REFERENCIAS

Abir, E. 2003a, *Content conversion method and apparatus*, US Pat. App. 20030061025.

Abir, E. 2003b, *Multilingual database creation system and method*, US Pat. App. 20030139920.

Abir, E. 2003c, *Word association method and apparatus*, US Pat. App. 20030171910.

Abir, E. 2004, *Knowledge system method and appparatus*, US Pat. App. 20040122656.

Abir, E., Klein, S., Miller, D. y Steinbaum, M. 2002, "Fluent Machines' EliMT System", en S.D. Richardson (Ed.): *AMTA 2002*, LNAI 2499, pp. 216-219.

Aduna B.V y Sirma AI Ltd. 2004, "The SeRQL query language", en *User Guide for Sesame*, [Online], openRDF.org, Disponible en: `<http://www.openrdf.org/doc/users/ch06.html>` [30 Junio 2005].

Aguillo, I.F. 2002, *Measuring informal scientific publication in the Web*, [Online], EICSTES, Disponible en: `<http://www.eicstes.org/EICSTES_PDF/PRESENTATIONS/Measuring%20informal%20scientifi c%20publication%20in%20the%20Web%20(Aguillo).PDF>` [30 Junio 2005].

Alfonseca, E. y Rodríguez, P. "Description of the UAM system for generating very short summaries at DUC-2003", *HLT-NAACL Text Summarization Workshop and Document Understanding Conference 2003*.

Alfonseca, E., Guirao, J.M. y Moreno Sandoval, A. 2004, "Description of the UAM system for generating very short summaries at DUC-2004", *NAACL Text Summarization Workshop and Document Understanding Conference 2004*.

Allan, J. 1995, "Relevance Feedback With Too Much Data", en *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*.

Androutsopoulos, I., Koutsias, J., Chandrinos, K.V., Paliouras, G. y Spyropoulos, C.D. 2000a, "An Evaluation of Naive Bayesian Anti-Spam Filtering", [Online], en Potamias, G., Moustakis, V. y van Someren, M. (Eds.): *Proceedings of the Workshop on Machine Learning in the New Information Age*, pp. 9-17, disponible en: `<http://www.aueb.gr/users/ion/docs/mlnet_paper.pdf>` [30 Junio 2005]

Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Sakkis, G., Spyropoulos, C.D. y Stamatopoulos, P. 2000b, "Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach", [Online], en Zaragoza, H., Gallinari, P. y Rajman, M. (Eds.): *Proceedings of the Workshop on Machine Learning and Textual Information Access*, pp. 1-13, disponible en: `<http://arxiv.org/abs/cs/ 0009009>` [30 Junio 2005]

Apté, C., Damerau, F. y Weiss, S.M. 1994, "Automated Learning of Decision Rules for Text Categorization", *ACM Transactions on Information Systems*, vol. 12, no. 3, pp. 233-251.

Apté, C., Damerau, F. y Weiss, S.M. 1998, "Maximizing Text-Mining Performance", *IEEE Intelligent Systems*, vol. 14, no. 4, pp. 63-69.

Attardi, G., Gullì, A. y Sebastiani, F. 1999, "Theseus: Categorization by context", en *Proceedings of WWW8*, pp. 136-137

Baclace, P.E. 1991, "Personal Information Intake Filtering", *Bellcore Information Filtering Workshop*.

Baclace, P.E. 1992, "Competitive agents for information filtering", *Communications of the ACM*, vol. 35, no. 12, p. 50.

Baeza-Yates, R. y Ribeiro-Neto, B. 1999, *Modern Information Retrieval*, ACM Press.

Balabanovic, M. 1998, "An interface for learning multi-topic user profiles from implicit feedback", en *Proceedings of AAAI Workshop on Recommender Systems*, pp. 6-10.

Balabanovic, M. y Shoham, Y. 1997, "Fab: Content-Based, Collaborative Recommendation", *Communications of the ACM*, vol. 40, no. 3, pp. 66-72.

Balabanovic, M., Shoham, Y. y Yun, Y. 1995, An adaptive agent for automated web browsing, Informe técnico, Stanford University.

Barton, I.J., Creasey, S.E., Lynch, M.F., Snell, M.J. 1974, "An Information-Theoretic Approach to Text Searching in Direct Access Systems", *Communications of the ACM*, vol. 17, no. 6, pp. 345-350.

Barzilay, R. y Elhadad, M. 1997, "Using Lexical Chains for Text Summarization", en *Proceedings of the Workshop on Intelligent Scalable Text Summarization*, pp. 10-17.

Barzilay, R. y Lee, L. 2002, "Bootstrapping Lexical Choice via Multiple-Sequence Alignment", en *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 164-171.

Barzilay, R. y Lee, L. 2003, "Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment", en *Proceedings of HLT-NAACL 2003*, pp. 16-23.

Bates, M.J. 1989, "The Design of Browsing and Berrypicking Techniques for the Online Search Interface.", *Online Review*, 13, pp. 407-424.

Bates, M.J. 2002, "After the Dot-Bomb: Getting Web Information Retrieval Right This Time", *First Monday*, vol. 7, no. 7, [Online], Disponible en: <http://firstmonday.org/issues/issue7_7/bates/index.html> [30 Junio 2005].

Baxendale, P.B. 1958, "Machine-made index for technical literature – an experiment", *IBM Journal*, pp. 354-361.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L, Patel Schneider, P.F. y Stein, L.A. 2004, *OWL Web Ontology Language Reference*, [Online], Recomendación W3C, Consorcio W3, Disponible en: <http://www.w3.org/TR/owl-ref/> [30 Junio 2005].

Beesley, K.R. 1988, "Language identifier: A computer program for automatic natural-language identification on on-line text", en *Proceedings of the 29th Annual Conference of the American Translators Association*, pp. 47-54.

Belkin, N. y Vickery, A. 1985, *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-based systems (LIR Report No 35)*, The British Library, Reino Unido.

Bennett, C.H., Gács, P., Li, M., Vitányi, M.B. y Zurek, W.H. 1998, "Thermodynamics of Computation and Information Distance", *IEEE Transactions in Information Theory*, vol. 44, no. 4, pp. 1407-1423.

Bergman, M.K. 2001, "The Deep Web: Surfacing Hidden Value", *The Journal of Electronic Publishing*, vol. 7, no. 1, Disponible en: <http://www.press.umich.edu/jep/07-01/bergman.html> [30 Junio 2005].

Berners-Lee, T. 1989, *Information Management: A Proposal*, Informe técnico, CERN.

Berners-Lee, T. 1990, *HyperText Design Issues: Topology*, [Online], W3 Consortium, Disponible en: <http://www.w3.org/DesignIssues/Topology.html> [30 Junio 2005].

Berners-Lee, T. 1992a, *Web of Indexes*, [Online], W3 Consortium, Disponible en: <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/DesignIssues/ManyIndexes.html> [30 Junio 2005].

Berners-Lee, T. 1992b, *Tracing Links*, [Online], W3 Consortium, Disponible en: <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/DesignIssues/TracingLinks.html> [30 Junio 2005].

Berners-Lee, T. 1992c, *World-Wide Web Servers*, [Online], W3 Consortium, Disponible en: <http://www.w3.org/History/19921103-hypertext/hypertext/DataSources/WWW/Servers.html> [30 Junio 2005].

Berners-Lee, T. 1998, *Semantic Web Road map*, [Online], Consorcio W3, Disponible en: <http://www.w3.org/DesignIssues/Semantic.html> [30 Junio 2005].

Berners-Lee, T., Hendler, J. y Lassila, O. 2001, "The Semantic Web", *Scientific American*, vol. 284, no. 5, pp. 34-43.

Bharat, K., y Henzinger, M. 1998, "Improved Algorithms for Topic Distillation in a Hyperlinked Environment", en *Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval*, pp. 104-111.

Billhardt, H., Borrajo, D. y Maojo, V. 2003, "Learning retrieval expert combinations with genetic algorithms", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 11, no. 1, pp. 87-114.

Blair-Goldensohn, S., Evans, D., Hatzivassiloglou, V., McKeown, K.R., Nenkova, A., Passonneau, R., Schiffman, B., Schlaikjer, A., Siddharthan, A. y Siegelman, S. 2004, "Columbia University at DUC 2004", en *Document Understanding Conference at HLT-NAACL*.

Bookstein, A. y Kraft, D. 1977, "Operations research applied to document indexing and retrieval decisions", *Journal of the ACM*, vol. 24, pp. 410-427 (citado por van Rijsbergen, C.J. 1979, Information Retrieval, 2nd Edition. Butterworth-Heinemann, EE.UU.)

Boser, B., Guyon, I. y Vapnik, V. 1992, "A training algorithm for optimal margin classifiers", en *Fifth Annual Workshop on Computational Learning Theory*, pp. 144-152.

Brandow, R., Mitze, K. y Rau, L.F. 1995, "Automatic condensation of electronic publications by sentence selection", *Information Processing and Management*, vol. 31, no. 5, pp. 675-685.

Bray, J.R. y Curtis, J.T. 1957, "An ordination of the upland forest communities of southern Wisconsin". *Ecological Monographs*, 27, pp. 325-349. Citado por Gauch, H.G. 1982, *Multivariate Analysis in Community Ecology*, Cambridge University Press, EE.UU.

Brickley, D. y Guha, R.V. 2004, *RDF Vocabulary Description Language 1.0: RDF Schema*, Recomendación W3C, Consorcio W3, [Online], Disponible en: `<http://www.w3.org/TR/rdf-schema>` [30 Junio 2005].

Brickley, D. y Miller, L. 2000, *RDF: Extending and Querying RSS channels*, [Online], Informe técnico, Universidad de Bristol, Disponible en: `<http://ilrt.org/discovery/2000/11/rss-query/>` [30 Junio 2005].

Brin, S. y Page, L. 1998, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 107-117.

Broadman, E. 1944, "Choosing physiology journals", *Bulletin of the Medical Library Association*, vol. 32, no. 4, pp. 479-483.

Broder, A. 2002, "A taxonomy of web search", *ACM SIGIR Forum*, vol. 36, no. 2, pp. 3-10.

