

6

Referências bibliográficas

BAEZA-YATES, R.; HURTADO, C.; MENDOZA, M. **Query Recommendation using Query Logs in Search Engines**. Center of Web Reseach, Dept. of Computer Science, Universidad de Chile, 2004.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. Association for Computing (ACM) and Addison Wesley Longman Limited, USA, 1999.

BAKER, D.; MCCALLUM, A. **Distributional clustering of words for text classification**. In W. Bruce Croft, Alistair Moffat, Cornelis J. van Rijsbergen, Ross Wilkinson, and Justin Zobel, editors, Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval, pages 96–103, Melbourne, AU, 1998. ACM Press, New York, US.

BASILI, R.; MOSCHITTI, A.; PAZIENZA, M. T. **Language sensitive text classification**. In Proceedings of 6th RIAO Conference (RIAO 2000), Content-Based Multimedia Information Access, Collge de France, Paris, France, 2000.

BENNETT, P. N.; DUMAIS, S. T.; HORVITZ, E. **Probabilistic combination of text classifiers using reliability indicators: models and results**. Proceedings of the 25th Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR02), pp. 207–214, 2002.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and Regression Trees**. Wadsworth Int, 1984.

CAROPRESO, M. F.; MATWIN, S.; SEBASTIANI, F. **A learner-independent evaluation of the usefulness of statistical phrases for automated text categorization**. In Amita G. Chin, editor, Text Databases and Document Management: Theory and Practice, pages 78–102. Idea Group Publishing, Hershey, US, 2001.

CHAI, K. M.; NG, H. T.; CHIEU, H. L. **Bayesian online classifiers for text classification and filtering**. In Micheline Beaulieu, Ricardo Baeza-Yates, Sung Hyon Myaeng, and Kalervo Järvelin, editors, Proceedings of SIGIR-02, 25th ACM International Conference on Research and Development in Information Retrieval, pages 97–104, Tampere, FI, 2002. ACM Press, New York, US.

DOMINGOS, P.; PAZZANI, M. **On the optimality of the simple Bayesian classifier under zero-one loss**. Machine Learning, 29(2/3), 103, 1997.

DUMAIS, S.; PLATT, J.; HECKERMAN, D.; SAHAMI, M. **Inductive learning algorithms and representations for text categorization**. In Georges Gardarin, James C. French, Niki Pissinou, Kia Makki, and Luc Bouganim, editors, Proceedings of CIKM-98, 7th ACM International Conference on Information and Knowledge Management, pages 148–155, Bethesda, US, 1998. ACM Press, New York, US.

GALAVOTTI, L.; SEBASTIANI, F.; SIMI, M. **Experiments on the use of feature selection and negative evidence in automated text categorization.** In José L. Borbinha and Thomas Baker, editors, Proceedings of ECDL-00, 4th European Conference on Research and Advanced Technology for Digital Libraries, pages 59–68, Lisbon, PT, 2000. Springer Verlag, Heidelberg, DE. Published in the “Lecture Notes in Computer Science” series, number 1923.

GILES, C. L.; HAN, H.; MANAVOGLU, E.; ZHA, H.; ZHANG, Z.; FOX, E. A. **Automatic document metadata extraction using support vector machines.** In JCDL '03: Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries, pages 37–48, Washington, DC, USA, IEEE Computer Society, 2003.

GRAHAM, P. **A plan for Spam.** Disponível em <<http://www.paulgraham.com/spam.html>>. Acesso em 22 jul. 2006.

HARMAN, D. **How effective is suffixing?** Journal of the American Society of Information Science 42(1), 7 – 16, 1991.

HAYES, P. J.; WEINSTEIN, S. P. **Construe/Tis: a system for content-based indexing of a database of news stories.** In Alain Rappaport and Reid Smith, editors, Proceedings of IAAI-90, 2nd Conference on Innovative Applications of Artificial Intelligence, pages 49–66. AAAI Press Menlo Park, US, 1990.

JOACHIMS, T. **Learning to classify text using support vector machines.** Kluwer Academic Publishers, 2002.

JOACHIMS, T. **Text categorization with support vector machines: learning with many relevant features.** In Machine Learning: ECML98, Tenth European Conference on Machine Learning, pages 137–142, 1998.

KOLLER, D.; SAHAMI, M. **Hierarchically classifying documents using very few words.** Proc. of the 14th International Conference on Machine Learning ICML97, pp. 170–178, 1997.

LAM, W.; LAI, K. **A meta-learning approach for text categorization.** In W. Bruce Croft, David J. Harper, Donald H. Kraft, and Justin Zobel, editors, Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval, pages 303–309, New Orleans, US, 2001. ACM Press, New York, US.

LEWIS, D. **An evaluation of phrasal and clustered representations on a text categorization task.** In Proceedings of SIGIR-92, 15th ACM International Conference on Research and Development in Information Retrieval, pages 37–50, Kobenhavn, DK, 1992.

