Blackhat Search Engine Optimization Techniques (SEO) and Counter Measures

Dr. R. D. Gaharwar[1], Prof. D. B. Shah[1]

G. H. Patel Department of Computer Science & Technology, Sardar Patel University, Vallabh Vidyanagar, Gujarat, India

ABSTRACT

Search engine are being used by the most of the World population as their basic information retrieval system for getting useful information from internet. As a service provider who uses internet for digital marketing it becomes mandatory to get high ranks from search engines. Search engines optimization (SEO) techniques are used for this purpose. Black hat SEO techniques are used for quick results but are prohibited by most of the search engines. Hence web space users or website developers should be well aware of SEO techniques and how to use them in optimal way. This paper presents some of the most commonly used black hat SEO techniques and the counter measures done by different search engines to prohibit them.

Keywords : Search engines, Search Engine Optimization (SEO), Spamdexing, Meta tags, Spamming

I. INTRODUCTION

A. Search engines and its classifications

Since the early days of internet evolution search engines are used by human beings as basic information retrival system to fetch desired information from internet. Search engines provides easiest and quickest way to browse useful information . This process of information retrival is divided following three basic tasks:

- User interfacing: User interface provides user-friendly interface for end users to enter their search query while using search engine.
- Searching: Any search engine engage web spiders or web crawlers to fetch maximum number of relevant web pages. These web crawlers crawls through internet searching thousands of web pages to get relevant web pages to end user search.
- Indexing: Every search engine uses some sort of unique algorithm to rank/ index web pages according to their relevance with end user search

query. All the web pages having higher rank/ index appears at top in the search result list [1].

- There are different classification for search engines but the boarder categories for search engines are as follows [2]:
- Crawler-based search engines: These search engines engage web crawlers for automatically indexing and searching web pages on internet.
- Human-powered directories: Indexing and searching of web pages is not aided by any robot or agent. These search engines allows to manually add the entries in the search result list.
- Hybrid search engines: These type of search engines allows both automatic and manual entries to be added in the indexing list.

B. Search Engine Optimization (SEO)

Search engines had worked effectively for decades but they had certain drawbacks as mentioned below [3]:

- Every search engine covers only certain area on internet
- Any search engine is insuffecient to address all the user searches
- Search engine output may be sometimes obsolete and irrelevant

Spamming is supposed to be a major issues with search engine nowadays. Every website owner wants its website to appear at top positions in search results. Search Engine Optimization makes effort towards optimizing website in order to improve its rank in search engine result list. It uses methodological approach for improving both website coding and details related to coding for improving its preferences and visibility for different search engines. The Search Engine Optimization tasks are being classified into two different ways as shown below:

[1] On-site optimization: During this process the focus is on developing website architecture and website content insuch a way that its visibility for search engine automatically increases. On-site optimization is done by carefully using appropriate domain name or website titles. Moreover special care is taken for properly using keywords in description and multiple way in web page title and meta tags of web pages. All the efforts are put into designing website in such a way that it can be easily crawled.

[2] Off-site optimization: During this process focuss is on rising the preference of website by increasing the web traffic towards website. Major Off-site optimization techniques include Back-linking and social network linking. During Back-linking links of the website is kept from the other reputed websites and during social network linking links of the website are kept on social site or social groups and the other trusted sources.

C. Spamdexing and its classifications

Spamdexing is one type of Search Engine Optimization techniques and become popular during mid-1990s. In spamdexing deliberate efforts are made to unethically increase the index of any website in search listing results. Different techniques used for spamdexing are link building, keyword stuffing etc. Different search engines keep a routine check for detecting the instances of spamdexing if any website is suspected of spamdexing it is outsed by search engine from result-listing or website may also be penalized. The basic focus of spamdexing is on finding the loop holes in search-engine promotion rules and guidelines and exploring their options with it.

Spamdexing can be classified into two broader categories as follows [4]:

[1] On-page content spamming:
This technique focus mainly on altering the content of website or webpage ino order to increase its visibility for different search engine. Different examples of on-page content spamming are keyword stuffing, hidden text, meta-tag stuffing etc.

[2] Off-page link spamming:
This technique focuses on increasing the links between web page without righteous reason. Any search engine assumes that if any web page is relevant to the end user search than the probalility that all the web pages having links to that previous web pages also increases. Off-page link spamming takes advantage of this and adds random links on the unrelated web page. Moreover if any web page have links to higher rank web page, the rank of previous web page automatically increases. Off-page link spamming focuses on taking the advantage of higher ranks website to increase its rank. Some common link spamming techniques are link farms, private blog networks, hidden links, spam blogs, cookie stuffing, etc.

