International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V71I3P103

Volume 71 Issue 3, 14-20, March 2023

Original Article

A Model to Detect Keyword Stuffing Spam on Webpages

Bodunde Odunola Akinyemi

Department of Computer Science and Engineering, Obafemi Awolowo University, Nigeria

Received: 28 January 2023 Revised: 04 March 2023 Accepted: 16 March 2023 Published: 28 March 2023

Abstract *- A well-designed website's dominant point and success depend on using keywords. Search engines heavily depend on the concept of keyword analysis to highlight results for search queries on web pages and to establish highly ranked websites. However, keyword stuffing evokes a spam issue with regard to the relevance of the content, so it becomes imperative that appropriate keywords are used to optimise web pages. This study developed a spam detection model to address the problem of keyword stuffing on a webpage. The model was developed by integrating three content analysis detection techniques: rates of compression ratio, average length, and keyword density. The Python programming language was used to implement the proposed approach. To evaluate the model's performance, twenty webpages were selected, out of which the contents of five sites were altered by including more keywords than usual. A simulation of the proposed model was tested on each webpage before and after the alteration of the keywords. The findings showed that before and after manipulation, the edited five sites' average identified keywords ranged from 2% to 3%. According to the results of the density of the pages' analysis, the average page density ranged from 3% to 5%. The study concluded that a keyword stuffing evaluation and detection model for webpages must be established to prevent online users from being misled and to increase trust between users and search engines.*

Keywords *- Content-based, Keyword density, Keyword Stuffing, Spam, Webpages.*

1. Introduction

The Keywords are key descriptive terms in a particular language that internet users employ as part of their search queries when using search engines to find information. The value of keywords varies depending on the domains on which they appear. Keywords are an essential component of website design since they play a significant role in how content is organised on websites. They serve as the foundation of website visibility and allow for web searches. Irrespective of an organisation's approach, keywords serve as the focal point of well-designed websites and are essential to their success.

A keyword analysis is the most critical component of Search Engine Optimization (SEO). Search engines heavily depend on the concept of keyword analysis to highlight results for search queries on web pages and to establish highly ranked websites. Using the right keywords is crucial [1], as doing so will improve the website's ranking in search results that contain the optimised keyword as well as the links it draws the most [2]. In addition, a high ranking might have a good impact on revenue growth. Keywords are essential SEO components for all search engines since these are what search terms are compared to. As a result, careful considerations should be made when employing them to prevent upsetting internet users rather than increasing contentment. Website keywords should generally be used by users to find the site when they search for it and should be able to describe the website's content.

Search engines typically give different fields different weights while ranking webpages to assess the relevancy of the content. Utilizing keywords in the text fields helps assess a web page's overall relevance in relation to a particular query. It is frequently observed that a keyword may be given greater weight if it occurs in the page title rather than the page body [3].

Keyword stuffing is considered a "black hat" or unethical technique. Keyword stuffing is used as using extreme keywords in text and meta tags or in certain parts of a web page to artificially boost a webpage's search engine rankings and draw visitors to the website [3-6]. The practise of keyword stuffing commonly referred to as "spamdexing," involves hiding significant and important words and phrases in web content. Some tactics, such as making the text the same colour as the page's background, changing the font size to zero, or hiding the phrase behind an image, are typically seen as kinds of keyword stuffing. Since the ranking of retrieved web pages in online search results is more important for various advertising objectives, many web pages make an effort to mislead search engines in order to

achieve high rankings [7]. Spammers use malicious links to increase the occurrence of specific keywords. Search engines give the phrase "spam pages" a high ranking when a query matches such a term [3]. This practise is regularly used to direct internet users to be fraudulent or hazardous websites. Ineffective SEO tactics, such as keyword stuffing as a ranking factor, are considered an attack or even cyber spam.

