

BIBLIOGRAPHY

- [1] D. Doran, “Detection, classification, and workload analysis of web robots,” 2014.
- [2] R. Cooley, B. Mobasher, and J. Srivastava, “Web mining: Information and pattern discovery on the world wide web,” in *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on*, pp. 558–567, IEEE, 1997.
- [3] P.-N. Tan and V. Kumar, “Discovery of web robot sessions based on their navigational patterns,” in *Intelligent Technologies for Information Analysis*, pp. 193–222, Springer, 2004.
- [4] “Google search engine.” [Online]. Available:<https://www.google.com/>.
- [5] “Yahoo search engine.” [Online]. Available:<https://www.yahoo.com/>.
- [6] “msn search engine.” [Online]. Available:<http://www.msn.com/en-in/>.
- [7] “bing search engine.” [Online]. Available:<http://www.bing.com/>.
- [8] “Four fake google haxbots hit your website every day.” [Online]. Available:http://www.theregister.co.uk/2014/07/25/four_fake_google_haxbots_hit_your_website_every_day/.
- [9] “Fake googlebot activity up 61%.” [Online]. Available:<http://searchenginewatch.com/article/2358345/Fake-Googlebot-Activity-up-61-Report>.
- [10] “Verifying googlebot.” [Online]. Available:<https://support.google.com/webmasters/answer/80553?hl=en>.
- [11] “Apache http server version 2.2.” [Online]. Available:<http://httpd.apache.org/docs/2.2/logs.html>.
- [12] “Gnu wget.” [Online]. Available:<http://www.gnu.org/software/wget/>.
- [13] “Httrack website copier.” [Online]. Available:<http://www.httrack.com/>.
- [14] “Rapidminer.” [Online]. Available:<http://rapidminer.com/>.
- [15] “Python.” [Online]. Available:<https://www.python.org/>.
- [16] “The perl programming language.” [Online]. Available:<https://www.perl.org/>.

- [17] "Java." [Online]. Available:<https://www.java.com/en/>.
- [18] L. Destailleur, "Awstats official web site." <http://www.awstats.org/>, 2010.
- [19] P. Huntington, D. Nicholas, and H. R. Jamali, "Web robot detection in the scholarly information environment," *Journal of Information Science*, vol. 34, no. 5, pp. 726–741, 2008.
- [20] M. Koster, *A standard for robot exclusion*. NEXOR., 1994.
- [21] D. Stevanovic, N. Vlajic, and A. An, "Detection of malicious and non-malicious website visitors using unsupervised neural network learning," *Applied Soft Computing*, vol. 13, no. 1, pp. 698–708, 2013.
- [22] M. D. Dikaiakos, A. Stassopoulou, and L. Papageorgiou, "An investigation of web crawler behavior: characterization and metrics," *Computer Communications*, vol. 28, no. 8, pp. 880–897, 2005.
- [23] L. Von Ahn, M. Blum, N. J. Hopper, and J. Langford, "Captcha: Using hard ai problems for security," in *Advances in Cryptology—EUROCRYPT 2003*, pp. 294–311, Springer, 2003.
- [24] A. Balla, A. Stassopoulou, and M. D. Dikaiakos, "Real-time web crawler detection," in *Telecommunications (ICT), 2011 18th International Conference on*, pp. 428–432, IEEE, 2011.
- [25] D. Stevanovic, N. Vlajic, and A. An, "Unsupervised clustering of web sessions to detect malicious and non-malicious website users," *Procedia Computer Science*, vol. 5, pp. 123–131, 2011.
- [26] N. Algiriyage, S. Jayasena, G. Dias, A. Perera, and K. Dayananda, "Identification and characterization of crawlers through analysis of web logs," in *Industrial and Information Systems (ICIIS), 2013 8th IEEE International Conference on*, pp. 150–155, Dec 2013.
- [27] N. Ye *et al.*, "A markov chain model of temporal behavior for anomaly detection," in *Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop*, vol. 166, p. 169, West Point, NY, 2000.
- [28] D. Zhang, D. Zhang, and X. Liu, "A novel malicious web crawler detector: Performance and evaluation," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 1, 2013.
- [29] G. Stermsek, M. Strembeck, and G. Neumann, "A user profile derivation approach based on log-file analysis.," in *IKE*, vol. 2007, pp. 258–264, Citeseer, 2007.
- [30] Y. Xie and V. V. Phoha, "Web user clustering from access log using belief function," in *Proceedings of the 1st international conference on Knowledge capture*, pp. 202–208, ACM, 2001.

