

Website Content Analyzer SEO Tool

Khushboo A. Sonawane

Bhagyshree K. Wagh

Meenakshi K. Shelar

Neha R. Yeole

N.D.M.V.P.S's KBT College of Engineering, Nashik

Abstract: - Search Engine Optimization is making your website easy for both users and search engine robots to understand. SEO helps the engines figure out what each page is about and how it may be useful for users. SEOs tend to use a lot of tools. Some of the most useful are provided by the search engines themselves. A Website Content Analyzer for SEO Tool is a replacement of the current tools. It not only works for some specific task but shows the overall analysis of the website pages and also presents it in the complete and pending percentage format. Through this the developer is able to approximate how much percent SEO is remaining and how much percent is completed for particular page. Website Content Analyzer for SEO tool is a component of SEO. To do SEO for a website means making use of correct SEO tools and make the website up to the mark for top page ranking. All the developer should have is a website URL and a correct tool to do SEO. Website Content Analyzer is going to use the Google's algorithm for page ranking and hence the SEO of the website is done by keeping Google's algorithm in mind.

Keywords: *SEO, Content Analyzer, Google Algorithm, SEO tool*

1. Introduction

Search engine optimization (SEO) is the process of affecting the visibility of a website or a web page in a search engine's "natural" or un-paid search results. In general, the earlier (or higher ranked on the search results page), and more frequently a site appears in the search results list, the more visitors it will receive from the search engine's users.[2] SEO may target different kinds of search, including image search, local search, video search, academic search, news search and industry-specific. As an Internet marketing strategy, SEO considers how search engines work, what people search for, the actual search terms or keywords typed into search engines and which search engines are preferred by their targeted audience. Optimizing a website may involve editing its content, HTML and associated coding to both increase its relevance to specific keywords and to remove barriers to the indexing activities of search engines. Promoting a site to increase the number of backlinks, or inbound links, is another SEO tactic. The plural of the abbreviation SEO can also refer to "search engine optimizers", those who provide SEO services. Google's official position on webmaster best practices really hasn't changed much over the years. What has changed is the search engine's ability to enforce these guidelines through improved algorithms. The implementation of Panda, Penguin, and Hummingbird has had a profound impact on the SEO landscape. The Google's Humming Bird algorithm is been used here to make the website's best SEO.

2. Literature Survey

The information present on the web is huge in amount, distributed, heterogeneous and dynamic. Due to the heterogeneity of network resources and the lack of structure

of web data, automated discovery of targeted knowledge retrieval mechanism is still facing many research challenges. With the huge amount of information available online, the World Wide Web is a fertile area for data mining research. The Web mining research is at the cross road of research from several research communities, such as database, information retrieval, and within AI, especially the sub-areas of machine learning and natural language processing. However, there are a lot of confusions when comparing research efforts from different point of views. [4]SEO is defined as the optimization of a website for search engines, so that the search engines views it in an optimal manner. Various techniques and methods are available to achieve high rankings and become visible on search engines. Various processes are gradually evolving for optimizing the website, by observing the working of search engines. Every major search engine has its own respective algorithm. All the power of ranking the websites on the results page is with these algorithms.[6] Website relevance and ranking are two important factors that are addressed by search algorithms. (Humayun, 2009) Search algorithms do have some weaknesses and many website owner do take this to their benefit. In order to address this problem, search engines keep their algorithms up to date. Google makes changes in its algorithm almost on a daily basis. So, it can be said that both SEO techniques and search algorithms are gradually developing in a vicious circle. Sean A. Gollither (2008) states that as search engines disapprove the manipulative strategies and punishes the websites by removing them from their index; therefore it is very important for webmasters to follow the guidelines of search engines for implementing only ethical SEO techniques.[7] There are various ethical techniques which can be implemented for improving

visibility in search engines. Although none of these techniques assure a particular ranking position in the search engine results pages, the implementation of these SEO techniques and strategies improve prospects of the websites to gain visibility on search engines. Ultimately this results in receiving visitors to the website and consequently leads to increased sales and revenue for the website.

3. Proposed system

Search Engine Optimization (SEO) is the Optimization which is very important for improving the volume as well as the quality of traffic to your site from search engines by using search results for your targeted keywords. So, you need to perform Optimization on your site. Search engines like Google, Yahoo and MSN etc. are the gateway to the Internet as these search engines are the tool that the customers use to find the products and services they require. That's why Search Engine Optimization is so important for your business as if your website is not visible to the customer then you may find difficulty in making any sales.[5] The use of Search Engine Optimization is easy. So, you use SEO to be the best in your niche. To attract and generate visitors income are also important for your site.

