

7 Bibliografia

[1] Lindemann, C. e Littig, L. *Coarse Grained Classification of Web Sites by Their Structural Properties*. In: *Proceedings of the eighth ACM international workshop on Web information and data management*, 2006

[2] Lindemann, C. e Littig, L. *Classifying Web Sites*. In: *Proceedings of the 16th international conference on World Wide Web*, 2007

[3] Mishne, G. e de Rijke, M. *Capturing Global Mood Levels using Blog Posts*. In: *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, 2006

[4] Java, A., Kolari, P., Finin, T., Joshi, A. e Martineau, J. *BlogVox: Separating Blog Wheat from Blog Chaff*. In: *Proceedings of the Workshop on Analytics for Noisy Unstructured Text Data, 20th International Joint Conference on Artificial Intelligence*, 2007

[5] Kolari, P., Finin, T. e Joshi, A. *SVMs for the Blogosphere: Blog Identification and Splog Detection*. In: *Proceedings of the AAAI Spring Symposium on Computational Approaches to Analysing Weblogs*. AAAI Press, Março 2006

[6] Bansal, N. e Koudas, N. *Searching the Blogosphere*. In: *Proceedings of the 10th International Workshop on Web and Databases (WebDB 2007)*, Junho 2007

[7] Glance, N. S., Hurst, M. Tomokiyo, T. *BlogPulse: Automated Trend Discovery for Weblogs*. In: *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, 2004

[8] Kolari, P., Finin, T., Java, A., Joshi, A., Martineau, J. e Mayfield, J. *Blog Track Open Task: Spam Blog Classification*. TREC 2006

[9] Nanno, T., Suzuki, Y., Fujiki, T. e Okumura, M. *Automatically Collecting, Monitoring, and Mining Japanese Weblogs*. In: *Proceedings of the 13th international World Wide Web conference*, 2004

[10] Attardi, G., Gulli, A. e Sebastiani, F. *Automatic Web Page Categorization By Link and Context Analysis*. In: *Proceedings of THAI'99, European Symposium on Telematics, Hypermedia and Artificial Intelligence*, 1999

[11] Page, L. *The PageRank Citation Ranking: Bringing Order to The Web*. In: *Proceedings of ASIS' 98, Annual Meeting of the American Society for Information Science*, 1998

[12] Mishne, G. e de Rijke, M. *A Study of Blog Search*. Em Lecture Notes In: *Computer Science*, no. 3936, 2006

[13] Yang, Y., Slaterry, S. e Ghani, R. *A Study of Approaches To Hypertext Categorization*. In: *Journal of Intelligent Information Systems*, 2002.

[14] (1) Fetterly, D., Manasse, M., Najork, M. e Wiener, J. *A Large-Scale Study of the Evolution of Web Pages*. In: *Proceedings of the 12th international conference on World Wide Web*, 2003

[15] www.robotstxt.org

[16] Blum, A. e Mitchell, T. *Combining Labeled and Unlabeled Data With Co-Training*. In: *Proceedings of the 1998 Conference on Computational Learning Theory*, 1998

- [17] Adar, E., Adamic, L. A., Zhang, L. e Lukose, R. M. *Implicit Structure and the Dynamics of Blogspace*. In: *Workshop on the Weblogging Ecosystem at the 13th International World. Wide Web Conference*, 2004
- [18] Cohen, W. W. *Improving a Page Classifier With Anchor Extraction and Link Analysis*. In: *Advances in Neural Processing Systems 15*, 2002
- [19] Cohen, W. W. *Learning and Discovering Structure in Web Pages*. In: *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2003
- [20] Elgersma, E. e de Rijke, M. *Learning to Recognize Blogs: A Preliminary Exploration*. In: *EACL 2006 Workshop on New Text-Wikis and Blogs and Other Dynamic Text Sources*, 2006
- [21] Pierre, J. M. *On The Automated Classification of Web Sites*. Linkoping University Electronic Press, 2001
- [22] Fürnkranz, J. *Using Links for Classifying Web Pages*, 1998
- [23] Amitay, E., Carmel, D., Darlow, A., Lempel, R. e Soffer, A. *The Connectivity Sonar: Detecting Site Functionality by Structural Patterns*. In: *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, 2003
- [24] Qi, X. e Davison, B. D. *Web Page Classification: Features and Algorithms*. In: *Technical Report LU-CSE-07-010*, 2007
- [25] Brin, S. e Page, L. *The Anatomy of a Large Scale Hypertextual Web Search Engine*. In: *WWW7 / Computer Networks 30*, 1998
- [26] Kumar, R., Novak, J., Raghavan, P. e Tomkins, A. *On the Bursty Evolution of Blogspace*. In: *Proceedings of the 12th international conference on World Wide Web*, 2003

[27] Pedersen, J., Yang, Y. *A Comparative Study on Feature Selection in Text Categorization*. In: *Proceedings of the 14th International Conference on Machine Learning*, 1997

[28] Mladenic, D. *Feature Subset Selection in Text-Learning*. In: *Proceedings of the 10th European Conference on Machine Learning*, 1998

[29] Chakrabarti, S. *Mining The Web – Discovering Knowledge from Hypertext Data*. Morgan Kaufmann Publishers, 2003. ISBN 1-55860-754-4

[30] Platt, J. C. *Fast Training of Support Vector Machines Using Sequential Minimal Optimization*. In: *Advances In Kernel Methods*, 1998