ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 5, Issue 11, November 2016**

# A Survey on Detecting Fake Pharmacy Sites

Shobhit Soni[1], Tejal Patil[1], Dinesh Patil[1], Professor S.G Phule[2]

Student, Department of Computer Engineering, Sinhgad College of Engineering Lonavala, Savitribai Phule Pune University, India[1]

Department of Computer Engineering, Sinhgad College of Engineering Lonavala, Savitribai Phule Pune University, India.[2]

**ABSTRAC**T: Fake online drug stores have turned out to be progressively inescapable, constituting more than 90% of on the web drug store sites. There is a requirement for fake site recognition systems equipped for distinguishing fake online drug store sites with a high level of precision. In this review, we thought about a few surely understood connections construct discovery methods in light of a substantial scale test bed with the hyperlink diagram enveloping more than 80 million connections between 15.5 million site pages, including 1.2 million known genuine and fake drug store pages. We found that the QoC and QoLclass propagation Algorithms accomplished an exactness of more than 90% on our dataset. The outcomes uncovered that calculations that join double class spread and additionally inlink and outlink data, on page-level or webpage level charts, are more qualified for recognizing fake drug store sites. Likewise, site-level investigation yielded essentially preferred outcomes over page-level examination for generally calculations assessed.

## I. INTRODUCTION

Online drug stores speak to a priceless wellspring of data and items for buyers and people looking for wellbeing related data. In like manner, there has been impressive development in the use of Internet drug stores. Lamentably, they have been joined by the multiplication of an expansive number of fake drug store sites (i.e., ones offering fake medications, giving invented data, as well as neglecting to deliver the settled upon products inside and out). As indicated by a Food and Drug Administration think about, under 10% of the 12,000 Internet drug stores analyzed were honest to goodness. The quantity of clients going to fake online drug stores has likewise tripled over the previous year. Fake drug store sites are probably going to offer a particular arrangement of difficulties to standard fake site recognition strategies because of an expansive number of inner connections inside a site. Despite the fact that such interior connections are normal inside an online store (like a drug store), they are not regular retained, bank, and other web based business destinations inspected by other parody website location contemplates. This gives the inspiration to concentrate the execution of best in class calculations on drug store destinations. Also, since fake drug store site designers burn through a huge number of dollars on dark cap site design improvement to make their sites more unmistakable we fight that they might be helpless to discovery by connection based strategies intended to battle web spam . Be that as it may, it is misty how powerful web spam discovery strategies, which make particular suppositions about the linkage amongst good and bad pages, would be on fake drug store sites.

## II. PROBLEM STATEMENT

Presently days, there has been progressively development in the use of Internet drug stores. Shockingly, they have been joined by the expansion of a substantial number of fake drug store sites i.e., ones offering fake medications, giving invented data, or potentially neglecting to deliver the settled upon merchandise inside and out. The quantity of clients going by fake online drug stores has likewise tripled over the previous year. The most effective method to distinguish the fake drug store is a major problem.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 5, Issue 11, November 2016**

## III. EXISTING SYSTEM

The existing system usesseveral link-based techniques for fake site identification. These techniques can be classified according to their propagation mechanisms, which depend on the presumptions they make with respect to how genuine and fake site pages are associated with each other. However, it is unclear as to how these link-based algorithms designed for web-spam detection wouldperform fake drug store sites. Moreover, there is a gap in the understanding of the impact of inlinks and outlinks (site page versus site) on these algorithms.

**Disadvantage of Existing System:-**

- The existing system does not detect appropriate fake website.
- Low performance.

## IV. PROPOSED SYSTEM

The proposed system uses several existing link-based techniques for fake site identification. The system first matches the URL then check QoC of page level and site level as well as finds the inlinks and outlinks of each page and site by sing the QoCandQoL algorithm.

**Advantage of Proposed System:-**

- The propose system detect appropriate fake website.
- Increase performance.

## V. LITERATURE SURVEY

Abbasi, A. and Chen, H.( 2007) have represented Detecting Fake Escrow Websites using Rich Fraud Cues and Kernel Based Methods[1].The author assessed the viability of different components and strategies for recognizing fake escrow websites. The examination incorporated a rich arrangement of elements separated from website page content, picture, and link information. The author also proposed a composite kernel tailored to represent the properties of fake websites, including content duplication and structural attributes. The mix of an expanded feature set and the composite kernel attained over 98% accuracy when differentiating fake sites from real ones, using the support vector machines algorithm. The results suggest that automated web-based information systems for detecting fake escrow sites could be feasible and may be utilized as authentication mechanisms.