SEO isn't just about getting the technical details of search-engine friendly web development correct. It's also about marketing. This is perhaps the most important concept to grasp about the functionality of search engines.[7] Making a perfect website can be done but making it competitive is a tough job. A SEO tool is the thing which can help the website to get the best SEO and rank at the top. And once the website is on internet the business definitely has a good future scope. To expand the business one has to be on internet and ranking of website can only be done through SEO TOOLS. Website Content Analyzer SEO Tool is made up from three modules and they are: 1] Machine Learning Algorithm, 2] Schema Creator, 3] Humming Bird Algorithm. By making use of these three modules the tool can work in most effective manner. The first phase is of Machine Learning Algorithm where the query which is entered by the user is being processed. When query is fired HTML parsing is done and after that text processing takes place where the words are divided into tokens, stemming on those words are performed, grammar is checked and finally stop word removal is done. Then after the chart of keywords and tags will be generated of URL to be processed and optimized URL and comparisons between the two charts is been done. Machine Learning Algorithm identifies the lacking of the URL to be processed and try to make it upto the level of the optimized URL. These are the activities done y machine learning algorithm. Then after schema creator is generated which will generate some tags automatically to the website to make it convenient to be ranked first. The

website URL then calls to Humming Bird function to cross check the corrections and generate the report.

Algorithm for Website Content Analyzer SEO Tool

Algorithm: Relevancy and keyword frequency based approach

Input: User's URL, Optimized URL, Keywords

Output: Suggestion generated for the website to be optimized and ranked high.

Step1: Enter the user query.

Step2: Perform pre-processing of user query.

Step3: Obtain keywords from processed query.

Step4: Find frequency of keywords

Step5: Generate the chart of website and optimized website.

Step6: Compare the charts.

Step7: Use of schema creator for making the content of website in a proper structured manner

Step8: Fulfilling Humming Bird Algorithm for better SEO.

Step9: Generate Report

Step10: Making corrections in the report. (unless the tool shows 100% SEO done)

Step11: Getting the website ranked highest position on Google's page.

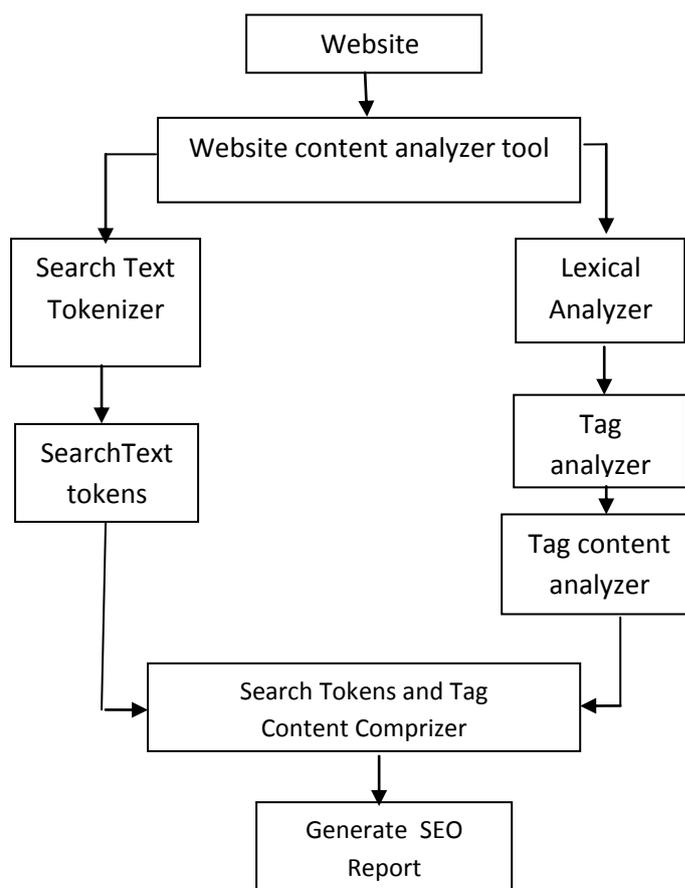


Fig: Block Diagram of Website Content Analyzer SEO Tool

3.1 Machine Learning Algorithm:-

Machine learning is used by the major search engines to detect patterns. Understanding many of the basics pertaining to machine learning can help SEO's get a grasp of how some of the algorithmic updates work. Typically, Google's algorithm is based on a set of signals and rules that are fixed. But after the Panda update searches are centered around machine learning techniques.[4]

