Referências bibliográficas

ADAMO, Jean-Marc. **Data Mining for Association Rules and Sequential Patterns: Sequencial and Parallel Algorithms**. Springer, 2000.

ARANHA, Christian Nunes. **Uma abordagem para o pré-processamento automático de mineração de textos em português – sob o enfoque da inteligência computacional.** Tese de doutorado, PUC-Rio, 2007.

BADR, Y., SAYAH, M., LAFOREST, F., FLORY, A. **Transformation rules from semi-structured XML documents to database model**. Computer Systems and Applications, ACS/IEEE International Conference, 2001.

BAEZA-YATES, Ricardo, RIBEIRO-NETO, Berthier. **Modern Information Retrieval**. ACM Press, 1999.

BALLY, C. **Linguistique générale et linguistique française**. A. Francke S. A., Berne, 1944.

BANG, S. L, YANG, J.D., YANG, H. J. **Hierarchical Document Categorization with k-NN and concept-based thesauri.** Information Processing & Management, 2006.

BEIL, F., ESTER, M., XU, X. **Frequent Term-Based Text Clustering**. Proceeding of the 8th International Conference on Knowledge Discovery and Data Mining (KDD 2002), Edmonton, Alberta, Canada, 2002.

BELLOT, P., EL-BEZE, M. **A Clustering Method for Information Retrieval**. Technical Report IR-0199, Laboratoire d'Informatique d'Avignon, France, 1999.

BEUSEKON, Joost Van, KEYSERS, Daniel, SHAFAIT, Faisal, BREUEL, Thomas M. **Distance Measures for Layout-Based Document Image Retrieval**. In Proceedings of the Second International Conference on Document Image Analysis for Libraries, 2006.

BOY, G. A., KUSS, P. M. **A fuzzy method for the modeling of human-computer interaction in information retrieval tasks**. In Karwowski, W. and Mital, A., eds. Applications of Fuzzy Set Theory in Human Factors, Elsevier, 1986.

BOYAN, Justin, FREITAG, Dayne, JOACHIMS, Thorsten. **A Machine Learning Architecture for Optimizing Web Search Engines**. Proceedings of the AAAI Workshop on Internet-Based Information Systems, 1996.

BRIN, Sergey, PAGE, Laurence. **The Anatomy of a Large-Scale Hypertextual Web Search Engine**. Computer Networks and ISDN Systems, Volume 30, 1998.

BRILL, Eric. **Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging**. Computacional Linguistics, 1995.

BURGES, Christopher J. C. **A Tutorial on Support Vector Machines for Pattern Recognition**. Data Mining and Knowledge Discovery Conference, Volume 2 Issue 2, 1998.

CAPPÉ, Olivier, MOULINES, Eric, RYDÉN, Tobias. **Inference in Hidden Markov Models.** Springer, 2005.

CARRILHO, João Ribeiro. **Desenvolvimento de uma Metodologia para Mineração de Textos**. Dissertação de mestrado, PUC-Rio, 2007.

CHAN, S. W. K., FRANKLIN, J. **Symbolic connectionism in natural language disambiguation**. Neural Networks, IEEE Transactions on Volume 9, Issue 5, Page(s):739 – 755, Sept. 1998.

CHANG, C.-H.; HSU, C.-C. **Hypertext information retrieval for short queries**. In Proceedings of the IEEE Knowledge and Data Engineering Exchange Workshop, 1998.

CHARKABARTI, Soumen. **Mining the Web: Discovery Knowledge from Hypertext Data**. Morgan Kaufmann Publishing, 2003.

CROSS, V. **Fuzzy information retrieval**. J. of Intelligent Information Systems, 1994.

DILIGENT, M., COETZEE, F. M., LAURENCE, S., GILES, C. L., GORI, M. **Focused crawlers using context graph**. In Proceedings of the 26th International Conference on Very Large Databases, VLDB 2000, Cairo, Egypt, pp. 527–534, 2000.

DIOU, C.; KATSIKATSOS, G.; DELOPOULOS, A.; **Constructing Fuzzy Relations from WordNet for Word Sense Disambiguation**. Semantic Media Adaptation and Personalization, SMAP '06. First International Workshop, page(s):135 – 140, 2006.

