

Referências bibliográficas

- AAS, K. e Eikvil, L. (1999). "Text categorisation: A survey.". <http://citeseer.ist.psu.edu/aas99text.html>
- ACM96A (1996). "ACM - Association for Computing Machinery, Inc.". In Memoriam: Gerard Salton. ACM Transactions on Information Systems. v.14.n.1. Janeiro.
- ACM96B (1996). "ACM - Association for Computing Machinery, Inc.". Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 18-22 Agosto. 1996. Zurich, Switzerland
- AGGARWAL, C. C., Gates, S. C., e Yu, P. S. (1999). "On the merits of building categorization systems by supervised clustering". in: Proceedings of the Fifth ACM Conference on Knowledge Discovery and Data Mining (KDD'99), San Diego, USA, August, pp.352-356
- AMORI, D. R. (1990). "A Natural Language Interface for Task-Oriented Activities". ACM 089791-372-8/90/0007/0553
- ANANYAN, S. (2006). "<http://www.megaputer.com/tech/wp/tm.php3>". Acessado em 7 de maio de 2006
- APOSTOLIC, A. e Giancarlo, R. (1986). "The Boyer-Moore-Galil string searching strategies revisited". SIAM J. Comput. 15, (1), 98-105
- APTÉ, C., Damerau, F., e Weiss, S. M. (1994). "Automated learning of decision rules for text categorization". ACM Transactions on Information Systems, 12(3), 233-251
- ARANHA, C., Freitas, M. C., Dias, M. C., e Passos, E. (2004). "Um modelo de desambigüização de palavras e contextos". TIL 2004: Workshop de Tecnologia da Informação e da Linguagem Humana
- ARANHA, C., Passos, C., Freitas, M. C., e Quental, V. (2005). "A Statistical Approach for Identifying Collocations". The X Ibero-American Conference on Artificial Intelligence (IBERAMIA). Submetido.
- BAEZA-YATES, B. e Ribeiro Neto, B. (1999). Modern Information Retrieval. Addison Wesley
- BANERJEE, S. e Pedersen, T. (2003). "The design, implementation and use of the ngram statistics package". In A. Gelbukh (Ed.), CICLing-2003, Lecture Notes in Computer Science 2588, pp.370-381. Springer-Verlag

- BARROS, F. A. e DeRoeck, A. (1994). "Resolving Anaphora in a Portable Natural Language Front End to Databases". In Proceedings of the 4th Conference on Applied Natural Language Processing (ANLP'94). Stuttgart, Germany.5p
- BATISTA, G. E. A. P. A. (2003). "Pré-processamento de dados em aprendizado de máquina supervisionado". Tese de Doutorado, ICMC-USP.
- BEKKERMAN, R., El-Yaniv, R., Tishby, N., e Winter, Y. (2003). "Distributional word clusters vs. words for text categorization". Journal of Machine Learning Research 3, 1183-1208
- BELL, T. C., Moffat, A., e Nevill-Manning, C. G. (1993). "Data Compression in full-text retrieval systems". Journal of American Society for Information Science, 44(9):508-531
- BIGGS, M. (2005). "Resurgent text-mining technology can greatly increase your firm's 'intelligence' factor". InfoWorld 11(2), 52
- BOOKSTEIN, A., Klein, S. T., e Raita, T. (1992). "Model based concordance compression". In J.A Storer and M.Cohn, editors, Proc.IEEE Data Compression Conference, pages, 82-91, Snowbird, Utah, IEEE Computer Society Press, Los Alamitos, California
- BORGURAEV, B. e Pustejovsky, J. (1996). Corpus Processing for Lexical Acquisition. The MIT Press
- BOYER, R. S. e Moore, J. S. (1977). "A fast string searching algorithm". Carom.ACM 20, (10), 262-272
- BREIMAN, L., Friedman, R., Olshen, R., e Stone, C. (1984). "Classification and Regression Trees". Wadsworth Monterrey, CA
- BRILL, E. (1993). "Automatic Grammar Induction and Parsing Free Text: A Transformation-Based Approach". In Proceedings of ACL 1993
- BRILL, E. (1995). "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging". Computational Linguistics, December
- BRILL, E. e Resnik, P. (1994). "A Rule-Based Approach To Prepositional Phrase Attachment Disambiguation". In Proceedings of COLING'94, Kyoto, Japan
- BRIN, S. e Page, L. (1998). "Anatomy of a large -scale hypertextual Web search engine". WWW7 Conf.Proceedings
- CARABALLO, S. A. (2001). "Automatic Acquisition of a Hypernym-Labeled Noun Hierarchy from Text". Brown University Ph.D.Thesis
- CARDIE, C. e Pierce, D. (1998). "*Error-driven pruning of treebank grammars for base noun-phrase identification*". In: Proceedings of the 36th Annual