Machine learning techniques used by Google rely heavily on some human input. But this is not the only thing used. They will use various factors like social signals, visitor data, links and bounce rate. These are all calculated as negative or positive influences on the algorithm. The many factors, both negative and positive are all mixed together and divided out into groups. From these groups predictions are made. Even if some of the group's entries share undesirable quality. By observing these patterns machine learning adapts its "behaviors". [8] Google basically made the switch to machine learning when they launched the Panda update that set everyone on edge. They changed the algorithm for ranking sites in search engine result pages and changed the way it looked at the quality of a site. Google began to use more of the human element (which sort of makes sense since the web 2.0 is a social or human interaction platform) to evaluate and rate sites. This was attempted to try and separate out the lower quality content on some sites.

3.2 Schema Creator:-

Structured data is away for search engine machines to make sense of content in your HTML. [7] Google and other search engines created a structured data standard called schema.org. If you're like most web designers or SEO's, you might find schema.org a little difficult to grasp at first. That's why we created schema creator-to help you quickly build and get started with schema.org microdata. On June 2, 2011, Google, Bing, Yahoo! Announced that they will be using a standard set of schema for data markup on web pages. Google webmaster Central Blog. To make things easy for web developers and the SEO community, a simple generator that allows you to input basic information and have that information converted into the standard schema.org markup structure.

3.3 Humming Bird Algorithm:-

Hummingbird it is also able to judge context – thereby judging the intent of a person carrying out a search, to determine what they are trying to find out. Google Inc has overhauled its search algorithm, the foundation of the Internet's dominant search engine, to better cope with the longer, more complex queries it has been getting from web users. Google is trying to keep pace with the evaluation of Internet usage. As search queries get more complicated,

traditional "Boolean" or keyword-based systems begin deteriorating because of the need to match concepts and meanings in addition to words. Google said that Hummingbird is paying more attention to each word in a query, ensuring that the whole query- the whole sentence or conversation or meaning- is taken into account. In a nutshell, think about why people are looking for something rather than what they are looking for. A content strategy should be designed to answer their needs, not just provide them with facts. "The age of semantic search." [6] Conversational search leverages natural language, semantic search, and more to improve the way search queries are parsed. Unlike previous search algorithms which would focus on each individual word in the search query, Hummingbird considers each word but also how each word makes up the entirety of the query- the whole sentence or conversation or meaning- is taken into account, rather than particular words. The goal is that pages matching the meaning do better, rather than pages matching just a few words. Much like an extension of Google's "Knowledge Graph", Hummingbird is aimed at making interactions more human- capable of understanding the concepts and relationships between keywords. Hummingbird places greater emphasis on page content making search result more relevant and pertinent and ensuring that Google delivers users to the most appropriate page of a website, rather than to a home page or top level page. [6] Search Engine Optimization received little changes with the addition of Hummingbird, though the more top ranking results are ones that provide natural content that reads conversationally. While keywords within the query still continue to be important, Hummingbird adds more strength to long-tailed keywords- effectively catering to the optimization of content rather than just keywords.

4. Conclusion

The increasing use of internet and online business is forcing the marketers and business people to expand more and more in their business- online and offline, and this can be possible by making use of relevant SEO tool- Website Content Analyzer SEO tool which will help the company holders to get their website on the top of the Google page and will ultimately bring more visitors to their page.

References

- [1] Gibson, J., Wellner, B., Lubar, S, "Adaptive web-page content identification", In WIDM '07: Proceedings of the 9th annual ACM international workshop on Web information and data management. New York, USA, 2007.
- [2] L. Page, S. Brin, R. Motwani, and T. Winograd. The Pagerank citation ranking: Bring order to the web. Technical report, Stanford University, 1998.

-
- [3] Brin, S., and Page, L., 1998. —The anatomy of a large-scale hyper textual Web search engine!, Computer Networks and ISDN Systems, Vol. 1-7, pp: 107-117.
 - [4] Cooley, R.; Mobasher, B.; Srivastava, J.; —Web mining: information and pattern discovery on the World Wide Web. Tools with Artificial Intelligence!, 1997.Proceedings., Ninth IEEE International Conference. Page(s):558 – 567 -3-8 Nov. 1997.
 - [5] <http://moz.com/beginners-guide-to-seo/>
 - [6] <http://www.devotol.com/>
 - [7] <http://blog.i2k2.com/seo/ Overall Description>
 - [8] <http://blog.w3school.com/seo/ Overall Description>