research work is carried out by creating an index array of Meta tags. This array will hold all the Meta tags. All the search keywords for the website from the users are stored in another array. The index array is matched and compared with the search keywords array. From this, hit to count is calculated for the analysis. Now the calculated hit count and the searched keywords will be analyzed to improve the performance of the website. The matched special keywords from the above comparison are included in the Meta tag to improve the performance of the website. Again all the Meta tags and newly specified keywords in the index array are matched with the SEO keywords. If this matches, then the matched keyword will be stored for improving the quality of the website. Metrics such as impressions, clicks, CTR, average positions are also measured along with the hit counts. The research is carried out under different types of browsers and different types of platforms. Queries about the website from different countries are also measured. In conclusion, if the number of the clicks for the website is more than the average number of clicks, then the quality of the website is good. This research helps in improvising the keywords using WRPA and SEO and thereby improves the quality of the website easily.

Keywords: Queries, Impression, Average position, Search Engine Optimization and WRPA

Copyright © 2016 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Search Engine Optimization (SEO) involves achieving the highest position or ranking in the natural or organic listings on main search engines – Google, Yahoo and Bing. Improving positions in the search engine listings are dependent on marketers understanding the crucial process of search engine marketing [1].

Most professional communication practitioners and researchers can point to some content on the web that they themselves have authored, such as on their employers' or clients' websites, or on sites they maintain for professional, personal, or community interests [2].

A simple experiment, might try to find the content using only a general web search engine. If, as is likely, they can compose a carefully worded search query by recalling very specific features of the content, such as its title, a distinctive key phrase, the name of the website or of the organization that owns it, and so forth, they stand a reasonable chance of success, with their content appearing on the first page of the search results [2].

The search engines are widely used in America and the world's most visited websites and, hence, provide a common intersection for the otherwise distinct interests of web users, search marketing practitioners, researchers, and of course the search engine companies themselves. These classes of search stakeholders contribute in distinct ways to search engine rankings and/or to research and insight into how to optimize sites to boost those rankings. Their contributions will become evident throughout this section, starting with this tutorial's theoretical

framework, which incorporates three classes of these stakeholders, and continuing with a description of how its literature was selected from a slightly different set of three classes of these stakeholders. After introducing basic search-related terminology, the section explains how, from these stakeholders, three classes of participants shape search engine rankings [3].

Users mainly use keywords with search engine to find the target website and meanwhile search engine also uses keywords as search criteria to retrieve website. Therefore the core keywords of a website need to be determined while designing a website, and then the work is expanded to generate keyword phrases and terms around those core keywords. After the core keywords are determined, these core keywords should be also appropriately scattered in the webpage text [4].

1.1. Keyword Identification:

Keyword identification is an important technique for document retrieval. It is a task to identify a small set of words, keys, key phrases, keywords or key segments from a document that describes the meaning of the document. For instance scientific articles, research papers are also often annotated with keywords and also web documents, especially multimedia resources can be already associated with tags. Keywords search is a usable and powerful tool which enables efficient scanning of large document collections. To achieve this, a simple correlation approach is used for deriving keywords.

1.2. Related Content

Keywords must be associated with the content of website. Even though many unnecessary keywords can bring in a lot of search hits the keywords are not well content related with the webpage and the traffic cannot be converted into the actual use of the webpage. This traffic is actually a burden to website because it consumes bandwidth.

1.3. Large Quantity of Searches with Less Competition

The best keywords are those of the most frequently searched but the least competed words. However, the majority of keywords being searched more are also the more competitive keywords. But those keywords with relative more searches and less competition can still be found through a great deal of detailed keyword study and extension. Both of Google Keyword Tool and Baidu Index provide data of search numbers on keywords.

1.4. Keywords should not be too special

If the words are too special or too long, the number of searches will be greatly reduced, even nobody searches the website so that these words cannot be used as website keyword [4].

