

Bibliography

- Abney, S. (1996). *Partial Parsing via Finite-State Cascades*. In Proceedings of Workshop on Robust Parsing, 8th European Summer School in Logic, Language, and Information. Prague, Czech Republic: 8–15.
- ACE (2004). *Annotation Guidelines for Entity Detection and Tracking (EDT)*. <http://www ldc.upenn.edu/Projects/ACE/>.
- Adam, C. K., Ng, H. T., and Chieu, H. L. (2002). *Bayesian Online Classifiers for Text Classification and Filtering*. In Proceedings of SIGIR-02, 25th ACM International Conference on Research and Development in Information Retrieval. Tampere, Finland, ACM Press, New York: 97–104.
- Adams, T. L., Dullea, J., Barrett, T. M., and Grubin, H. (2001). “Technology Issues Regarding the Evolution to a Semantic Web.” *ISAS-SCI* 1: 316–322.
- Aggarwal, C. C., Gates, S. C., and Yu, P. S. (1999). *On the Merits of Building Categorization Systems by Supervised Clustering*. In Proceedings of EDBT-00, 7th International Conference on Extending Database Technology. Konstanz, Germany, ACM Press, New York: 352–356.
- Agrawal, R., Bayardo, R. J., and Srikant, R. (2000). *Athena: Mining-based Interactive Management of Text Databases*. In Proceedings of EDBT-00, 7th International Conference on Extending Database Technology. Konstanz, Germany, Springer-Verlag, Heidelberg: 365–379.
- Agrawal, R., Imielinski, T., and Swami, A. (1993). *Mining Association Rules between Sets of Items in Large Databases*. In Proceedings of the ACM SIGMOD Conference on Management of Data. Washington, DC, ACM Press, New York: 207–216.
- Agrawal, R., and Srikant, R. (1994). *Fast Algorithms for Mining Association Rules*. In Proceedings of the 20th International Conference on Very Large Databases (VLDB-94). Santiago, Chile, Morgan Kaufmann Publishers, San Francisco: 487–499.
- Agrawal, R., and Srikant, R. (1995). *Mining Sequential Patterns*. In Proceedings of the 11th International Conference on Data Engineering. Taipei, Taiwan, IEEE Press, Los Alamitos, CA: 3–14.
- Agrawal, R., and Srikant, R. (2001). *On Integrating Catalogs*. In Proceedings of WWW-01, 10th International Conference on the World Wide Web. Hong Kong, ACM Press, New York: 603–612.
- Ahlberg, C., and Schneiderman, B. (1994). *Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays*. In Proceedings of the International Conference on Computer-Human Interaction. Boston, ACM Press, New York: 313–317.

- Ahlberg, C., and Wistrand, E. (1995). *IVEE: An Information Visualization and Exploration Environment*. In Proceedings of Information Visualization '95 Symposium. Atlanta, GA, IEEE, Los Alamitos, CA: 66–73.
- Aho, A., Hopcroft, J., and Ullman, J. (1983). *Data Structures and Algorithms*. Reading, MA, Addison-Wesley.
- Ahonen-Myka, H. (1999). *Finding All Frequent Maximal Sets in Text*. In Proceedings of the 16th International Conference on Machine Learning, ICML-99 Workshop on Machine Learning in Text Data Analysis. Ljubljana, AAAI Press, Menlo Park, CA: 1–9.
- Ahonen, H., Heinonen, O., Klemettinen, M., and Verkamo, A. (1997a). *Applying Data Mining Techniques in Text Analysis*. Helsinki, Department of Computer Science, University of Helsinki.
- Ahonen, H., Heinonen, O., Klemettinen, M., and Verkamo, A. (1997b). *Mining in the Phrasal Frontier*. In Proceedings of Principles of Knowledge Discovery in Databases Conference. Trondheim, Norway, Springer-Verlag, London.
- Aitken, J. S. (2002). *Learning Information Extraction Rules: An Inductive Logic Programming Approach*. In Proceedings of the 15th European Conference on Artificial Intelligence. Lyon, France, IOS Press, Amsterdam.
- Aizawa, A. (2000). *The Feature Quantity: An Information-Theoretic Perspective of TFIDF-like Measures*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. Athens, ACM Press, New York: 104–111.
- Aizawa, A. (2001). *Linguistic Techniques to Improve the Performance of Automatic Text Categorization*. In Proceedings of NLPRS-01, 6th Natural Language Processing Pacific Rim Symposium. Tokyo, NLPRS, Tokyo: 307–314.
- Al-Kofahi, K., Tyrrell, A., Vachher, A., Travers, T., and Jackson, P. (2001). *Combining Multiple Classifiers for Text Categorization*. In Proceedings of CIKM-01, 10th ACM International Conference on Information and Knowledge Management. Atlanta, ACM Press, New York: 97–104.
- Albert, R., Jeong, H., and Barabasi, A.-L. (1999). “Diameter of the World-Wide Web.” *Nature* **401**: 130–131.
- Alias, F., Iriondo, I., and Barnola, P. (2003). *Multi-Domain Text Classification for Unit Selection Text-to-Speech Synthesis*. In Proceedings of ICPHS-03, 15th International Congress on Phonetic Sciences. Barcelona.
- Allen, J. (1995). *Natural Language Understanding*. Redwood City, CA, Benjamin Cummings.
- Amati, G., and Crestani, F. (1999). “Probabilistic Learning for Selective Dissemination of Information.” *Information Processing and Management* **35**(5): 633–654.
- Amati, G., Crestani, F., and Ubaldini, F. (1997). *A Learning System for Selective Dissemination of Information*. In Proceedings of IJCAI-97, 15th International Joint Conference on Artificial Intelligence. M. E. Pollack, ed. Nagoya, Japan, Morgan Kaufmann Publishers, San Francisco: 764–769.
- Amati, G., Crestani, F., Ubaldini, F., and Nardis, S. D. (1997). *Probabilistic Learning for Information Filtering*. In Proceedings of RIAO-97, 1st International Conference “Recherche d’Information Assistée par Ordinateur.” Montreal: 513–530.
- Amati, G., D’Aloisi, D., Giannini, V., and Ubaldini, F. (1996). *An Integrated System for Filtering News and Managing Distributed Data*. In Proceedings of PAKM-96, 1st International Conference on Practical Aspects of Knowledge Management. Basel, Switzerland, Springer-Verlag, London.
- Amati, G., D’Aloisi, D., Giannini, V., and Ubaldini, F. (1997). “A Framework for Filtering News and Managing Distributed Data.” *Journal of Universal Computer Science* **3**(8): 1007–1021.