- [31] J. Xu and H. Liu, "Web user clustering analysis based on kmeans algorithm," in *2010 International Conference on Information Networking and Automation (ICINA)*, vol. 2, pp. V2–6, IEEE, 2010.
- [32] A. Sharma, A. Goel, P. Gulati, *et al.*, "A novel approach for clustering web user sessions using rst," in *Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT'09. International Conference on*, pp. 657–659, IEEE, 2009.
- [33] O. Nasraoui, H. Frigui, A. Joshi, and R. Krishnapuram, "Mining web access logs using relational competitive fuzzy clustering," in *Proceedings of the Eight International Fuzzy Systems Association World Congress*, vol. 1, pp. 195–204, Citeseer, 1999.
- [34] B. S. Suryavanshi, N. Shiri, and S. P. Mudur, "An efficient technique for mining usage profiles using relational fuzzy subtractive clustering," in *Web Information Retrieval and Integration, 2005. WIRI'05. Proceedings. International Workshop on Challenges in*, pp. 23–29, IEEE, 2005.
- [35] G. Castellano, F. Mesto, M. Minunno, and M. A. Torsello, "Web user profiling using fuzzy clustering," in *Applications of Fuzzy Sets Theory*, pp. 94–101, Springer, 2007.
- [36] J. Xiao, Y. Zhang, X. Jia, and T. Li, "Measuring similarity of interests for clustering web-users," in *Proceedings of the 12th Australasian database conference*, pp. 107–114, IEEE Computer Society, 2001.
- [37] J. D. Velásquez, H. Yasuda, and R. WEBER, "A new similarity measure to understand visitor behavior in a web site," *IEICE TRANSACTIONS on Information and Systems*, vol. 87, no. 2, pp. 389–396, 2004.
- [38] B. De Finetti and B. de Finetti, "Theory of probability, volume i," *Bull. Amer. Math. Soc.* 83 (1977), 94-97 DOI: <http://dx.doi.org/10.1090/S0002-9904-1977-14188-8> PII, pp. 0002–9904, 1977.
- [39] "Computational biology." [Online]. Available:<http://www.cs.hunter.cuny.edu/~saad/courses/compbio/lectures/lecture9.pdf>.
- [40] S. Langhnoja, M. Barot, and D. Mehta, "Pre-processing: Procedure on web log file for web usage mining," *International Journal for Emerging Technology and advanced engineering*, vol. 2, no. 12, 2012.
- [41] R. Cooley, B. Mobasher, and J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowledge and information systems*, vol. 1, no. 1, pp. 5–32, 1999.
- [42] "Hypertext transfer protocol – http/1.1." [Online]. Available:<https://www.ietf.org/rfc/rfc2616.txt>.

- [43] “Spamhaus database.” [Online]. Available:<http://www.spamhaus.org/zen/>.
- [44] “Abuseat database.” [Online]. Available:<http://cbl.abuseat.org/>.
- [45] “Project honeypot database.” [Online]. Available:<https://www.projecthoneypot.org/index.php>.
- [46] R. Durbin, *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998.
- [47] “Google crawlers.” [Online]. Available:<https://support.google.com/webmasters/answer/1061943?hl=en>.
- [48] “Encoding categorical features.” [Online]. Available:<http://scikit-learn.org/stable/modules/preprocessing.html>.
- [49] H. Jiawei and M. Kamber, “Data mining: concepts and techniques,” *San Francisco, CA, itd: Morgan Kaufmann*, vol. 5, 2001.
- [50] “scipy.cluster.hierarchy.dendrogram.” [Online]. Available:<http://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.cluster.hierarchy.dendrogram.html>.