Keyword stuffing is considered a sort of Web spam. Creating spam websites using this method is a common practice [8]. Knowing the appropriate number of keywords to use without sacrificing relevancy is crucial since keyword stuffing compromises the quality of the content. In order to make a webpage relevant for some common search queries, the content is enhanced with a number of prominent keywords. The fact that the optimizers could use excessive repetition on the page is also a drawback. Stuffing can occasionally go too far and ruin the integrity of the content by making it unnatural [5,9]. Keyword stuffing can be seen on a website in the following ways:

- Keyword redundancy: Repetition of some keywords is a useful strategy for spamming a page's content. The likelihood of the page being spammed increases with the amount of keyword redundancy.
- Invisible keywords: By simply making the typeface of a webpage the same colour as the backdrop, spam keywords on a page can be rendered invisible to web users. The likelihood of the page being considered spam increases with the increasing invisibility rate.
- Webpage URL spam: URL spam describes spam tactics in which some spam keywords are included in the page's URL. Some spammers might wish to generate lengthy URLs with spammy keywords since search engines consider the keywords in URL addresses. The likelihood of keyword spam on a page increases with the size of the URL keyword utility.

Web spam is a nuisance since it makes people browse websites with information that frequently has nothing to do with the search query they entered. It has been characterized as one of the key issues that Web search engines need to resolve since it not only degrades the quality of search results but also undermines user confidence in the search engine provider and uses up a large portion of the search engine's processing capability [10]. The three common components of web search engines are web crawlers, document indexers, and document retrievers. It was acknowledged that indexing depends on how often crawlers visit a page. Crawlers will give websites with higher keyword densities first preference over those with lower keyword densities because they place a stronger emphasis on the significance of the content published. Email spam focuses on misleading people, whereas spamdexing targets search engine crawlers in an effort to score highly in search engine results. This is the key distinction between the two types of spam. One of the key issues that web search engines need to solve is spamdexing since it damages user confidence in the search engine more than just lowering the quality of results.

Many methods for identifying and preventing web spam have been put forth, but they are ineffective at identifying spam in web content. There is currently no conclusive quantitative evidence that shows where spamdexing and keyword-rich website text transcend over into keyword stuffing. The majority of approaches do not identify spam from the search engine's web crawling point prior to indexing and document fetching to reduce or eliminate keyword-stuffed pages from outranking legitimate pages. As a result, the problem of keyword stuffing, which has repeatedly misled online visitors, has not been sufficiently addressed [11]. Web content purposely published with the objective of causing some web pages to become unjustifiably relevant or significant is still how web spam manifests itself today. Therefore, it is necessary to investigate the issue of deceptive practises (spam) in online material, with a particular emphasis on the phenomena brought on by users who attempt to get some sort of profit by interfering with the regular operation of the processes.

In this study, an attempt was made to use the construct of content analysis to develop a keyword stuffing detection model that will uphold fairness in search engine ranking of websites and allow legitimate sites to acquire what they merit, prevent web users from being misled, and use fewer resources when crawling processing, and indexing websites.

The rest of this paper is organised as follows: Section 2 discusses related works, Section 3 describes the modelling process, Section 4 discusses the results, and Section 5 discusses the conclusions.

2. Related Works

There has been a continuing study into web spam detection for a while now. Distinct web spam detection algorithms have been proposed to examine the different kinds of spam that emerge on the web. The state-of-the-art the various methods used for detecting web spam was comprehensively analysed in [12-15]. Web spam is discovered either by examining the features of either or both web page content and the link.

Some research has used content-based features to detect web spam [16-22]. The researchers attempted to increase the likelihood of a webpage being returned as a search result and to raise its rating by stuffing the website with relevant keywords. By including keywords frequently used as search terms on a webpage, the page will appear in the list of results for the searches; selecting effective query term combinations will increase the percentage of the relevance score based on textual features. Search engines typically use the proposed content-based web detection algorithms to combat

spammers, who usually create content spam by stringing together a variety of well-known search phrases because web pages with a high concentration of topically irrelevant keywords or with poor grammar will show statistical differences from typical web pages. Similarly, spammers who generate millions of target web pages, each supplemented with just a few common search keywords, can also be detected using content-based web spam detection algorithms.