Meeting of the Association for Computational Linguistics, Ithaca-NY: 1998.p.218-224.

CAVNAR, W. B. (1994). "Using An N-Gram-Based Document Representation With a Vector Processing Retrieval Model". In Proceedings Of TREC-3 (Third TextRetrieval Conference).Gaithersburg, Maryland, USA

CEDERBERG, S. e Widdows, D. (2003). "Using LSA and Noun Coordination Information to Improve the Precision and Recall of Automatic Hyponymy Extraction". In: Proceedings of CoNLL-2003, pp.111-118

CHARNIAK, E. (1996). Statistical Language Learning. The MIT Press

CHEN, H. (1994). "A Textual Database/Knowledge-Base Coupling Approach to Creating Computer-Supported Organizational Memory". MIS Department.University of Arizona.5

CHEN, H. (2001). "*Knowledge management systems: a text mining perspective*". University of Arizona (Knowledge Computing Corporation), Tucson, Arizona.

CHOUÉKA, Y., Fraenkel, A. S., e Klein, S. T. (1988). "Compression of concordance in full text retrieval systems". In Proc.ACM-SIGIR International Conference on Research and Development in Information Retrieval, pages 597-612, Grenoble, France

COHEN, W. W. e Hirsh, H. (1998). "*Joins that generalize: text classification using Whirl*". Proceedings of KDD-98, 4th International Conference on Knowledge Discovery and Data Mining

COHEN, W. W. e Singer, Y. (1996). "Context-sensitive learning methods for text categorization". Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval.Zurich.Pages 307-315

COLE, R. (1990). "Tight bounds on the complexity of the Boyer-Moore pattern matching algorithm". Technical Report 512, Computer Science Dept, New York University

COLUSSI, L., Galil, Z., e Giancarlo, R. (1990). "The exact complexity of string matching". 31st Symposium an Foundations of Computer Science I, 135-143, IEEE

COWIE, J. e Wilks, Y. (2000). Information Extraction. A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text, Editors: Robert Dale, Hermann Moisl, and Harold Somers, Marcel Dekker Inc.

CRAIN, S. e Thornton, R. (1998). Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics. The MIT Press

- CROW, L. R. e Shadbolt, N. R. (2001). "Extracting Focused Knowledge from the Semantic Web". *International Journal of Human- Computer Studies* 54(1):pp.155-84
- CUCERZAN, S. e Brill, E. (2004). "Spelling correction as an iterative process that exploits the collective knowledge of web users". *Proceedings of EMNLP 2004* 293-300
- DI FELIPPO, A. e Dias-da-Silva, B. C. (2004). "Inclusão de informação semântica dos adjetivos na base da rede Wordnet para o Português do Brasil". In *Proceedings of the IX Iberoamerican Conference on Artificial Intelligence (IBERAMIA).Workshop "Taller de Herramientas y Recursos Linguisticos para el Espanol y el Português" (HyRL).Puebla/México*, p.147-154
- DIAS-DA-SILVA, B. C. (2003). "O TeP: construção de um thesaurus eletrônico para o português do Brasil". *Boletim da Associação Brasileira de Lingüística (ABRALIN).Fortaleza: Imprensa Universitária*, v.26, número especial, p.86 - 89
- DIAS-DA-SILVA, B. C. e Moraes, H. R. (2003). "A construção de thesaurus eletrônico para o português do Brasil". *Alfa* , v.47, n.2, p.101 - 115
- DIAS-DA-SILVA, B. C., Oliveira, M. F., e Moraes, H. R. (2002). "Groundwork for the development of the Brazilian Portuguese Wordnet". *Advances in natural language processing*.Berlin, Springer-Verlag, p.189-196
- DINI, Lucas, Di Tomaso, V., e Segond, F. (1998). "Error-Driven Word sense disambiguation". In *Proceedings of ACL-COLING-98*, Montreal
- DÖRRE, J., Gerstl, P., e Seiffert, R. (1999). "Text mining: Finding nuggets in mountains os textual data". San Diego, CA, pp.398-401.ACM
- DUDA, R. O., Hart, P. E., e Stork, D. G. (2000). *Pattern Classification*. Hardcover
- DUMAIS, S., Platt, J., Heckerman, D., e Sahami, M. (1998). "Inductive learning algorithms and representations for text categorization". In *Proceedings of the 7th International Conference on Information and Knowledge Management (ACM CIKM 98)*, pp.148-155
- FELLBAUM, C. (1998). *WordNet: An Electronic Lexical Database and Some of its Applications*. MIT Press
- FERNEDA, E. e Smit, J. (2003). "Recuperação de Informação: Análise sobre a contribuição da ciência da computação para a ciência da informação". Tese de Doutorado - USP - Ciência da Informação
- FOX, E. A., Ingwersen, P., e Fidel, R. (1995). "Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval". Seattle, Washington.USA.9-13, Julho.1995.