Bruce, T.R. 1993, *Resource discovery and the Web*, [Online], University of Calgary, Disponible en: `<http://ksi.cpsc.ucalgary.ca/archives/WWW-TALK/www-talk-1993q2.messages/36.html>` [30 Junio 2005].

Brunn, M., Chali, Y. y Pinchak, C.J. 2001, "Text Summarization Using Lexical Chains", en *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Buckley, C., Salton, G., Allan, J. y Singhal, A. 1994, "Automatic Query Expansion Using SMART: TREC-3", en *Text REtrieval Conference*, pp. 69-80.

Burke, R. 1999, "Integrating Knowledge-based and Collaborative-filtering Recommender Systems", en *Proceedings of the AAAI Workshop on AI and Electronic Commerce*, pp. 69-72.

Busemann, S. 2001, "Language Generation for Cross-Lingual Document Summarisation", en *Proceedings of the International Workshop on Innovative Language Technology and Chinese Information Processing (ILT&CIP'01)*.

Cancedda, N., Déjean, H., Gaussier, É., Renders, J.M. y Vinokourov, A. 2003, "Report on CLEF-2003 Experiments: Two Ways of Extracting Multilingual Resources from Corpora", *CLEF 2003*, pp. 98-107.

Carpineto, C., y Romano, G. 2000, "Order-theoretical ranking", *Journal of American Society for Information Science*, vol. 51, no. 7, pp. 587-601.

Cauwenberghs, G. y Poggio, T. 2000, "Incremental and decremental support vector machine learning", en *Advances in Neural Information Systems*, pp. 409-415.

Cavnar, W.B. 1994, "Using an *n*-gram-based document representation with a vector processing retrieval model", en *Proceedings of TREC-3*, pp. 269-277.

Cavnar, W.B. y Trenkle, J.M. 1994, "N-Gram-Based Text Categorization", en *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, pp. 161-175.

Chakrabarti, S. 2003, *Mining the Web. Discovering Knowledge from Hypertext Data*, Morgan-Kaufmann Publishers.

Chakrabarti, S., Dom, B.E., Gibson, D., Kleinberg, J., Raghavan, P. y Rajagopalan, S. 1998a, "Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text", en *Proceedings of the 7th World-Wide Web conference*, pp. 65-74.

Chakrabarti, S., Dom, B.E., Gibson, D., Kumar, R., Raghavan, P., Rajagopalan, S. y Tomkins, A. 1998b, "Experiments in topic distillation", en *Proceedings of the ACM SIGIR Workshop on Hypertext Information Retrieval on the Web*, pp. 185-193.

Chakrabarti, S., Dom, B.E., Agrawal, R. y Raghavan, P. 1998c, "Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies", *The VLDB Journal*, no. 7, pp. 163-178.

Chan, L.M. 1994, *Cataloging and Classification: An Introduction, 2nd Edition*, McGraw-Hill, Nueva York.

Chang, B. 1998, "In-place editing of web pages: Sparrow community-shared documents", *Computer Networks and ISDN Systems*, vol. 30, pp.489-498.

Chekuri, C., Goldwasser, M.H., Raghavan, P. y Upfal, E. 1997, "Web Search Using Automatic Classification", en *Sixth International World Wide Web Conference*.

Chen, X., Kwong, S. y Li, M. 1999, "A Compression Algorithm for DNA Sequences and its Applications in Genome Comparison", en *Proceedings of the 10th Workshop on Genome Informatics (GIW'99)*, pp. 51-61.

China Internet Network Information Center (CNNIC). 2003, *12th Statistical Survey on the Internet Development in China (2003/7)*.

Clark, K.G. (ed), 2004, *RDF Data Access Use Cases and Requirements*, [Online], Borrador de trabajo, Consorcio W3, Disponible en: <http://www.w3.org/TR/rdf-dawg-uc/> [30 Junio 2005].

Cleverdon, C., Mills, J. y Keen, E.M. 1966, *Aslib Cranfield Research Project Volume 2: Factors Determining the Performance of Indexing Systems*, ASLIB, Reino Unido.

Cleverdon, C.W. 1962, *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems*, College of Aeronautics, Reino Unido (citado por Harman, D. 1993, "Overview of the First Text REtrieval Conference (TREC-1)", en *NIST Special Publication 500-207: The First Text REtrieval Conference TREC-1*).

Cohen, J.D. 1995, "Highlights: Language and Domain-Independent Automatic Indexing Terms for Abstracting", *JASIS*, 46(3), pp. 162-174.

Cohen, W. 1996, "Learning rules that classify email", en *Proceedings of the IEEE Spring Symposium on Machine Learning for Information Access*.

Cohen, W. y Hirsh, H. 1998, "Joins that generalize: Text categorization using WHIRL", en *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, pp. 169-173.

Collobert, R., Bengio, S. y Bengio, Y. 2002, "Parallel mixture of SVMs for very large scale problems", *Neural Computation*, vol. 14, no. 5, pp. 1105-1114.

Conroy, J.M., Schlesinger, J.D., O'Leary, D.P. y Okurowski, M.E. 2001, "Using HMM and Logistic Regression to Generate Extract Summaries for DUC", en *Proceedings of the 1st Document Understanding Conference*.

Cowie, J., Mahesh, K., Nirenburg, S. y Zajac, R. 1998, "MINDS - Multi-lingual INteractive Document Summarization", en *AAAI 1998 Spring Symposium Series*.

Cranor, L.F. y LaMacchia, B.A. 1998, "Spam!", *CACM*, vol. 41, no. 8, pp. 74-83.

Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K. y Slattery, S. 1998, "Learning to Extract Symbolic Knowledge from the World Wide Web", en *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI98)*.

Crestani, F. y van Rijsbergen, C.J. 1998, "A study of probability kinematics in information retrieval", *ACM Transactions on Information Systems*, vol. 16, no. 3, pp. 225-255.

Croft, W.B. y Harper, D.J. 1979, "Using probabilistic models of document retrieval without relevance information", *Journal of Documentation*, vol. 35, pp. 285-295.

Croft, W.B. y Thompson, R.H. 1987, "I3R: A new approach to the design of document retrieval systems", *Journal of the American Society for Information Science*, vol. 38, no. 6, pp. 389-404.

Cross, V. 1994, "Fuzzy Information Retrieval", *Journal of Intelligent Information Systems*, vol. 3, pp. 29-56.

D'Amore, R., Mah, C.P. 1985, "One-time complete indexing of text: Theory and practice", en *Proceedings of SIGIR 1985*, pp. 155-164.

Damashek, M. 1995, *Gauging Similarity via N-Grams: Language-Independent Sorting, Categorization, and Retrieval of Text*. Departamento de Defensa, EE.UU.

David, C. 2003, *Information Society Statistics – PCs, Internet and mobile phone usage in the EU*, [Online], Eurostat, Disponible en: `<http://epp.eurostat.cec.eu.int/cache/ITY_OFFPUB/KS-NP-03-015/EN/KS-NP-03-015-EN.PDF>` [30 Junio 2005].

Davison, B.D. 2000a, "Recognizing Nepotistic Links on the Web", en *Proceedings of AAAI-2000 Workshop on Artificial Intelligence for Web Search*, pp. 23-28.

Deerwester, S., Dumais, S.T., Landauer, T.K., Furnas, G.W. y Harshman, R.A. 1990, "Indexing by latent semantic analysis", *Journal of the Society for Information Science*, vol. 41, no. 6, pp. 391-407.

DeJong, G. 1982, "An Overview of the FRUMP System", en Lehnert, W.G. y Ringle, M.H. (Eds.) *Strategies for Natural Language Processing*, pp. 149-176.

Denker, G., Hobbs, J.R., Martin, D., Narayanan, S. y Waldinger, R. 2001, "Accessing Information and Services on the DAML-Enabled Web", en *Proceedings of the Second International Workshop Semantic Web*.

Denning, P.J. 1982, "Electronic junk", *CACM*, vol. 25, no. 3, pp. 163-165.

Domingos, P. y Pazzani, M. 1997, "On the optimality of the simple Bayesian classifier under zero one loss", *Machine Learning*, vol. 29, no. 2-3, pp. 103-130.

Doran, W., Stokes, N., Carthy, J. y Dunnion, J. 2004, "Assessing the impact of lexical chain scoring methods and sentence extraction schemes on summarization", *Proceedings of the 5th International conference on Intelligent Text Processing and Computational Linguistics CICLing-2004*.

Doyle, L.B. 1959, "Programmed interpretation of text as a basis for information retrieval systems", en *Proceedings of the Western Joint Computer Conference*, San Francisco, pp. 60-63 (citado por van Rijsbergen, C.J. 1979, *Information Retrieval, 2nd Edition*. Butterworth-Heinemann, EE.UU.)

Doyle, L.B. 1965, "Some compromises between word grouping and document grouping", en Stevens *et al.* (Eds.): *Statistical Association Methods for Mechanized Documentation*, National Bureau of Standards, EE.UU., pp. 15-24 (citado por van Rijsbergen, C.J. 1979, *Information Retrieval, 2nd Edition*. Butterworth-Heinemann, EE.UU.)

Drucker, H., Wu, D. y Vapnik, V. 1999, "Support vector machines for spam categorization", *IEEE transactions on Neural Networks*, vol. 10, no. 5, pp. 1048-1055.

Duda, R.O. y Hart, P.E. 1973, *Pattern Classification and Scene Analysis*, John Wiley and Sons, Inc. Citado en Jain, A.K., Murty, M.N. y Flynn, P.J. 1999, "Data Clustering: a review", *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323.

Dumais, S.T., Furnas, G.W., Landauer, T.K., Deerwester, S. y Harshman, R. 1988, "Using Latent Semantic Analysis to improve access to textual information", en *Human Factors in Computing Systems, CHI'88 Conference Proceedings,* pp. 281-285.

Dumais, S.T., Platt, J., Heckerman, D. y Sahami, M. 1998, "Inductive learning algorithms and representations for text categorization", en *Proceedings of ACM-CIKM98*, pp. 148-155.