- Amir, A., Aumann, Y., Feldman, R., and Fresko, M. (2003). “Maximal Association Rules: A Tool for Mining Associations in Text.” *Journal of Intelligent Information Systems* **25**(3): 333–345.

- Amir, A., Aumann, Y., Feldman, R., and Katz, O. (1997). *Efficient Algorithm for Association Generation*. Department of Computer Science, Bar-Ilan University.
- Anand, S. S., Bell, D. A., and Hughes, J. G. (1995). *The Role of Domain Knowledge in Data Mining*. In Proceedings of ACM CIKM'95. Baltimore, ACM Press, New York: 37–43.
- Anand, T., and Kahn, G. (1993). *Opportunity Explorer: Navigating Large Databases Using Knowledge Discovery Templates*. In Proceedings of the 1993 Workshop on Knowledge Discovery in Databases. Washington, DC, AAAI Press, Menlo Park, CA: 45–51.
- Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., and Spyropoulos, C. D. (2000). *An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-mail Messages*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. Athens, ACM Press, New York: 160–167.
- Aone, C., and Bennett, S. (1995). *Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies*. In Proceedings of Meeting of the Association for Computational Linguistics. Cambridge, MA, Association for Computational Linguistics, Morristown, NJ: 122–129.
- Appelt, D., Hobbs, J., Bear, J., Israel, D., Kameyama, M., Kehler, A., Martin, D., Meyers, K., and Tyson, M. (1993). *SRI International FASTUS System: MUC-6 Test Results and Analysis*. In Proceedings of 16th MUC. Columbia, MD, Association for Computational Linguistics, Morristown, NJ: 237–248.
- Appelt, D., Hobbs, J., Bear, J., Israel, D., Kameyama, M., and Tyson, M. (1993). *FASTUS: A Finite-State Processor for Information Extraction from Real-World Text*. In Proceedings of the 13th International Conference on Artificial Intelligence (IJCAI). Chambéry, France, Morgan Kaufmann Publishers, San Mateo, CA: 1172–1178.
- Appiani, E., Cesarini, F., Colla, A., Diligenti, M., Gori, M., Marinai, S., and Soda, G. (2001). “Automatic Document Classification and Indexing in High-Volume Applications.” *International Journal on Document Analysis and Recognition* 4(2): 69–83.
- Apte, C., Damerau, F., and Weiss, S. (1994a). *Towards Language Independent Automated Learning of Text Categorization Models*. In Proceedings of ACM-SIGIR Conference on Information Retrieval. Dublin, Springer-Verlag, New York: 23–30.
- Apte, C., Damerau, F. J., and Weiss, S. M. (1994b). “Automated Learning of Decision Rules for Text Categorization.” *ACM Transactions on Information Systems* 12(3): 233–251.
- Apte, C., Damerau, F. J., and Weiss, S. M. (1994c). *Towards Language-Independent Automated Learning of Text Categorization Models*. In Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval. Dublin, Springer-Verlag, Heidelberg: 23–30.
- Arning, A., Agrawal, R., and Raghavan, P. (1996). *A Linear Method for Deviation Detection in Large Databases*. In Proceedings of the 2nd International Conference on Knowledge Discovery in Databases and Data Mining. Portland, OR, AAAI Press, Menlo Park, CA: 164–169.
- Ashish, N., and Knoblock, C. A. (1997). *Semi-Automatic Wrapper Generation for Internet Information Sources*. In the Proceedings of the 2nd IFCIS International Conference on Cooperative Information Systems. Charleston, SC, IEEE Press, Los Alamitos, CA: 160–169.
- Attardi, G., Gulli, A., and Sebastiani, F. (1999). *Automatic Web Page Categorization by Link and Context Analysis*. In Proceedings of THAI-99, 1st European Symposium on Telematics, Hypermedia and Artificial Intelligence. Varese, Italy: 105–119.
- Attardi, G., Marco, S. D., and Salvi, D. (1998). “Categorization by Context.” *Journal of Universal Computer Science* 4(9): 719–736.
- Aumann, Y., Feldman, R., Yehuda, Y., Landau, D., Liphstat, O., and Schler, Y. (1999). *Circle Graphs: New Visualization Tools for Text-Mining*. In Proceedings of the 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases, (PKDD-99). Prague, Czech Republic, Springer-Verlag, London: 277–282.

- Avancini, H., Lavelli, A., Magnini, B., Sebastiani, F., and Zanolini, R. (2003). *Expanding Domain-Specific Lexicons by Term Categorization*. In Proceedings of SAC-03, 18th ACM Symposium on Applied Computing. Melbourne, FL, ACM Press, New York: 793–797.
- Azzam, S., Humphreys, K., and Gaizauskas, R. (1998). *Evaluating a Focus-Based Approach to Anaphora Resolution*. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Quebec, Morgan Kaufmann Publishers, San Francisco: 74–78.
- Backer, F. B., and Hubert, L. G. (1976). “A Graphtheoretic Approach to Goodness-of-Fit in Complete-Link Hierarchical Clustering.” *Journal of the American Statistical Association* **71**: 870–878.
- Baeza-Yates, R., and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York, ACM Press.
- Bagga, A., and Biermann, A. W. (2000). *A Methodology for Cross-Document Coreference*. In Proceedings of the 5th Joint Conference on Information Sciences (JCIS 2000). Atlantic City, NJ: 207–210.
- Bairoch, A., and Apweiler, R. (2000). “The Swiss-Prot Protein Synthesis Database and Its Supplement TrEMBL in 2000.” *Nucleic Acids Research* **28**: 45–48.
- Baker, L. D., and McCallum, A. K. (1998). *Distributional Clustering of Words for Text Classification*. In Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval. Melbourne, Australia, ACM Press, New York: 96–103.
- Baldwin, B. (1995). *CogNIAC: A Discourse Processing Engine*. Ph.D. thesis, Department of Computer and Information Sciences, University of Pennsylvania.
- Baluja, S., Mittal, V. O., and Sukthankar, R. (2000). “Applying Machine Learning for High-Performance Named-Entity Extraction.” *Computational Intelligence* **16**(4): 586–596.
- Bao, Y., Aoyama, S., Du, X., Yamada, K., and Ishii, N. (2001). *A Rough Set-Based Hybrid Method to Text Categorization*. In Proceedings of WISE-01, 2nd International Conference on Web Information Systems Engineering. Kyoto, Japan, IEEE Computer Society Press, Los Alamitos, CA: 254–261.
- Bapst, F., and Ingold, R. (1998). “Using Typography in Document Image Analysis.” *Lecture Notes in Computer Science* **1375**: 240–260.
