International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(01), Nov 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Site Content Analyzer for Analysis of Web Contents and Keyword Density

Bharat Bhushan

Asstt. Professor, Government National College, Sirsa, Haryana, (India)

ABSTRACT

Web searching has become a daily behavior for many people, with the Web now the first choice for many people seeking information. In recent years, Internet has emerged as the most powerful medium for storage and retrieval of information. The web has major activity known as Searching and the Search Engines is the tool through which the user can find the information. The web has vast amounts of information; the number of results for most queries is usually in thousands but the user browse through the first few results only. Thus results ranking is crucial to the success of a Search Engine. Here we present the brief description of the "Site Content Analyzer" which we have used in web content analysis and keyword density. Site Content Analyzer is software that helps you improve website search engine ratings, increase traffic flow and optimize site content. The site content analyzer tool is the perfect tool which is used to measure the content quality. Objective of this research is to study and analyze the keyword density which has a great impact on increasing the relevance of page of website.

Keywords *–Web Content, Search Engine, Site Content Analyzer, Keyword Density.*

I. INTRODUCTION

The Web is both an excellent medium for sharing information as well as an attractive platform for delivering products and services. The Web has become a worldwide source of information and a mainstream business tool. It is changing the way people conduct the daily business of their lives. In practice, search engines usually combine relevance and importance, computing a combined rank score that is used to order query results presented to the user. When a user enter a query at a search engine site, then the user input is checked against the search engine's index of all the web pages it has analyzed. The best URL's are then returned to user as hits, ranked in order with the best results at the top. Internet search engines are special sites on the Web that are designed to help people find information stored on other sites. There are differences in the ways various search engines work, but they all perform three basic tasks:

• They search the Internet or select pieces of the Internet based on important words.

• They keep an index of the words they find, and where they find them.

• They allow users to look for words or combinations of words found in that index.

International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(01), Nov 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

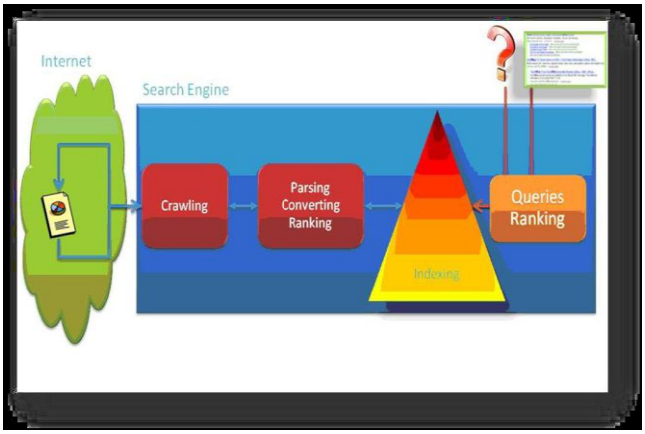

Fig. 1. The simplified view of how Search Engine Works

The figure above shows the simplified view of how search engine works. A Web search engine is a search engine designed to search for information on the World Wide Web. The most important measure for a search engine is the search performance, quality of the results and ability to crawl, and index the web efficiently [1].Site Content Analyzer is software that helps you improve website search engine ratings, increase traffic flow and optimize site content. Site Content Analyzer examines HTML pages on and offline to provide you a detailed report about keywords density, their weight and relevance. In addition, program uses a smart algorithm that counts in tags and keyword location, as well as font properties. If your business is linked with search engines optimization, you'll find Site Content Analyzer useful and time-saving, especially when doing keywords research for your clients. You don't have to be a SEO expert to get benefits of using Site Content Analyzer. It's a powerful and handy tool for webmasters too [2].

II. SITE CONTENT ANALYZER

Content Analyzer for Web gives any enterprise the ability to build a total view of its corporate web estate. First, Content Analyzer for Web utilizes Vamosa's web crawl technology to provide a total view of all content that is actually being published. Secondly Content Analyzer for Web will access the source system where all content is managed to provide a view of all content that is stored, and finally web usage statistics are used to provide insight into what content is being used [3]. It is well-known that the web ranking of a website mostly depends on two main factors: the number and the quality of inbound links and the amount and the quality of website's content. While determining the quality of the content seems to be one of the most complex tasks in search engine optimization .The site content analyzer tool is the perfect tool which is used to measure the content quality. The Site content analyzer tool can list such parameters as: keyword density [4], keyword weight [4,5], keyword distribution, discover the most relevant keyword and key phrases, find out the quality of in-site links, overview the whole site-wide picture and many more [6].

International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(01), Nov 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Fig. 2. The Site Content Analyzer Tool.

Professional website content parser: Site Content Analyzer 3 is new generation in website analysis software. It parses website on- and offline for keywords, suggests the most relevant and weighty phrases, analyzes link structure, and many more. With it, you can quickly discover the most relevant keywords and key phrases for each page of your site, learn if the parameters of your site (that is keyword density, keyword weight etc) meet the guidelines of search engines [7].

A. Input for site content analyzer tool

The input for Site content analyzer tool is the Webpage URL. The Offline URL is stored in the system which is the input of the Site content analyzer tool [6].