Dunlavy, D.M., Conroy, J.M., Schlesinger, J.D., Goodman, S.A., Okurowski, M.E., O'Leary, D.P. y van Halteren, H. 2003, "Performance of a three-stage system for multi-document summarization", en *Proceedings of the Document Understanding Conference (DUC)*.

Dunning, T. 1993, "Accurate methods for the statistics of surprise and coincidence", en *Computational Linguistics*, vol. 19, no. 1, pp. 61-74.

Dunning, T. 1994, *Statistical identification of language*, Informe técnico, New Mexico State University.

Dyen, I., Kruskal, J.B. y Black, P. 1992, "An Indoeuropean Classification: a lexicostatistical experiment", *Transactions of the American Philosophical Society*, vol. 82, no. 5.

EAGLES, 1996, *Preliminary recommendations on corpus typology. EAG-TCWG-CTYP/P*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

Edmundson, H.P. 1969, "New methods in automatic abstracting", *Journal of the ACM*, vol. 16, no. 2, pp. 264-285.

Erkan, G. y Radev, D. 2004a, "LexPageRank: Prestige in multi- document text summarization", en *Proceedings of EMNLP*.

Erkan, G. y Radev, D. 2004b, "The University of Michigan at DUC 2004", en *DUC 2004*.

Evans, D.K. y Klavans, J.L. 2003, *A platform for multilingual news summarization*, [Online], Informe Técnico, University of Columbia, Disponible en : `<http://www.cs.columbia.edu/~library/TR-repository/reports/reports-2003/cucs-014-03.pdf>` [30 Junio 2005]

Evans, D.K., Klavans, J.L. y McKeown, K.R. 2004, "Columbia Newsblaster: Multilingual News Summarization on the Web", en *Proceedings of HLT-NAACL 2004, Demonstrations*, pp. 1-4.

Evans, S.N., Ringe, D. y Warnow, T. 2004, "Inference of divergence times as a statistical inverse problem", *Phylogenetic Methods and the Prehistory of Languages.* Cambridge, Reino Unido.

Fan, W., Gordon, D. y Pathak, P. 2004a, "A generic ranking function discovery framework by genetic programming for information retrieval", *Information Processing and Management*, vol. 40, no. 4, pp. 587-602.

Fan, W., Gordon, D. y Pathak, P. 2004b, "Discovery of context-specific ranking functions for effective information retrieval using genetic programming", *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 4, pp. 523-527.

Fensel, D., Angele, J., Decker, S., Erdmann, M., Schnurr, H.P., Staab, S., Studer, R. y Witt, A. 1999, "On2broker: Semantic-based access to information sources at the WWW", en *Proceedings of the World Conference on the WWW and Internet (WebNet 99)*.

Fensel, D., Decker, S., Erdmann, M. y Studer, R. 1998, "Ontobroker: Or How to Enable Intelligent Access to the WWW", en *Proceedings of the 11th Workshop on Knowledge Acquisition, Modeling, and Management.*

Ferreira da Silva, J., Pereira Lopes, G. 1999, "A Local Maxima method and a Fair Dispersion Normalization for extracting multi-word units from corpora", en *Proceedings of MOL6*, pp. 369-381.

Filo, D. y Yang, J. 1994, *Yahoo!*, [Online], Yahoo! Inc., Disponible en: `<http://www.yahoo.com>` [30 Junio 2005].

Flake, G.W., Pennock, D.M. y Fain, D.C. 2003, "The Self-Organized Web: The Yin to the Semantic Web's Yang", *IEEE Intelligent Systems*, vol. 18, no. 4, pp. 75-77.

Fletcher, J. 1994, *Internet Robots – Structure from Anarchy?*, [Online], Internet Archive, Disponible en: `<http://web.archive.org/web/19990116235019/http://lorne.stir.ac.uk/~jf1/papers/signidr.html>` [30 Junio 2005].

Florescu, D., Levy, A. y Mendelzon, A. 1998, "Database Techniques for the World-Wide Web: A Survey", *ACM SIGMOD Record*, vol. 27, no. 3, pp. 59-74.

Foltz, P.W. 1990, "Using Latent Semantic Indexing for Information Filtering", en *Proceedings of the ACM Conference on Office Information Systems*, pp. 40-47.

Foltz, P.W. y Dumais, S.T. 1992, "Personalized Information Delivery: An Analysis of Information Filtering Methods", *Communications of the ACM*, vol. 35, no. 12, pp. 51-60.

Fox, E. 1983. *Characteristics of Two New Experimental Collections in Computer and Information Science Containing Textual and Bibliographic Concepts*, Informe técnico, Cornell University.

Fuentes, M., Massot, M., Rodríguez, H. y Alonso, L. 2003, "Mixed approach to headline extraction for DUC 2003", en *Proceedings of DUC 2003*.

Fukumoto, F., Suzuki, Y. y Fukumoto, J. 1997, "An Automatic Extraction of Key Paragraphs Based on Context Dependency", en *Proceedings of the 5th International on Applied Natural Language Processing.*

Fum, D., Guida, G. y Tasso, C. 1985, "Evaluating Importance: A Step Towards Text Summarization", en *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pp. 840-844.

Furnas, G.W., Landauer, T.K., Gómez, L.M. y Dumais, S.T. 1987, "The vocabulary problem in human-system communication", *Communications of the ACM*, vol. 30, no. 11, pp. 964-971.

Fussler, H.H. 1949, "Characteristics of the research literature used by chemists and physicists in the United States" en *The origins of information science*, 1987, ed. A.J. Meadows, Taylor Graham Publishing.

Galliers, J.R. y Spärck-Jones, K. 1993, *Evaluating Natural Language Processing Systems*, Informe Técnico, Computer Laboratory, University of Cambridge.

Garfield, E. 1972, "Citation Analysis as a Tool in Journal Evaluation", *Science*, vol. 178, pp. 471-479.

Gawronska, B. 2002, "Employing Cognitive Notions in Multilingual Summarization of News Reports", en *Proceedings of NLULP-02 (The 7th International Workshop on Natural Language Understanding and Logic Programming)*.

Gayo Avello, D. y Álvarez Gutiérrez, D. 2002, "The Cooperative Web: A Complement to the Semantic Web", en *Proceedings of the 26th Annual International Computer Software & Applications Conference (COMPSAC'02)*, pp. 179-183.

Gayo Avello, D., Álvarez Gutiérrez, D. y Gayo Avello, J. 2004a, "Naïve Algorithms for Keyphrase Extraction and Text Summarization from a Single Document Inspired by the Protein Biosyntehis Process", en A.J. Ijspeert *et al.* (Eds.): *BioADIT 2004*, LNCS 3141, pp. 440-455.

Gayo Avello, D., Álvarez Gutiérrez, D. y Gayo Avello, J. 2004b, "One Size Fits All? A Simple Technique to Perform Several NLP Tasks", en J.L. Vicedo *et al.* (Eds.): *EsTAL 2004*, LNAI 3230, pp. 267-278.

Gayo Avello, D., Álvarez Gutiérrez, D. y Gayo Avello, J. 2004c, "Application of variable length *n*-grams to monolingual and bilingual information retrieval", en Peters, C. y Borri, F. (Eds.): *Working Notes for the CLEF 2004 Workshop*.

Goddard, I. y Campbell, L. 1994, "The History and Classification of American Indian Languages: What are the Implications for the Peopling of the Americas?", en Bonnichsen, R. y Steele, D.G. (Eds.): *Method and Theory for Investigating the Peopling of the Americas*.

Goldberg, D., Nichols, D., Oki, B.M. y Terry, D. 1992, "Using Collaborative Filtering to Weave an Information Tapestry", *Communications of the ACM*, vol. 35, no. 12, pp. 61-70.

Goldman, C.V., Langer, A. y Rosenschein, J.S. 1996, "Musag: an agent that learns what you mean", en *Proceedings of The First International Conference on The Practical Application of Intelligent Agents and Multi Agent Technology (PAAM '96)*, pp. 311-329.

Graham, L. y Metaxas, P.T. 2003, "Of Course it's True; I Saw It on the Internet! Critical Thinking in the Internet Era", *Communications of the ACM*, vol. 46, no. 5, pp. 71-75.

Gray, M. 1995, *Measuring the Growth of the Web – June 1993 to June 1995*, [Online], MIT, Disponible en: <http://www.mit.edu/people/mkgray/growth> [30 Junio 2005].

Gray, R.D. y Atkinson, Q.D. 2003, Language-tree divergence times support the Anatolian theory of Indo-European origin, *Nature*, 426, pp. 435-439.

Greenberg, J.H. 1966, *The Languages of Africa*, Bloomington, Indiana University.

Greenberg, S. y Roseman, M. 1996, "GroupWeb: A WWW Browser as Real Time Groupware", en *Proceedings of ACM SIGCHI'96 Conference on Human Factors in Computing System*.

Grefenstette, G. 1995, "Comparing two language identification schemes", en *Proceedings of 3rd International Conference on Statistical Analysis of Textual Data*.

Griffith, B.C. 1980, *Key Papers in Information Science*, American Society for Information Science, EE.UU.

Gross, P.L.K. y Gross, E.M. 1927, "College libraries and chemical education", en *The origins of information science*, 1987, ed. A.J. Meadows, Taylor Graham Publishing.

Gruber, T.R. 1993, "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, vol. 5, no. 2, pp. 199-220.

Guha, R.V., Lassila, O., Miller, E. y Brickley, D. 1998, "Enabling inferencing", [Online], en *W3C Query Languages Workshop (QL'98)*, Disponible en: <http://www.w3.org/TandS/QL/QL98/pp/enabling.html> [30 Junio 2005].

Han, E., Boley, D., Gini, M., Gross, R., Hastings, K., Karypis, G., Kumar, V., Mobasher, B. y Moore, J. 1998, "WebACE: A Web Agent for Document Categorization and Exploration", en *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)*, pp. 408-415.

Han, E., Karypis, G. y Kumar, V. 1999, *Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification*, Informe técnico, Universidad de Minnesota.

Hardy, H., Shimizu, N., Strzalkowski, T., Ting, L., Zhang, X. y Wise, B. 2001, "Cross-Document Summarization by Concept Classification", *Document Understanding Conference, a SIGIR'01 workshop*.