2. Problem Definitions

2.1. Dynamic Webpage

The search engine assumes that a file generated by the dynamic page has more variable factors and a lot of uncertainty in content. Therefore, the static web pages have a higher priority when indexing and ranking web pages, so it is better to use the static webpage in developing SEO. URL needs to have some minor changes and remove special characters, such as "?", "=", "&", and "%" when using SEO [5].

2.2. Link Optimization

Incoming Link is essential for website ranking to have as much high quality external links (also known as incoming links) as possible. Therefore, a link analysis tool can be used to analyze the situation of how the website is linked.

2.3. Outgoing Links

Outgoing link means link to other websites. The search engine not only analyzes incoming links but also analyzes outgoing links to other websites. If the content of outgoing linked website is associated with original website, it will also be good for search engine rankings.

2.4. Internal Links

The webpages within a website should also be linked to each other, like the "related articles" list at the end of a regular article, which links all other articles under the same subject on this website. In this way it not only facilitates the visitors, and increases traffic, but also is conducive to search engine, to index webpages, and links multiple webpages to the same internal webpage, as well as helps improve ranking for that internal page [5].

Web-server-performance monitoring is an increasingly important field of research. Our review focuses mostly on existing software tools. Monitoring a Web server is usually performed by analyzing its log files, which holds information about each single HTTP request and the consequent response and errors. Three Classes of Participants are Shaping, Search and Ranking. The search engine rankings are directly and indirectly shaped by the three classes of interdependent participants.

1. Search engine companies and programmers
2. Webmasters and SEO practitioners
3. Search engine users

A web application can have many thousands of more users than a conventional, non-Web application. The immediacy of the Web creates an immediate expectation of quality and rapid application delivery, but the technical complexities of a website and variances in the browser make testing and quality control more difficult, and in some ways, more subtle. Automated testing of Websites is both an opportunity and a challenge. A Website is like any piece of software that is single, but all inclusive quality measure applies, and even multiple quality metrics may not apply.

2.5. Website Quality Status

This is used to test the website quality and also to find whether it contains any errors or not. This can be checked in web-tool. The web tool consists of three categories such as Crawl errors, crawl and status. Site maps, Crawl errors are used to find the errors that are in website. The crawl errors are displayed for past 90days like DNS error and Server error [6].

DNS Error-DNS error is those errors that are displayed when there is a DNS server problem. This is displayed for the last three months. The graph reading is taken based on months and on the number of errors occurred. In this error, some URL errors are occurred due to Server error, Access denied error, File not found error and other errors [6]. Website's visibility in search engines is vital factor for high traffic rank. A well-designed site will not be successful if the potential users cannot find the site [7].

Extract all the uncommon words from the web pages downloaded by the crawler and record the URL where each word has occurred. The result is stored in a large table containing URLs pointing to pages in the repository where a given word occurs [8]. SEO reports-By connecting Google Webmaster Tools account to Analytics account that can track the number of impressions listings are getting and even click-through rates. First, log into Google Analytics account and click on website URL to view report. The three reports are: queries, landing pages and geographical summary. A web search engine often returns thousands of pages in response to a big query, making it very difficult for users to browse or to identify the relevant information which the user needs [9].

3. Methodology

Table 1. Keyword table from documents

| S.No | Keyword1 | Keyword2 | Page |
|------|--|---|--------------|
| 1 | sriramakrishna mission vidyalaya college of arts and science | ramakrishna mission | 3,6,9 |
| 2 | ramakrishna mission coimbatore | ramakrishna mission vidyalaya coimbatore | 3,6,9 |
| 3 | srkv.org | www.srkv.org | 5,6 |
| 4 | srkv | sriramakrishna mission vidyalaya college of education | 1,3,6,8,9,10 |
| 5 | srkvcas | maruthi college of physical education | 1,3,6,8,9 |

Queries: The keywords are used as a query and each query returns a page. The total number of queries and other information can be obtained from web master tools.

Impressions: This shows the number of times the pages appeared in search result. The number of search result can be increased /decreased for each query.