LEWIS, D. **Naive Bayes at forty: The independence assumption in information retrieval.** Conference proceedings of European Conference on Machine Learning (pp. 4–15), 1998.

LEWIS, D.; RINGUETTE, M. **A comparison of two learning algorithms for text categorization.** The Third Annual Symposium on Document Analysis and Information Retrieval, pp. 81–93, 1994.

LOVINS, J. B. **Development of a stemming algorithm.** Mechanical Translation and Computacional Linguistics 11 (1–2), 22–31, 1968.

MARON, M. **Automatic indexing: an experimental inquiry.** Journal of the Association for Computing Machinery, 8(3):404–417, 1961.

- MCCALLUM, A. K. **Multi-label text classification with a mixture model trained by EM**. In AAAI 99 Workshop on Text Learning, 1999.
- MCCALLUM, A. K.; NIGAM, K. **A comparison of event models for naive Bayes text classification**. In Proceedings of the 1st AAAI Workshop on Learning for Text Categorization, pages 41–48, Madison, US, 1998.
- MCCALLUM, A. K.; ROSENFELD, R.; MITCHELL, T., NG, A. **Improving text classification by shrinkage in a hierarchy of classes**. Machine Learning: Proceedings of the Fifteenth International Conference, pp. 359–367, 1998.
- MITCHELL, T. **Machine Learning**. Boston, USA: McGraw-Hill, 1997.
- MLADENIĆ, D.; GROBELNIK, M. **Word sequences as features in text-learning**. In Proceedings of ERK-98, the Seventh Electrotechnical and Computer Science Conference, pages 145–148, Ljubljana, SL, 1998.
- MOSCHITTI, A. **A study on optimal parameter tuning for Rocchio text classifier**. In proceedings of the 25th European Conference on Information Retrieval Research (ECIR), Pisa, Italy, 2003a.
- MOSCHITTI, A. **Natural Language Processing and Text Categorization: a study on the reciprocal beneficial interactions**. PhD thesis, University of Rome Tor Vergata, Rome, Italy, 2003b.
- MOSTELLER, F.; WALLACE, D. **Inference and Disputed Authorship: The Federalist**. Addison-Wesley, Reading, Massachusetts, 1964.
- NIGAM, K.; MCCALLUM, A.; THRUN, S.; MITCHELL, T. **Text classification from labeled and unlabeled documents using EM**. Machine Learning, 39(2/3): 103–134, 2000.
- PORTER, M. **An algorithm for suffixing stripping**. Program 14(3), 130–137, 1980.
- ROBERTSON, S. E.; SPARCK JONES, K. **Relevance weighting of search terms**. Journal of the American Society for Information Science, 27(3), 129–146, 1976.
- SALTON, G. **The SMART Retrieval System**. Experiment in Automatic Document Processing, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.
- SCHAPIRE, R.; SINGER, Y. **BOOSTEXTER: A boosting-based system for text categorization**. Machine Learning, Vol.39, No. 2/3, 2000.
- SCOTT, S.; MATWIN, S. **Feature engineering for text classification**. In Ivan Bratko and Saso Dzeroski, editors, Proceedings of ICML-99, 16th International Conference on Machine Learning, pages 379–388, Bled, SL, 1999. Morgan Kaufmann Publishers, San Francisco, US.
- SEBASTIANI F.; SPERDUTI A.; VALDAMBRINI, N. **An improved boosting algorithm and its application to automated text categorization**. In Arvin Agah, Jamie Callan, and Elke Rundensteiner, editors, Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management, pages 78–85, McLean, US, 2000. ACM Press, New York, US.
- SEBASTIANI, F. **Machine learning in automated text categorization**, Tech. Rep. IEI-B4-31-1999, Consiglio Nazionale delle Ricerche, Pisa, Italy, 1999.

SEBASTIANI, F.; DEBOLE, F. **An analysis of the relative hardness of Reuters-21578 subsets.** In Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation, pages 971–974, Lisbon, PT, 2004.

SHAVLIK, J.; ELIASSI-RAD, T. **Intelligent agents for web-based tasks: An advice-taking approach.** In: Proceedings of the Fifteenth National Conference on Artificial Intelligence: Workshop on Learning for Text Categorization. Madison, WI, pp. 63–70, 1998.

TONG, S.; KOLLER, D. **Support vector machine active learning with applications to text classification.** Journal of Machine Learning Research, 2:45–66, 2001.

VAPNIK, V.N. **Statistical learning theory.** John Wiley & Sons, Inc., N.Y., 1998.

VILAR, D.; JOSÉ CASTRO, M.; SANCHIS, E. **Multi-label text classification using multinomial models.** Computer Science Department, Aachen University (Germany), Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València (Spain), 2004.

YANG, Y.; LIU, X. **A re-examination of text categorization methods.** In Marti A. Hearst, Fredric Gey, and Richard Tong, editors, Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval, pages 42–49, Berkeley, US, 1999. ACM Press, New York, US.