



Analysis of the Search Engine Optimization level of an academic website

Sherin Babu

Assistant Professor

Department of Computer Science

Assumption College Autonomous, Changanacherry

Kottayam, Kerala- 686101

Abstract

The exponential growth rate of Internet is greatly making the task of search engines more complex and complicated. Getting relevant and reliable information at the right time is an important matter to be noted. Search Engines have to analyze and sort through millions of web pages and then select and display only relevant and accurate information to the users. Every academic institution must have an effective and reliable website through which its potential users (particularly students) can benefit from. These websites should convey the right information to its stakeholders in a useful manner. It should contain complete details of the institution including details about that institution, departments, courses offered, fee details, faculty, research, admission and examination related matters, contact details etc. This work aims to check the Search engine optimization level of an academic website ie, the website of University of Kerala with the www.keralauniversity.ac.in/ using Google search engine.

Keywords: academic website, Internet, ranking, search engines, search engine optimization.

1. Introduction

In the era of Internet, users are highly depending on search engines to retrieve information of their choice. Since Internet has overruled all the geographic barriers and communication hindrances, people from any part of the world can search for information and they all wish to get their search result without having any obscurity. Thousands of websites are being created every day around the globe, in addition to the currently existing websites. The owners of the websites have to work with the aim of bringing more visitors to their website and convert those visitors to satisfied customers. If the visitors are unable to find the website useful, the chances of survival in the technology world will get diminished. Here arises the relevance of Search Engine Optimization (SEO). The objective of SEO is to make a website accessible to a maximum number of potential users, thus boosting the goodwill and profits of an organization/institution.

SEO is a set of techniques and practices to make a website and its contents noteworthy and thus making it visible to search engines and eventually to the web users. SEO can be defined as an organic method of making a search engine friendly website. SEO aims to increase the search engine visibility of a particular website from among other websites. A correctly optimized website will be ranked higher in the search process by the search engines. The optimized websites will always appear at the top of search results page. This helps the users to easily locate a particular website, when he is looking for information that is obtainable in that website. Majority of



the web users never go beyond the first result page while they perform their search activity i.e., the chance of visiting the less ranked websites are very low. A search engine optimized website can improve its volume of quality traffic to that website. This ensures goodwill, global reach of the organization name and a high return on investment.

1.1 Working of Search Engines

Search Engines (SE) are online software programs used by Internet users for searching and finding information on the World Wide Web. The aim of SE is to help the user in obtaining appropriate websites on the WWW that contain the search information or keywords entered by the web user. The results for a search are displayed line by line, based on its significance in a page wise manner called search engine results pages or SERP. At the top of the SERP, the most significant and relevant websites will appear. Search engines use a data structure called indexed table or indexed database to store information about the web pages. The main property of SEs is that they index only textual data, i.e., they cannot index graphical or pictorial data. The prominent search engines available on the web are Google, Bing and Yahoo. The following are the various steps involved in the working of search engines.

- **Crawling**- SEs frequently visits or scans through textual data contained on the web pages to mark the information contained in it using a software program known as Spider or Crawler. The frequency with which a SE crawls a website is not predictable and usually the crawl frequency depends on factors like Page Rank, frequency of updating page content, page size, hyperlinks to a page etc. Each of the search engines available on the WWW has its own spider program which differs from one another. Spider program of Google is called Googlebot. Yahoo Slurp is the name of the Yahoo web crawler and Bingbot is the name of Microsoft's Bing webcrawler.
- **Indexing** - The information about the web pages are stored in an index database. The keywords contained in the web page are identified first and then those keywords along with the respective webpage URL are added to the index table. If a website is not indexed by the search engine, then that website will not get displayed in the search results page. SEs constantly builds and updates the index table as it crawls through the website.
- **Processing**- When a user searches for particular information, this step will be performed. During this step, the Search Engines does the searching process in the index table i.e., by comparing the search query entered by the user with the contents in the index table or index database SEs do not examine the entire web when a user enters a search keyword. The most relevant matching web pages are retrieved as search result.
- **Ranking or Relevancy Calculation**-The relevance or significance of a searched webpage is calculated based on ranks assigned by the page ranking algorithms. Page ranking algorithms use mathematical equations to rank the web pages. The factors measured in ranking algorithm involve the location and frequency of keywords on a web page, number of broken links, HTML W3C validations, quality of page content, hyperlinks etc.



- **Retrieving the result**– In the SERP, the web pages are displayed in the descending order of the page rank obtained, as mentioned in the previous step. I.e., the most relevant web pages are displayed first and least relevant web sites are shown at the bottom, in a page wise manner.