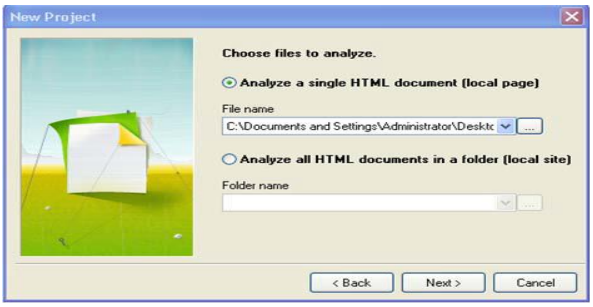

Fig. 3. The URL Address is input for the Site Content Analyzer tool.

B. Site Content Analyzer Description

The Site content analyzer is used to measure the different quality terms:

Keywords mode: It defines the no. of keywords in counted in different tags. Keyword is a single word in the text of a web page. This is a term search engines operate with. Every word on a page is in fact a keyword; it alters the overall rank of a page in search engine's index. The way it does that depends on its weight, distribution and density [6].

Key phrases mode: In key phrases mode the calculation of key phrases could be done. The weight, count and density of key phrases in a website are calculated in key phrases mode.

Keyword density and keyword weight mode: In the keyword density and keyword weight mode, the calculation of important keywords is done [4]. The terms which are calculated here are- Keyword density, Keyword weight, Keyword distribution.

Keyword cloud mode: The keyword cloud defines the appearances of important keywords. The keyword cloud is a visual depiction of keywords used on a website; keywords having higher density are depicted in large fonts. The main keywords of a website appear in large fonts.

Linked Mode: In linked mode, the list of linked web pages can be displayed. The linked web page is the interlinking web page in a website.

C. Report Generation

The Site content analyzer tool generates report of all contents. This report is about the web page's contents. The different calculation of factors that can be included in the website is defined in this report [7]. The report includes the different information detail:

File information: File summary, Keyword density report, Key phrases report, Keyword weight report, Links report.

Keywords information: Keywords summary, Keywords details.

Key phrases information: Key phrases summary, Key phrases details.

III. WORKING AND EXPERIMENT

Here we describe the working of Site Content Analyzer that is used under the process of simulation, analysis and experimental results.

A. Working of Site Content Analyzer

This paper presents the analysis of the keyword density which has impact on increasing the relevance of page of website with the help of Site Content Analyzer tool. It gives the output in the form of report of the input web site. The present research also analyzes the link popularity of a webpage. Fig 4 shows Data Flow Diagram for Evaluation/ analysis of web contents.

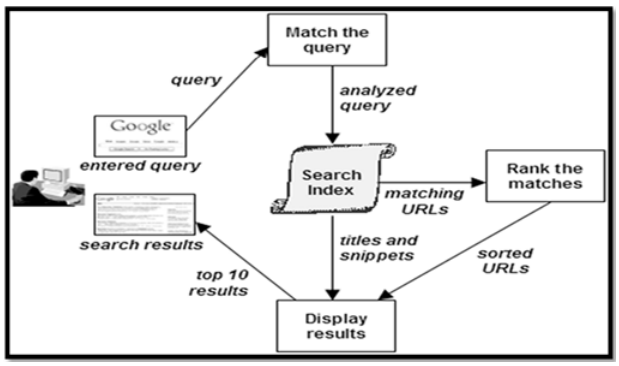

Fig. 4. The Overview of Organic search results

B. Experimental Evaluation

Content Analysis: The ranking of websites is not an easy task. The keyword strategy can have a great impact on improving the ranking of websites.

Keyword Analysis: The keyword is the important content of every website. In keyword analysis the different measurement can be done. The different factors related to key strategy can be calculated by the Site Content Analyzer which describes the page size, title, words count, average count, external links, internal links, keyword density report, key phrases report that specifies which keyword has high density.

IV. CONCLUSION

At last the summary of this paper has defined the tool which has major importance in our work. The research shows that site content analyzer tool is the perfect tool which is used to measure the content quality. The Site Content Analyzer tool is used to define the keyword density, weight of key phrases.

REFERENCES

[1] Cai, D.,Yu, S., Wen, J., Ma, W., (2003), "Extracting Content Structure for Web Pages Based on Visual Representation", *AP Web*, pp. 406-417.

[2] file:///D:/Site Content Analyzer - Free download and software reviews - CNET Download.com.htm.

[3] Content Analyzer for web content – V Product Datasheet.

[4] Hulbert, W., (2005), "Keyword Density: SEO Considerations", www.webpronews.com/news/ebusinessnews/wpn-520050501KeywordDensitySEOconsiderations.html

[5] Curran, K., (2004), "Tips for achieving high positioning in the results pages of the major search engines".

[6] Yadav, Y., and Yadav, P. K., (2011), "Site Content Analyzer in Context of Keyword density and Key Phrase", Int. *J. Comp. Tech. Appl., Vol 2 (4)*, pp. 860-872.

[7] file:///D:/Site Content Analyzer - keywords and phrases suggestion and analysis software.htm.

[8] Cavtat Crotia, (2010), "Proceeding of the
ITI", *Int.conf.on Information technology
Interfaces.*

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, (1999), "The page rank citation ranking: Bringing order to the web", *Technical report, Stanford Digital Libraries SIDL-WP-1999-0120.*