

Mining Search Log for Keyword Stuffing

Ahir Bhoomika A., Bhalala Urvashi A., Pathan Azmeena R., Gangani Pinal L.,
Patel Suhani D., Poonam Godhwani

Abstract— The core fundamental concept behind keyword stuffing is to avail with combination of relevant and irrelevant keyword where the semantics of keywords are not matching which ultimately leads to poor ranking of website. Search engine spider interprets and detects places in website where the keywords are stuffed whereby considers it as a negative design element. Websites should be designed with high quality, well-written content included with keyword stuffing in limited region of website. Even though keyword stuffing is unlikely to lead to search engine penalties but if penalized then webmasters can recover, it could deter human visitors and reduce website value.

Keywords— Content Optimization, Keyword Stuffing, SEO, Spamdexing.

1. INTRODUCTION

The increasing use of web sites as marketing tools has resulted in major competition for high rankings amongst web sites. Search Engine Optimization (SEO) is a technique which helps search engine find & rank your site higher than the millions of other websites [3]. It is noted that the “value” of a web site can be determined by the page on which it ranks. For this reason SEO has become a priority for most web sites [1]. In SEO terminology, black hat SEO technique used for quickly high rank of your site, but usually does not obey search engine guidelines. It means black hat SEO refers to the use of aggressive SEO strategies & techniques that are unethical & unauthorized. One of these techniques is Keyword stuffing. Keyword stuffing is the practice of inserting a large number of keywords into Web page content and meta tags in the attempt to artificially increase the page's ranking in search results. There are number of popular keywords is being stuffed in web pages even if that are irrelevant to the site. It was so frustrating for users as they were not getting what they were searching for instead they were landing on pages with irrelevant words.

1.1 Keyword Stuffing Examples

There are some ways in which web developer tries to do keyword stuffing in their site are described below:

A. Repeating different keywords related to business

E.g. <meta name=“description” content=“call us today for contracting, contractors, remodelling, remodelers, home design, living room design, bathroom remodelling, home construction, living room construction, and bathroom

construction in los angeles, irvine, anaheim, orange country, san diego, san francisco, san jose, sacramento, and more.”/>

In the above example, we have a website that is keyword stuffing their meta description. Although the potential customer does not see all this content, the search engine can still crawl this text. This business owner is hoping that by listing various keywords that pertain to their business along, their website rank will well on search engines.

B. Using keywords not related to business

E.g.<imgsrc=http://content/uploads/2014/12/Our-Work-Background.png, alt=”contractors, Insurance, Insurance agency, agent, desk, bowl, notebook, cups, building, remodelling, remodelers, remodelling, kitchens, bathrooms, living rooms, computers, computer desks, office space, medicine, medical prescriptions, cheap medicine, gyms, gym, restaurant, air conditioners, heating repair, tax planners, tax planning, office supplier, water bottle delivery, bulk water delivery ” />

In the above example, we have a website that is keyword stuffing their alt text with keywords not related to their business. From “contractors” to “medical prescriptions” to “restaurant”, this business owner is hoping that their website will show up for various categories regardless of whether or not it is actually related to their business. Their main purpose is to gain as much traffic as possible.

C. Repeating the same keyword but making the text invisible to the reader

E.g. <p style=”color: white”>We provide the best dentists in los angeles. Call us today to get the best deals for our dentists. We service all of los angeles with dentists. Our dentists are certified and we provide teeth cleaning and teeth whitening that no other dentists compete with. If you need a new dentist, our dentist office is friendly and professional. Get to know our dentists today. </p>

In the above example, we have a website that is keyword stuffing their website content with the word “dentists”. When you read the content, you will notice that it does not provide useful information to the reader. You will also notice that the colour of the text is white. Since the background of the website is also white, this means that the reader cannot actually see this text on the page, but search engines still can. This is a business owner’s attempt at manipulating search engines to crawl keywords while hiding it from the reader.

1.2 Places where Keyword can be Stuffed

Spammers repeat the keywords in various HTML tags such as follow:

- title
- meta
- body
- head
- anchor
- alt

Also keywords are stuffed by spammers in URL [2].