- Barbu, C., and Mitkov, R. (2001). *Evaluation Tool for Rule-Based Anaphora Resolution Methods*. In Proceedings of Meeting of the Association for Computational Linguistics. Toulouse, France, Morgan Kaufmann Publishers, San Mateo, CA: 34–41.
- Basili, R., and Moschitti, A. (2001). *A Robust Model for Intelligent Text Classification*. In Proceedings of ICTAI-01, 13th IEEE International Conference on Tools with Artificial Intelligence. Dallas, IEEE Computer Society Press, Los Alamitos, CA: 265–272.
- Basili, R., Moschitti, A., and Pazienza, M. T. (2000). *Language-Sensitive Text Classification*. In Proceedings of RIAO-00, 6th International Conference “Recherche d’Information Assistée par Ordinateur.” Paris: 331–343.
- Basili, R., Moschitti, A., and Pazienza, M. T. (2001a). *An Hybrid Approach to Optimize Feature Selection Process in Text Classification*. In Proceedings of AI*IA-01, 7th Congress of the Italian Association for Artificial Intelligence. F. Esposito, ed. Bari, Italy, Springer-Verlag, Heidelberg: 320–325.
- Basili, R., Moschitti, A., and Pazienza, M. T. (2001b). *NLP-Driven IR: Evaluating Performances over a Text Classification Task*. In Proceedings of IJCAI-01, 17th International Joint Conference on Artificial Intelligence. B. Nebel, ed. Seattle, IJCAI, Menlo Park, CA: 1286–1291.
- Basu, S., Mooney, R., Pasupuleti, K., and Ghosh, J. (2001). *Evaluating the Novelty of Text-Mined Rules Using Lexical Knowledge*. In Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining (KDD-01). San Francisco, CA, ACM Press, New York: 233–239.
- Batagelj, V. (1997). “Notes on Blockmodeling.” *Social Networks* **19**: 143–155.

- Batagelj, V., Doreian, P., and Ferligoj, A. (1992). "An Optimization Approach to Regular Equivalence." *Social Networks* **14**: 121–135.
- Batagelj, V., Ferligoj, A., and Doreian, P. (1999). "Generalized Blockmodeling." *Informatica* **23**: 501–506.
- Batagelj, V., and Mrvar, A. (2003). *Pajek – Analysis and Visualization of Large Networks*. Graph Drawing Software. Springer-Verlag, Berlin.
- Batagelj, V., Mrvar, A., and Zaversnik, M. (1999). *Partitioning Approach to Visualization of Large Networks*. Graph Drawing '99. Castle Stirin, Czech Republic.
- Batagelj, V., and Zaversnik, M. (2001). *Cores Decomposition of Networks*. Presented at Recent Trends in Graph Theory, Algebraic Combinatorics, and Graph Algorithms. Bled, Slovenia. <http://vlado.fmf.uni-lj.si/pub/networks/doc/cores/pCores.pdf>.
- Bayer, T., Kressel, U., Mogg-Schneider, H., and Renz, I. (1998). "Categorizing Paper documents. A Generic System for Domain and Language-Independent Text Categorization." *Computer Vision and Image Understanding* **70**(3): 299–306.
- Becker, B. (1998). *Visualizing Decision Table Classifiers*. In Proceedings of IEEE Information Visualization (InfoVis '98). North Carolina, IEEE Computer Society Press, Washington, DC: 102–105.
- Beeferman, D., Berger, A., and Lafferty, J. D. (1999). "Statistical Models for Text Segmentation." *Machine Learning* **34**(1–3): 177–210.
- Beil, F., and Ester, M. (2002). *Frequent Term-Based Text Clustering*. In Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD) 2002. Edmonton, Canada, ACM Press, New York: 436–442.
- Bekkerman, R., El-Yaniv, R., Tishby, N., and Winter, Y. (2001). *On Feature Distributional Clustering for Text Categorization*. In Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval. New Orleans, ACM Press, New York: 146–153.
- Bel, N., Koster, C. H., and Villegas, M. (2003). *Cross-Lingual Text Categorization*. In Proceedings of ECDL-03, 7th European Conference on Research and Advanced Technology for Digital Libraries. Trodheim, Norway, Springer-Verlag, Heidelberg: 126–139.
- Benkhalifa, M., Bensaid, A., and Mouradi, A. (1999). *Text Categorization Using the Semi-Supervised Fuzzy C-means Algorithm*. In Proceedings of NAFIPS-99, 18th International Conference of the North American Fuzzy Information Processing Society. New York, IEEE Press, New York: 561–565.
- Benkhalifa, M., Mouradi, A., and Bouyakhf, H. (2001a). "Integrating External Knowledge to Supplement Training Data in Semi-Supervised Learning for Text Categorization." *Information Retrieval* **4**(2): 91–113.
- Benkhalifa, M., Mouradi, A., and Bouyakhf, H. (2001b). "Integrating WordNet Knowledge to Supplement Training Data in Semi-Supervised Agglomerative Hierarchical Clustering for Text Categorization." *International Journal of Intelligent Systems* **16**(8): 929–947.
- Bennett, P. N. (2003). *Using Asymmetric Distributions to Improve Text Classifier Probability Estimates*. In Proceedings of SIGIR-03, 26th ACM International Conference on Research and Development in Information Retrieval. Toronto, ACM Press, New York: 111–118.
- Bennett, P. N., Dumais, S. T., and Horvitz, E. (2002). *Probabilistic Combination of Text Classifiers Using Reliability Indicators: Models and Results*. In Proceedings of SIGIR-02, 25th ACM International Conference on Research and Development in Information Retrieval. Tampere, Finland, ACM Press, New York: 207–214.
- Berendt, B., Hotho, A., and Stumme, G. (2002). *Towards Semantic Web Mining*. In Proceedings of the International Semantic Web Conference (ISWC02). Sardinia, Italy, Springer, Berlin/Heidelberg: 264–278.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). "The Semantic Web." *Scientific American*, May 2001. <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>.
- Berry, M. (1992). "Large-Scale Sparse Singular Value Computations." *International Journal of Supercomputer Applications*. **6**(1): 13–49.

- Bettini, C., Wang, X., and Jojodia, S. (1996). *Testing Complex Temporal Relationships Involving Multiple Granularities and Its Application to Data Mining*. In Proceedings of the 15th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS-96). Montreal, Canada, ACM Press, New York: 68–78.
- Biebricher, P., Fuhr, N., Knorz, G., Lustig, G., and Schwantner, M. (1988). *The Automatic Indexing System AIR/PHYS. From Research to Application*. In Proceedings of SIGIR-88, 11th ACM International Conference on Research and Development in Information Retrieval. Y. Chiaramella, ed. Grenoble, France, ACM Press, New York: 333–342.