Harman, D. 1993, "Overview of the First Text REtrieval Conference (TREC-1)", en *NIST Special Publication 500-207: The First Text REtrieval Conference TREC-1*.

Harman, D. 1996, "Overview of the Fourth Text REtrieval Conference (TREC-4)", en *NIST Special Publication 500-236: The Fourth Text REtrieval Conference (TREC-4).*

Harper, D. y van Rijsbergen, C.J. 1978, "An evaluation of feedback in document retrieval using co-occurrence data", *Journal of Documentation*, vol. 34, pp. 189-216.

Hearst, M.A. 1994, "Multi-Paragraph Segmentation of Expository Text", en *Proceedings of the 32nd Meeting of the Association for Computational Linguistics*, pp. 9-16.

Hearst, M.A. y Karadi, C. 1997, "Cat-a-Cone:An Interactive Interface for Specifying Searching and Viewing Retrieval Results using a Large Category Hierarchy", en *Proceedings of the 20th Annual International ACM/SIGIR Conference.*

Henderson, M.M. 1999, "Examples of Early Nonconventional Technical Information Systems", en *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, eds. Bowden, M.E., Hahn, T.B. y Willians R.V., Information Today Inc., pp. 169-176.

Hersh, W., Buckley, C., Leone, T. y Hickam, D. 1994, "OHSUMED: An interactive retrieval evaluation and new large test collection for research", en *Proc. of SIGIR'94*, pp. 192-201.

Hirao, T., Sasaki, Y., Isozaki, H. y Maeda, E. 2002, "NTT's Text Summarization System for DUC 2002", en *Proceedings of the Workshop on Multi-Document Summarization Evaluation of the 2nd Document Understanding Conference at the 40th Meeting of the Association for Computational Linguistics.*

Hirao, T., Suzuki, J., Isozaki, H. y Maeda, E. 2003, "NTT's Multiple Summarization System for DUC2003", en *Proc. of the 3th Document Understanding Conference (DUC).*

Honkela, T., Kaski, S., Lagus, K., y Kohonen, T. 1996, *Newsgroup exploration with WEBSOM method and browsing interface*, Informe técnico, Universidad Tecnológica de Helsinki.

Horrocks, I., Fensel, D., Broekstra, J., Decker, S., Erdmann, M., Goble, C., van Harmelen, F., Klein, M., Staab, S., Studer, R. y Motta, E. 2000, *The Ontology Inference Layer OIL. Technical report*, [Online], Informe técnico, On-To-Knowledge, Disponible en: `<http://www.ontoknowledge.org/oil/TR/oil.long.html>` [30 Junio 2005].

Hovy, E. (Ed.) 1999, "Cross-lingual Information Extraction and Automated Text Summarization", en *Multilingual Information Management: Current Levels and Future Abilities* (Hovy, E. *et al*, eds.), [Online], US National Science Fundation, Disponible en: `<http://www-2.cs.cmu.edu/~ref/mlim/>` [30 Junio 2005]

Hovy, E. y Lin, C.Y. 1997, "Automated Text Summarization in SUMMARIST", en *Proceedings of the Workshop on Intelligent Scalable Text Summarization*, pp. 18-24.

Hovy, E. y Lin, C.Y. 1998, "Automated Text Summarization and the SUMMARIST System", en *Proceedings of the TIPSTER Text Program Phase III final report*, pp. 197-214.

Huffman, S.M. 1998, *The Genetic Classification of Languages by n-Gram Analysis: A Computational Technique*, tesis doctoral, Universidad de Georgetown.

Hull, D.A. 1994, "Improving text retrieval for the routing problem using latent semantic indexing", en Croft, W.B. y van Rijsbergen, C.J. (Eds.): *Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval*, pp. 282-289.

HyperText Mark-up Language, 1992, *HyperText Mark-up Language*, [Online], W3 Consortium, Disponible en: `<http://www.w3.org/History/19921103-hypertext/hypertext/WWW/MarkUp/MarkUp.html>` [30 Junio 2005].

IBM, 1994-2004, *IBM 1301 disk storage unit*, [Online], IBM Corporation, Disponible en: `<http://www.ibm.com/ibm/history/exhibits/storage/storage_1301.html>` [30 Junio 2005].

IBM, 2002, *Fifty Years Of Storage Innovation: Magnetic Tape & Beyond*, [Online], IBM Corporation, Disponible en: `<http://www.ibm.com/ibm/history/exhibits/storage/storage_fifty.html>` [30 Junio 2005].

Ingwersen, P. 1992, *Information Retrieval Interaction*, Taylor Graham, Reino Unido, Disponible en: `<http://www.db.dk/pi/iri/>` [30 Junio 2005]

IPA (*International Phonetic Association*) 1999, *Handbook of the International Phonetic Association : A Guide to the Use of the International Phonetic Alphabet*, Cambridge University Press, Reino Unido.

Ittner, D.J., Lewis, D.D. y Ahn, D.D. 1995, "Text categorization of low quality images", en *Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval*, pp. 301-315.

Iwayama, M. y Tokunaga, T. 1995, "Hierarchical Bayesian Clustering for Automatic Text Classification", en *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95*, pp. 1322-1327.

Jain, A.K. y Dubes, R.C. 1988, *Algorithms for Clustering Data*, Prentice Hall. Citado en Zhao, Y. y Karypis, G. 2002, *Criterion Functions for Document Clustering: Experiments and Analysis*, Informe técnico, Universidad de Minnesota.

Jain, A.K., Murty, M.N. y Flynn, P.J. 1999, "Data Clustering: a review", *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323.

Jalam, R. y Teytaud, O. 2001, "Kernel-based text categorization", en *International Joint Conference on Neural Networks (IJCNN'2001)*.

Jansen, B.J. y Spink, A. 2003, "An Analysis of Web Documents Retrieved and Viewed", *The 4th International Conference on Internet Computing*, pp. 65-69.

Jansen, B.J., Spink, A., Bateman, J. y Saracevic, T. 1998, "Real life information retrieval: A study of user queries on the web", *SIGIR Forum*, vol. 32, no. 1, pp. 5-17.

Jaoua, M. y Ben Hamadou, A. 2003, "Automatic Text Summarization of Scientific Articles Based on Classification of Extract's Population", *CICLing 2003* (citado por Alfonseca, E. *et al.* 2004, "Description of the UAM system for generating very short summaries at DUC-2004", *NAACL Text Summarization Workshop and Document Understanding Conference 2004*).

Jardine, N. y van Rijsbergen, C.J. 1971, "The use of hierarchic clustering in information retrieval", *Information Storage and Retrieval*, vol. 7, pp. 217-240.

Järvelin, K. y Vakkari, P. 1992, "Content Analysis of research articles in LIS 1965-1985", en Cronin, B. y Vakkari, P. (Eds.): *Conceptions of Library and Information Science. Proceedings of First CoLIS Conference* (citado por *Ingwersen*, P. 1992, Information Retrieval Interaction, Taylor Graham, Reino Unido).

Jarvis, R.A. y Patrick, E.A. 1973, "Clustering Using a Similarity Measure Based on Shared Near Neighbors", *IEEE Transactions on Computers*, pp. 1025-1034.

Jing, H. 2000, "Sentence reduction for automatic text summarization", en *Proceedings of the First Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-2000)*.

Jing, H. y McKeown, K.R. 2000, "Cut and Paste-Based Text Summarization", en *Proceedings of the 6th Applied Natural Language Processing Conference and the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 178-185.

Jing, H., Barzilay, R., McKeown, K. y Elhadad, M. 1998, "Summarization evaluation methods: Experiments and analysis", en *Working Notes of the AAAI-98 Spring Symposium on Intelligent Text Summarization*, pp. 60-68.

Joachims, T. 1997, "A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization", en *Proceedings of the 14th International Conference on Machine Learning*, pp.143-151.

Joachims, T. 1998, "Text categorization with support vector machines: Learning with many relevant features", en *Proceedings of the European Conference on Machine Learning*, pp. 137-142.

Johnson, T. 1993, *Re: WWW Information Discovery Tools*, [Online], University of Calgary, Disponible en: `<http://ksi.cpsc.ucalgary.ca/archives/WWW-TALK/www-talk-1993q2.messages/30.html>` [30 Junio 2005].

Kando, N. 1999, "Preface", en *Proceedings of the First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition*, [Online], Disponible en: `<http://research.nii.ac.jp/ntcir/workshop/ OnlineProceedings>` [30 Junio 2005]

Kantor, P.B., Boros, E., Melamed, B., Menkov, V., Shapira, B. y Neu, D.J. 2000, "Capturing human intelligence in the Net", *Communications of the ACM*, vol. 43, no. 8, pp. 112-115.

Kantor, P.B., Voorhees, E.M. 2000, "The TREC-5 Confusion Track: Comparing Retrieval Methods for Scanned Text", *Information Retrieval*, vol. 2, no. 3, pp. 165-176.

Karvounarakis, G., Christophides, V., Plexousakis, D. y Alexaki, S. 2001, "Querying RDF Descriptions for Community Web Portals", *The French National Conference on Databases*.

Karypis, G. y Han, E. 2000, *Concept indexing: A fast dimensionality reduction algorithm with applications to document retrieval and categorization*, Informe técnico, University of Minnesota.

Kaufman, L. 1998, "Solving the Quadratic Programming Problem Arising in Support Vector Classification", en *Advances in Kernel Methods - Support Vector Learning*, pp. 147-167. Citado por Platt, J.C. 1999, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines", *NIPS 11*, pp. 557-563.

Kearns, M. 1988, *Thoughts on Hypothesis Boosting*, manuscrito no publicado, [Online], Disponible en: <http://www.cis.upenn.edu/~mkearns/papers/boostnote.pdf> [30 Junio 2005].

Kessler, B. 1995, "Computational Dialectology in Irish Gaelic", en *Proceedings of the European Association for Computational Linguistics*, pp. 60-67.

King, B. 1963, "Step-wise clustering procedures", *Journal of the American Statistical Association*, 69, pp. 86-101.

Kleinberg, J.M. 1998, "Authoritative sources in a hyperlinked environment", en *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pp. 668-677.