DOM, Byron, CHAKRABARTI, Soumen, BERG, Martin van den. **Focused crawling: a new approach to topic-specific Web resource discovery**. Computer Networks 31, pp. 1623-1640, 1999.

DRAEGER, Vitor H. P. **Extração de Informações em Patentes**. Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, 2007.

DUDA, Richard O., HART, Peter E., STORK, David G. **Pattern Classification (Second Edition)**. Willey-Iterscience, 2000.

DUNN, J. C. **A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters.** Journal of Cybernetics 3: 32-57, 1973.

ELIS, R., SIMPSON, R., CULVERHOUSE, P. F., PARISINI, T., **Committees, Collectives and Individuals: Expert Visual Classification by Neural Network**. Neural Computing & Applications Journal, Volume 5, Pages 99-105, 1997.

EMPSON, William. **Seven Types of Ambiguity**. New Directions Publishing Corporation, 1965

FELLBAUM, Christiane. **WordNet: An Electronic Lexicon Database**. The MIT Press, 1998.

FELDMAN, Rone, SANGER, James. **The Text Mining Handbook – Advanced Approaches in Analyzing Unstructured Data**. Cambridge University Press, 2007.

FLORIAN, Radu, NGAI, Grace. **Multidimensional Transformation-Based Learning**. In Proceedings of the fifth Conference on Computacional Natural Language Learning CoNLL, 2001.

FRAKES, W. B. **Stemming algorithms, Information retrieval: data structures and algorithms.** Prentice-Hall, Inc., 1992.

FUNG, B. C. M. **Hierarchical Document Clustering Using Frequent Itemsets**. Master Thesis, Dept. Computer Science, Simon Fraser University, Canada, 2002.

FURNAS, G. W., DEERWESTER, S., DUMAIS, S. T., LAUDAUER, T. K., HARSHMAN, R. A., STREETER, L. A., LOCHBAUM, K. E. **Information retrieval using a singular value decomposition model of latent semantic structure**. In Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1988.

GIUDICI, Paolo. **Applied Data Mining – Statistical Methods for Business and Industry**. John Wiley & Sons Ltd, 2003.

GOLDSCHMIDT, Ronaldo, PASSOS, Emmanuel. **Data Mining – Um Guia Prático**. Editora Campus, 2005.

GONÇALVES, Teresa, SILVA, Cassiana, QUARESMA, Paulo, VIEIRA, Renata. **Analyzing Part-of-Speech for Portuguese Text Classification**. CICLing - Computational Linguistics and Intelligent Text Processing (551-562), 2006.

HAN, J. and KAMBER, M. **Data Mining: Concepts and Techniques**. Morgan Kaufmann Publishers, 2001.

HAMMOUDA, K. M. **Web Mining: Identifying Document Structure for Web Document Clustering**. Master's Thesis, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada, 2002.

HARTIGAN, J. A., WONG, M. A. **A K-Means Clustering Algorithm**. Applied Statistics Journal, Volume 28, 1979.

HAIKIN, Simon. **Neural Network – A Comprehensive Foudation**. Prentice Hall, 1999.

HOPCROFT, Jeffrey D., ULLMAN, John E. **Formal Languages and Their Relation to Automata**. Addison Wesley Publishing Company, 1969.

HUNG, Chihli, WERMTER, Stefan, SMITH, Peter. **Hybrid Neural Document Clustering Using Guided Self-Organization and Wordnet**. IEEE Intelligent Systems, 2003.

ILIOPOULOS, I., ENRIGHT, A. J., OUZOUNIS, C. A. **Textquest: Document Clustering of Medline Abstracts for Concept Discovery in Molecular Biology**. Proceedings of the Sixth Annual Pacific Symposium on Biocomputing (PSB 001), 2001.

IDE, N. VEROMIS, J. **Word Sense Disambiguation: The State of the Art**. Computational Linguistics, 1998.

JEREBKO, Anna K., MALLEY, James D., FRANASZEK, Marek, SUMMERS, Ronald M. **Support Vector Machines Committee Classification Method for Computer-aided Polyp Detection in CT Colonography**. Academic Radiology Journal, April-12, 2005.