- Bigi, B. (2003). *Using Kullback–Leibler Distance for Text Categorization*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Berlin/Heidelberg: 305–319.
- Bikel, D. M., Miller, S., Schwartz, R., and Weischedel, R. (1997). *Nymble: A High-Performance Learning Name-Finder*. In Proceedings of ANLP-97. Washington, DC, Morgan Kaufmann Publishers, San Francisco: 194–201.
- Bikel, D. M., Schwartz, R. L., and Weischedel, R. M. (1999). “An Algorithm that Learns What’s in a Name.” *Machine Learning* **34**(1–3): 211–231.
- Blake, C., and Pratt, W. (2001). *Better Rules, Fewer Features: A Semantic Approach to Selecting Features from Text*. In Proceedings of the 2001 IEEE International Conference on Data Mining. San Jose, CA, IEEE Computer Society Press, New York: 59–66.
- Blanchard, J., Guillet, F., and Briand, H. (2003). *Exploratory Visualization for Association Rule Rummaging*. In Proceedings of the 4th International Workshop on Multimedia Data Mining MDM/KDD2003. Washington, DC, ACM Press, New York: 107–114.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* **3**: 993–1022.
- Bloedorn, E., and Michalski, R. S. (1998). “Data-Driven Constructive Induction.” *IEEE Intelligent Systems* **13**(2): 30–37.
- Blosseville, M. J., Hebrail, G., Montell, M. G., and Penot, N. (1992). *Automatic Document Classification: Natural Language Processing and Expert System Techniques Used Together*. In Proceedings of SIGIR-92, 15th ACM International Conference on Research and Development in Information Retrieval. Copenhagen, ACM Press, New York: 51–57.
- Blum, A., and Mitchell, T. M. (1998). *Combining Labeled and Unlabeled Data with Co-Training*. *COLT*. Madison, WI, ACM Press, New York: 92–100.
- Bod, R., and Kaplan, R. (1998). *A Probabilistic Corpus-Driven Model for Lexical-Functional Analysis*. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Montreal, Morgan Kaufmann Publishers, San Francisco: 145–151.
- Bonacich, P. (1972). “Factoring and Weighting Approaches to Status Scores and Clique Identification.” *Journal of Mathematical Sociology* **2**: 113–120.
- Bonacich, P. (1987). “Power and Centrality: A Family of Measures.” *American Journal of Sociology* **92**: 1170–1182.
- Bonnema, R., Bod, R., and Scha, R. (1997). *A DOP Model for Semantic Interpretation*. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics. Somerset, NJ, Morgan Kaufmann Publishers, San Francisco: 159–167.
- Borgatti, S. P., and Everett, M. G. (1992). “Notions of Positions in Social Network Analysis.” In *Sociological Methodology*, P. V. Marsden, ed. San Francisco, Jossey Bass: 1–35.
- Borgatti, S. P., and Everett, M. G. (1993). “Two Algorithms for Computing Regular Equivalence.” *Social Networks* **15**: 361–376.
- Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2002). *Ucinet 6 for Windows*, Cambridge, MA, Harvard: Analytic Technologies. <http://www.analytictech.com>.
- Borko, H., and Bernick, M. (1963). “Automatic Document Classification.” *Journal of the Association for Computing Machinery* **10**(2): 151–161.
- Borko, H., and Bernick, M. (1964). “Automatic Document Classification. Part II: Additional Experiments.” *Journal of the Association for Computing Machinery* **11**(2): 138–151.

- Borner, K., Chen, C., and Boyack, K. (2003). "Visualizing Knowledge Domains." *Annual Review of Information Science and Technology* **37**: 179–255.
- Borthwick, A. (1999). *A Maximum Entropy Approach for Named Entity Recognition*. Computer Science Department, New York University.
- Brachman, R., and Anand, T. (1996). In "The Process of Knowledge Discovery in Databases: A Human Centered Approach." *Advances in Knowledge Discovery and Data Mining*. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds. Menlo Park, CA, AAAI Press and MIT Press: 37–58.
- Brachman, R., Selfridge, P., Terveen, L., Altman, B., Borgida, A., Halper, F., Kirk, T., Lazar, A., McGuinness, D., and Resnick, L. (1993). "Integrated Support for Data Archeology." *International Journal of Intelligent and Cooperative Information Systems*. **2**(2): 159–185.
- Bradley, P. S., Fayyad, U., and Reina, C. (1998). *Scaling Clustering Algorithms to Large Databases*. In Proceedings of the Knowledge Discovery and Data Mining Conference (KDD '98). New York, AAAI Press, Menlo Park, CA: 9–15.
- Brank, J., Grobelnik, M., Milic-Frayling, N., and Mladenic, D. (2002). *Feature Selection Using Support Vector Machines*. In Proceedings of the 3rd International Conference on Data Mining Methods and Databases for Engineering, Finance, and Other Fields. Bologna, Italy.
- Brill, E. (1992). *A Simple Rule-Based Part of Speech Tagger*. In Proceedings of the 3rd Annual Conference on Applied Natural Language Processing. Trento, Italy, Morgan Kaufmann Publishers, San Francisco: 152–155.
- Brill, E. (1995). "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging." *Computational Linguistics* **21**(4): 543–565.
- Brin, S. (1998). *Extracting Patterns and Relations from the World Wide Web*. In Proceedings of WebDB Workshop, EDBT '98. Valencia, Spain, Springer, Berlin: 172–183.
- Brown, R. D. (1999). *Adding Linguistic Knowledge to a Lexical Example-Based Translation System*. In Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99). Chester, UK: 22–32.
- Bruckner, T. (1997). *The Text Categorization System TEKLIS at TREC-6*. In Proceedings of TREC-6, 6th Text Retrieval Conference. Gaithersburg, MD, National Institute of Standards and Technology, Gaithersburg, MD: 619–621.
- Cai, L., and Hofmann, T. (2003). *Text Categorization by Boosting Automatically Extracted Concepts*. In Proceedings of SIGIR-03, 26th ACM International Conference on Research and Development in Information Retrieval. Toronto, ACM Press, New York: 182–189.
- Caldon, P. (2003). *Using Text Classification to Predict the Gene Knockout Behaviour of S. Cerevisiae*. In Proceedings of APBC-03, 1st Asia-Pacific Bioinformatics Conference. Y.-P. P. Chen, ed. Adelaide, Australia, Australian Computer Society: 211–214.
- Califf, M. E., and Mooney, R. J. (1998). *Relational Learning of Pattern-Match Rules for Information Extraction*. In Working Notes of AAAI Spring Symposium on Applying Machine Learning to Discourse Processing. Menlo Park, CA, AAAI Press, Palo Alto, CA: 6–11.