Clicks: This is the number of times a user clicked the search results for a particular query. The number of clicks for a particular query increases /decreases each time [7].

Average position: The number of the clicks for each keyword is calculated. Then average of keyword clicks is calculated. This shows which keywords are used mostly to search a particular page.

Any search engine can be used to search web content from the website. The content of the website can be assessed from different websites. SRKV website was selected for the analysis. Search Engine optimization (SEO) is used to analyses the search results. A keyword table is maintained for each website. To refer each page, two keywords can be used. In Table1 the list of keywords to access the website is maintained in the keyword.

This also shows the quality of the website. If the number of clicks for the webpage is more than the average number of clicks, then the quality of the webpage is good. Otherwise the quality of webpage needs improvement.

The collection of all the Meta tag is stored in an indexed array of Meta tags which is described in the Algorithm 1. Store all the search keywords from the search engine in an array whenever the page is redirected to our webpage. Indexed array values are compared and matched with the stored search keyword and then that count will be calculated for analysis. All the search keywords and its hit count will be analyzed for improving the performance and SEO search results of the webpage. Visualizing the position and average hit counts for the performance of the web page is based on keyword search.

The special search keywords are included in the Meta tag for increasing the performance of the search results. In this algorithm, we search and match the SEO keywords from that Meta tag. All the specified Meta tag keywords are saved in the index array of the domain. All the SEO keywords are matched and compared with the Meta tag words. If it is matched, then the search keyword will be stored for improving the quality of the website based on Meta tag based on the Algorithm 1.

```

//Collect Meta Tag list from Website
if (Redirect from Search Engine)
m[] ← Get_MetaKeywords
//Get Keyword from search Engine
s[] ← Get_Search_Keyword;
for each (s[] in temps)
{
//initialize count value as zero
count ← 0;
for each (m[] in tempm)
{
if (temps ⇔ tempm )
count = count + 1;
}
//sorting No.of Hits
index [tempm] = count;
}
cur_clicks ← 0;
//Total No.of Clicks
Impressions ← cal_total_clicks;
rclick ← 0;
//Recent Visits
clicks ← Recent_visits;
//Average Value for Position
Avg_Position ←  $\frac{\text{impressions}+4}{\text{clicks}}$ 
//function to calculate clicks
cal_total_clicks()

```

```

{
  for each click
  cur_clicks ← cur_clicks + 1
  return cur_clicks;
}
//function to calculate recent visits
Recent_visits()
{
  for each click
  rclick ← rclick + 1
  return rclick;
}

```

Algorithm 1. Web Rank Position Algorithm (WRPA)

4. Result and Discussion

This is given in Meta tag for easy search. The quality of a site depends on the list of words that is given to be searched. This is given by the user and then uploaded by the site developer. In Table 2 giving the site information or the particular words the site will take up the first position in the search engine.

Table 2. Summary of Top Search Queries

| Query | Impressions | Clicks | CTR | Avg. position |
|--|-------------|--------|-----|---------------|
| srkv.org | 220 | 182 | 83% | 1 |
| srkv | 447 | 114 | 26% | 2.2 |
| www.srkv.org | 165 | 91 | 55% | 1 |
| ramakrishna mission coimbatore | 237 | 71 | 30% | 1 |
| ramakrishna mission | 1095 | 45 | 4% | 8 |
| sriramakrishna mission vidyalaya college of arts and science | 167 | 41 | 25% | 1 |
| ramakrishna mission vidyalaya | 316 | 40 | 13% | 1.7 |
| ramakrishna mission vidyalaya,coimbatore | 126 | 30 | 24% | 1 |
| maruthi college of physical education | 126 | 29 | 23% | 1.1 |
| arts and science result 2014 | 69 | 28 | 41% | 1.5 |
| srkvcas | 104 | 26 | 25% | 1 |
| sriramakrishna mission vidyalaya college of education | 87 | 19 | 22% | 1 |
| sriramakrishna mission vidyalaya | 60 | 17 | 28% | 1 |
| srkvcoimbatore | 21 | 17 | 81% | 1 |
| srmv college of arts and science | 84 | 16 | 19% | 1 |