JESPERSEN, Otto. **Language: Its Nature, Development and Origin**. Allen and Unwin, 1922.

JIANWU, Yang, XIAOOU, Chen. **A semi-structured document model for text mining**. Journal of Computer Science and Technology, Volume 17, Issue 5, 2002.

JOACHIMS, T. **Text Categorization with Support Vector Machines: Learning with Many Relevant Features**. In Proceedings of the European Conference on Machine Learning, Springer, 1998.

JOHNSON, S. C. **Hierarchical Clustering Schemes.** Psychometrika, 2:241-254, 1969.

JOLLIFFE, I. T. **Principal Component Analysis**. Springer Series in Statistics, 1986.

KANDARTZIC, Mehmed. **Data Mining: Concepts, Models, Methods and Algorithms**. John Wiley & Sons, 2003.

KAO, Anne, POTEET, Steve. **Text Mining and Natural Language Processing – Introduction for the Special Issues**. Special Interest Group on Knowledge Discovery and Data Mining - SIGKDD Explorations, Volume 7, Issue 1, 2004.

KONCHADY, Manu. **Text Mining Application Programing**. Charles River Media, 2006.

KOOJI, Jan G. **Ambiguity in natural language : an investigation of certain problems in its linguistic description**. North Holland, 1971.

KUDO, Taku, MATSUMOTO, Yuji. **A Boosting Algorithm for Classification of Semi-Structured Text**. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2004.

LAY, David C. **Linear Algebra and Its Applications, Third Edition**. Addison Wesley, 2005.

LETSCH, Todd A., BERRY, Michael W. **Large-Scale Information Retrieval with Latent Semantic Indexing.** Journal of Information Sciences, 1997.

LIN, K., KONDADADI, R. **A Word-Based Soft Clustering Algorithm for Documents**. Proceedings of 16th International Conference on Computers and Their Applications, Mar. 2001.

LITTMAN, M. L., DUMAIS, S. T. and LANDAUER, T. K. **Automatic cross-linguistic information retrieval using Latent Semantic Indexing**. To appear in G. Gredenstette (Ed.), Cross Language Information Retrieval, 1997.

LOVINS, J. B. **Development of a Stemming Algorithm**. Mechanical Translation and Computational Linguistics, vol. 11, 1968,

MANNING, Christopher D., RAGHAVAN, Prabhakar, SCHÜTZE, Hinrich. **An Introduction to Information Retrieval**. Cambridge University Press, 2007.

MENDELSON, Elliott. **Introduction to Mathematical Logic, Fourth Edition**. Chapman & Hall/CRC, 1997.

MEYER, Paul L. **Probabilidade – Aplicações à Estatística**. Livros Técnicos e Científicos Editora S.A., 1976.

MILLER, G. A., BECKWITH, R. FELLBAUM, C. D. GROSS, D. **Five Papers on WordNet**. Techinical Report, Princetom University, 1993.

MITCHELL, Tom M. **Machine Learning**. McGRaw-Hill, 1997.

MONTGOMERY, Douglas C., RUNGER, George C. **Applied Statistics and Probability for Engineers, Fouth Edition**. John Wiley & Sons, 2006.

MOSCALEWSKI, Maria Angela. **Extraindo Regras de Assosiação a Partir de Textos**. Dissertação de Mestrado, PUC-PR, 2002.

NIST. **Information Extraction Definitions**. http://itl.nist.gov/iaui/894.02/ related_projects/muc/inf/definitions.html, 2003. [capturado em 20 de junho de 2008].

OGAWA, Y., MORITA, T., KOBAYASHI, K. **A fuzzy document retrieval system using the keyword connection matrix and a learning method**. Fuzzy Sets and Systems, 1991

OLIVEIRA, Francisco, WONG, Fai, LI Yiping, ZHENG Jie. **Unsupervised word sense disambiguation and rules extraction using non-aligned bilingual corpus**. Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2005.

PORTER, M. F., RIJSBERGEN, C. J. Van, ROBERTSON, S.E. **New models in probabilistic information retrieval**. London: British Library - British Library Research and Development Report, nº. 5587, 1980.

POWEL, Gavin. **Beginning XML Databases**. Wiley Publishing, Inc., 2007.

