

Referências Bibliográficas

ALPAYDIN, E. (2004). *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

AMITAY, E. (1998). Using Common Hypertext Links to Identify the Best Phrasal Description of Target Web Documents. *In Proceedings of the ACM Special Interest Group on information Retrieval*.

ARANHA, C. N. (2007). *Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional, Tese de Doutorado, Departamento de Engenharia Elétrica, PUC-Rio*.

ARANHA, C. N., & PASSOS, E. P. (2006). A Tecnologia de Mineração de Textos. *Revista Eletrônica de Sistemas de Informação*, 2, 2.

ARDÖ, A. (2005). Focused Crawling in the ALVIS Semantic Search Engine. *In Proceedings of 2nd European Semantic Web Conference (ESWC)*. Heraklion, Greece.

BAEZA-YATES, R., & BERTIER, R. N. (1999). *Modern Information Retrieval*. Harlow: Addison-Wesley.

BARROS, C. D. (2010). *Antonímia nos adjetivos descritivos do português do Brasil: uma proposta de análise e representação*. 2010. 89 f. *Dissertação (Mestrado em Linguística) – Universidade Federal de São Carlos, São Carlos, 2010.*

BASTOS, V. M. (2006). *Ambiente de Descoberta de Conhecimento na Web para a Língua Portuguesa, Tese de Doutorado, Departamento de Engenharia Civil, UFRJ*. Rio de Janeiro.

BENNETT, P. N., DUMAIS, S. T., & HORVITZ, E. (2005). *The combination of text classifiers using reliability indicators. Information Retrieval, Kluwer Academic Publishers*. v. 8, n. 1, p. 67-100. MA, USA.

BERGMARK, D., LAGOZE, C., & SBITYAKOS, A. (2000). Focused Crawls, Tunneling, and Digital Libraries. *In Proceedings of the 6th European Conference on Digital Libraries*.

BICK, E. (2000). *The parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Denmark: Aarhus University Press.

BISHOP, C. M. (2007). *Pattern Recognition and Machine Learning*. Springer.

BORKO, H., & BERNICK, M. (1963). *Automatic document classification*. *J.Assoc.Comput.Mach.*10, 2, 151–161.

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

BRIN, S., & PAGE, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, Vol. 30.

BUCKLAND, M., & GEY, F. (1944). The relationship between Recall and Precision. *Journal of the American Society for Information*, Vol. 45, 12-19.

BURGES, C. J. (1998). A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery Conference*, Vol. 2.

CAMARGO, Y. B. (2007). *Abordagem Linguística na Classificação Automática de Textos em Português*. *Dissertação de Mestrado*. Departamento de Engenharia Elétrica, UFRJ.

CARDOSO, O. N. (2000). *Recuperação de Informação*. Departamento de Ciência da Computação, Universidade Federal de Lavras.

CARRILHO, J. (2007). *Desenvolvimento de uma Metodologia para Mineração de Textos*, *Dissertação de Mestrado*, Departamento de Engenharia Elétrica, PUC-Rio.

CASTILLO, C. (2004). *Effective Web Crawling*. *Ph.D. Thesis*, Dept. of Computer Science, University of Chile. Santiago, Chile.

CEGALLA, D. P. (2005). *Novíssima Gramática da Língua Portuguesa* (46 ed.). IBEP.

CHAGAS, F. (2009). *Variações do Método kNN e suas Aplicações na Classificação Automática de Textos*. *Dissertação de Mestrado*, Instituto de Informática, UFG.

CHAKRABARTI, S. (2003). *Mining the web: Discovering knowledge from hypertext data*. San Francisco, CA: Morgan Kaufmann Publishers.

CHAVES, A. C. (2006). *Extração de regras fuzzy para máquinas de vetores suporte (SVM) para classificação em múltiplas classes*. Tese de Doutorado. Departamento de Engenharia Elétrica. PUC-Rio.

COTHEY, V. (2004). Web-crawling reliability. *Journal of the American Society for Information Science and Technology*, Vol. 55, 1228-1238.

DATE, C. J. (2005). *Introdução a Sistemas de Bancos de Dados* (8 ed.). São Paulo: Campus.

DAVISON, B. D. (2000). Topical Locality in the Web. *In Proceedings of the 23th Annual international Conference on Research and Development in Information Retrieval*.

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

DOM, B., CHAKRABARTI, S., & BERG, M. (1999). Focused crawling: a new approach to topic-specific Web resource discovery. *Computer Networks*, Vol. 31, 1623-1640.