- Carbonell, J., Cohen, W. W., and Yang, Y. (2000). "Guest Editors' Introduction to the Special Issue on Machine Learning and Information Retrieval." *Machine Learning* **39**(2/3): 99–101.
- Card, S., MacKinlay, J., and Shneiderman, B. (1998). *Readings in Information Visualization: Using Vision to Think*. San Francisco, Morgan Kaufmann Publishers.
- Cardie, C. (1994). *Domain Specific Knowledge Acquisition for Conceptual Sentence Analysis*. Department of Computer Science, University of Massachusetts, Amherst, MA.
- Cardie, C. (1995). "Embedded Machine Learning Systems for Natural Language Processing: A General Framework." In *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*. S. Wermter, E. Riloff, and G. Scheler, eds. Berlin, Springer: 315–328.
- Cardie, C. (1997). "Empirical Methods in Information Extraction." *AI Magazine* **18**(4): 65–80.
- Cardie, C. (1999). "Integrating Case-Based Learning and Cognitive Biases for Machine Learning of Natural Language." *JETAI* **11**(3): 297–337.

- Cardie, C., and Howe, N. (1997). *Improving Minority Class Prediction Using Case-Specific Feature Weights*. In Proceedings of 14th International Conference on Machine Learning. Nashville, TN, Morgan Kaufmann Publishers, San Francisco: 57–65.
- Cardoso-Cachopo, A., and Oliveira, A. L. (2003). *An Empirical Comparison of Text Categorization Methods*. In Proceedings of SPIRE-03, 10th International Symposium on String Processing and Information Retrieval. Manaus, Brazil, Springer-Verlag, Heidelberg: 183–196.
- Carlis, J., and Konstan, J. (1998). *Interactive Visualization of Serial Periodic Data*. In Proceedings of the 11th Annual Symposium on User Interface Software and Technology (UIST '98). San Francisco, ACM Press, New York: 29–38.
- Caropreso, M. F., Matwin, S., and Sebastiani, F. (2001). “A Learner-Independent Evaluation of the Usefulness of Statistical Phrases for Automated Text Categorization.” In *Text Databases and Document Management: Theory and Practice*. A. G. Chin, ed. Hershey, PA, Idea Group Publishing: 78–102.
- Carpineto, C., and Romano, G. (1996). “Information Retrieval through Hybrid Navigation of Lattice Representations.” *International Journal of Human-Computer Studies* 45(5): 553–578.
- Carreras, X., and Marquez, L. (2001). *Boosting Trees for Anti-Spam Email Filtering*. In Proceedings of RANLP-01, 4th International Conference on Recent Advances in Natural Language Processing. Tzigrav Chark, Bulgaria.
- Carroll, G., and Charniak, E. (1992). *Two Experiments on Learning Probabilistic Dependency Grammars from Corpora*. Technical Report CS-92-16.
- Cattoni, R., Coianiz, T., Messelodi, S., and Modena, C. (1998). *Geometric Layout Analysis Techniques for Document Image Understanding: A Review*. Technical Report. Trento, Italy, ITC-IRST I-38050.
- Cavnar, W. B., and Trenkle, J. M. (1994). *N-Gram-Based Text Categorization*. In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval. Las Vegas, UNLV Publications/Reprographics, Las Vegas: 161–175.
- Ceci, M., and Malerba, D. (2003). *Hierarchical Classification of HTML Documents with Web-ClassII*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Berlin: 57–72.
- Cerny, B. A., Okseniuk, A., and Lawrence, J. D. (1983). *A Fuzzy Measure of Agreement between Machine and Manual Assignment of Documents to Subject Categories*. In Proceedings of ASIS-83, 46th Annual Meeting of the American Society for Information Science. Washington, DC, American Society for Information Science, Washington, DC: 265.
- Chai, K. M., Ng, H. T., and Chieu, H. L. (2002). *Bayesian Online Classifiers for Text Classification and Filtering*. In Proceedings of SIGIR-02, 25th ACM International Conference on Research and Development in Information Retrieval. Tampere, FI, ACM Press, New York: 97–104.
- Chakrabarti, S., Dom, B. E., Agrawal, R., and Raghavan, P. (1997). *Using Taxonomy, Discriminants, and Signatures for Navigating in Text Databases*. In Proceedings of VLDB-97, 23rd International Conference on Very Large Data Bases. Athens, Morgan Kaufmann Publishers, San Francisco: 446–455.
- Chakrabarti, S., Dom, B. E., Agrawal, R., and Raghavan, P. (1998). “Scalable Feature Selection, Classification and Signature Generation for Organizing Large Text Databases into Hierarchical Topic Taxonomies.” *Journal of Very Large Data Bases* 7(3): 163–178.
- Chakrabarti, S., Dom, B. E., and Indyk, P. (1998). *Enhanced Hypertext Categorization Using Hyperlinks*. In Proceedings of SIGMOD-98, ACM International Conference on Management of Data. Seattle, ACM Press, New York: 307–318.
- Chakrabarti, S., Dom, B. E., Kumar, S. R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D., and Kleinberg, J. (1999). “Mining the Web’s Link Structure.” *IEEE Computer* 32(8): 60–67.

- Chakrabarti, S., Roy, S., and Soundalgekar, M. (2002). *Fast and Accurate Text Classification via Multiple Linear Discriminant Projections*. In Proceedings of VLDB-02, 28th International Conference on Very Large Data Bases. Hong Kong: 658–669.
- Chalmers, M., and Chitson, P. (1992). *Bead: Exploration in Information Visualization*. In Proceedings of the 15th Annual ACM/SIGIR Conference. Copenhagen, ACM Press, New York: 330–337.
- Chandrinou, K. V., Androutsopoulos, I., Paliouras, G., and Spyropoulos, C. D. (2000). *Automatic Web Rating: Filtering Obscene Content on the Web*. In Proceedings of ECDL-00, 4th European Conference on Research and Advanced Technology for Digital Libraries. Lisbon, Springer-Verlag, Heidelberg: 403–406.
- Chang, S.-J., and Rice, R. (1993). “Browsing: A Multidimensional Framework.” *Annual Review of Information Science and Technology* **28**: 231–276.
- Charniak, E. (1993). *Statistical Language Learning*. Cambridge, MA, MIT Press.
- Charniak, E. (2000). *A Maximum-Entropy-Inspired Parser*. In Proceedings of the Meeting of the North American Association for Computational Linguistics. Seattle, ACM Press, New York: 132–139.