1. **s[]** - Array variable is for the search keyword.
2. **Hits** - Total site hits counter value.
3. **imp** - Number of times appeared in the search engine for the specific keyword.
4. **Clicked** - Number of times used or clicked the site.
5. **CTR** - Percentage of impressions (Click Through Rate).
6. **Avgpos** - Division of Keyword based on top ranking position with number of times searched with different keywords.

If the site is appeared in search engine at the third top position for one keyword and the seventh top position for another keyword then average position is based on Table 3 is Summary of TopSearch Page.

1. $3 + 7 / 2$
2. $10 / 2$
3. 5

Table 3. Summary of Top Search Page

| Page | Impressions | Clicks | CTR | Avg. position |
|---------------------------|-------------|--------|-----|---------------|
| http://www.srkv.org/ | 9110 | 1483 | 16% | 7.1 |
| /semresults.aspx | 1086 | 190 | 17% | 4.6 |
| /songs1.aspx | 2054 | 147 | 7% | 5.8 |
| /collegeofarts.aspx | 1675 | 140 | 8% | 4.4 |
| /missionlinks.aspx | 3439 | 120 | 3% | 16 |
| /contactus.aspx | 1190 | 115 | 10% | 5.1 |
| /iti.aspx | 752 | 98 | 13% | 10 |
| /collegeofeducation.aspx | 888 | 93 | 10% | 7.4 |
| /polytechnic.aspx | 1225 | 85 | 7% | 6.8 |
| /maruthicollege.aspx | 415 | 77 | 19% | 4.7 |
| /SSHSS_EM.aspx | 655 | 63 | 10% | 11 |
| /vidyalaya.aspx | 318 | 47 | 15% | 2 |
| http://www.srkv.org | 1632 | 36 | 2% | 10 |
| /vidyalayahighschool.aspx | 440 | 30 | 7% | 10 |
| http://www.srkv.org/live/ | 65 | 28 | 43% | 3.7 |

4.1. Search Queries:

The Search Query page shows the web search queries that have been returned from the URLs of a particular site. One can also view the information about the pages which were searched most often in the search results (top pages). Data is compared between the daily averages for the selected period and to the daily averages for the previous period. We have shown the data for the last 30 days by default, but one can also set the date range for the report using the calendar selector which appears below in the Figure 1.

The data is also shown for the different countries from where the queries are coming from. This is important to know where the most queries are coming from and if it comes from different countries then ranking can be improved. The graph showing the queries from different countries are given in the Figure 2.

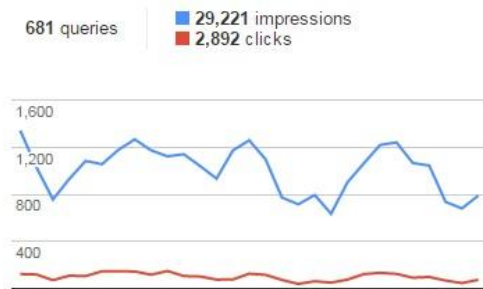


Figure 1. Number of page s (Hits) chart

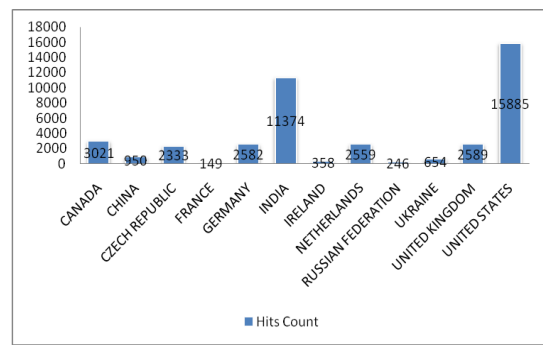


Figure 2. To Find the Location

Queries to view the SRKV website from different web browsers are taken into account. The result shows that the maximum number of user hit are from the unknown browsers, as the mobile users are higher than the other users like Tablets, iPhone, etc. Table 4 shows the data from different browsers.