FELDMAN, R., & SANGER, J. (2007). *The Text Mining Handbook – Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

FERRO, M., & LEE, H. D. (2001). O Processo de KDD – Knowledge Discovery in Database para Aplicações na Medicina. *SEMINC - Semana de Informática de Cascavel, Universidade Estadual do Oeste do Paraná*. Cascável.

FINATTO, M. J. (2005). *Análise Textual Assistida por computador: Reconhecimento Lingüístico-Terminológico do texto técnico-científico de Química em Português - Da coesão à Enunciação (TextQuim), Relatório Final de Atividades, CNPQ, Instituto de Letras, UFRGS*. Porto Alegre.

FONSECA, B. M., & REIS, D. C. (2002). *O fantástico mundo da distância de edição*.

FONSECA, F., & FIDALGO, R. (2002). *Gerenciamento de Dados e Informação, Centro de Informática, Universidade Federal de Pernambuco*.

FOUCAULT, M. (2002). *A ordem do discurso* (8 ed.). São Paulo: Loyola.

FREUND, Y., & SCHAPIRE, R. E. (1999). *A short introduction to boosting*. In *Journal of Japanese Society for Artificial Intelligence*, v.5, n.14, p.771-780.

FUNREDES. (2007). *Línguas e Culturas na Web*. Paris, França: Terminologia e Indústrias da Língua - DTIL.

GDS PUBLISHING. (2008). *Managing the Data Explosion*. *Business Management* .

GIBSON, D., KLEINBERG, J., & RAGHAVAN, P. (1998). *Inferring Web Communities from Link Topology*. In *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*.

GOLDSCHMIDT, R., & PASSOS, E. (2005). *Data Mining: um guia prático*. Rio de Janeiro: Campus.

GOMES, R. M. (2008). *Mineração de Textos na Desambiguação, Dissertação de Mestrado*. Rio de Janeiro: Departamento de Engenharia Elétrica, PUC-Rio.

GONÇALVES, T., SILVA, C., QUARESMA, P., & VIEIRA, R. (2006). *Analyzing Part-of-Speech for Portuguese Text Classification*. *Proceedings of Computational Linguistics and Intelligent Text Processing (CICLing)*, (pp. 551-562). Mexico City, Mexico.

HALL, M., FRANK, E., HOLMES, G., FAHRINGER, B., REUTEMANN, P., & WITTEN, I. H. (2009). *The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1*.

HARPER, M. P., & THEDE, S. M. (1999). *A Second-Order Hidden Markov Model for Part-of-Speech Tagging*. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, (pp. 175-182). College Park.

HARTIGAN, J. A., & WONG, M. A. (1979). *A k-means clustering algorithm*. *Applied Statistics* , 100-108.

HEARST, M. (1999). *Untangling Text Data Mining*. In *proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland*.

HEATON, J. (2002). *Programming Spiders, Bots, and Aggregators in Java*. Sybex.

HERLOCKER, J., KONSTAN, J., TERVEEN, L., & RIEDL, J. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 5–53.

IBM. (13 de Fevereiro de 2008). *Unstructured Information – The Knowledge Rush*. Acesso em 03 de Maio de 2008, disponível em The Knowledge Rush:

http://domino.research.ibm.com/comm/research_projects.nsf/pages/uima.knowledgeRush.html

INSITE. (2001). *Grupo de Lingüística da Insite Processamento de Linguagem Natural*. Acesso em 01 de 04 de 2008, disponível em Processamento de Linguagem Natural: <http://linguistica.insite.com.br/nlp.phtml>

JARGAS, A. M. (2006). *Expressões Regulares - Uma Abordagem Divertida*. Novatec.

JOACHIMS, T. (1998). *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*.

JOHNSON, R., & FOOTE, B. (1988). *Designing reusable classes*. *Journal of Object-Oriented Programming* v. 1, n. 5, pp. 22-35, Jun 1988.

KANTARDZIC, M. (2002). *Data Mining: Concepts, Models, Methods, and Algorithms*. Wiley-IEEE Press.

KIM, S.-B., HAN, K.-S., RIM, H.-C., & MYAENG, S. H. (2006). Some Effective Techniques for Naive Bayes Text Classification. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, pp. 1457-1466.

KLEINBERG, J. M. (1998). Authoritative Sources in a Hyperlinked Environment. . *In Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*.

KONCHADY, M. (2006). *Text Mining Application Programming* (1 ed.). Charles River Media.

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

LAUDON, K. C., & LAUDON, J. P. (2002). *Management Information Systems: managing the digital firm* (7ª ed.). New Jersey: Prentice-Hal.