II. METHODS AND MATERIAL

The authors studied search engine working in detail. They mentioned that any website may fail to appear in search results because they are not following proper search engine optimization guidelines. Search Engine Optimization (SEO) basically ensures that any webpage/website having related content to end user search query appears at top in search results. The authors penned down that SEO technique are sufficient enough to be used for single search engine and the cost effectiveness can also easily be justified, but as long as multiple search engines are concerned there is a need of an adaptive search engine which provides the end user flexibility to consider multiple search engine optimization dynamically from single user interface. The authors pinpointed that generally any search engine decide relevance of any webpage/website by allocating them ranks. The webpage with higher ranks are considered to be most relevant. But sometimes even non relevant web pages make their way in the final output. Most of the search engine uses propriety ranking algorithms for allocating ranks to millions of web pages appearing on internet. These ranking algorithms generally consider frequency and location of keywords that appear on webpage. Hence the higher appearance of keyword either on webpage content or webpage meta content assures its higher rank. The authors mentioned that any search engine architecture consider two concepts for SEO perspectives:

[1] Crawling
[2] Indexation

Hence web pages in any website should be designed in such a way that they are easily crawled and indexed [5].

Laria et al. studied different search engines and proposed a new information retrieval model having its own searches, filtering and indexing schemes. The authors analyzed that most of the current search engines simply focus on matching the keywords from user search with the content on the web page. This process is very naïve and semantic content of the web document is simply ignored. The proposed information retrieval model mainly focus on semantics of the document which depends on different factors like common links, connected links and cross document similarities. This model assumes that documents having shared return links may be connected, the documents pointing to same links may also be connected and different cross-referenced documents showing some similarities in cross document comparison may be connected. All these points aid in understanding the semantics of the web document. The authors clustered the output of the search model in different groups depending on their similarities and differences. The clustering parameters used by the authors were size of the indexing vocabulary, depth of indexes, the data slot size, minimum similarity values etc. in the end the authors concluded that the newly proposed information retrieval model works fine for small group of web documents but its consistency might decrease if the sample size of domain is increased [6].

The author in her research article titled "Link analysis algorithm for web mining" discussed the importance of effectively and efficiently retrieving information from the World Wide Web. Some of the challenges faced by researchers in web mining are as follows:-

i. WWW is humongous
ii. Most of the web pages are semi-structured or unstructured
iii. The amount of information extracted from any web page and its concluding the meaning of it.

The author showed that any type of web mining can be categorized as web content mining, web structure mining or web usage mining. The retrieval of information from the World Wide Web can easily be done by understanding the content of web page which is also known as web content mining. The author mentioned that web content mining is different from any other type of data mining process as it considers the web documents which are generally semi-structured or unstructured in nature. The extraction of useful and relevant information from web document requires different technologies like natural language processing and sophisticated information retrieval model and algorithms [7].

III.SPAMDEXING TECHNIQUES

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A. Keyword Stuffing

Whenever search engine checks for the relevance of any web page, it counts the number of times keywords from the user search query appears on it. Any search engine assumes that if web pages contain the keyword repeatedly they are more likely to be relevant. Keyword stuffing in good old days was simply restricted to stuffing hidden text in the web page content but newer versions of the indexing algorithms checks whether the count of keyword is consistent with the other sites or not.

Older versions of keyword stuffing become obsolete soon. Hence new keyword stuffing techniques dynamically inserts the keyword randomly on any web pages irrespective of whether it is relevant with user search or not. Whenever end user searches any keyword java script of website runs and keyword is dynamically inserted in the web page content as shown below:

Source: https://tectrick.org/spamdexing-its-impact-on-website/

B. Hidden/invisible text

Unrelated hidden text is inserted in web page content to illegitimately increase the accessability of any website. The older version of hidden text technique simply hides the keywords or any unrelated text as

normal hidden tag field or making its color same as background color as shown below [8]:

bgcolor="#000000"

and the font color is black

 this is hidden text

OR by making it invisible as shown below [9]:

display:none, *visibility:hidden*, *height:0*, *width:0*, *text-spacing:-1000*

With this technique it became very easy for search engines to detect invisible text on the web pages hence new hidden text technique evolved. In this technique the unrelated text is hidden in the HTML code in "no frame" sections, "no script" sections, in alt attributes or in header tags etc.