Also, link structure analysis was used in various studies [17-28]. By establishing numerous hyperlinks connecting to a particular webpage, the researchers attempted to improve its link-based score. The spammer's own and controlled websites, partner websites, or independent third-party websites may be where the links come from. In order to combat link spam, search engines typically use the proposed link-based web spam detection algorithms, which comb over the online graph looking for suspicious elements and spreading mistrust from spam webpages backwards along the web graph.

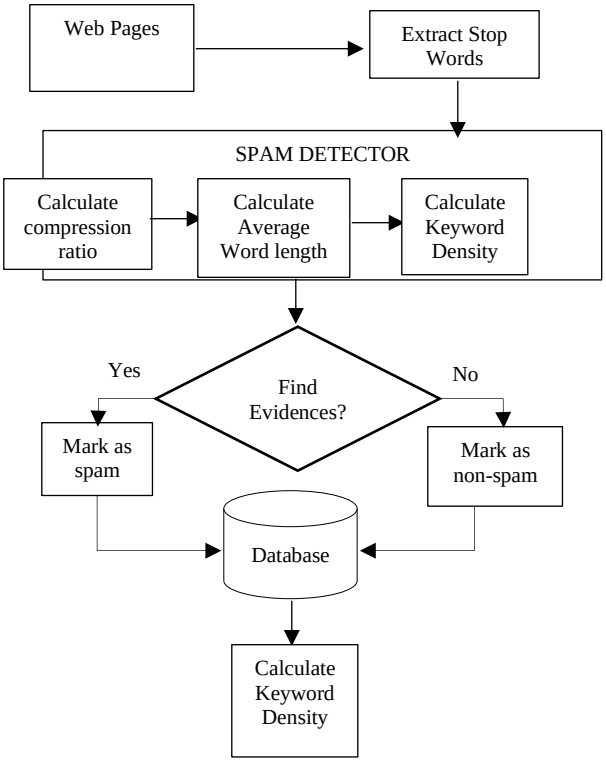
Some research had to combine the use of content-based and link-based features [19, 29-31]. The inclusion of additional elements, such as click-based and posting-based features for Spam URL Detection, has been embraced by certain studies [32-33].

The necessity of using keywords often in website text is stressed by numerous scholars, according to previous studies. However, the issue with the keyword density to content count ratio has not yet been resolved. However, most web spam detection algorithms do not explicitly state how keyword stuffing in a webpage is evaluated. There is no in-depth assessment of the factors crawlers take into account when determining if a website uses keyword stuffing, and spamdexing is regarded in a variety of different ways. Additionally, the algorithms' interpretations of spamdexing vary but do not provide any conclusive quantitative proof of keyword-stuffed websites.

3. Methodology

This study aimed to design a content analysis model to detect web spam in the form of keyword stuffing to optimise search results. Content analysis is used in this study to identify the existence of specific phrases in texts or collections of texts on a webpage. The concept was employed to quantify and analyse the presence, meanings, and relationships of such words and then make inferences about the messages within the texts. The proposed model is shown in Figure 1. This is described as follows:

- Website collection: this involves the gathering of web documents that can be displayed in a web browser.
- Extraction of stopwords: This involves the collection of foundational words like pronouns, articles, prepositions, and conjunctions like those present in the contents of the web pages.
- Keyword stuffing detector: Three sets of factors are used in this study to determine how a webpage is keyword-densified. The three major components are described as follows:
  (i.) Compression Ratio: Compression ratios are fractions, percentages, or ratios that express the difference between a web page's original size and its final size after



No	Compressibility	Ave Word Length	Keyword Density	Spam Status
1	…	…	…	0
2	…	…	…	1
…	…	…	…	…

Fig. 1 Proposed Keyword stuffing detection Model

compression. This ratio depends on a number of variables, primarily the initial file's condition and the compression algorithm employed. The more resources required to compress or decompress the webpage in question, the higher the compression ratio. This feature helps in determining the level of redundancy in a webpage. It is expressed in Equation 1 as follows:

$$Compression\ ratio = \frac{Size\ of\ normal\ web\ page}{Size\ of\ compressed} \quad (1)$$