The platform which is used under different browsers is identified. This shows that the maximum number of user hit are from the unknown platforms which are used under mobile phones and other devices. Table 5 shows the data from the different platforms.

Table 4. Summary of browser used for Webpage

| Browser(Hits Count) | srmvcas | valuesol | Nilessoftware | AES |
|----------------------|---------|----------|---------------|------|
| BlackBerry | 1 | 4 | 1 | 6 |
| Chrome | 4027 | 263 | 311 | 1033 |
| Firefox | 1910 | 142 | 63 | 965 |
| IE | 1230 | 65 | 68 | 475 |
| IEMobile | 32 | 3 | 98 | 65 |
| InternetExplorer | 107 | 44 | 37 | 74 |
| Mozilla | 2399 | 66 | 652 | 102 |
| Opera | 708 | 5 | 36 | 36 |
| Safari | 256 | 3 | 66 | 0 |
| Unknown | 33071 | 42 | 685 | 691 |

Table 5. Summary of platforms used for view

| Platform Hits Count | srmvcas | valuesol | Nilessoftware | AES |
|---------------------|---------|----------|---------------|------|
| UNIX | 174 | 66 | 120 | 896 |
| Unknown | 36735 | 265 | 152 | 1023 |
| Win16 | 3 | 2 | 0 | 4 |
| Win2000 | 83 | 56 | 65 | 102 |
| Win95 | 57 | 5 | 8 | 201 |
| Win98 | 25 | 4 | 42 | 23 |
| WinCE | 12 | 3 | 10 | 6 |
| WinNT | 4415 | 677 | 235 | 365 |
| WinXP | 2237 | 678 | 632 | 963 |

5. Conclusion

This research evaluates the quality of the website using a series of online diagnostic tools and many dimensions of quality of website were examined. The result of this study confirmed that the quality of website is based on the keywords which are important for the SEO (Search Engine Optimization). The SEO will find the best keywords for the better use of the website which will leads to the improvement in the website quality.

References

- [1] Bubanja M. *Search Engine Marketing As Key Factor for generating quality online visitors*. MIPRO, 2010 Proceedings of the 33rd International Convention. 2010: 1193-1196.
- [2] Vignesh J, Deepa V. Search Engine Optimization to Increase Website. *International Journal of Science and Research (IJSR)*. 2014: 425-430.
- [3] Killoran JB. How to Search Engine Optimization Techniques to Increase Website Visibility. *IEEE Transactions on Professional Communication*. 2013; 56(1): 50-54.
- [4] Shigang ZHQ, Jiuhua L, Jianli C. *Study on Website Search Engine Optimization*. International Conference on Computer Science and Service System IEEE Computer Society. 2012: 930-933.
- [5] Zhou Hui, Liu Jinhua. Website Promotion Strategy Based on Search Keyword Analysis and Selection. *Advances in information Sciences and Service Sciences (AISS)*. 2012; 22(4): 61-68.
- [6] Article title <https://www.quicksprout.com/2013/12/30/how-3-simple-google-analytics-reports-will-increase-your-search-engine-traffic/>. Date accessed: 05/06/2016.
- [7] Aljumah A, Kouchay SA. Global Ranking, Web Visibility and Accessibility of Quranic Websites. *Indian Journal of Science and Technology*. 2015; 8(30): 1-7.
- [8] Kumar V, Singhal N, Dixitand AA, Sharma K. Architecture of Perception Oriented Web Search Engine Based on Decision Theory. *Indian Journal of Science and Technology*. 2015; 8(7): 1-7.
- [9] Ramesh N, Andrews J. Search Engine using Social Networking Activity. *Indian Journal of Science and Technology*. 2015; 8(4): 1-6.