Abbasi, A., Zhang, Z., and Chen, H(2008) have discovered A Statistical Learning Based System for Fake Website Detection[2].Existing fake site recognition frameworks can't adequately distinguish fake sites. In this review, the author advocate the improvement of fake site recognition frameworks that utilize that employ classification methods grounded in statistical learning theory (SLT). Trial comes about uncover that a model framework created utilizing SLT-based techniques outflanks seven existing fake website detection systems on a test bed encompassing900 real and fake websites.

Abbasi, A. and Chen, H. (2009) has represent A Comparison of Fraud Cues and Classification Methods for Fake Escrow Website Detection[3]. The author evaluated the effectiveness of various features and techniques for detecting fake escrow websites. The analysis included a rich set of fraud cues extracted from web page text, image, and link information. The author also compared several machine learning algorithms, including support vector machines, neural networks, decision trees, naïve bayes, and principal component analysis. The study confirms the feasibility of using automated methods for detecting fake escrow websites. The results may also be useful for informing existing online escrow fraud resources and communities of practice about the plethora of fraud cues pervasive in fake websites.

Zhang, L., Zhang, Y., and Zhang, Y., and Li, X.(2006) has discovered Exploring Both Content and Link Quality for Anti-Spamming[4]. The author has proposed a new page importance metric, which takes both the content quality and the link quality into consideration. Based on this metric, we can judge the trust scores of all the web pages using the web link graph.

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker Jr., J. F.( 2010) has proposed Detecting Fake Websites: The Contribution of Statistical Learning Theory[5]. The author has proposed the development of a new class of fake website detection systems that are based on statistical learning theory (SLT). Using a design science approach, a prototype system was developed to demonstrate the potential utility of this class of systems.

Krishnan, V. and Raj, R. (2006) has discover Web Spam Detection with Anti-Trust Rank [6]. The author proposes a method of selecting a seed set of pages to be evaluated by a human. We then use the link structure of the web and the manually labeled seed set, to detect other spam pages. Our experiments on the Web Graph dataset show that our approach is very effective at detecting spam pages from a small seed set and achieves higher precision of spam page detection than the Trust Rank algorithm, apart from detecting pages with higher page ranks on an average.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 5, Issue 11, November 2016**

Abbasi, A. and Chen, H. (2009b), A Comparison of Tools for Detecting Fake Websites [7].As fake website developers become more innovative, so too must the tools used to protect Internet users. A proposed system combines a support vector machine classifier and a rich feature set derived from website text, linkage, and images to better detect fraudulent sites.

Gyongyi, Z. and Garcia-Molina, H. (2005) has represents Spam: It's not Just for Inboxes Anymore [8].E-mail spam is a nuisance that every user has come to expect. But Web spammers prey on unsuspecting users and undermine search engines by subverting search results to increase the visibility of their pages.

Gyongyi, Z., Garcia-Molina, H., and Pedersen, J. (2004) has represents Combating Web Spam with Trust Rank [9].Web spam pages utilize different procedures to accomplish higher-than-merited rankings in a web crawler's outcomes. While human specialists can recognize spam, it is excessively costly, making it impossible to physically assess countless. Rather, we propose methods to semi-naturally isolate trustworthy, great pages from spam. We first select a little arrangement of seed pages to be assessed by a specialist. When we physically distinguish the legitimate seed pages, we utilize the connection structure of the web to find different pages that are probably going to be great. In this paper we talk about conceivable approaches to actualize the seed choice and the disclosure of good pages. We exhibit consequences of investigations keep running on the World Wide Web filed by AltaVista and assess the execution of our procedures. Our outcomes demonstrate that we can successfully sift through spam from a huge portion of the web, in view of a decent seed set of under 200 destinations.

Liu, W., Deng, X., Huang, G., and Fu, A. Y.(2006) have discovered An Antiphishing Strategy Based on Visual Similarity Assessment[10].The authors' Proposedantiphishing methodology utilizes visual attributes to distinguish potential phishing sites and measure suspicious pages closeness to actual sites registered with the system. The first of two successive procedures in the SiteWatcher system keeps running on nearby email servers and screens messages for catchphrases and suspicious URLs. The second procedure then thinks about the potential phishing pages against genuine pages and surveys visual similitude's between them as far as key sites, page designs, and general styles. The approach is intended to be a piece of an endeavor antiphishingsolution.