Knight, K. y Marcu, D. 2000, "Statistics-based summarization – step one: Sentence compression", en *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI)*, pp. 703-710.

Koehn, P., *Europarl: A Multilingual Corpus for Evaluation of Machine Translation*, manuscrito no publicado, [Online], Disponible en: <http://people.csail.mit.edu/~koehn/publications/europarl.ps> [30 Junio 2005]

Kohonen, T. 1982, "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*, 43, pp. 59-69.

Kolda, T.G. y O'Leary, D.P. 1998, "A Semidiscrete Matrix Decomposition for Latent Semantic Indexing in Information Retrieval", *ACM Trans. Information Systems*, vol. 16, pp. 322-346.

Koller, D. y Sahami, M. 1997, "Hierarchically classifying documents using very few words", en *Proceedings of ICML-97, 14th International Conference on Machine Learning*, pp. 170-178.

Koller, D. y Sahami, M. 1997, "Hierarchically classifying documents using very few words", en *Proc. of the 14th Int. Conf. on Machine Learning*, pp. 170-178.

Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. y Riedl, J. 1997, "GroupLens: Applying Collaborative Filtering to Usenet News", *Communications of the ACM*, vol. 40, no. 3, pp. 77-87.

Koster, M. 1994, "ALIWEB – Archie-Like Indexing in the WEB", *Computer Networks and ISDN Systems*, vol. 27, no. 2, pp. 175-182.

Kovács, L. y Micsik, A. 2000, "The Collaborative Web", *ERCIM News*, no. 41.

Kraaij, W., Spitters, M. y Hulth, A. 2002, "Headline extraction based on a combination of uni- and multidocument summarization techniques", en *Proceedings of the ACL workshop on Automatic Summarization/Document Understanding Conference (DUC 2002)*.

Kraaij, W., Spitters, M. y van der Heijden, M. 2001, "Combining a mixture language model and Naive Bayes for multi-document summarisation", en *Proceedings of the DUC 2001 workshop (SIGIR 2001)*.

Kraft, D.H. y Buell, D.A. 1983, "Fuzzy sets and generalized Boolean retrieval systems", *International Journal of Man-Machine Studies*, vol. 19, pp. 45-56.

Kruijff, G.M., 2002, *Learning linearization rules from treebanks*, [Online], Disponible en: <http://www.cs.haifa.ac.il/~shuly/fg02/kruijff.pdf> [30 Junio 2005].

Kuhner, M.K. y Felsenstein, J. 1994, "A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates", *Molecular Biology and Evolution*, vol. 11, no. 3, pp. 459-468.

Kupiec J., Pedersen J. y Chen F. 1995, "A Trainable Document Summarizer", en *Proceedings of the ACM SIGIR conference*, pp. 68-73.

Kwok, J.T.Y. 1998, "Automated text categorization using support vector machine", en *Proceedings of the International Conference on Neural Information Processing (ICONIP)*, pp. 347-351.

Lal, P. y Ruger, S. 2002, "Extract-based Summarization with Simplification", en *Proceedings of the ACL Workshop on Text Summarisation / DUC 2002*.

Lam, S. y Lee, D. 1999, "Feature reduction for neural network based text categorization", en *6th Conference On Database Systems For Advanced Applications*, pp. 195-202.

Lancaster, F.W. 1968, *Information Retrieval Systems: Characteristics, Testing and Evaluation*, Wiley, EE.UU. (citado por van Rijsbergen, C.J. 1979, *Information Retrieval, 2nd Edition*, Butterworth-Heinemann, EE.UU.)

Langer, S. 2001, "Natural languages on the Word Wide Web", [Online], en *Bulag. Revue annuelle*, pp. 89-100, disponible en: `<http://www.cis.uni-muenchen.de/people/langer/veroeffentlich ungen/bulag.pdf>` [30 Junio 2005]

Larkey, L.S. y Croft, W.B. 1996, "Combining classifiers in text categorization", en *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 289-297.

Larocca Neto, J., Santos, A.D., Kaestner, C.A.A. y Freitas, A.A. 2000, "Document Clustering and Text Summarization", en *Proceedings of the 4th International Conference Practical Applications of Knowledge Discovery and Data Mining*, pp. 41-55.

Lawrence, S. y Giles, C.E. 1998, "Searching the World Wide Web", *Science*, vol. 280, no. 3, pp. 98-100.

Lees, R.B. 1953, "On the basis of glottochronology", *Language*, 29, pp. 113-127.

Lenci, A., Bartolini, R., Calzolari, N., Agua, A., Busemann, S., Cartier, E., Chevreau, K. y Coch, J. "Multilingual summarization by integrating linguistic resources in the mlis-musi project", en *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC'02)*.

Lenhart, A., Fallows, D. y Horrigan, J. 2004, *Content Creation Online*, Informe, Pew Internet & American Life Project.

Lerman, I.C. 1970, *Les Bases de la Classification Automatique*, Gauthier-Villars, Paris. Citado en van Rijsbergen, C.J. 1979, *Information Retrieval, 2nd Edition*. Butterworth-Heinemann, EE.UU.

Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals, (traducción inglesa del ruso), *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707-710.

Lewis, D. y Catlett, J. 1994. "Heterogeneous uncertainty sampling for supervised learning", en *Proceedings of the 11th International Conference on Machine Learning*, pp. 148-156.

Lewis, D.D. 1991, "Evaluating text categorization", en *Proceedings of Speech and Natural Language Workshop*, pp. 312-318.

Lewis, P.A.W, Baxendale, P.B. y Bennett, J.L. 1967, "Statistical Discrimination of the Synonymy/Antonymy Relationship Between Words", *Journal of the ACM*, vol. 14, no. 1, pp. 20-44.

Li, H. y Yamanishi, K. 1999, "Text classification using esc-based stochastic decision lists", en *8th ACM International Conference on Information and Knowledge Management (CIKM99)*, pp. 122-130.

Li, M., Chen, X., Li, X., Ma, B. y Vitányi, P. 2004, "The Similarity Metric", *IEEE Transactions on Information Theory*, vol. 50, pp. 3250-3264.

Li, W.S., Vu, Q., Chang, E., Agrawal, D., Hara, Y. y Takano, H. 1999, "PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management", en *Proceedings of the Eighth International World-Wide Web Conference*.

Lieberman, H. 1995, "Letizia: An Agent That Assists Web Browsing", en *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp. 924-929.

Lin, C.Y. 1999, "Machine Translation for Information Access across the Language Barrier: the MuST System", en *Proceedings of the Machine Translation Summit VII*.

Lin, C.Y. 2003, "Improving Summarization Performance by Sentence Compression – A Pilot Study", en *Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages*.

Lin, C.Y. 2004a, "Rouge: A Package for Automatic Evaluation of Summaries", en *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pp. 74-81.

Lin, C.Y. 2004b, "Looking for a Few Good Metrics: Automatic Summarization Evaluation – How Many Samples Are Enough?", en *Proceedings of NTCIR Workshop 4*.

Lin, C.Y. y Hovy, E.H. 1997, "Identifying Topics by Position", en *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP)*.

Lin, C.Y. y Hovy, E.H. 2000, "The automated acquisition of topic signatures for text summarization", en *Proceedings of the 18th COLING Conference*.

Lin, C.Y. y Hovy, E.H. 2001, "NeATS: A multidocument summarizer", en *Proceedings of the Document Understanding Conference (DUC01)*.

Lin, C.Y. y Hovy, E.H. 2002a, "NeATS in DUC 2002", en *Proceedings of the Document Understanding Conference (DUC02)*.

Lin, C.Y. y Hovy, E.H. 2002b, "Manual and Automatic Evaluation of Summaries", en *Proceedings of the Workshop on Automatic Summarization (including DUC 2002)*, pp. 45-51.

Lin, C.Y. y Hovy, E.H. 2003, "Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics", en *Proceedings of the Human Technology Conference 2003 (HLT-NAACL-2003)*.

Lin, C.Y. y Hovy, E.H. 2003, "Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics", en *Proceedings of the Human Technology Conference 2003 (HLT-NAACL-2003)*.

Lotka, A.J. 1926, "The frequency distribution of scientific productivity", en *The origins of information science*, 1987, ed. A.J. Meadows, Taylor Graham Publishing.

Luhn, H.P. 1957, "A Statistical Approach to Mechanized Encoding and Searching of Literary Information", *IBM Journal of Research and Development*, vol. 1, no. 4, pp. 309-317.

Luhn, H.P. 1958, "The Automatic Creation of Literature Abstracts", *IBM Journal of Research and Development*, vol. 2, no. 2, pp. 159-165.

Luhn, H.P. 1959, *Keyword-in-context index for technical literature (KWIC)*, IBM Advanced System Development Division, RC-127 (citado por Ohlman, H. 1999, "Mechanical Indexing: A Personal Remembrance", en Bowden, M.E., Bellardo Hahn, T. y Williams, R.V. (Eds.): *Proc. 1998 Conf. History and Heritage of Science Information Systems*, ASIS Monograph Series, Information Today, EE.UU., pp. 187-192).

Luke, S., Spector, L. y Rager, D. 1996, "Ontology-Based Knowledge Discovery on the World-Wide Web", en *Working Notes of the Workshop on Internet-Based Information Systems at the 13th National Conference on Artificial Intelligence (AAAI96)*.

Maarek, Y.S. y Ben Shaul, I.Z. 1996, "Automatically Organizing Bookmarks per Contents", en *Proceedings of the 5th International World Wide Web Conference*.

Maes, P. 1994, "Agents that Reduce Work and Information Overload", *Communications of the ACM*, vol. 37, no. 7, pp. 811-821.

Mani, I. y Maybury, M.T. (Eds.) 1999, *Advances in Automatic Text Summarization*, The MIT Press.

Mani, I., Gates, B. y Bloedorn, E. 1999, "Improving Summaries by Revising Them", en *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pp. 558-565.