- FRAKES, W. B. e Baeza-Yates, R. (1992). *Information Retrieval: Data Structures and Algorithms*. Prentice Hall
- FREGE, G. (1975). Sobre o sentido e a referência. In *Lógica e Filosofia da Linguagem* São Paulo: Cultrix
- FREITAS, M. C. (2004). "Elaboração automática de uma ontologia baseada em corpus da língua portuguesa". Proposta de tese de doutorado (qualificação) apresentada no curso de Pós-Graduação em Letras da PUC-Rio
- FREITAS, M. C., Aranha, C., e Quental, V. (2005). "Aprendendo Nomes Próprios". IV Congresso Internacional da ABRALIN (Associação Brasileira de Linguística)
- FUKUDA, F. (1999). "Sistemas inteligentes para textos da web". Tese de Mestrado.PUC-Rio
- GAWRYSIAK, P. (1999). "Using Data Mining methodology for text retrieval ". DIBS'99 conf.proceedings
- GE, N., Hale, J., e Charniak, E. (1998). "A statistical approach to anaphora resolution". In *Proceedings of the Sixth Workshop on Very Large Corpora*, pages 161-170, Montreal, Canada
- GEAN, C. C. e Kaestner, C. A. A. (2004). "Classificação Automática de Documentos usando Subespaços Aleatórios e Conjuntos de Classificadores". In: TIL 2004 - 2º WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA, Salvador.Anais do SBC 2004 v.1, p.1-8.
- GIBSON, D., Kleinberg, J., e Raghavan, P. (1998). "Inferring Web communities from link topology". *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*
- GIRJU, R., Badulescu, A., e Moldovan, D. (2003). "Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations". In: *Proceedings of HLT-2003*
- GOLDSCHMIDT, R. e Passos, E. (2005). *Data Mining - Um Guia Prático*. Campus
- GRISHMAN, R. (1997). "Information Extraction: Techniques and Challenges". in *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, J.G.Carbonell ,Ed.Frascati, Italy: Springer, pp.10-26
- GRUBER, T. R. (1993). "A translation approach to portable ontologies". *Knowledge Acquisition*, 5(2):199-220, 1993
- GRUNE, D. e Jacobs, C. J. H. (1991). *Parsing Techniques: A Practical Guide*. Ellis Horwood Ltd