(ii.) Average Word Length: The average word length is determined by dividing the total number of characters by the total number of words on the webpage, excluding the HTML tags. This is done to combat composite keyword spamming. It is expressed in Equation 2 as follows:

$$Average\ Word\ Length = \frac{Total\ amount\ of\ characters}{Total\ amount\ of\ words} \quad (2)$$

(iii.) Keyword Density: The keyword density is the frequency with which a term or phrase appears on a webpage. A page is more relevant to the search query and penalised if it has a high keyword density. It is expressed in Equation 3 as follows:

$$Keyword\ Density = \frac{Total\ amount\ of\ time\ a\ word\ appear}{Total\ amount\ of\ word\ in\ a\ page} \quad (3)$$

The detection component, which manages the overall detection of keyword stuffing, includes each of these three components. The model predicts that the details of the pages will be recorded to the database designated as spam if evidence of keyword stuffing was found and marked as non-spam if no evidence of keyword stuffing was found, based on the outcomes of the spam detector.

4. Results and Discussions

The proposed approach was transcribed into an algorithm, implemented, and evaluated in a Python Programming Language environment. Twenty (20) websites were chosen randomly, and the source code was downloaded from various domains. The selected source codes were compared with the proposed model. These websites' findings revealed that no evidence of keyword stuffing was found. After that, a few HTML pages were constructed and loaded with keywords to test the system's functionality and ensure it was operating as intended. The model detects keyword-stuffed web pages flawlessly. Five (5) of the sites downloaded were altered by adding more keywords to their sources to test the system further. Too many distinct terms were repeated across the web pages, so there are too few distinct terms on the web pages. This is accomplished by repeatedly using certain distinct terms (or a specific different term) throughout the content, in meta tags, alt attributes, and comment tags, among other places. Consequently, when the webpages were tested with the proposed model, it was able to identify the inclusion of keyword stuffing in the various webpages while pretending to adhere to the benchmark compression ratio of 4.0% and average word length of 8% [4], and the recommended keyword density of Google, which is 2%, and Yahoo, which is between 2% and 3% [34].

Table 1 shows the results of evaluating the keyword stuffing of web pages before and after manipulating their contents. Table 2 shows the part of the system database after the detection check, while Table 3 shows the keywords extracted from some of the web pages.

5. Conclusion

Numerous unethical strategies have been employed to harm web users' interests by having fake websites appear at the top of search results. Choosing the ideal keywords that potential buyers might use to locate the site is one of the pillars of a successful search engine marketing campaign. The study developed a balanced use of keywords to prevent keyword stuffing, which could lead to over-ranking of websites. This study aimed to develop a system to detect keyword stuffing in a web page supporting web spam detection. The proposed model does not deal with detecting all the spamming techniques but adequately deals with examining web page content for keyword stuffing.

Table 2. System database after detection check

ID	Title	Amount Of Words	Status ID
1043	Software Development	10425	1
1044	Home - BCX	14893	1
1045	Paga: Send and Receive Money	1597	1
1046	Axxess - Fibre I...	12679	2
1047	Strong authentication	8081	2
1048	CSS Template	121	1
1049	VoIP IHosted P...	7449	2
1050	Afrihost - Fibre,...	10757	2
1051	The 8 Best Alric...	10096	2
1052	Liquid Telecom...	2850	2

Table 3. keywords extracted from some pages

ID	Page_lD	Keywords
4176	1037	Service
4177	1037	Internet
4178	1037	Service
4179	1037	Providers
4180	1037	Isp
4181	1037	Telkom
4183	1037	Telkom
4186	1037	Telkom
4187	1037	Internet
4189	1037	Uncapped
4191	1037	Website
4192	1037	Hosting
4193	1037	Cpanel
4239	1043	Andela
4248	1043	Software
4249	1043	Developers
4250	1043	Hire
4251	1043	Developers
42SS	1043	Team
4256	1043	Andela
4262	1043	Invests
4264	1043	Africa
4267	1043	Talented