## VI. SYSTEM ARCHITECTURE

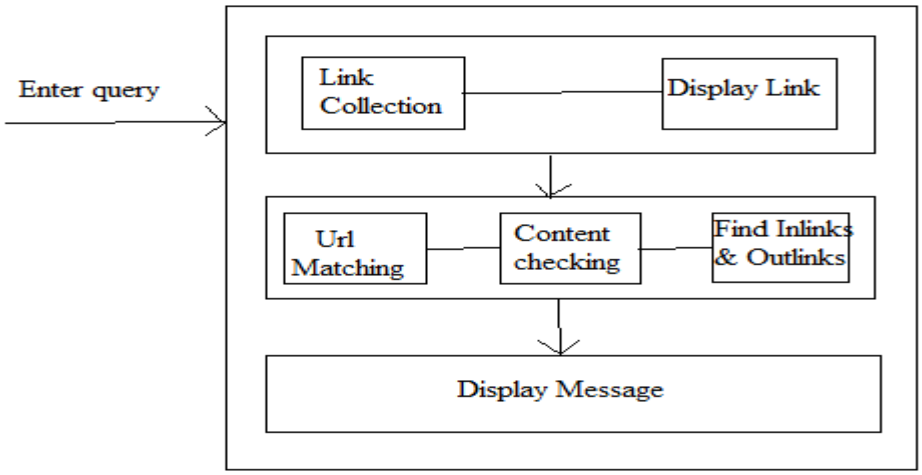


**Fig.System Architecture**

## VII. CONCLUSION

In this preparatory review, we examined the effectiveness of a few existing link-based algorithms for the recognition of fake online drug store sites. The outcomes uncovered that algorithms and chart qualities play an important role in the exact identification of fake drug store sites.Dual propagation algorithms that considered inlink and outlink data for good and bad seed pages given the best execution. Facilitate, site-level diagrams yielded extensively preferable outcomes over page level ones for most algorithms assessed.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 5, Issue 11, November 2016**

## REFERENCES

[1] Zhang, L., Zhang, Y., and Zhang, Y., and Li, X. "Exploring Both Content and Link Abbasi, A. and Chen, H. "Detecting Fake Escrow Websites using Rich Fraud Cues and Kernel Based Methods," In the 17th Annual Workshop on Information Technologies and Systems, Montreal, Canada, December 8-9, 2007, pp. 55-60.
[2] Abbasi, A., Zhang, Z., and Chen, H. "A Statistical Learning Based System for Fake Website Detection," In the Workshop on Secure Knowledge Management, Dallas, Texas, November 3-4, 2008.
[3] Abbasi, A. and Chen, H. "A Comparison of Fraud Cues and Classification Methods for Fake Escrow Website Detection," Information Technology and Management, 10(2), 2009a, pp. 83-101.
[4] Quality for Anti- Spamming," In Proceedings of the 6th IEEE International Conference on Computer and Information Technology, 2006, pp. 37-42.
[5] Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker Jr., J. F. "Detecting Fake Websites: The Contribution of Statistical Learning Theory," MIS Quarterly, 2010, forthcoming.
[6] Krishnan, V. and Raj, R. "Web Spam Detection with Anti-Trust Rank," In Proceedings of the 2nd International Workshop on Adversarial Information Retrieval on the Web, 2006, pp. 37-40.
[7] Abbasi, A. and Chen, H. "A Comparison of Tools for Detecting Fake Websites," IEEE Computer, 42(10),2009b, pp. 78-86.
[8] Gyongyi, Z. and Garcia-Molina, H. "Spam: It's not Just for Inboxes Anymore," IEEE Computer, 38(10), 2005, pp. 28-34.
[9] Gyongyi, Z., Garcia-Molina, H., and Pedersen, J. "Combating Web Spam with Trust Rank," In Proceedings of the 13th International Conference on Very Large Data Bases, 2004, pp. 576-587.
[10] Liu, W., Deng, X., Huang, G., and Fu, A. Y. "An Antiphishing Strategy Based on Visual Similarity Assessment," IEEE Internet Computing, 10(2), 2006, pp. 58-65.