1.2 SEO Strategies

In making a website search engine optimized, each webpage should be properly structured so that it can be correctly indexed by search engines in the most effective way. SEO cannot be considered as a onetime procedure rather it is a continual activity for the existence of organization. The SEO strategies can be broadly divided into two: - On-page SEO and Off-page SEO. Both of them are crucial for improving the search engine rankings.

1.2.1 On-page SEO Techniques

On-page SEO means optimizing the content and format of each webpage of a website so as to get higher page rank and utmost visibility within a particular search engine. In other words on-page SEO means web page structure optimization and content optimization. It includes various techniques that are applied on the HTML code part and the content part of the webpage. On-page SEO includes factors that have direct control over the website; namely title of the page, headings, URLs of the web pages, keywords, hyperlinks, content of the page, etc. On-page SEO helps to make the website as search engine friendly. The various On-page optimization techniques are listed below.

- **Keyword Optimization** – The most important factor in optimizing a webpage is the usage of relevant keywords. Keyword means words entered by the users in their searching process to a search engine. Each web page must contain right amount of keywords in right places and in right amount, in order to drive the users to the website. Keywords can be included in the following sections of an HTML page - content, titles, headings, meta tags, URL, anchor text, page footer etc. Extensive and profound research should be done before selecting and placing keywords into the page content. It should be noted that the SEs penalizes websites that have extensive use of keywords. When a page is created entirely of keywords, then it is called keyword stuffing.
- **Page Content** – While designing a webpage, the developer or designer should follow the golden rule “Content is the King”. All the web pages should contain unique, relevant and useful information, and the content should be easily readable, understandable and should be in a simple language, without spelling and grammatical mistakes. The desirable property of each web page is that it can contain about 250 to 500 unique words of text with the keywords appearing 4 or 5 times.
- **Listing of items in a webpage** should be done using ordered or unordered lists.
- **Web page content updation** – The content of a web page should be updated frequently, whenever it is possible to do so. In any case, the content duplication within the same domain or across the domains should not be allowed.



- Title tag – It is the most significant part on the web page scanned by the SEs for finding what a web page is all about. Every web page must have unique and appropriate page titles. It should be brief but can be descriptive, up to 70 characters.
- Headings Tags <h1> to <h6> -The headings inside a web page should be labeled using the heading tags. The most significant and primary visible page heading should be in <h1> tag. For sub headings, use <h2> to <h6> tags based on their prominence.
- Meta Tags –These tags contain additional information about a web page. I.e., metadata about the html page. The meta tag information will not be displayed on the webpage. Different meta tags are:
 - Description meta tag –specifies a tiny and unique description about a web page. The content of this tag is displayed in the search results page.
 - Author meta tag – provides information about the author or designer of a particular web page.
 - Keyword meta tag - specify about 7 to 10 keywords contained in that web page.
 - Viewport meta tag –gives information about the page's dimensions and scaling to the browser, so that the web page can be displayed correctly in all type devices.
 - Revise meta tag –tells the last revision of the webpage.
- Hyperlink Text– While creating hyperlinks, the clickable text should be simple and understandable and it should clearly describe what the linked pages are about. If possible, hyper link text should contain keywords for that page.
- Image tag optimization–In the tag, the image file name should reflect what is shown in the image. It is mandatory that the "Alt" attribute of tag must be filled. Alt attribute specifies the image substitute text if the browser could not display images. Give short and meaningful values for image filenames and alt text.
- Image types - It is advisable to use commonly supported image file types like jpeg, png, gif, bmp etc only in the web page. There must be a separate directory in which all the images used in the website should be stored.
- Personalized 404 error page – It is strict that every web site must contain customized error pages, rather than showing the server error messages.
- Frame tags– Avoid using HTML <frame> tags in designing web pages since it is difficult for search engines to follow the frame links.
- Internal Linking using Breadcrumbs– navigation tool that helps users to easily move back and forth around a website and find all the web content. Bread crumb has the format : Home page > Section page > Subsection page.
- HTML W3C Validation – Always ensure that the website's HTML code is validated using the W3C HTML validator. SEs will not give high ranking to websites with HTML errors.



- JavaScript (JS) and CSS files –Instead of writing inline JS and styling, it is better to use external JS or CSS files, that helps the website to get more visibility to the SE. The number of JS and CSS files should not exceed 3.
- Robots.txt –This file tells the spider program of SE how to crawl through the website. It is placed in the root directory of the site.
- Sitemap.xml file – It is an XML file that contains the URLs of all the pages in a website. This file also contains additional information like which all web pages are changed recently, what is the frequency of page updations etc.
- Website directory structure – Create a simple directory structure with a maximum of 6 levels of hierarchy.
- Cross - browser compatibility- ensures that the website is displayed consistent across all the browsers.
- Avoid the usage of <table> tag in the webpages. Instead we can use <div> tag.
- Content highlighting – can be done using bold, strong, em, underline or italics tags.
- Keyword rich Page footer – this helps the website to get easily indexed.