Mani, I., House, D., Klein, G., Hirschman, L., Obsrt, L., Firmin, T., Chrzanowski, M. y Sundheim, B. 1998, *The TIPSTER SUMMAC Text Summarization Evaluation: Final Report*, Informe técnico, [Online], Disponible en: <http://www.itl.nist.gov/iaui/894.02/related_projects/tipster_summac/summac-final-report-part2.ps> [30 Junio 2005]

Marchiori, M. y Saarela, J. 1998, "Query + Metadata + Logic = Metalog", en *Proceedings of Query Languages Workshop*.

Marchiori, M. y Saarela, J. 1999a, *Let's Reason on the Web: Metalog*, [Online], Consorcio W3C, Disponible en: <http://www.w3.org/RDF/Metalog/metalog990224.html> [30 Junio 2005].

Marchiori, M. y Saarela, J. 1999a, *Towards the Semantic Web: Metalog*, [Online], Consorcio W3C, Disponible en: <http://www.w3.org/RDF/Metalog/CIKM-050299.html> [30 Junio 2005].

Maron, M.E. y Kuhns, K.L. 1960, "On relevance, probabilistic indexing and information retrieval", *Journal of the ACM*, vol. 7, no. 3, pp. 216-244.

Mauldin, M.L. y Leavitt, J.R.R. 1994, "Web Agent Related Research at the Center for Machine Translation", [Online], en *Proceedings of the ACM Special Interest Group on Networked Information Discovery and Retrieval*, Disponible en: <http://web.archive.org/web/19970607125802/http://fuzine.mt.cs.cmu.edu/mlm/signidr94.html> [30 Junio 2005].

McBryan, O.A. 1994, "GENVL and WWWW: Tools for taming the Web", *Computer Networks and ISDN Systems*, vol. 27, no. 2, p. 308.

McCallum, A.K., Rosenfeld, R., Mitchell, T.M. y Ng, A.Y. 1998, "Improving text classification by shrinkage in a hierarchy of classes", en *Proceedings of ICML-98, 15th International Conference on Machine Learning*, pp. 359-367.

McCulloch, W.S. y Pitts, W. 1943, "A logical calculus of ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, 5, pp. 115-133.

McGuinness, D.L., Fikes, R., Stein, L.A. y Hendler, J. 2000, "DAML-ONT: An Ontology Language for the Semantic Web", en *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, Fensel, D., Hendler, J., Lieberman, H. y Wahlster, W. (eds), pp. 65-94.

McKeown, K.R., Barzilay, R., Blair-Goldensohn, S., Evans, D., Hatzivassiloglou, V., Klavans, J., Nenkova, A., Schiffman, B. y Sigelman, S. 2002, "The Columbia Multi-Document Summarizer", en *Proceedings of the Workshop on Multi-Document Summarization Evaluation of the 2nd Document Understanding Conference at the 40th Meeting of the Association for Computational Linguistics*.

McKeown, K.R., Barzilay, R., Evans, D., Hatzivassiloglou, V., Teufel, S., Kan, Y.M. y Schiffman, B. 2001, "Columbia Multi-Document Summarization: Approach and Evaluation", en *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

McNamee, P. y Mayfield, J. 2003, "JHU/APL Experiments in Tokenization and Non-Word Translation", *Working Notes for the CLEF 2003 Workshop*. 21-22 August, Trondheim, Norway.

McNamee, P. y Mayfield, J. 2004, "Cross-Language Retrieval Using HAIRCUT for CLEF 2004", en *Working Notes for the CLEF 2004 Workshop (Volume I: Papers)*, pp. 31-37.

Menczer, F. y Belew, R.K. 1998, "Adaptive Information Agents in Distributed Textual Environments", en *Proceedings of the 2nd International Conference on Autonomous Agents*, pp. 157-164.

Menczer, F., Belew, R.K. y Willuhn, W. 1995, "Artificial Life Applied to Adaptive Information Agents", en *Proceedings of AAAI Spring Symposium on Information Gathering*.

Merkl, D. y Rauber, A. 1998, "CIA's View of the World and What Neural Networks Learn from It: A Comparison of Geographical Document Space Representation Metaphors", en Quirchmayr, G. *et al.* (Eds.): *Database and Expert Systems Applications, 9th International Conference*, LNCS 1460, pp. 816-825

Miller, L., Seaborne, A. y Reggiori, A. 2002, "Three Implementations of SquishQL, a Simple RDF Query Language", en *Proceedings of 1st International Semantic Web Conference. ISWC2002*.

Miller, M. 1998, *The GnuHoo BooBoo*, [Online], Slashdot, Disponible en: `<http://slashdot.org/ article.pl?sid=98/06/23/0849239&mode=nested&tid=95>` [30 Junio 2005].

Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT). 2001, *Number of Internet Users in CY2000 Reaches 47,080,000*.

Minsky, M. y Papert, S. 1969, *Perceptrons: an introduction to computational geometry*. MIT Press. Citado en Elkan, C. 1997, *Boosting and Naive Bayesian learning*, Informe Técnico, University of California, San Diego.

Mitra, M., Singhal, A. y Buckley, C. 1997, "Automatic text summarization by paragraph extraction", en *Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization*, pp. 31-36.

Mitra, M., Singhal, A. y Buckley, C. 1998, "Improving automatic query expansion", en *ACM SIGIR'98*, pp. 20-27.

Morita, M. y Shinoda, Y. 1994, "Information filtering based on user behavior analysis and best match text retrieval", en *Proceedings of the 17th Annual International Retrieval*, pp. 272-281.

Mosteller, F. y Wallace, D.L. 1964, "Inference and Disputed Authorship: The Federalist Series", en *Behavioral Science: Quantitative Methods*, Citado en Fung, G. 2003, "The Disputed Federalist Papers: SVM and Feature Selection via Concave Minization", [Online], en *Proceedings of the 2003 Conference of Diversity in Computing*, pp. 42-46, disponible en: `<http://www.cs.wisc.edu/~gfung/federalist.pdf>` [30 Junio 2005]

Moulinier, I. y Ganascia, J.G. 1996, "Applying an existing machine learning algorithm to text categorization", en *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, pp. 343-354.

Muthukrishnan, S. y Sahinalp, S.C. 2000, "Approximate nearest neighbors and sequence comparison with block operations", en *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pp. 416-424.

Nenkova, A. y Passonneau, R. 2004, "Evaluating Content Selection in Summarization: The Pyramid Method", en *Proceedings of HLT/NAACL 2004*.

Nenkova, A., Schiffman, B., Schlaiker, A., Blair-Goldensohn, S., Barzilay, R., Sigelman, S., Hatzivassiloglou, V. y McKeown, K.R. 2003, "Columbia at the DUC 2003", en *DUC03*.

Nerbonne, J. y Heeringa, W. 1997, "Measuring Dialect Distance Phonetically", en *Proceedings of the third meeting of the ACL special interest group in computational phonology*, pp. 11-18.

Ng, H., Goh, W. y Low, K. 1997, "Feature selection, perceptron learning, and a usability case study for text categorization", en *Proc. 20th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR'97)*, pp. 67-73.

NIST, 2004. *DUC 2004: Documents, Tasks, and Measures*, [Online], Disponible en: `<http://duc.nist.gov/duc2004/tasks.html>` [30 Junio 2005].

Ohlman, H. 1957, *Permutation Indexing: Multiple-Entry Listing by Electronic Accounting Machines*, Rand Corp. System Development Div., EE.UU.

Ohlman, H. 1999, "Mechanical Indexing: A Personal Remembrance", en Bowden, M.E., Bellardo Hahn, T. y Williams, R.V. (Eds.): *Proc. 1998 Conf. History and Heritage of Science Information Systems*, ASIS Monograph Series, Information Today, EE.UU., pp. 187-192.

Ontrup, J. y Ritter, H. 2001, "Text categorization and semantic browsing with self-organizing maps on non-euclidean spaces", en *Proceedings of PKDD-01*, pp. 338-349.

Osuna, E., Freund, R. y Girosi, F. 1997, *Support vector machines: Training and applications*, informe técnico, MIT.

Page, L., Brin, S., Motwani, R. y Winograd, T. 1998, *The PageRank Citation Ranking: Bringing Order to the Web*, [Online], Stanford University, Disponible en: `<http://dbpubs.stanford.edu/pub/1999-66>` [30 Junio 2005].

Paice, C.D. y Jones, P.A. 1993, "The identification of important concepts in highly structured technical papers", en Korfhage, R., Rasmussen, E. y Willet, P. (Eds.) *Proceedings of the 16th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 69-75.

Pantel, P. y Lin, D. 1998, "SpamCop: A Spam Classification & Organization Program", en *Learning for Text Categorization: Papers from the 1998 Workshop*.

Papineni, K., Roukos, S., Ward, T. y Zhu, W., 2002, "Bleu: a Method for Automatic Evaluation of Machine Translation", en *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 311-318.

Peinado, V., Artiles, J., López-Ostenero, F., Gonzalo, J. y Verdejo, F. 2004, "UNED at Image CLEF 2004: Detecting Named Entities and Noun Phrases for Automatic Query Expansion and Structuring", en Peters, C. y Borri, F. (Eds.): *Working Notes for the CLEF 2004 Workshop*.

Perry, W.M. 1993, *Re: WWW Information Discovery Tools*, [Online], University of Calgary, Disponible en: `<http://ksi.cpsc.ucalgary.ca/archives/WWW-TALK/www-talk-1993q2.messages/27.html>` [30 Junio 2005].

Peters, C. 2001, "Introduction", en Peters, C. (Ed.): *CLEF 2000*, LNCS 2069, pp. 1-6.

Peters, C., Clough, P.D., Jones, G.J.F., Gonzalo, J., Kluck, M. y Magnini, B. (Eds.) 2005, *Multilingual Information Access for Text, Speech and Images: Results of the 5th CLEF Evaluation Campaign*, LNCS 3491.

Pinkerton, B. 1994, "Finding what people want: Experiences with the WebCrawler", [Online], Internet Archive, en *Electronic Proceedings of the "Second World Wide Web Conference '94: Mosaic and the Web"*, NCSA, Disponible en: `<http://archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/pinkerton/WebCrawler.html>` [30 Junio 2005].

