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
2006.
- Dai, N., Davison, B. D., & Qi, X. "Looking into the past to better classify web
spam", In Proceedings of the 5th international workshop on adversarial information
- 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.
- Erdélyi, M., Garzó, A., & Benczúr, A. A. "Web spam classification: a few features
worth more", In Proceedings of the 2011 Joint WICOW/AIRWeb ACM Workshop on Web
Quality, 2011, March, pp. 27-34.
- 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.
- Fetterly, D., Manasse, M., & Najork, M. "Detecting phrase-level duplication on
the world wide web", In Proceedings of the 28th annual international ACM SIGIR
conference on Research and development in information retrieval, 2005, August, pp. 170-177.
- Fetterly, D., Manasse, M., & Najork, M. "Spam, damn spam, and statistics: Using
statistical analysis to locate spam web pages", In Proceedings of the 7th International
Workshop on the Web and Databases: colocated with ACM SIGMOD/PODS 2004, pp. 1-6,
ACM.
- Friedman, J., Hastie, T., & Tibshirani, R. "Additive logistic regression: A
- Ghiam, Shekoofeh, and Alireza Nemaney Pour. "A Survey on Web Spam
- Gyongyi, Z., & Garcia-Molina, H. "Web spam taxonomy", In First international
workshop on adversarial information retrieval on the web AIRWeb, 2005.
- K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song. "Design and evaluation
of a real-time URL spam filtering service", In IEEE Symposium on Security and Privacy,
2011.
- Liu, Y., Chen, F., Kong, W., Yu, H., Zhang, M., Ma, S., & Ru, L. "Identifying
Web Spam with the Wisdom of the Crowds", ACM Transactions on the Web (TWEB),
(6:1), 2012, pp. 2-12.
- Liu, Y., Zhang, M., Ma, S., & Ru, L. "User behavior oriented web spam
April, pp. 1039-1040. ACM.
- Lu, L., Perdisci, R., & Lee, W. "SURF: detecting and measuring search
poisoning", In Proceedings of the 18th ACM conference on Computer and
communications security, 2011, October, pp. 467-476. ACM.
- Martin, A., Anutthamaa, N., Sathyavathy, M., Francois, M. M. S., & Venkatesan, P.
Approaches for Web Spam Detection

- Spirin, Nikita, and Jiawei Han. “Survey on web spam detection: principles and algorithms.” ACM SIGKDD Explorations Newsletter 13. 2 (2012): 50-64.
Approaches for Web Spam Detection

IEEE.
- 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