Table 1. Output before and after websites manipulation

Websites	Web Pages	Output before websites manipulation					Output after websites manipulation				
		CR (%)	AWL (%)	KWD (%)	AD (%)	S	CR (%)	AWL (%)	KWD (%)	AD (%)	S
Btechinstutute.edu.lr	Btechinsitutte.edu.lr/home	3	3	1	2	N	3	3	7	4	Y
	Btechinsitutte.edu.lr/about	4	3	1	3	N	7	3	5	5	Y
	Btechinsitutte.edu.lr/admission	4	3	1	3	N	4	3	6	4	Y
	Btechinstutute.edu.lr/administration	5	3	1	3	N	5	3	5	4	Y
	Btechinstutute.edu.lr/academic	5	3	1	3	N	5	3	6	5	Y
Afrihost.com	www.afrihost.com/landing/mobile	4	4	1	3	N	4	4	5	4	Y
	afrihost.com/site/product/windiws_hosting	3	3	1	2	N	3	3	5	4	Y
	afrihost.com/fibre	2	3	1	2	N	7	3	4	5	Y
	afrihost.com/site/product/domain_registration	3	3	1	2	N	4	3	4	4	Y
	afrihost.com/site/product/cloud_hosting	3	3	1	2	N	3	3	4	3	Y
axxess.co.za	axxess.co.za	4	3	1	3	N	7	3	5	5	Y
	axxess.co.za/fibre	3	3	1	2	N	6	3	4	4	Y
	axxess.co.za/mobile	4	3	1	3	N	7	3	3	4	Y
	axxess.co.za/hosting	3	3	1	2	N	3	3	4	3	Y
	axxess.co.za/voice	3	3	1	2	N	6	3	6	5	Y
privateproperty.com	privateproperty.com	3	3	1	2	N	3	3	7	4	Y
	privateproperty.com.ng/houses-for-sale	4	4	1	3	N	4	4	6	5	Y
	privateproperty.com.ng/land-for-sale	3	3	1	2	N	6	3	4	4	Y
	privateproperty.com.ng/flats-apartments-for-sale	3	3	1	2	N	3	3	7	4	Y
	privateproperty.com.ng/commercial-property-for-sale	3	4	1	3	N	4	4	4	4	Y
mwetana.com.lr	mwetana.com.lr/index.php	5	3	1	3	N	5	3	5	4	Y
	mwetana.com.lr/pages1.php?pgID=59	4	3	1	3	N	5	3	4	4	Y
	mwetana.com.lr/pages1.php?pgID=61	3	3	1	2	N	4	3	4	4	Y
	mwetana.com.lr/pages1.php?pgID=63	5	3	1	3	N	5	3	4	4	Y
	mwetana.com.lr/pages1.php?pgID=56	3	3	1	2	N	3	3	4	3	Y

Legend: Compression Ratio – CR | Average Word Length – AWL | Keyword Density – KWD | Average Detection – AD | Status - S
N-No │ Y-Yes

References

[1] Ahmad Al-Ananbeh et al., "Website Usability Evaluation and Search Engine Optimization for Eighty Arab University Websites," *Basic Science & Engineering*, vol. 21, no. 1, pp. 107-122, 2012. [Google Scholar] [Publisher link]

[2] Meenakshi Bansal, and Deepak Sharma, "Improving Webpage Visibility in Search Engines by Enhancing Keyword Density Using Improved On-Page Optimization Technique," *International Journal of Computer Science and Information Technologies*, vol. 6, no. 6, pp. 5347-5352, 2015. [Google Scholar] [Publisher link]

[3] Bin Zhou, Jian Pei, and Zhaohui Tang, "A Spamicity Approach to Web Spam Detection," *In Proceedings of the 2008 SIAM International Conference on Data Mining (SDM)*, pp. 277-288, 2008. [Google Scholar] [Publisher link]

[4] Alexandros Ntoulas et al., "Detecting Spam Web Pages Through Content Analysis," *In Proceedings of the ACM 15th international conference on World Wide Web*, pp. 83–92, 2006. [CrossRef] [Google Scholar] [Publisher link]