- Chen, C. (2002). “Visualization of Knowledge Structures.” In *Handbook of Software Engineering and Knowledge Engineering*. S. K. Chang, ed. River Edge, NJ, World Scientific Publishing Co.: 201–238.
- Chen, C., and Paul, R. (2001). “Visualizing a Knowledge Domain’s Intellectual Structure.” *Computer* **34**(3): 65–71.
- Chen, C. C., Chen, M. C., and Sun, Y. (2001). *PVA: A Self-Adaptive Personal View Agent*. In Proceedings of KDD-01, 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, ACM Press, New York: 257–262.
- Chen, C. C., Chen, M. C., and Sun, Y. (2002). “PVA: A Self-Adaptive Personal View Agent.” *Journal of Intelligent Information Systems* **18**(2/3): 173–194.
- Chen, H., and Dumais, S. T. (2000). *Bringing Order to the Web: Automatically Categorizing Search Results*. In Proceedings of CHI-00, ACM International Conference on Human Factors in Computing Systems. The Hague, ACM Press, New York: 145–152.
- Chen, H., and Ho, T. K. (2000). *Evaluation of Decision Forests on Text Categorization*. In Proceedings of the 7th SPIE Conference on Document Recognition and Retrieval. San Jose, CA, SPIE – The International Society for Optical Engineering, Bellingham, WA: 191–199.
- Chenevoy, Y., and Belaïd, A. (1991). *Hypothesis Management for Structured Document Recognition*. In Proceedings of the 1st International Conference on Document Analysis and Recognition (ICDAR’91). St.-Malo, France: 121–129.
- Cheng, C.-H., Tang, J., Wai-Chee, A., and King, I. (2001). *Hierarchical Classification of Documents with Error Control*. In Proceedings of PAKDD-01, 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Hong Kong, Springer-Verlag, Heidelberg: 433–443.
- Cheung, D. W., Han, J., Ng, V. T., and Wong, C. Y. (1996). *Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Technique*. In Proceedings of the 12th ICDE, New Orleans, IEEE Computer Society Press, Los Alamitos, CA: 106–114.
- Cheung, D. W., Lee, S. D., and Kao, B. (1997). *A General Incremental Technique for Maintaining Discovered Association Rules*. In Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA). Melbourne, Australia: 185–194.
- Chinchor, N., Hirschman, L., and Lewis, D. (1994). “Evaluating Message Understanding Systems: An Analysis of the Third Message Understanding Conference (MUC-3).” *Computational Linguistics* **3**(19): 409–449.
- Chouchoulas, A., and Shen, Q. (2001). “Rough Set-Aided Keyword Reduction for Text Categorization.” *Applied Artificial Intelligence* **15**(9): 843–873.
- Chuang, W. T., Tiyyagura, A., Yang, J., and Giuffrida, G. (2000). *A Fast Algorithm for Hierarchical Text Classification*. In Proceedings of DaWaK-00, 2nd International Conference

- on Data Warehousing and Knowledge Discovery. London, Springer-Verlag, Heidelberg: 409–418.
- Ciravegna, F. (2001). *Adaptive Information Extraction from Text by Rule Induction and Generalization*. In Proceedings of the 17th IJCAI. Seattle, Morgan Kaufmann Publishers, San Francisco: 1251–1256.
- Ciravegna, F., Lavelli, A., Mana, N., Matiassek, J., Gilardoni, L., Mazza, S., Black, W. J., and Rinaldi, F. (1999). *FACILE: Classifying Texts Integrating Pattern Matching and Information Extraction*. In Proceedings of IJCAI-99, 16th International Joint Conference on Artificial Intelligence. T. Dean, ed. Stockholm, Morgan Kaufmann Publishers, San Francisco: 890–895.
- Clack, C., Farrington, J., Lidwell, P., and Yu, T. (1997). *Autonomous Document Classification for Business*. In Proceedings of the 1st International Conference on Autonomous Agents. W. L. Johnson, ed. Marina Del Rey, CA, ACM Press, New York: 201–208.
- Cleveland, W. S. (1994). *The Elements of Graphing Data*. Summit, NJ, Hobart Press.
- Clifton, C., and Cooley, R. (1999). *TopCat: Data Mining for Topic Identification in a Text Corpus*. In Proceedings of the 3rd European Conference on Principles of Knowledge Discovery and Data Mining. Prague, Springer, Berlin: 174–183.
- Cockburn, A. (2004). *Revisiting 2D vs 3D Implications on Spatial Memory*. In Proceedings of the 5th Conference on Australasian User Interface, Volume 28. Dunedin, New Zealand, Australian Computer Society, Inc.: 25–31.
- Cohen, W., and Singer, Y. (1996). *Context-Sensitive Learning Methods for Text Categorization*. In Proceedings of SIGIR-96, 19th ACM International Conference on Research and Development in Information Retrieval. H.-P. Frei, D. Harman, P. Schauble and R. Wilkinson, eds. Zurich, Switzerland, ACM Press, New York, 307–315.
- Cohen, W. W. (1992). *Compiling Prior Knowledge into an Explicit Bias*. In Proceedings of the 9th International Workshop on Machine Learning. D. Sleeman and P. Edwards, eds. Morgan Kaufmann Publishers, San Francisco: 102–110.
- Cohen, W. W. (1995a). “Learning to Classify English Text with ILP Methods.” In *Advances in Inductive Logic Programming*. L. D. Raedt, ed. Amsterdam, IOS Press: 124–143.
- Cohen, W. W. (1995b). *Text Categorization and Relational Learning*. In Proceedings of ICML-95, 12th International Conference on Machine Learning. Lake Tahoe, NV, Morgan Kaufmann Publishers, San Francisco: 124–132.
- Cohen, W. W., and Hirsh, H. (1998). *Joins that Generalize: Text Classification Using Whirl*. In Proceedings of KDD-98, 4th International Conference on Knowledge Discovery and Data Mining. New York, AAAI Press, Menlo Park, CA: 169–173.
- Cohen, W. W., and Singer, Y. (1996). *Context-Sensitive Learning Methods for Text Categorization*. In Proceedings of SIGIR-96, 19th ACM International Conference on Research and Development in Information Retrieval. Zurich, ACM Press, New York: 307–315.
- Cohen, W. W., and Singer, Y. (1999). “Context-Sensitive Learning Methods for Text Categorization.” *ACM Transactions on Information Systems* 17(2): 141–173.
- Collins, M. (1997). *Three Generative, Lexicalized Models for Statistical Parsing*. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics. Madrid, ACM Press, New York: 16–23.
- Collins, M., and Miller, S. (1998). *Semantic Tagging Using a Probabilistic Context Free Grammar*. In Proceedings of the 6th Workshop on Very Large Corpora. Montreal, Morgan Kaufmann Publishers, San Francisco: 38–48.