1.2.2 Off-page SEO Techniques

Off Page SEO refers techniques that are primarily focused in increasing the online reputation of a website via various web relation modes. In other terms it includes all the stuff that we are not doing directly on our website to increase the quality traffic. I.e., Off-page SEO is aimed at having techniques for promoting the website. This includes techniques like social networking, link building, article submission, forum & blog marketing etc to get higher page ranking. Various Off-page SEO techniques are given below.

- Promotion through Social Networking Sites – The institution or organization must start a page in the most popular social networking websites such as Facebook, Twitter, Google, LinkedIn etc. This method helps to extend the online network of contacts, interaction with customers and most importantly in the promotion of website and thereby growing the online reputation of the organization.
- Web Blogs - Blog is a best way for promoting online existence. Make sure that the blog is kept up to date with latest posts. Another method is to post comments on other blogs and there include a link to the organization's website.
- Social Bookmarking - Submit recent and relevant pages to well-liked bookmarking sites like Delicious, Digg, Stumble Upon, Reddit, etc. It is evident that these are updated frequently and therefore there is greater possibility for search engines to crawl through the organization's website via these bookmarking sites.
- Building up of Quality Links – Link Building means of receiving incoming links for a website. The organization can try various methods to get high quality links, especially from their partners, stakeholders, customers etc. Obtain back links from high-ranking



websites only. Back links from other websites can be obtained to organization's home page as well as internal pages.

- Image Sharing – post images in Image Sharing websites like photobucket, flickr, etc. and there provide information about the institution's website, and thereby helping to increase the visitors count.
- Online Press Releases – This method is to distribute online press releases, in popular news media websites with links to the institution's web page.
- Article distribution – means offering the expertise of the organization/institution to other web sites by providing guest articles for them and in that page provide link back to the organization's site.
- Video sharing – This can be done by submitting free promotional videos in websites like YouTube, Vimeo etc.
- Use of satisfied customers - Another important matter is to request satisfied customers to give a link back to the organization's site.
- Duration of domain registration period - Domains that are already registered for years will be valued higher. So use the same domain name throughout the business period, without altering it regularly.

2. Methodology

To begin with the project, the HTML code of the website of the www.keralauniversity.ac.in was extracted. The web browsers used in this project are Google Chrome with Version 64.0.3282.167 and Mozilla Firefox Quantum with version 58.0.2. The HTML code extraction was done from the web browser by right clicking on the page and selecting "View Page Source" option. The Search Engine chosen for SEO is Google. Then SEO checking was performed in two stages. First, the HTML source code was analyzed manually for all the On-page SEO techniques mentioned in the previous chapter. Second, online SEO analysis tools like website.grader.com, www.seoscores.com and www.seoptimer.com are used for analyzing Off-page SEO performance.

3. Observations

The on-page and off-page SEO techniques are tested in detail. The SEO techniques that passed the test are given below.

SEO Feature Checked	Finding
Meta Title	The meta title has a length of 20 characters. (University of Kerala)
JS Error	No JS error in the webpage.
Nested Tables	The page does not use nested tables. This speeds up page loading time and optimizes the user experience.



Frameset	This webpage does not use frames.
Doctype	The website has a doctype declaration:www.w3.org It is an instruction to the web browser about what version of the markup language the page is written in.
URL Redirects	The URL doesn't have any redirects, thus saves site-loading speed.
HTTPS	This website is successfully using https, a secure communication protocol over the Internet.

The features that have failed during SEO testing are given below.

SEO Feature Checked	Finding
Meta Description	The meta description tag is missing from page. This tag is included in order to provide a brief description of the webpage which is used by search engines for optimization.
Keyword Usage	There is likely no optimal keyword density. The keywords on the web page should appear within natural sounding and grammatically correct.
Keyword Consistency	The website's main keywords are not distributed well across the important HTML Tags. The page content should be focused around particular keywords that would be liked to rank for. Ideally these keywords should also be distributed across Tags such as the Title, Meta and Header Tags.
<h1> Headings Status and <h2> to <h6> tag status	The page does not contain any H1 headings. H1 headings help indicate the important topics of the web page to search engines. Sub heading tag <h4> is present without having <h2> and <h3> tags.
Poor keyword identification	Eg. good research university in kerala Output – Ranked 11 in the second SERP Eg. statistics research centre in kerala Output –ranked 8 in the first SERP
Robots.txt	The site lacks a "robots.txt" file. This file can protect private content from appearing online, save bandwidth, and lower load time on the server. A missing "robots.txt" file also generates additional errors in apache log whenever robots request one.
Internal links	Found 146 links. An optimized site should not have more than 100 internal links.