QIN Ying; WANG Xiaojie; **Rules selection in word sense disambiguation using Adaboost**. Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE, page(s):26 – 29, 2005.

REDFEARN, Judy. **Text Minig – Brief Paper**. National Centre of Text Minig – JISC Communication Team, 2006.

RIBEIRO-NETO, Berthier A., MUNTZ, Richard. **A belief network model for IR**. Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1996.

RIJSBERGEN, C. J. van. **Information Retrieval (Second Edition)**. Publicado pelo Departamento de Ciências da Computação da Universidade de Glasgow e disponível em http://www.dcs.gla.ac.uk/~iain/keith/, 1999.

ROBERTSON, S. E., JONES, K. Sparck. **Relevance of weighting of search terms**. Journal of the American Society for Information Sciences, 1976.

ROSSO, P.; MASULLI, F.; BUSCALDI, D.; **Word sense disambiguation combining conceptual distance, frequency and gloss**. Proceedings of International Conference on Natural Language Processing and Knowledge Engineering, 26-29, page(s):120 – 12, 2003.

SADAT, F. MAEDA, A. YOSHIKAWA, M. UEMURA, S. **A combined statistical query term disambiguation in cross-language information retrieval**. Database and Expert Systems Applications. Proceedings. 13th International Workshop, Page(s):251 – 255 on 2-6 Sept. 2002.

SALTON, G., LESK, M. E. **Computer evaluation of indexing and text processing**. Journal of the ACM, 1968.

SALTON, Gerard, FOX, Edward A., WU, Harry. **Extended Boolean information retrieval**. Communications of the ACM, 1983.

SALTON, Gerard and BUCKLEY, Chris. **Term-weighting approaches in automatic text retrieval**. Information Processing & Management, 1988.

SEYMORE, Kristie, MCCALUM, Andrew, ROSENFELD, Roni. **Learning Hidden Markov Model Structure for Information Extraction**. AAAI 99 Workshop on Machine Learning for Information Extraction, 1999

SHAFER, S. M., ROGERS, D. F. **Similarity and Distance Measures for Cellular Manufacturing – A Survey**. International Journal of Production Research, 1993.

SHOLOM, M. Weiss, INDURKHYA, Nitin, ZHANG, Tong, DAMERAU, Fred J. **Text Mining – Predictive Methods for Analyzing Unstructured Information**. Springer Science+Business Media, Inc., 2005.

SOARES, Fábio. **Mineração de Textos na Coleta Inteligente de Dados na Web**.Dissertação de mestrado, PUC-Rio, 2008.

SONG, Sa Kwang; JIN, Yun; MYAENG, Sung Hyon; **Abbreviation disambiguation using semantic abstraction of symbols and numeric terms**. Proceedings IEEE International Conference on Natural Language Processing and Knowledge Engineering, Page(s):14 – 19, 2005.

STEINBACH, M., KARYPIS, G., KUMAR, V. **A Comparison of Document Clustering Techniques.** KDD Workshop on Text Mining, 2000.

TURTLE, Howard, CROFT, W. Bruce. **Inference networks for document retrieval**. In Proceedings of the 13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1990.

WANG, P.F.; MOSLEY, C. **Associative memory neural networks for information retrieval of text word pairs**. Neural Information Processing, 2002.

WALL, M. E., RECHTSTEINER, A., ROCHA, L. M. **Singular Value Decomposition and Principal Component Analysis**. ArXiv Physics e-prints, 2002.

WEISS, S., WHITE, H., APT'E, E. **Lightweight Document Clustering**. Proceedings of PKDD-2000, Springer, 2000.

WILSON, C.; SRINIVASAN, B.; INDRAWAN, M. **A general inference network based architecture for multimedia information retrieval**. Multimedia and Expo, 2000.

WONG, S. K. M., ZIARKO, W., WONG P. C. N. **Generalized vector space model in information retrieval**. In Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1985.

ZAMIR, Oren Eli. **Clustering Web Documents: A Phrase-Based Method for Group Search Engine Results**. Ph.D. dissertation, Dept. Computer Science & Engineering, Univ. of Washington, 1999.

ZANASI, A. **Text Mining and its Applications to Intelligence, CRM and Knowledge Management (Advances in Management Information)**. WIT Press, 1997.