[5] Ashish Chandra, Mohammad Suaib, and Rizwan Beg, "Google Search Algorithm Updates Against Web Spam," *Informatics Engineering- an International Journal (IEIJ)*, vol. 3, no. 1, 2015. [CrossRef] [Google Scholar] [Publisher link]

[6] Santiago Villasenor et al., "Scalable Spam Classifier for Web Tables," *2017 IEEE International Conference on Big Data (Big Data)*, pp. 4849-4851, 2017. [CrossRef] [Google Scholar] [Publisher link]

[7] Tyler Moore, Nektarios Leontiadis, and Nicolas Christin, "Fashion Crimes: Trending Term Exploitation on the Web," *In Proceedings of the 18th ACM Conference on Computer and Communications Security*, pp. 455-466, 2011 [CrossRef] [Google Scholar] [Publisher link]

[8] Z. Gyongyi, and H. Garcia-Molina, "Web Spam Taxonomy," *First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, pp. 10-14, 2005.

[9] Sachin Kumar, and Pratishtha Gupta, "A Survey of Techniques and Applications for Search Engine Optimization," *Research Journal of Science and Technology*, vol. 8, no. 2, 2016. [CrossRef] [Google Scholar] [Publisher link]

[10] F. Javier Ortega, "Detection of Dishonest Behaviors in On-Line Networks Using Graph-Based Ranking Techniques," *AI Communications*, vol. 26, no. 3, pp. 327-329, 2013. [Google Scholar] [Publisher link]

[11] Cherukuri Kiranmai, and Gandi Satyanarayana, "Multi-Top Keyword Search Over Outsourced Data Files," *International Journal of Computer and Organization Trends*, vol. 8, no. 4, pp. 9-12, 2018. [Publisher link]

[12] Nikita Spirin, and Jiawei Han, "Survey on Web Spam Detection: Principles and Algorithms," *ACM SIGKDD Explorations Newsletter*, vol.13, no. 2, pp. 50-64, 2011. [CrossRef] [Google Scholar] [Publisher link]

[13] Carlos Castillo, and Brian Davison, "Adversarial Web search," *Foundations and Trends in Information Retrieval Journal*, vol. 4, no.5, pp. 377-486, 2011 [CrossRef] [Google Scholar] [Publisher link]

[14] Mugdha Kolhe, and Disha Bhukte, "Data Mining for Web Spam Detection Analysis of Techniques," *International Journal of Science and Research (IJSR)*, vol.5, no.10, pp. 1395 – 1399, 2015. [Publisher link]

[15] K. Jino Abisha et al., "Detection of Twitter Spam's using Machine Learning Algorithm," *SSRG International Journal of Computer Science and Engineering*, vol. 6, no. 3, pp. 10-13, 2019. [CrossRef] [Publisher link]

[16] Dennis Fetterly, Mark Manasse, and Marc Najork, "Spam, Damn Spam, And Statistics: Using Statistical Analysis to Locate Spam Webpages," *In Proceedings of the 7th International Workshop on the Web and Databases*, pp. 1–6, 2004. [CrossRef] [Google Scholar] [Publisher link]

[17] Gilad Mishne, David Carmel, and Ronny Lempel, "Blocking Blog Spam with Language Model Disagreement," *In Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web*, 2005. [Google Scholar] [Publisher link]

[18] Krysta Marie Svore et al., "Improving Web Spam Classification Using Rank-time Features," *In Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web (AIRWeb'07)*, 2007. [CrossRef] [Google Scholar] [Publisher link]

[19] Yiqun Liu et al., "Identifying Web Spam with User Behavior Analysis," *In Proceedings of the 4th international workshop on Adversarial information retrieval on the web (AIRWeb '08)*, pp. 9-16, 2008. [CrossRef] [Google Scholar] [Publisher link]

[20] Miklos Erd´elyi, Andras Garz´o, and Andras A. Bencz´ur, "Web Spam Classification: A Few Features worth More," *In Proceedings of the 2011 Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality'11)*, pp. 27-34, 2011. [CrossRef] [Google Scholar] [Publisher link]