C. Meta tag stuffing

Meta tag is a special tag use to give a short description of the contents of the website. Different types of meta tags are used to describe different web page properties. Any search engine checks meta tags as a quick reference to web page. If meta tag contents matches with the end user search query the web page is considered relevant. Meta tag stuffing generally adds more keywords than necessary in the tag. This technique is not always considered amongst Black hat SEO techniques but still can be used for this purpose. Meta tag has the limitation of 100 characters hence website developers should use keywords wisely in meta tag as shown below:

<meta name="Keywords" content="Arunachal Pradesh, Timeline - 2005">
<meta   name="description"   content="Arunachal Pradesh Timeline - 2005">

D. URL redirection

Some sites contain the web pages with higher ranks just to redirect the user to some other web pages with less rank. Generally the links like "click here to enter" are keep on the web page to take user to some completely unrelated web pages. Many well known company websites are being block for this type of spamdexing by Google in past. Nowdays some pages

contains redirection links on scroll bar of the web pages which complusorily takes user to other web pages as soon as user tries to scroll the web pages

E. Trap

As soon as end user reaches on any unrelated web pages he/she tries to return to prevoius page as known as bounce back. Search engines considers bounce back rate as one of the relevance check factors. Some web developers created trap for end user to deliberately decrease the bounce back rate. Web developers provides a web page with no back button working or whenever user tries to click on close button new page opens. This things keeps end user forcefully stuck on a particular web page.

IV.COUNTERMEASURES

Google introduced page ranking system to overcome keyword stuffing. If any website unethically introduces unrelated keyword on webpage, its relevance and consistency can be cross-checked from other website.

Appearance of multiple header tags in the source code of any web page leads to suspecting it with occurrence of meta tag stuffing. Search engines searches for unnecessary multiple header tags in the source code of any web document to detect meta tag stuffing.

Hidden text can be detected by disabling the CSS code for any web page. This method also aids in ousting the problem of invisible text on the web document.

Google proposed the use of "nofollow" tag in the links. When any link has "nofollow" tag embedded in it, any link-based search engine will not increase the score of next linked website. Most of the popular websites like Blogger, Wikipedia etc are using this tag as shown below:

```
<meta
    name="robots" content="noindex,nofollow">
```

V. CONCLUSION

People view whatever search engine shows to them. Hence in the era of digital marketing it is very important for any website to have high visibility and preferences. Apart from designing good website architecture it becomes mandatory to use effective SEO techniques for the same. SEO techniques can help to instantly increase the visibility of the website. Black-hat SEO techniques are generally prohibited by most of the search engines hence web developers should also be aware of these techniques to avoid ousting by search engines. This paper shows different black-hat SEO techniques and its counter measures.

VI. REFERENCES

[1]   Rehman, K., & Khan, M. N. A. (2013). The foremost guidelines for achieving higher ranking in search results through Search Engine Optimization. International Journal of Advanced Science and Technology,52, 101-110.

[2]   Danko, A. The concept of web search engines and metacrawlers within the Internet environment.

[3]   Selberg, E., & Etzioni, O. (1995, December). Multi-service search and comparison using the MetaCrawler. In Proceedings of the Fourth Int'l WWW Conference, Boston

[4]   Dr. B. V. Patel and Dr. R D Gaharwar.(2018) Search Engine Optimization (SEO) using HTML Meta-Tags. International Journal of Scientific Research in Science and Technology(IJSRST) , vol. 4 , Issue 9, page: 298-302.

[5]   Dr. B. V. Patel and Dr. Dipti Shah.(2011) Adaptive Systems and Search Engine Optimization. International Journal of Information and Computing Technology, ISSN: 0976-5999, vol. 1 , Issue 2, page: 14-15.

[6]   Laria, V.G., Griffiths, R., Winstanley, G.: Application of a clustering algorithm to recover topic content in an unstructured text-based environment.
https://www.researchgate.net/profile/ Graham_Winstanley/publication/2504349_App lication_of_a_Clustering_Algorithm_to_Recov er_Topic_Content_in_an_Unstructured_TextB ased_Environment/links/02e7e52f231abb5cac0 00000.pdf

[7]   T. Bhatia (2011). Link Analysis Algorithm For Web Mining. International Journal of Computer Science And Technology, Vol. 2, Issue 2, pp. 243-246.

[8]   http://www.seochat.com/c/a/search-engine-optimization-help/hidden-text-in-websites/

[9]   https://www.searchenginejournal.com/hidden-text-seo/208866/