Sitemap	The site lacks a sitemap file. Sitemaps can help robots to index the content more thoroughly and quickly.
Broken Links	www.research.keralauniversity.ac.in/viewGuides.php This page is missing (Error- page not found)
URL	Some unfriendly URLs are present. Eg. www.keralauniversity.ac.in/pdfs/news/MBA_Nottification1517656936.pdf
Image Alt	Out of 58 tags and 50 of them are missing the required 'alt' attribute. Eg. <imgsrc='images/downs.png' title='Click here for brochure'>
Inline CSS	The webpage is using 41 inline CSS style, which is very high in number.
Social Media Check	This website is not connected with social media using the API's provided by Facebook, Google +, Twitter etc
HTML validation	This site is not HTML W3C validated.110 Errors, 6 warning(s) are reported by W3C validator.
No Use of Mobile Viewports	The page either does not have a Meta Viewport specified, or it is configured incorrectly. Viewports are important to ensure that the web page content sizes appropriately for different devices, in particular mobile and tablet.Ie, mobile page rendering of the web site is not proper.

4. Result Analysis

The findings of the study were analyzed and it is evident that the website of University of Kerala with URL has the Search Engine Optimization on an average level only. The Global Page rank of this site using Alexa is 51777. This is because many of the On-page and Off-page techniques are not properly taken care while designing the website. The major errors that are to be rectified are listed as follows. If appropriate keywords are selected and added to the page, it can create immense changes in the SEO performance. Properly selected keywords and appropriate distribution of the keywords in right places can greatly increase page rankings. Another major problem identified was the absence of validated HTML code. In the HTML code, all deprecated keywords should be either eliminated or to be replaced. If the HTML code of a webpage is W3C validated, it can surely increase page rankings. Broken links always reduce page rankings. So utmost care should be taken to ensure that the website should not contain any broken links. Another major error discovered is the absence of sitemap.xml file, which is a crucial factor in search engine optimization. The absence of robots.txt file also reduces the SEO performance. The other errors listed also lessen the SEO performance of the website, but the degree to which it adversely affects SEO is low. If these major errors are rectified with high priority and immediate effect, the website can be made more optimized to Google Search Engine.



5. Conclusion

The Internet has turned out to be an inevitable part of our day-to-day life. SEO is a technique which enables a website to get simply recognized among thousands of websites used for various purposes, across the web. Hence making the website SE optimized has become an extreme requirement for all the organizations or institutions. SEO ensures online presence and helps in grabbing target audience and expanding the customer base. In this project, the SEO level of the website of University of Kerala is analyzed using Google Search Engine. The findings pointed out that the SEO performance level could be termed as "Average". There are several factors that adversely affect SEO level, which needs immediate attention. If those factors are taken care within a short span of time, it is guaranteed that the SEO performance level of the website will surely boost up as "Good".

6. Future Plans

The present study can be extended using Microsoft Bing and Yahoo as the Search Engines. Since each Search Engine has its own page ranking algorithms, while conducting SEO testing using the above said Search Engines, it may yield different results. Then a comparative study of the SEO performance of the website on the search engines Google, Bing and Yahoo can be done.

Another future plan is to carry out SEO performance checking of another academic website (say www.mgu.ac.in of Mahatma Gandhi University, Kerala) in the above said three Search Engines. Then a comparative study on the SEO performance of both websites can also be carried out.

References

- [1] Edgar Damian Ochoa (2012), An Analysis Of The Application Of Selected Search Engine Optimization (SEO) Techniques And Their Effectiveness On Google's Search Ranking Algorithm, California State University, Northridge.
- [2] Burdon, D. (June 2005): The Basics of Search Engine Optimisation, Simply Clicks.
- [3] Callen, B: Search Engine Optimization Secrets Made Easy.
- [4] Dover, D; Dafforn, E: Search Engine Optimization Secrets, Wiley publishing.
- [5] Engne, E; Fishkin, R; Spencer, S (2009): The Art of Seo, O'Reilly Media, Inc.
- [6] Fleischner, M. H. (2009): SEO Made Simple, 2nd edition, Lightning Press.
- [7] Harris, S (2009): 52 SEO Tips, Big Oak SEO.
- [8] Jones, K. B: Search Engine Optimization, 2nd edition, John Wiley & sons.
- [9] Ledford, J. L. (2009): Search Engine Optimization Bible, John Wiley & sons.
- [10] Ryals, K (2008): SEO 100 Success Secrets, EmereoPty Ltd.
- [11] SherinBabu (2012), Relevance Of Search Engine Optimization In Promoting Online Business, Indian Journal of Computer Science and Engineering, Jun- Jul 2012, Vol.3, No.3, Pages 453-456