Abstract

Spam is a major threat to web security. The web of trust is being abused by the spammers through their ever evolving new tactics for their personal gains. In fact, there is a long chain of spammers who are running huge business campaigns under the web. Spam causes underutilization of search engine resources and creates dissatisfaction among web community. Web Security being a prime challenge for search engines has motivated the researchers in academia and industry to devise new techniques for web spam detection. In this paper we present a comprehensive survey of techniques for detection of web spam and discuss their applicability and performance in various scenarios where they outperformed the others. We have categorized web spam detection with the primary focus on the approaches used for spam detection. The paper also gives the possible directions for future work.

References

- Abernethy, J., Chapelle, O., & Castillo, C. "Graph regularization methods for
- Caferrella M. & Cutting, &quot;Building Nutch: Open source search;&quot;, Queue, (2:2), 2004, pp. 54-61.
- Cohen, W. W. & Kou, Z. &quot;Stacked graphical learning: approximating learning in markov random fields using very short inhomogeneous markov chains;&quot;, Technical report,
2006.  
- Dudley, J. , Barone, L. , & While, L. "Multi-objective spam filtering using an evolutionary algorithm".  In Evolutionary Computation, IEEE World Congress on Computational Intelligence, 2008, June, pp. 123-130.  
- Fuad, M. M. , Deb, D. , & Hossain, M. S. "A trainable fuzzy spam detection system".  In Proc. of the 7th Int. Conf. on Computer and Information Technology, 2004, December  
- Gyongyi, Z. , & Garcia-Molina, H. "Web spam taxonomy".  In First international workshop on adversarial information retrieval on the web AIRWeb, 2005.  
Approaches for Web Spam Detection

- Spriin, Nikita, and Jiawei Han. "Survey on web spam detection: principles and algorithms". ACM SIGKDD Explorations Newsletter 13. 2 (2012): 50-64.
- Sobek, M. "Pr0-google's pagerank 0 penalty. Badrank". 2002.
IEEE.
- Wu, B. , & Davison, B. D. "Identifying link farm spam pages". In Special interest tracks and posters of the 14th international conference on World Wide Web, 2005, May, pp. 820-829. ACM.
- Zhang, Y. , Li, H. , Niranjan, M. , & Rockett, P. "Applying cost-sensitive multiobjective genetic programming to feature extraction for spam e-mail filtering". Genetic Programming, Springer Berlin

**Index Terms**

Computer Science

Security
Keywords
  Anti-Spam  web security  spam detection  approaches  search engines