- GUARINO, N. (1998). "Formal ontology and information systems". In Nicola Guarino, editor, *Formal Ontology and Information Systems*, (FOIS'98).IOS Press, 1998
- HEARST, M. A. (1992). "Automatic Acquisition of Hyponyms from Large Text Corpora". In: *Proceedings of the Fourteenth International Conference on Computational Linguistics*.Nantes, France
- HEARST, M. A. (1999). "Untangling Text Data Mining". *Proceedings of the ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics*.College Park: University of Maryland
- HIRSH-PASEK, K. e Golinkoff, R. (1999). *The Origins of Grammar: Evidence from Early Language Comprehension*. The MIT Press
- HO, T. K. (1998). "The Random Subspace Method for Constructing Decision Forests". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20, Nº 8, pp.832-844
- HU, X. R. e Atwell, E. (2003). "*A survey of machine learning approaches to analysis of large corpora*". School of Computing, University of Leeds, U.K.LS2 9JT
- INOKI, S. (1992). "Uma Gramática de um Fragmento do Português Baseado na Lógica Illocutória". IME - Dissertação de Mestrado
- JOACHIMS, T. (1997). "*A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization*". In *Proceedings of International Conference on Machine Learning (ICML)*
- JOACHIMS, T. (2002). "*Learning to Classify Text Using Support Vector Machines*". Kluwer Academic Publishers.32, 36, 39, 44, 131
- JOACHIMS, T., Freitag, D., e Mitchell, T. (1997). "WebWatcher: A Tour Guide for the World Wide Web". In *IJCAI (1)*, pp.770-777
- KANTROWITZ, M., Mohit, B., e Mittal, V. O. (2000). "Stemming and its effects on TFIDF ranking". *SIGIR 2000: 357-359*
- KAO, A. e Poteet, S. (2004). "*Can Natural Language Processing Help Text Mining?*". Report on KDD Conference 2004 Panel Discussion.SIGKDD Explorations.Volume 6, Issue 2 - Page 132
- KARP, R. e Rabin, M. (1987). "Efficient randomized pattern-matching algorithms". *IBM J.RES.DEVELOP*.Vol 31 No 2
- KIMBALL, R. (1996). *The Data Warehouse Toolkit*. John Wiley & Sons
- KNUTH, D. E., Morris, J. H., e Pratt, V. R. (1977). "Fast pattern matching in strings". *SIAM J.Comput*.6, (2), 323-350

- KOELING, R. (2000). "Chunking with maximum entropy models". CoNLL.Lisbon, Portugal
- KOENIG, M. E. D. (2000). "Information Driven Management: The New, but Little Perceived, Business Zeitgeist". *International Journal of Libraries and Information Services* Vol 50, No 3, pages 137-221
- KOHONEN, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V., e Saarela, A. (2000). "*Self organization of a massive text document collection*". *IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery*, vol.11, pp.574-585
- KRISTA, L. (2000). "Text Mining with the WEBSOM". *Acta Polytechnica Scandinavica, Mathematics and Computing Series* no.110, Espoo 2000, 54 pp. Published by the Finnish Academy of Technology
- KROEZE, H. J., Matthee, C. M., e Bothma, J. D. T. (2003). "Differentiating Data- and Text-Mining Terminology". *Proceedings of SAICSIT 2003*, Pages 93 -101
- LAPPIN, S. e Leass, H. (1994). "An algorithm for pronominal anaphora resolution". *Computational Linguistics*, 20(4):535-561
- LEBART, L., Salem, A., e Berry, L. (1998). *Exploring textual data*. Kluwer Academic Publishers, Dordrecht
- LENAT, D. (1995). "CYC: A large-scale investment in knowledge infrastructure". In: *Communications of the ACM*, 38(11):33-38
- LEWIS, D. D. (1992). "An evaluation of phrasal and clustered representations on a text categorization task". In *Proceedings of the 15th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.37-50
- LI, H. e Yamanishi, K. (2002). "Text Classification Using ESC-based Stochastic Decision Lists". *Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management*
- LI, Y. e Jain, A. (1998). "*Classification of text documents*". *The Computer Journal*, 41, 537-546
- LUCAS, M. (2000). "Mining in textual mountains, an interview with Marti Hearst". *Mappa Mundi Magazine, Trip-M*, 005, 1-3.
- LUHN, H. P. (1958). "The automatic creation of literature abstracts". *IBM Journal of Research and Development* 2(2), 159-165.
- MACCALLUM, A. K. (1996). "Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering". <http://www.cs.cmu.edu/~mccallum/bow>