Pirkola, A, Keskustalo, Heikki, Leppänen, Erkka, Känsälä, Antti-Pekka and Järvelin, Kalervo, 2002, "Targeted *s*-gram matching: a novel *n*-gram matching technique for cross- and monolingual word form variants", *Information Research*, vol. 7, no. 2.

Platt, J.C. 1998, "How to implement SVMs", *IEEE Intelligent Systems*, vol. 13, no. 4.

Platt, J.C. 1999, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines", *NIPS 11*, pp. 557-563.

Porter, M.F. 1980, "An algorithm for suffix stripping", *Program*, vol. 14, no. 3, pp. 130-137.

Poser, W.J. y Campbell, L. 1992, "Indo-European Practice and Historical Methodology", en *Proceedings of the Eighteenth Annual Meeting of the Berkeley Linguistics Society*, pp. 214-236.

Postel, J. 1975, *On the Junk Mail Problem*, RFC-0706, [Online], disponible en: `<http://www.faqs.org/rfcs/rfc706.html>` [30 Junio 2005]

Poutsma, A. 2002, "Applying Monte Carlo Techniques to Language Identification", en *CLIN 2001 (Twelfth Meeting of Computational Linguistics in the Netherlands)*.

Prager, J.M. 1999, "Linguini: Recognition of Language in Digital Documents", en *Proceedings of the 32nd Hawaii International Conference on System Sciences*.

Prud'hommeaux, E. y Seaborne, A. 2004, *BRQL - A Query Language for RDF*, [Online], Borrador de trabajo, Consorcio W3, Disponible en: `<http://www.w3.org/2001/sw/DataAccess/rq23/>` [30 Junio 2005].

Putz, S. 1993, *Re: WWW Information Discovery Tools*, [Online], University of Calgary, Disponible en: `<http://ksi.cpsc.ucalgary.ca/archives/WWW-TALK/www-talk-1993q2.messages/31.html>` [30 Junio 2005].

Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Çelebi, A., Dimitrov, S., Drabek, E., Hakim, A., Lam, W., Liu, D., Otterbacher, J., Qi, H., Saggion, H., Teufel, S., Topper, M., Winkel, A. y Zhu, Z. "MEAD – A platform for multidocument multilingual text summarization", en *Proceedings of LREC 2004*.

Radev, D., Blair-Goldensohn, S. y Zhang, Z. 2001, "Experiments in single and multi-document summarization using MEAD", en *First Document Understanding Conference*.

Radev, D., Otterbacher, J., Qi, H. y Tam, D. 2003, "MEAD ReDUCs: Michigan at DUC 2003", en *DUC03*.

Radev, D., Teufe, S., Saggion, H., Lam, W., Blitzer, J., Çelebi, A., Qi, H., Drabek, E. y Liu, D. 2002, *Evaluation of text summarization in a cross-lingual information retrieval framework*, [Online], Informe Técnico, Johns Hopkins University, Disponible en: `<http://www.clsp.jhu.edu/ws2001/groups/asmd/jhu01summ-finalreport.ps>` [30 Junio 2005]

Rapp, R. 1999, "Automatic identification of word translations from unrelated English and German corpora", en *Proc. of the 37th Annual Meeting of the ACL*, pp. 1-17.

Rauber, A. y Merkl, D. 1999, "The SOMLib Digital Library System", en *European Conference on Digital Libraries*, pp. 323-342.

Reimer, U. y Hahn, U. 1988, "Text condensation as knowledge base abstraction", en *Proceedings of the Fourth Conference on Artificial Intelligence Applications*, pp. 338–344.

Real Academia Española, 2001, *Diccionario de la Lengua Española*.

Rennie, J. y Rifkin, R. 2001, *Improving multiclass text classification with the support vector machine*, informe técnico, MIT.

Riloff, E. 1995, "Little words can make a big difference for text classification", en *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 130-136.

Robertson, S.E. 2004, "Understanding inverse document frequency: on theoretical arguments for IDF", *Journal of Documentation*, vol. 60, pp. 503-520.

Robertson, S.E. y Spärck-Jones, K. 1976, "Relevance weighting of search terms", *Journal of the ASIS*, vol. 27, no. 3, pp. 129-146.

Robertson, S.E., Walker, S., Jones, S., Hancock-Beaulieu, M. y Gatford, M. 1994, "Okapi at TREC-2", en *Text REtrieval Conference*, pp. 21-34.

Rocchio, J.J. 1966, *Document retrieval systems - optimization and evaluation*, Disertación doctoral, Harvard University.

Rocchio, J.J. 1971, "Relevance Feedback in Information Retrieval", en Salton, G. (Ed.): *The SMART Retrieval System: Experiments in Automatic Document Processing*, pp. 313-323. Prentice-Hall.

Rogati, M. y Yang, Y. 2001, "Cross-Lingual Pseudo-Relevance Feedback using a Comparable Corpus: CLEF Working Notes", *CLEF 2001*, pp. 151-157.

Rosenblatt, F. 1958, "The Perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, 65, pp. 386-408.

Roussinov, D. y Chen, H. 1998, "A scalable self-organizing map algorithm for textual classification: A neural network approach to thesaurus generation", *Communication and Cognition*, vol. 15, no. 1-2, pp. 81-112.

Rubner, Y., Tomasi, C. y Guibas, L.J. 2000, "The Earth Mover's Distance as a Metric for Image Retrieval", *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99-121.

Rucker, J. y Marcos, J.P. 1997, "Siteseer: Personalized Navigation for the Web", *Communications of the ACM*, vol. 40, no. 3, pp. 73-75.

Rudman, J. 1998, "The State of Authorship Attribution Studies: SomeProblems and Solutions", *Computers and the Humanities*, 31, pp. 351-365.

Ruiz, M. E. y Srinivasan, P. 1997, "Automatic Text Categorization Using Neural Networks", en *Proceedings of the 8th ASIS SIG/CR Classification Research Workshop*, pp. 59-72.

Ruiz, M.E. y Srinivasan, P. 1999, "Hierarchical neural networks for text categorization", en *Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99)*, pp. 281-282.

Saggion, H. y Gaizauskas, R. 2004, "Multi-Document Summarization by Cluster/Profile Relevance and Redundancy Removal", *DUC 2004*.

Sahami, M., Dumais, S., Heckerman, D. y Horvitz, E. 1998, "A bayesian approach to filtering junk e-mail", en *AAAI-98 Workshop on Learning for Text Categorization*.

Sahr, R. 2004, *Inflation Conversion Factors for Dollars 1665 to Estimated 2014*, [Online], Oregon State University, Disponible en: <http://oregonstate.edu/dept/pol_sci/fac/sahr/sahr.htm> [30 Junio 2005].

Salton, G. 1968, *Automatic Content Analysis in Information Retrieval*, Informe técnico, Cornell University.

Salton, G. 1968, *Automatic Information Organisation and Retrieval,* McGraw-Hill. Citado en Jardine, N. y van Rijsbergen, C.J. 1971, "The use of hierarchic clustering in information retrieval", *Information Storage and Retrieval*, vol. 7, pp. 217-240.

Salton, G. 1972, *Automatic processing of current afairs queries*, Informe Técnico, Cornell University.

Salton, G. y Crouch, D.B. 1989, *User-System Interaction in Automatic Information Retrieval*, Informe Técnico, Cornell University.

Salton, G. y Lesk, M.E. 1965, "The SMART Automatic Document Retrieval System – An Illustration", *Communications of the ACM*, vol. 8, no. 6, pp. 391-398.

Salton, G. y Singhal, A. 1994, *Automatic text theme generation and the analysis of text structure*, Informe Técnico, Cornell University.

Salton, G., Fox, E.A. y Wu, H. 1983, "Extended Boolean information retrieval", *Communications of the ACM*, vol. 26, no. 11, pp. 1022-1036.

Salton, G., Singhal, A., Buckley, C. and Mitra, M. 1996, "Automatic Text Decomposition Using Text Segments and Text Themes", en *Proceedings of the Seventh ACM Conference on Hypertext*, pp.53-65.

Salton, G., Singhal, A., Mitra, M. y Buckley, C. 1997, "Automatic Text Structuring and Summarization", *Information Processing and Management*, vol. 33, no. 2, pp. 193-207.

Santos Jr., E., Mohamed, A.A. y Zhao, Q. 2004, "Automatic Evaluation of Summaries Using Document Graphs", en *Proceedings of the ACL-04 Workshop*, pp. 66-73.

Schapire, R.E. 1990, "The strength of weak learnability", *Machine Learning*, vol. 5, no. 2, pp. 197-227.

Schapire, R.E. y Singer, Y. 2000, "BoosTexter: a boosting-based system for text categorization", *Machine Learning*, vol. 39, no. 2-3, pp. 135-168.

Schapire, R.E., Singer, Y. y Singhal, A. 1998, "Boosting and Rocchio applied to text filtering", en *Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval*, pp. 215-223.

Schohn, G. y Cohn, D. 2000, "Less is more: Active learning with support vector machines", en *Proceedings of the 17th International Conference on Machine Learning*, pp. 839-846.

Schütze, H., Hull, D.A. y Pedersen, J.O. 1995, "A comparison of classifiers and document representations for the routing problem", en *Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval*, pp. 229-237.

Seaborne, A. 2004, *RDQL - A Query Language for RDF*, [Online], Propuesta al W3C, Hewlett-Packard, Disponible en: <http://www.w3.org/Submission/RDQL/> [30 Junio 2005].

Sebastiani, F. 2002, "Machine learning in automated text categorization", *ACM Computing Surveys*, vol. 34, no. 1, pp. 1-47.

Selberg, E. y Etzioni, O. 1995, "Multi-Service Search and Comparison Using the MetaCrawler", en *Proceedings of the 4th World Wide Web Conference*.

Shardanand, U. y Maes, P. 1995, "Social Information Filtering: Algorithms for Automating «Word of Mouth»", en *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, pp. 210-217.

Sibun, P. y Reynar, J.C. 1996, "Language Identification: Examining the Issues", en *5th Symposium on Document Analysis and Information Retrieval*, pp. 125-135.

Silverstein, C., Henzinger, M., Marais, H. y Moricz, M. 1998, *Analysis of a Very Large AltaVista Query Log*, Informe técnico, Digital Systems Research Center.