[21] Odukoya Oluwatoyin et al., "An Improved Machine Learning-Based Short Message Service Spam Detection System," *International Journal of Computer Network and Information Security(IJCNIS)*, vol. 10, no.12, pp. 40-48, 2019. [CrossRef] [Google Scholar] [Publisher link]

[22] Brain D. Davison, "Recognizing Nepotistic Links on the Web," *In Proceedings of the AAAI-2000 Workshop on Artificial Intelligence for Web Search*, pp. 23–28, 2000. [Google Scholar] [Publisher link]

[23] Einat Amitay et al., "The Connectivity Sonar: Detecting Site Functionality by Structural Patterns," *In Proceedings of the 14th ACM Conference on Hypertext and Hypermedia*, pp. 38–47, 2003. [CrossRef] [Google Scholar] [Publisher link]

[24] James Caverlee, and Ling Liu, "Countering Web Spam with Credibility-Based Link Analysis," *In Proceedings of the Twenty-sixth Annual ACM Symposium on Principles of Distributed Computing (PODC'07)*, pp. 157-166, 2007. [CrossRef] [Google Scholar] [Publisher link]

[25] Andras A. Bencz´ur, Karoly Csalog´any, and Tamas Sarl´os, "Link-Based Similarity Search to Fight Web Spam," *In Proceedings of the Second Workshop on Adversarial Information Retrieval on the Web (AIRWeb'06)*, 2006. [Google Scholar] [Publisher link]

[26] Luca Becchetti et al., "Using Rank Propagation and Probabilistic Counting for Link-Based Spam Detection," *In Proceedings of the Workshop on Web Mining and Web Usage Analysis (WebKDD'06)*, 2006. [Google Scholar] [Publisher link]

[27] Xiaofei Niu, Guangchi Liu, and Qing Yang, "Trustworthy Website Detection Based on Social Hyperlink Network Analysis," *IEEE Transactions on Network Science and Engineering*, pp. 1-12, 2018. [Google Scholar] [Publisher link]

[28] Dorit S. Hochbaum, Quico Spaen, and Mark Velednitsky, "Detecting Aberrant Linking Behavior in Directed Networks," *In Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, pp. 72-82, 2019. [CrossRef] [Google Scholar] [Publisher link]

[29] Oluwatoyin Odukoya et al., "Performance Evaluation of User-Behavior Techniques of Web Spam Detection Models," *Network and Complex Systems*, vol.10, 2019. [CrossRef] [Google Scholar] [Publisher link]

[30] Zoltan Gyongyi, Hector Garia-Molina, and Jan Pedersen, "Combating Web Spam with TrustRank," *In Proceeding of the Thirtieth International Conference on Very Large Data Bases - VLDB '04*, vol. 30, pp. 576-587, 2004. [Google Scholar] [Publisher link]

[31] Carlos Castillo et al., "Know Your Neighbors: Web Spam Detection Using the Web Topology," *In Proceedings of the 30th Annual International ACM SIGIR Conference (SIGIR)*, pp. 423–430, 2007. [CrossRef] [Google Scholar] [Publisher link]

[32] Chao Wei et al., "Fighting Against Web Spam: A Novel Propagation Method Based On Click-Through Data," *in Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '12)*, pp. 395-404, 2012. [CrossRef] [Google Scholar] [Publisher link]

[33] B. Wu, and B. Davison, "Detecting Semantic Loaking On the Web," *In Proceedings of the 15th International Conference on World Wide Web (WWW'06),* pp. 819-828, 2006.

[34] Sean Si, Keyword Density Tutorial, SEO Hacker School Series. [Online]. Available: https://seo-hacker.com/keyword-density-tutorial/

[35] Jacob Abernethy, Olivier Chapelle, and Carlos Castillo, "Web Spam Identification Through Content and Hyperlinks," *In Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web*, pp. 41-44, 2008. [CrossRef] [Google Scholar] [Publisher link]