A. URL keyword Stuffing

It is designed to give pages a higher rank than they truly deserve. The technique relies on mainly two assumptions about ranking algorithms:

1. Keyword matching is used
2. Matching against the URL is especially valuable

Taking advantage of this fact, spammers firstly use keyword rich domain names. Since their goal is to maximize impressions, they tend to go after high value/ frequency keywords (e.g. loan, payday, free, etc).

Spammers take a variety of approaches when implementing this concept. Here are some of the most common ones:

1. Multiple hosts, with keyword-rich hostnames. e.g.
<http://account.free.online.savings.samedaypaydayloansusa.com>
2. Host/ domain names with repeating keywords. e.g.
<http://loan.payday.paydayloanspaydayloansusa.com>
3. URL cluster across same domain, but varied hostnames comprised of keyword permutations.
 - a. <http://contososhoes.shoesonsale.com/>
 - b. <http://bestsrunningsneakers.shoesonsale.com/>
 - c. <http://discountrunningapparelforwomen.shoesonsale.com/>
4. URL squatting: This is a little different as the spammer is playing on a human tendency to misspell keywords of typically high profile/ traffic sites.
E.g. <http://ebey.com> (misspelling of <http://ebay.com>).

2. LITERATURE REVIEW

2.1 Search Engines interpreting keyword stuffing

SEs have been defined as programs that offer users interaction with the internet through a front end, where users can specify search terms. The SE will then compare the search term against an index file, which contains web page content. Any matches found are returned to the user via the front end [1].

SE optimizes the web pages through a piece of software, called a crawler or a spider.

Some search engines & their crawlers are as follows:

Table 1 Search Engine crawler

Search Engine	Crawler
Google	Googlebot
Yahoo!	Slurp
Bing	Bingbot
MSN	Msnbot
Ask	Teoma

With respect to three widely used search engines, Google, Yahoo & Bing, how do they interpret keyword stuffing?

1. Google went on to describe keyword stuffing under its quality guidelines, however, nothing was mentioned about the point at which a site is blacklisted or removed from its index for repeating keywords [1].
2. Yahoo! and Bing did not go into detail about keyword stuffing even though they stated some points in passing as part of the techniques that must be avoided [1].

2.2 Detection of keyword stuffing

Search engines are concerned with individual web pages, rather than whole websites. There are various algorithms to calculate relevancy of keywords used on webpage in response to a user's query. Each of algorithms has different relative weights for common factors.

Web site owners should make use of programs such as Webmaster Tools and other open source and commercial software for monitoring web page indexing. These tools also offer detailed information regarding the status of the submitted web pages [1]. One can check whether his webpage is blocked or spam indexing by search engine or not using webmaster tools.

Using these tools, one should analyse the submitted webpage to see whether there are errors preventing indexing, otherwise the webpage should be resubmitted.

E.g.

The screenshot displays the Google Webmaster Tools interface for the website utu.ac.in. At the top, it shows site health information: URL (http://www.utu.ac.in), Language (en, 89.9% Confidence), and various blocked status (No). Below this is a banner for 'webseo ANALYTICS' with the text 'A POWERFUL TOOLBOX BUILT BY SEO EXPERTS!' and a 'FREE TRIAL' button. The main content is divided into two columns: 'ON-PAGE STATS' and 'TITLE & METATAGS'.

ON-PAGE STATS		TITLE & METATAGS	
TITLE LENGTH	32	Title:	Uka Tarsadia University, Bardoli
TITLE RELEVANCE	100%	author:	srinca.edu.in
DESCRIPTION LENGTH	23	description:	Uka Tarsadia University
DESCRIPTION RELEVANCE	100%	distribution:	Global
KEYWORDS LENGTH	67	keywords:	Uka Tarsadia University, Uka Tarsadia, UTU University, UTU, college
KEYWORDS RELEVANCE	100%		
AVERAGE HTML HEADINGS RELEVANCE	85.71%		

Fig. 1 Google Webmaster tool Demo

In above Example, webmaster tool provide information related keywords used in site (utu.ac.in) & also gives some more information. It also shows relevancy of used keywords to your site.

2.3 Identification of site been penalized for keyword stuffing

Spammers have knowledge to evolve their techniques and cheat the system. However, there is some number of ways to check whether the webpage contain keyword stuffing or not discussed as below:

Remember keyword stuffing can also be due to the use of the ALT tag on pictures.