- Cooper, J. (1997). *What Is Lexical Navigation?* IBM Thomas J. Watson Research Center. http://www.research.ibm.com/people/j/jwcnmr/LexNav/lexical_navigation.htm.
- Cover, T. M., and Thomas, J. A. (1991). *Elements of Information Theory*. New York, John Wiley and Sons.
- Cowie, J., and Lehnert, W. (1996). “Information Extraction.” *Communications of the Association of Computing Machinery* 39(1): 80–91.

- Crammer, K., and Singer, Y. (2002). *A New Family of Online Algorithms for Category Ranking*. In Proceedings of SIGIR-02, 25th ACM International Conference on Research and Development in Information Retrieval. Tampere, Finland, ACM Press, New York: 151–158.
- Craven, M., DiPasquo, D., Freitag, D., McCallum, A. K., Mitchell, T. M., Nigam, K., and Slattery, S. (1998). *Learning to Extract Symbolic Knowledge from the World Wide Web*. In Proceedings of AAAI-98, 15th Conference of the American Association for Artificial Intelligence. Madison, WI, AAAI Press, Menlo Park, CA: 509–516.
- Craven, M., DiPasquo, D., Freitag, D., McCallum, A. K., Mitchell, T. M., Nigam, K., and Slattery, S. (2000). “Learning to Construct Knowledge Bases from the World Wide Web.” *Artificial Intelligence* **118**(1/2): 69–113.
- Craven, M., and Kumlien, J. (1999). *Constructing Biological Knowledge-Bases by Extracting Information from Text Sources*. In Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology (ISMB-99). Heidelberg, AAAI Press, Menlo Park, CA: 77–86.
- Craven, M., and Slattery, S. (2001). “Relational Learning with Statistical Predicate Invention: Better Models for Hypertext.” *Machine Learning* **43**(1/2): 97–119.
- Creedy, R. M., Masand, B. M., Smith, S. J., and Waltz, D. L. (1992). “Trading MIPS and Memory for Knowledge Engineering: Classifying Census Returns on the Connection Machine.” *Communications of the ACM* **35**(8): 48–63.
- Cristianini, N., Shawe-Taylor, J., and Lodhi, H. (2001). *Latent Semantic Kernels*. In Proceedings of ICML-01, 18th International Conference on Machine Learning. Williams College, MA, Morgan Kaufmann Publishers, San Francisco: 66–73.
- Cristianini, N., Shawe-Taylor, J., and Lodhi, H. (2002). “Latent Semantic Kernels.” *Journal of Intelligent Information Systems* **18**(2/3): 127–152.
- Cutting, C., Karger, D., and Pedersen, J. O. (1993). *Constant Interaction-Time Scatter/Gather Browsing of Very Large Document Collections*. In Proceedings of ACM–SIGIR Conference on Research and Development in Information Retrieval. Pittsburgh, ACM Press, New York: 126–134.
- Cutting, D. R., Karger, D. R., Pedersen, J. O., and Tukey, J. W. (1992). *Scatter/Gather: A Cluster-Based Approach to Browsing Large Document Collections*. In Proceedings of the 15th Annual International ACM–SIGIR Conference on Research and Development in Information Retrieval. Copenhagen, ACM Press, New York: 318–329.
- Cyram Company, Ltd. (2004). NetMiner Webpage <http://www.netminer.com>.
- D’Alessio, S., Murray, K., Schiaffino, R., and Kershenbaum, A. (1998). *Category Levels in Hierarchical Text Categorization*. In Proceedings of EMNLP-98, 3rd Conference on Empirical Methods in Natural Language Processing. Granada, Spain, Association for Computational Linguistics, Morristown, NJ.
- D’Alessio, S., Murray, K., Schiaffino, R., and Kershenbaum, A. (2000). *The Effect of Using Hierarchical Classifiers in Text Categorization*. In Proceedings of RIAO-00, 6th International Conference “Recherche d’Information Assistée par Ordinateur.” Paris: 302–313.
- Daelemans, W., Buchholz, S., and Veenstra, J. (1999). *Memory-Based Shallow Parsing*. In Proceedings of CoNLL. Bergen, Norway, Association for Computational Linguistics, Somerset, NJ: 53–60.
- Dagan, I., Feldman, R., and Hirsh, H. (1996). *Keyword-Based Browsing and Analysis of Large Document Sets*. In Proceedings of SDAIR-96, 5th Annual Symposium on Document Analysis and Information Retrieval. Las Vegas, UNLV Publications/Reprographics, Las Vegas: 191–207.
- Dagan, I., Karov, Y., and Roth, D. (1997). *Mistake-Driven Learning in Text Categorization*. In Proceedings of EMNLP-97, 2nd Conference on Empirical Methods in Natural Language Processing. Providence, RI, Association for Computational Linguistics, Morristown, NJ: 55–63.
- Dagan, I., Pereira, F., and Lee, L. (1994). *Similarity-Based Estimation of Word Cooccurrence Probabilities*. In Proceedings of the Annual Meeting of the Association for Computational

- Linguistics. Las Cruces, NM, Association for Computational Linguistics, Morristown, NJ: 272–278.
- Damashek, M. (1995). “Gauging Similarity with N-Grams: Language-Independent Categorization of Text.” *Science* **267**(5199): 843–848.
- Dasigi, V., Mann, R. C., and Protopopescu, V. A. (2001). “Information Fusion for Text Classification: An Experimental Comparison.” *Pattern Recognition* **34**(12): 2413–2425.
- Davidson, G. S., Hendrickson, B., Johnson, D. K., Meyers, C. E., and Wylie, B. N. (1999). “Knowledge Mining with VxInxight: Discovery through Interaction.” *Journal of Intelligent Information Systems* **11**(3): 259–285.
- Davidson, R., and Harel, D. (1996). “Drawing Graphs Nicely Using Simulated Annealing.” *ACM Transactions on Graphics* **15**(4): 301–331.
- de Buenaga Rodriguez, M., Gomez Hidalgo, J. M., and Diaz-Agudo, B. (2000). *Using WordNet to Complement Training Information in Text Categorization. Recent Advances in Natural Language Processing II*. Amsterdam, J. Benjamins: 189.
- De Nooy, W., Mrvar, A., and Batagelj, V. (2004). *Exploratory Social Network Analysis with Pajek*. New York, Cambridge University Press.
- De Sitter, A., and Daelemans, W. (2003). *Information Extraction via Double Classification*. International Workshop on Adaptive Text Extraction and Mining. Catvat-Dubroknik, Croatia, Springer, Berlin: 66–73.
