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
- Aburrous, M. , Hossain, M. A. , Dahal, K. , & Thabtah, F. &quot;Intelligent phishing
detection system for e-banking using fuzzy data mining&quot;, Expert systems with
applications, (37:12), 2010, 7913-7921.
- Agichtein, E. , Brill, E. , & Dumais, S. &quot;Improving web search ranking by
incorporating user behaviour information&quot;, In Proceedings of the 29th annual international
ACM SIGIR conference on Research and development in information retrieval, 2006 August,
- Akoglu, L. , & Faloutsos, C. &quot;Anomaly, event, and fraud detection in large network
datasets&quot;, In Proceedings of the sixth ACM international conference on Web search and
- Almeida, T. A. , & Yamakami, A. &quot;Facing the spammers: A very effective
6557-6561.
- Amitay, E. , Carmel, D. , Darlow, A. , Lempel, R. , & Soffer, A. &quot;The connectivity
sonar: detecting site functionality by structural patterns&quot;, In Proceedings of the fourteenth
ACM conference on Hypertext and hypermedia, 2003, August, pp. 38-47
- Anagnostakis, K. G. , Sidiroglou, S. , Akritidis, P. , Xinidis, K. , Markatos, E. , &
- Blei, D. M. , Ng, A. Y. , & Jordan, M. I. &quot;Latent dirichlet allocation&quot;, The
&amp;quot;Link-Based Characterization and Detection of Web Spam&quot;, In international
1-8.
rank propagation and probabilistic counting for link-based spam detection&quot;, In
Proceedings of WebKDD (Vol. 6), 2006, August.
54-61.
- Castillo, C. , Donato, D. , Gionis, A. , Murdock, V. , & Silvestri, F. &quot;Know your
neighbors: Web spam detection using the web topology&quot;, In Proceedings of the 30th
annual international ACM SIGIR conference on Research and development in information
- Cohen, W. W. & Kou, Z. &quot;Stacked graphical learning: approximating learning in
markov random fields using very short inhomogeneous markov chains&quot;, Technical report,
Approaches for Web Spam Detection

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.
- Spirin, Nikita, and Jiawei Han. "Survey on web spam detection: principles and algorithms." ACM SIGKDD Explorations Newsletter 13. 2 (2012): 50-64.
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