- To see what your alt tags contain open a page using Firefox, right hand click on an image, select 'View Image Info' and then see what is written next to the 'Associated text' label & check whether it is descriptive or spammy.

A good way to check whether a site is stuffed is to look at the *Content Keywords* page under *Google Index* in Google Webmaster Tools.

- Google Webmaster Tools (GWT) are used to monitor the health of your site on a regular basis. Once you've logged into GWT, navigate to the "Manual Actions" tab on the left hand side—you'll find it under the tab that says "Search Traffic".
- If Google has flagged you for keyword stuffing, you will see a message here that says "Hidden text and/or keyword stuffing".

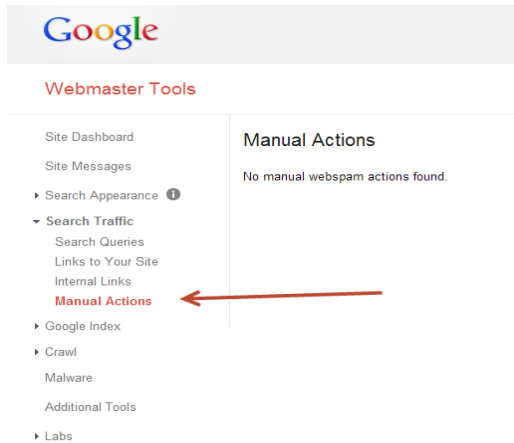


Fig.2 Google Webmaster Tool

Bing did also reveal a number of signals they use that could suggest possible use of URL keyword stuffing.

1. Site size
2. Number of hosts
3. Number of words in host name
4. % of the site cluster comprised of top frequency host/ domain name keywords
5. Host/ domain names containing certain lexicons/ pattern combinations (e.g. ["year", "event | product name"], <http://www.turbotaxonline2014.com>)
6. Site/page content quality & popularity signals.

3. Experiments and Analysis

3.1 Warning signs of got keyword stuffing problem

There are possible over-optimization or webspam signals that SEOs and webmasters should be aware of. It's worth noting that most of these factors also cause a poor user experience and reduce conversion rates as well.

- Spammy Page Titles
- Meta Keyword Tag Stuffing
- Content Keyword Stuffing

A. Spammy Page Titles

This includes titles that are consistently too long (over 64 characters), that include multiple different keywords, or use the same keyword more than twice. An

example of a bad title would be “Buy Blue Widgets | Blue Widget Distributor | Wholesale Discount Blue Widgets – Acme, Inc”. The ranking strength that titles give makes it a likely target for over-optimization and webspam.

B. Meta Keyword Stuffing

Many of the spammier sites stuff meta tag with dozens of keywords, and sometimes the same dozens on every page of the site. There’s nothing wrong with using the meta keywords tag, but it should hold no more than 5 or so keywords, and the keyword selection should be different from page to page.

C. Content Keyword Stuffing

Excessive keyword use in the body text of a page is one of the clearest overly-optimized signals on a page. There is a fine line between being sure to use your keywords often (which is good) and using them too often.

3.2 Problems with keyword stuffing

What is certain is that using a keyword or set of keywords or phrases too often begins to look like keyword stuffing to a search engine and that can negatively impact the ranking of your site. It begins to lose its effectiveness, and in fact can cause a search engine crawler to rank your site lower in the SERPs.

If there are too many keywords in your WebPages, mainly two problems can be occur:

1. effectiveness of your keywords will be reduced.
2. being delisted or ignored.

If you are targeting the keywords unrelated to your business, and there is a huge lack of quality content to support those keywords, it is possible to outrank your website.

If keywords are extreme, it might face registration cancellation of website.

Search engine also penalise a page if it contained text in the same colour as the background, considered in Keyword Stuffing.

Sometimes keyword stuffing causes spam indexing.

3.3 Right keywords to implement keyword stuffing

Choosing the right keywords to optimize for is thus the first and most crucial step to a successful SEO campaign [3]. The higher the compactness of keywords in webpage, the closer significance between the webpage and the keywords, and the more forward place the page ranked in the search results. Keywords compactness

is the ratio between keywords number and webpage words number [5]. A useful tool for discovering potential keywords is Google's Keyword Tool, which returns not only the estimated search volume of a keyword but also hundreds of related keywords and phrases [4].