- MANNING, C. e Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press.
- MARTINS, B. e Silva, M. J. (2004). "Spelling Correction for Search Engine Queries". *EsTAL - España for Natural Language Processing*, Alicante, Spain
- MARTINS, C. A. (2003). "Uma abordagem para pré-processamento de dados textuais em algoritmos de aprendizado". Tese de Doutorado, ICMC-USP - São Carlos
- MATHIAK, B. e Eckstein, S. (2004). "Five Steps to Text Mining in Biomedical Literature". In *Proceedings of the Second European Workshop on Data Mining and Text Mining for Bioinformatics, held in Conjunction with ECML/PKDD in Pisa, Italy*
- MATSUBARA, E. T., Martins, C. A., e Monard, M. C. (2003). "PreTexT - uma ferramenta para pré-processamento de textos utilizando a abordagem *bag-of-words*". Relatório Técnico 209, ICMC-USP
- MATURANA, H. e VARELA, F. J. (1980). *Autopoiesis and cognition; the organization of the living*. Boston: Reidel
- MICHAEL W.EYSENCK (1990). *Psicologia Cognitiva: Um Manual Introductório*. Porto Alegre: Artes Médicas
- MICHIE, D., Spiegelhalter, D. J., e Taylor, C. C. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood
- MOULINIER, I. e Ganascia, J.-G. (1996). "*Applying an existing machine learning algorithm to text categorization*". In S.Wermter, E.Riloff, and G.Scheler, editors, *Connectionist, statistical, and symbolic approaches to learning for natural language processing*, pages 343--354.
- MURADAS, A. (1995). "Interface em Linguagem Natural para Consulta à Bancos de Dados Relacionais". IME - Dissertação de Mestrado
- NASCIMENTO, M. R. (1993). *Pesquisa e Desenvolvimento de uma Interface em Linguagem Natural*. Instituto Militar de Engenharia. Dissertação de Mestrado
- NIRENBURG, S. e Raskin, V. (2004). *Ontological Semantics*. Cambridge, MA: MIT Press
- OLIVEIRA, M. F., Freitas, M. C., Garrão, M. O., Aranha, C., e Santos, C. N. (2004). "A extração de expressões multi-vocabulares: uma abordagem estatística". *Revista PaLavra*, no.12, páginas 172-192
- OWEI, V. (2002). "An Intelligent Approach To Handling Imperfect Information In Concept-Based Natural Language Queries". *ACM Transactions on Information Systems*, Vol.20, No.3, Pages 291-328.

- PANTEL, P. e Ravichandran, D. (2004). "Automatically Labeling Semantic Classes". In: Proceedings of NAACL-2004
- PÉREZ, C. e Vieira, R. (2003). "Extração Semi-automática de Conhecimento a partir de Textos". ENIA - Encontro Nacional de Inteligência Artificial
- PETERSON, R. E. (1997). "*Eight internet search engines compared*". First Monday. http://www.firstmonday.dk/issues/issue2_2/peterson/
- PHILLIPS, W. e Riloff, E. (2002). "Exploiting Strong Syntactic Heuristics and Co-training to learn semantic lexicons". In: Proceedings of EMNLP-2002, pp.125-132
- PINKER, S. (2002). *O Instinto da Linguagem: Como a Mente Cria a Linguagem*. Martins Fontes, SP
- PYLE, D. (1999). *Data Preparation for Data Mining*. San Francisco, CA: Morgan Kaufmann.
- RABINER, L. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition". Proc.IEEE, 77 (2), 257-286
- RAM, A. e Moorman, K. (1999). *Understanding Language Understanding: Computational Models of Reading*. The MIT Press
- RAMSHAW, L. A. e Marcus, M. P. (1995). "Text Chunking using Transformation-Based Learning". In Proceedings of the ACL Third Workshop on Very Large Corpora, June 1995, pp.82-94. In: Proceedings of the 18th International Conference on Computational Linguistics, Saarbrücken, 2000 , p.857-863
- RATNAPARKKHI, A. (1998). "Unsupervised Statistical Models for Prepositional Phrase Attachment". In Proceedings of the Seventeenth International Conference on Computational Linguistics, Aug.10-14, 1998.Montreal, Canada.
- REIS, M. (2005). "Descoberta de Conhecimento em Documentos XML". Tese de Doutorado, DEE PUC-Rio
- SAEED, J. L. (1997). *Semantics*. Oxford: Blackwell
- SALTON, G., Wong, A., e Yang, C. S. (1997). "*A Vector Space Model for Automatic Indexing*". in Readings in Information Retrieval, K.Sparck Jones and P.Willet, eds.,Morgan Kaufmann Publishers, Inc., San Francisco
- SALTON, G. (1983). *Introduction to Moder Information Retrieval*. McGraw-Hill
- SANTOS, D. (2001). "Introdução ao processamento de linguagem natural aqui através das aplicações". in Elisabete Ranchhod (ed.), *Tratamento das Línguas por Computador.Uma introdução à linguística computacional e suas aplicações*, Lisboa: Caminho, pp.229-259.