Singhal, A, Mitra, M. y Buckley, C. 1997, "Learning routing queries in a query zone", en *Proceedings of SIGIR'97*, pp. 25-32.

Singhal, A., Buckley, C. y Mitra, M. 1996, "Pivoted Document Length Normalization", en *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 21-29.

Smalheiser, N.R. y Swanson, D.R. 1998, "Using Arrowsmith: a computer-assisted approach to formulating and assessing scientific hypotheses", *Computer Methods and Programs in Biomedicine*, vol. 57, pp. 149-153.

Sneath, P.H.A. y Sokal, R.R. 1973, *Numerical Taxonomy*, Freeman, Londres.

Soergel, D. 1999. "The Rise of Ontologies or the Reinvention of Classification", *Journal of the American Society for Information Science*, vol. 50, no. 12, pp. 1119-1120.

Spärck-Jones, K. 1972, "A statistical interpretation of term specificity and its application in retrieval", *Journal of Documentation*, vol. 28, no. 1, pp. 11-21.

Spärck-Jones, K. 1974, "Automatic indexing", *Journal of Documentation*, vol. 30, no. 4, pp. 393-432. Citado en Rasmussen, E. 2002, "Evaluation in Information Retrieval", en *Proceedings of the ISMIR 2002*, pp. 45-49.

Spärck-Jones, K. 1979, "Experiments in relevance weighting of search terms", *Information Processing and Management*, vol. 15, no. 3, pp. 133-144.

Spärck-Jones, K. 1981, *Information Retrieval Experiment*, Butterworth-Heinemann, Reino Unido.

Spärck-Jones, K. 1999, "Automatic summarizing: Factors and directions", en Mani, I. y Maybury, M. (Eds.), *Advances in automatic text summarization*, pp. 2-12.

Spärck-Jones, K. y van Rijsbergen, C.J. 1975, *Report on the Need for and Provision of an "Ideal" Information Retrieval Test Collection*, Informe técnico, University of Cambridge.

Spärck-Jones, K. y Webster, C.A. 1979, *Research in Relevance Weighting*, Informe técnico, University of Cambridge.

Spärck-Jones, K., van Halteren, H., Moens, M-F., Lapalme, G., Radev, D., Dorr, B., Over, P., Hovy, E., McKeown, K. y Harman, D. 2004, *Road Map for DUC 05-DUC 07 Automatic Summarising Evaluation Programme*, [Online], Disponible en: <http://www-nlpir.nist.gov/projects/duc/RM0507/rm.html> [30 Junio 2005].

Speyer, B. y Allen, W. 1994, *New Directory Services*, [Online], Google Inc., Disponible en: <http://groups.google.com/groups?selm=9401211534.AA26997%40faith.mcc.com&output=gplain> [30 Junio 2005].

Staab, S., Erdmann, M., Mädche, A. y Decker, S. 2000, "An extensible approach for Modeling Ontologies in RDF(S)", en *Proceedings of the ECDL-2000 Workshop "Semantic Web: Models, Architectures and Management"*.

Stamatos, E., Fakotakis, N. y Kokkinakis, G. 2001, "Computer-based Authorship Attribution without lexical Measures", [Online], *Computers and the Humanities*, vol. 35, no. 2, pp. 193-214, disponible en: <http://slt.wcl.ee.upatras.gr/papers/stamatatos6.pdf> [30 Junio 2005]

Steinbach, M., Karypis, G. y Kumar, V. 2000, *A Comparison of Document Clustering Techniques*, Informe Técnico, Universidad de Minnesota.

Steinberg, S.G. 1996, "Seek and Ye Shall Find (Maybe)", *Wired*, [Online], vol. 4, no. 5, Disponible en: <http://www.wired.com/wired/archive/4.05/indexweb.html> [30 Junio 2005].

Stevens, M.E. 1970, *Automatic Indexing: A state-of-the-art report*, US National Bureau of Standards, Washington.

Stiles, H.E. 1961, "The association factor in information retrieval", *Journal of the ACM*, vol. 8, pp. 271-279 (citado por van Rijsbergen, C.J. 1979, *Information Retrieval, 2nd Edition*. Butterworth-Heinemann, EE.UU.)

Swadesh, M. 1950, "Salish internal relationships", *International Journal of American Linguistics*, 16, pp. 157-167.

Swanson, D.R. 1977, "Information retrieval as a trial-and-error process", *Library Quarterly*, vol. 47, no. 2.

Swanson, D.R. 1986, "Undiscovered public knowledge", *Library Quarterly*, vol. 56, no. 2, pp. 103-118.

Swanson, D.R. 1991, "Complementary structures in disjoint science literatures", en Bookstein, A., Chiaramella, Y., Salton, G. y Raghavan, V.V. (Eds.): *SIGIR'91*, pp. 280-289.

Swanson, D.R. y Smalheiser, N.R. 1997, "An interactive system for finding complementary literatures: A stimulus to scientific discovery", *Artificial Intelligence*, vol. 91, no. 2, pp. 183-203.

Tait, J.I. 1983, *Automatic Summarizing of English Texts*, disertación doctoral, University of Cambridge.

Tombros, A., Villa, R. y van Rijsbergen, C.J. 2002, "The effectiveness of query-specific hierarchic clustering in information retrieval", *Information Processing and Management: an International Journal*, vol. 38, no. 4, pp. 559-582.

Tong, S. y Koller, D. 2000, "Support vector machine active learning with applications to text classification", en *Proceedings of the 17th International Conference on Machine Learning*, pp. 999- 1006.

U.S. Department of Commerce, 2002. *A Nation Online: How Americans are Expanding Their Use of the Internet*, Washington, D.C.

Ursing, B.M. y Arnason, U. 1998, "Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade", *Proceedings of the Royal Society of London*. Series B, Biological Sciences, vol. 265, pp. 2251-2255.

Valiant, L. 1984, "A theory of the learnable", *Communications of the ACM*, vol. 27, no. 11, pp. 1134-1142.

van Halteren, H. 2002, "Writing Style Recognition and Sentence Extraction", en *DUC'02 Conference Proceedings*.

van Harmelen, F., Patel-Schneider, P.F. y Horrocks, I. 2001, *Reference Description of the DAML+OIL (March 2001) Ontology Markup Language*, Informe técnico, Disponible en: <http://www.daml.org/2001/03/reference.html> [30 Junio 2005].

van Rijsbergen, C.J. 1977, "A theoretical basis for the use of co-occurrence data in information retrieval", *Journal of Documentation*, vol. 33, pp. 106-119.

van Rijsbergen, C.J. 1979, *Information Retrieval, 2nd Edition*. Butterworth-Heinemann, EE.UU.

Vanderwende, L., Banko, M. y Menezes, A. 2004, "Event-centric summary generation", en *Document Understanding Conference at HLT-NAACL*.

Varré, J.S., Delahaye, J.P. y Rivals, É. 1999, "The Transformation Distance: A Dissimilarity Measure Based on Movements of Segments", *Bioinformatics*, vol. 15, no. 3, pp 194-202.

Verdaguer, P. 1999, *Grammaire de la langue catalane. Les origines de la langue*, Curial.

Wang, L., Fan, W., Yang, R., Xi, W., Luo, M., Zhou, Y. y Fox, E.A. 2004, "Ranking Function Discovery by Genetic Programming for Robust Retrieval", en *Proceedings of the Twelfth TREC Conference*.

Warnow, T., Evans, S.N., Ringe, D. y Nakleh, L. 2004, *Stochastic models of language evolution and an application to the Indo-European family of languages*, Informe técnico, Universidad de California en Berkeley.

Watanabe, Y., Murata, M., Takeuchi, M. y Nagao, M. 1996, "Document Classification Using Domain Specific Kanji Characters Extracted by χ^2 Method", en *Proceedings of the 16th COLING*, pp. 794-799.

Weigend, A.S., Wiener, E.D. y Pedersen, J.O. 1999, "Exploiting Hierarchy in Text Categorization", *Information Retrieval*, vol. 1, no. 3, pp. 193-216.

Weiss, S.M., Apté, C., Damerau, F.J., Johnson, D.E., Oles, F.J., Goetz, T. y Hampp, T. 1999, "Maximizing text-mining performance", *IEEE Intelligent Systems*, vol. 14, no. 4, pp. 63-69.

Whitworth, B. y Whitworth, E. 2004, "Spam and the Social-Technical Gap", [Online], *Computer*, vol. 37, no. 10, pp. 38-45, disponible en: `<http://www.computer.org/computer/homepage/1004/whitworth/rx038.pdf>` [30 Junio 2005]

Wiener, E.D., Pedersen, J.O. y Weigend, A.S. 1995, "A neural network approach to topic spotting", en *Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval*, pp. 317-332.

Wong, S.K.M., Ziarko, W. y Wong, P.C.N. 1985, "Generalized Vector Space Model in Information Retrieval", en *Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 18-25.

Yang, Y. y Chute, C.G. 1994, "An example-based mapping method for text categorization and retrieval", *ACM Transactions on Information Systems*, vol. 12, no. 3. Citado por Han, E., Karypis, G. y Kumar, V. 1999, *Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification*, Informe técnico, Universidad de Minnesota.

Yang, Y. y Liu, X. 1999, "A re-examination of text categorization methods", en *Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval*, pp. 42-49.

Zechner, K. 1996, "Fast generation of abstracts from general domain text corpora by extracting relevant sentences", en *Proceedings of COLING-96*, pp. 986-989.

Zhang, L. y Yao, T. 2003, "Filtering Junk Mail with A Maximum Entropy Model", [Online], en *Proceedings of 20th International Conference on Computer Processing of Oriental Languages (ICCPOL03)*, disponible en: `<http://homepages.inf.ed.ac.uk/s0450736/paper/junk.pdf>` [30 Junio 2005]

Zhang, L., Zhu, J. y Yao, T. 2004, "An evaluation of statistical spam filtering techniques", *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 3, no. 4, pp. 243-269.

Zhao, Y. y Karypis, G. 2002, *Criterion Functions for Document Clustering: Experiments and Analysis*, Informe técnico, Universidad de Minnesota.