The site should be designed to inform or convince your site visitor to reach a goal conversion. Some keywords that are proved to be right to your website are:

- A. Long-tail keywords
 - B. Implementing synonyms
 - C. Generic keywords
- A. Long-tail keywords

It defines the non-competitive keywords- usually three-to-five word phrases that site visitors use to find web sites & are very, very specific to whatever you are selling.

To capture the Long Tail, you need to create content that not only is relevant to your product, but is also related to your product.

- B. Implementing Synonyms

Another tip is to try implementing synonyms (words with the same meaning). SEs recognize that words like "bat" are homonyms; they are words that have multiple meanings. For this reason, there's a database index of Google synonyms to help it differentiate between word meanings. Google knows that if a site is talking about "clubs" and "bats," they are probably talking about sports equipment & not flying mammals.

- C. Generic keywords

Generic keywords are all the other keywords that are not directly associated with your company brand.

E.g. If your web site, TeenFashions.com, sells teen clothing, then keywords such as clothing, tank tops, cargo pants, and bathing suits might be generic keywords that you could use on your site.

3.4 Keyword stuffing Penalty & its recovery

Google's algorithm, Penguin, was designed to penalize what Google calls web spam. The goal of this is to target spammy SEO tactics.

Excessive keyword stuffing can also lead to a Panda penalty.

Keyword Stuffing Penalty Recovery

1. Re-write your content and/or your ALT contents for images.
2. Go the extra mile (show Google you are serious) and re-write the content of pages which Google has not flagged but which, when you think about it, contain keyword stuffing.

4. CONCLUSIONS

This document studies page ranking algorithms, search engine optimization techniques, black SEO techniques. Website ranking in search result is strongly depends on how SEO is implemented. Black hat SEO techniques provide quick results but for short time and if search engine find out the unethical activities of the site, then site can also penalized. The goal of document is provide awareness and detection of keyword stuffing & how one can apply keyword stuffing in right way in website.

5. REFERENCES

- [1] H. Zuze and M. Weideman, "Keyword stuffing and the big three search engines," Emerald Group Publishing Limited, vol. 37, no. 2, pp. 19, May 2013.
- [2] S. P. Patil, A. S. Patil, and B. V. Pawar, "Search Engine Optimization: A Study," Research Journal of Computer and Information Technology Sciences, vol. 1, no. 1, pp. 4, February 2013.
- [3] Dr. S. O. Khanna, "CONCEPT OF SEARCH ENGINE OPTIMIZATION IN WEB SEARCH ENGINE," International Journal of Advanced Engineering Research and Studies, vol. 1, no. 1, pp. 3, October-December 2011.
- [4] S. Saberi, G. Saberi, and M. Mohd, "What Does The Future of Search Engine Optimization Hold?," International Journal of New Computer Architectures and their Applications, vol. 3, no. 4, pp. 7, 2013.
- [5] R. Kumar and S. Saini, "A Study on SEO Monitoring System Based on Corporate Website Development," International Journal of Computer Science, Engineering and Information Technology (IJCSUIT), vol. 1, no. 2, pp. 8, June 2011.

6. AUTHORS' PROFILE

Ahir Bhoomika is studying in third year MCA programme of Shrimad Rajchandra Institute of Management and Computer Application affiliated to Uka Tarsadia University-Bardoli

Bhalala Urvashi is studying in third year MCA programme of Shrimad Rajchandra Institute of Management and Computer Application affiliated to Uka Tarsadia University-Bardoli

Pathan Azmeena is studying in third year MCA programme of Shrimad Rajchandra Institute of Management and Computer Application affiliated to Uka Tarsadia University-Bardoli.

Gangani Pinal is studying in third year MCA programme of Shrimad Rajchandra Institute of Management and Computer Application affiliated to Uka Tarsadia University-Bardoli.

Patel Suhani is studying in third year MCA programme of Shrimad Rajchandra Institute of Management and Computer Application affiliated to Uka Tarsadia University-Bardoli.

Poonam Godhwani is working as an Assistant Professor at SRIMCA. Her area of interests is SEO and Natural Language Processing. She has attended many seminars.