- SAUSSURE, F. (1982). Curso de Linguística Geral. São Paulo: Cultrix
- SCHAFFER, C. (1994). "A Conservation Law for Generalization Performance". 11th International Conference on Machine Learning, New Brunswick, New Jersey, pp.259-265. Morgan Kaufmann
- SEYMORE, K., McCallum, A., e Rosenfeld, R. (1999). "Learning Hidden Markov Model Structure for Information Extraction". In Proceedings of the AAAI-99 Workshop on ML for IE.
- SILVA, C., Pérez, C., Osório, F., Vieira, R., e Goulart, R. (2003). "Informações sintáticas na mineração de textos". TIL 2003: Workshop de Tecnologia da Informação e da Linguagem Humana
- SNOW, R., Jurafsky, D., e Andrew, Y. (2004). "Learning syntactic patterns for automatic hypernym discovery". Advances in Neural Information Processing Systems 17
- SPARK-JONES, K. e Willet, P. (1997). Readings in Information Retrieval. Morgan Kaufmann
- SUBASIC, P. e Huettner, A. (2001). "Affect Analysis of Text Using Fuzzy Semantic Typing". IEEE Transactions on Fuzzy Systems, Special Issue
- SULLIVAN, D. (2000). The need for text mining in business intelligence. DM Review, Dec. 2000. <http://www.dmreview.com/master.cfm>
- TAN, A.-H. (1999). "Text mining: The state of the art and the challenges". In Proceedings, PAKDD'99 workshop on Knowledge Discovery from Advanced Databases, Beijing, pp.65-70
- THOMAS, J. e Sycara, K. (1999). "Integrating Genetic Algorithms and Text Learning for Financial Prediction". Proceedings of the GECCO-2000 Workshop on Data Mining with Evolutionary Algorithms
- THURASINGHAM, B. (1999). Data mining: technologies, techniques, tools, and trends. CRC Press, Boca Raton, Florida.
- TJONG, E. F. (2000). "*Noun Phrase Recognition by System Combination*". Appears in Proceedings of ANLP-NAACL 2000, Seattle, Washington, USA. Morgan Kaufman Publishers
- TJONG, E. F. (2002). "*Memory-Based Shallow Parsing*". Journal of Machine Learning Research 2 559-594.
- VAN RIJSBERGEN, C. J. (1979). "*Information Retrieval, 2nd edition*". Dept of Computer Science, University of Glasgow.
- VAN RIJSBERGEN, C. J. (1992). "*Probabilistic retrieval revisited*". The Computer Journal, 35(3):291-298

- VITERBI, A. J. (1967). "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm". IEEE Transactions on Information Processing, 13:260-269
- WARTIK, S. (1992). "Boolean Operations". in [Frakes et al.1992], pp.264-292
- WEISS, S. M. e Indurkha, N. (1998). "*Predictive Data Mining*". California, USA: Morgan Kaufmann.
- WIDDOWS, D. (2003). "Unsupervised methods for developing taxonomies by combining syntactic and statistical information". In: Proceedings of HLT/NAACL 2003.pp.276-283
- WIENER, N. (1948). "Cybernetics, or Control and Communication in the Animal and the Machine". Paris: Hermann and Co., Cambridge, MA: The Technology Press, and New York, NY: John Wiley and Sons
- WITTGENSTEIN, L. (1979). *Investigações Filosóficas*. Coleção Os Pensadores, São Paulo: Arril Cultural
- WOLPERT, D. H. (1992). "Stacked Generalization". Neural Networks 5, 241-259
- WU, S., Tsai, T., e Hsu, W. (2003). "Text Categorization using Automatically Acquired Domain Ontology". In: Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages.pp.138-145
- YANG, Y. e Liu, X. (1999). "*A re-examination of text categorization methods*". In Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99), pages 42--49, Berkley
- ZIPF, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley

10

Anexo I:

Principais distribuições de frequência dos significados

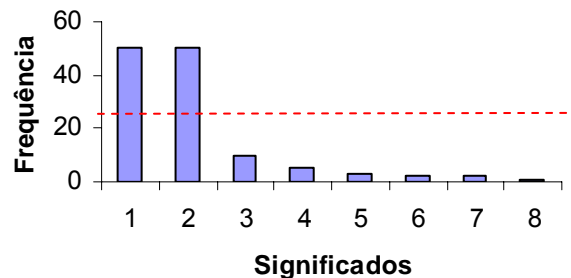
Bimodal, com dois

significados fortes. Ex.

“manga”.

1. manga = fruta
2. manga = trecho da camisa

Força do Significado



Dois sentidos relevantes,

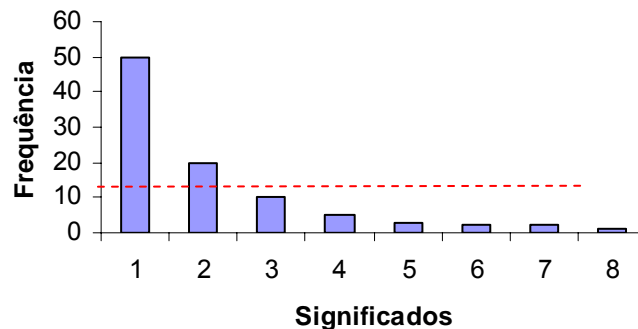
porém com um sentido

mais forte que o outro.

Ex. “prova”.

1. Prova = teste
2. Prova = competição
3. Prova = evidência
4. Prova = provar
5. Prova = imune (à prova de)

Força do Significado



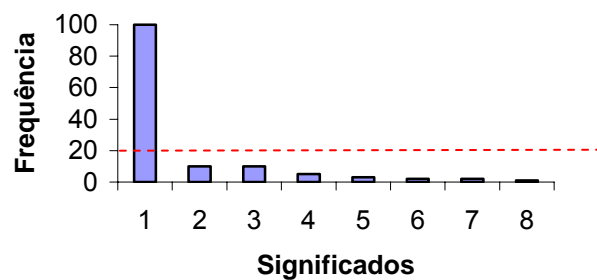
Com apenas um sentido

forte.

Ex. “livro”.

1. livro = texto
2. livro = verbo

Força do Significado



11

Anexo II: Exemplo de Mineração de Textos por Perguntas

Inicializando interface com banco de dados...

```
15:00:16,049 INFO [DbConnectionManager] Registrando driver JDBC.  
(com.microsoft.jdbc.sqlserver.SQLServerDriver)  
15:00:16,059 INFO [DbConnectionManager] Iniciando Fila:  
connection  
15:00:16,370 INFO [DbConnectionManager] Criando conexão para o  
pool de conexões. (connection)  
LexicoDB: 2734
```

Sistema pronto:

Olá, o que deseja saber?

Pergunta: Que locais existem na base?

```
ABC  
Afeganistão  
África  
África do Sul  
Alagoas  
Alemanha  
AM  
Amazônia  
América Latina  
Arábia Saudita  
Araçatuba  
Argentina  
Ásia  
BA  
Bagdá  
Bahia  
Bauru  
BELO HORIZONTE  
BERLIM  
...  
Venezuela  
WASHINGTON  
Zhuhai
```

Pergunta: Quais pessoas foram citadas?

```
Abelardo de Oliveira Filho  
Abram Szajman  
Ahmed Yassin  
Alan Greenspan  
Alberto Fernández  
Alexandre Vasarhelyi  
Allen Johnson  
Aloizio Mercadante  
...  
Vantuil Abdala  
Vargas  
Vladimir Putin
```

Wagner Canhedo
Wagner Victor
Waldomiro Diniz

Pergunta: quais empresas?

Abrafix
Acer Inc.
ADM do Brasil
Agência Brasil
Agência Espacial Brasileira
Agência Nacional de Energia Elétrica
Agência Nacional de Telecomunicações
Agência Nacional de Transportes Terrestres
...
Varig
Vasp
Volkswagen
Votorantim
Votorantim Celulose
Warner Music
Yahoo

Pergunta: tchau
Adeus.