LEE, Y. K., NG, H. T., & CHIA, T. K. (2004). *Supervised Word Sense Disambiguation with SVM and Multiple Knowledge Sources*. *In International*

Workshop on the Evaluations of Systems for the Semantic Analysis of Text. Barcelona.

LINDEN, G. S. (2008). *Combinação de classificadores na categorização de textos. Dissertação de Mestrado. Faculdade de Informática. PUC-RS.*

LINGUATECA. (19 de 07 de 2007). *Atomização e separação de frases.* Acesso em 07 de 03 de 2008, disponível em Projecto AC/DC: Linguateca: <http://acdc.linguateca.pt/acesso/atomizacao.html>

LOPES, M. C. (2004). *Mineração de Dados Textuais utilizando técnicas de Clustering para o idioma Português. Tese de Doutorado. Departamento de Engenharia Civil. UFRJ. Rio de Janeiro.*

LOVINS, J. B. (1968). Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics* , Vol. 11, 22-31.

MANDEL, A., SIMON, I., & DELYRA, J. (1997). Informação: computação e comunicação. *Revista da USP* , 35, 11-45.

MANNING, C. D., RAGHAVAN, P., & SCHÜTZE, H. (2007). *Introduction to Information Retrieval.* Cambridge University Press.

MARKOV, Z., & LAROSE, D. T. (2007). *Data Mining the Web: Uncovering patterns in Web content, structure, and usage.* New Jersey: Wiley-Interscience.

MCCALLUM, A., & NIGAM, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. *ICML-Workshop on Learning for Text Categorization.* AAAI Press.

MCCALLUM, A., & NIGAM, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. *AAAI/ICML-98 Workshop on Learning for Text Categorization.* AAAI Press.

MCCALLUM, A., NIGAN, K., RENNIE, J., & SEYMORE, K. (1999). Building Domain-Specific Search Engines With Machine Learning Techniques. *In AAAI Spring Symposium on Intelligent Agents in Cyberspace.*

MELO, L. B. (2007). *Reconhecimento de padrões textuais para Categorização Automática de Documentos. Dissertação de Mestrado. Departamento de Engenharia Elétrica, UFRJ.*

MLAENIC, D., & GROBELNIK, M. (1998). *Feature Selection for Classification Based on Text Hierarchy. In Working Notes of Learning from Text*

and the Web, Conf. Automated Learning and Discovery (CONALD-98). Pittsburgh.

MODESTO, M., PEREIRA, A. R., ZIVIANI, N., CASTILHO, C., & BAEZA-YATES, R. (2005). Um novo retrato da Web brasileira. *Proceedings of XXXII SEMISH*, (pp. 2005-2017). São Paulo.

MONTEIRO, L., GOMES, I., & OLIVEIRA, T. (2006). Etapas do Processo de Mineração de Textos – uma abordagem aplicada a textos em Português do Brasil. *Anais do XXVI Congresso da SBC, I Workshop de Computação e Aplicações*, (pp. 78-81). Campo Grande, MS.

NAJORK, M., & WIENER, J. L. (2001). Breadth-first search crawling yields high-quality pages. *In Proceedings of the 10th International World Wide Web Conference*. Hong-Kong.

NEVES, P. I. (2012). *Um estratégia para apoiar a decisão baseada em Mineração de Textos Livres. Dissertação de Mestrado. Departamento de Ciência e Tecnologia, Instituto Militar de Engenharia.*

NG, H. T., GOH, W. B., & LOW, K. L. (1997). Feature Selection, Perception Learning and a Usability Case Study for Text Categorization. *Proceedings of SIGIR-97, 20th ACM International Conference on Research and Development in Information Retrieval*. Philadelphia, PA, USA.

NISBET, R., ELDER, J., & MINER, G. (2009). *Handbook of Statical Analysis & Data Mining Applications*. California: Elsevier.

ORENGO, V. M., & HUYCK, C. R. (2001). A Stemming Algorithm for the Portuguese Language, in *8th International Symposium on String Processing and Information Retrieval (SPIRE)*. 2001: Laguna de San Raphael, Chile. p. 183-193.

ORENGO, V., & HUYCK, C. (2001). A Stemming Algorithm for the Portuguese Language. *In Proceedings of the SPIRE Conference*. Laguna de San Raphael, Chile.

PEIXOTO, M. D., BATISTA, M. D., & CAPELO, M. J. (s.d.). Categorização de Textos. *Departamento de Informática, Universidade da Beira Interior*.

PINKER, S. (1998). *Como a mente funciona*. São Paulo: Companhia das Letras.

PORTER, M. (1980). An algorithm for suffixing stripping. *Program: electronic library and information systems*, Vol. 14 (3), 130-137.

POWEL, G. (2007). *Beggining XML Databases*. Wiley Publishing.

REIS, C. C. (2011). *Carla Corrêa Tavares dos Reis Julho. Tese de Doutorado. Departamento de Engenharia de Sistemas e Computação, UFRJ.*

REZENDE, S. O. (2005). *Sistemas inteligentes: fundamentos e aplicações*. Barueri, SP: Manoele.

RICOTTA, F. C. (2007). *Como os search engines funcionam? Projeto Final de Graduação, Departamento de Matemática e Computação, Universidade Federal de Itajubá. Itajubá, MG.*

RIJSBERGEN, C. J. (1979). *Information Retrieval*. London: University of Glasgow.

RINO, L. H., & PARDO, T. A. (2003). A Sumarização Automática de Textos: Principais Características e Metodologias. NILC – Núcleo Interinstitucional de Lingüística Computacional.

RUSSELL, N., & NORVIG, P. (2004). *Inteligência Artificial* (2 ed.). Rio de Janeiro: Elsevier.

SALTON, G., & BUCKLEY, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, Vol. 24, 513–523.

SEBASTIANI, F. (2002). *Machine Learning in Automated Text Categorization*. Pisa, Italy.

SETIONO, R., & LEOW, W. K. (1998). *FERNN: An Algorithm for Fast Extraction of Rules from Neural Networks*. National University of Singapore.

SEYMORE, K., MCCALLUM, A., & ROSENFELD, R. (1999). Learning Hidden Markov Model Structure for Information Extraction. *AAAI 99 Workshop on Machine Learning for Information Extraction*.

SHKAPENYUK, V., & SUEL, T. (2002). Design and implementation of a high performance distributed web crawler. *In Proceedings of the 18th International Conference on Data Engineering (ICDE)* (pp. 357-368). San Jose, California: IEEE CS Press.

SHOLOM, M. W., INDURKHYA, N., ZHANG, T., & DAMERAU, F. J. (2005). *Text Mining – Predictive Methods for Analyzing Unstructured Information*. Springer.

SILVA, A. A. (2007). *Aiuri: Um portal para Mineração de Textos integrado a Grids, Dissertação de Mestrado, Engenharia Civil, UFRJ.*

SILVA, F. R. (2007). *GEODISCOVER – Mecanismo de busca especializado em Dados Geográficos. Tese de Doutorado, Departamento de Computação Aplicada, INPE. São José dos Campos.*

SINGH, H. S. (2001). *DATA WAREHOUSE: Conceitos, tecnologias, implementação e gerenciamento. São Paulo: Makron Books.*

SOARES, F. A. (2008). *Mineração de Textos na Coleta Inteligente de Dados da Web. Dissertação de Mestrado. Engenharia Elétrica. PUC-Rio.*

SPIEGEL, M. R. (2003). *Estatística. Makron Books.*

SPINK, A., WOLFRAM, D., JANSEN, M. B., & SARACEVIC, T. (2001). Searching the web: the public and their queries. *Journal of the American Society for Information Science and Technology*, Vol. 52, 226–234.

SULLIVAN, D. (Dezembro de 2000). The need for Text Mining in Business Intelligence. *DM Review*.

SUTTON, C., & MCCALLUM, A. (2006). *An introduction to conditional random fields for relational learning. In Introduction to Statistical Relational Learning. MIT Press.*

TAN, A.-H. (1999). Text Mining: The state of the art and the challenges. *PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases*, (pp. 65-70). Beijing.

TAN, P.-N., STEINBACH, M., & KUMAR, V. (2005). *Introduction to Data Mining. Pearson Addison Wesley.*

TANENBAUM, A. (2003). *Redes de Computadores. Rio de Janeiro: Campus.*

TOYODA, M., & KITSUREGAWA, M. (2001). Creating a Web Community Chart for Navigation Related Communities. *In Proceedings of ACM Conference on Hypertext and Hypermedia.*

WEN, J.-R. (2006). *Search Engine Overview. Microsoft Research Asia.*

WIENER, E., PEDERSEN, L. O., & WEIGEND, A. S. (1995). Neural Network Approach to Topic Spotting. *Proceeding of the Symposium on Document Analysis and Information Retrieval*, (pp. 317-322). Las Vegas, US.

XU, Q., & ZUO, W. (2007). First-order Focused Crawling. *In Proceedings of World Wide Web Conference.*

ZHU, X., & DAVIDSON, I. (2007). *Knowledge Discovery and Data Mining: Challenges and Realities. New York: Hershey.*