- Debole, F., and Sebastiani, F. (2003). *Supervised Term Weighting for Automated Text Categorization*. In Proceedings of SAC-03, 18th ACM Symposium on Applied Computing. Melbourne, FL, ACM Press, New York: 784–788.
- Decker, S., Melnik, S., Harmelen, F. V., Fensel, D., Klein, M. C. A., Broekstra, J., Erdmann, M., and Horrocks, I. (2000). “The Semantic Web: The Roles of XML and RDF.” *IEEE Internet Computing* **4**(5): 63–74.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). “Indexing by Latent Semantic Analysis.” *Journal of the American Society of Information Science* **41**(6): 391–407.
- Denoyer, L., and Gallinari, P. (2003). *A Belief Networks–Based Generative Model for Structured Documents. An Application to the XML Categorization*. In Proceedings of MLDM-03, 3rd International Conference on Machine Learning and Data Mining in Pattern Recognition. Leipzig, Springer-Verlag, Heidelberg: 328–342.
- Denoyer, L., Zaragoza, H., and Gallinari, P. (2001). *HMM-Based Passage Models for Document Classification and Ranking*. In Proceedings of ECIR-01, 23rd European Colloquium on Information Retrieval Research. Darmstadt, Germany, Springer, Berlin: 126–135.
- Dermatas, E., and Kokkinakis, G. (1995). “Automatic Stochastic Tagging of Natural Language Texts.” *Computational Linguistics* **21**(2): 137–163.
- Dhillon, I., Mallela, S., and Kumar, R. (2002). *Enhanced Word Clustering for Hierarchical Text Classification*. In Proceedings of KDD-02, 8th ACM International Conference on Knowledge Discovery and Data Mining. Edmonton, Canada, ACM Press, New York: 191–200.
- Di-Nunzio, G., and Micarelli, A. (2003). *Does a New Simple Gaussian Weighting Approach Perform Well in Text Categorization?* In Proceedings of IJCAI-03, 18th International Joint Conference on Artificial Intelligence. Acapulco, Morgan Kaufmann Publishers, San Francisco: 581–586.
- Diao, Y., Lu, H., and Wu, D. (2000). *A Comparative Study of Classification-Based Personal E-mail Filtering*. In Proceedings of PAKDD-00, 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Kyoto, Japan, Springer-Verlag, Heidelberg: 408–419.
- Dickerson, J., Berleant, D., Cox, Z., Qi, W., and Syrkin Wurtele, E. (2003). *Creating and Modeling Metabolic and Regulatory Networks Using Text Mining and Fuzzy Expert Systems*. In Proceedings of Computational Biology and Genome Informatics Conference. World Scientific Publishing, Hackensack, NJ: 207–238.

- Diederich, J., Kindermann, J., Leopold, E., and Paass, G. (2003). "Authorship Attribution with Support Vector Machines." *Applied Intelligence* **19**(1/2): 109–123.
- Ding, Y., Fensel, D., Klein, M. C. A., and Omelayenko, B. (2002). "The Semantic Web: Yet Another Hip?" *DKE* **41**(2–3): 205–227.
- Dixon, M. (1997). "An Overview of Document Mining Technology." Unpublished manuscript.
- Doan, A., Madhavan, J., Domingos, P., and Halevy, A. Y. (2002). "Learning to Map between Ontologies on the Semantic Web." In Proceedings of WWW'02, 11th International Conference on World Wide Web. Honolulu, ACM Press, New York: 662–673.
- Domingos, P. (1999). "The Role of Occam's Razor in Knowledge Discovery." *Data Mining and Knowledge Discovery* **3**(1999): 409–425.
- Domingos, P., and Pazzani, M. (1997). "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss." *Machine Learning* **29**: 103–130.
- Dorre, J., Gerstl, P., and Seiffert, R. (1999). *Text Mining: Finding Nuggets in Mountains of Textual Data*. In Proceedings of KDD-99, 5th ACM International Conference on Knowledge Discovery and Data Mining. San Diego, ACM Press, New York: 398–401.
- Dou, D., McDermott, D., and Qi, P. (2003). *Ontology Translation on the Semantic Web*. In Proceedings of the International Conference on Ontologies, Databases and Applications of Semantics. Catania (Sicily), Italy, Springer, Berlin: 952–969.
- Doyle, L. B. (1965). "Is Automatic Classification a Reasonable Application of Statistical Analysis of Text?" *Journal of the ACM* **12**(4): 473–489.
- Drucker, H., Vapnik, V., and Wu, D. (1999). "Support Vector Machines for Spam Categorization." *IEEE Transactions on Neural Networks* **10**(5): 1048–1054.
- Duffet, P. L., and Vernik, R. J. (1997). *Software System Visualisation: Netmap Investigations*. Technical Report, DSTO-TR-0558, Defense Science and Technology Organization, Government of Australia.
- Dumais, S. T., and Chen, H. (2000). *Hierarchical Classification of Web Content*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. Athens, ACM Press, New York: 256–263.
- Dumais, S. T., Platt, J., Heckerman, D., and Sahami, M. (1998). *Inductive Learning Algorithms and Representations for Text Categorization*. In Proceedings of 7th International Conference on Information and Knowledge Management. Bethesda, MD, ACM Press, New York: 148–155.
- Dzbor, M., Domingue, J., and Motta, E. (2004). *Magpie: Supporting Browsing and Navigation on the Semantic Web*. In Proceedings of International Conference on Intelligent User Interfaces (IUI04), Madeira, Funchal, Portugal, ACM Press, New York: 191–197.
- Eades, P. (1984). "A Heuristic for Graph Drawing." *Congressus Numerantium* **44**: 149–160.
- El-Yaniv, R., and Souroujon, O. (2001). *Iterative Double Clustering for Unsupervised and Semi-Supervised Learning*. In Proceedings of ECML-01, 12th European Conference on Machine Learning. Freiburg, Germany, Springer-Verlag, Heidelberg: 121–132.
- Elworthy, D. (1994). *Does Baum-Welch Re-estimation Help Taggers?* In Proceedings of the 4th Conference on Applied Natural Language Processing. Stuttgart, Germany, Morgan Kaufmann Publishers, San Francisco: 53–58.
- Escudero, G., Márquez, L., and Rigau, G. (2000). *Boosting Applied to Word Sense Disambiguation*. In Proceedings of ECML-00, 11th European Conference on Machine Learning. Barcelona, Springer-Verlag, Heidelberg: 129–141.
- Esteban, A. D., Rodriguez, M. D. B., Lopez, L. A. U., and Vega, M. G. (1998). *Integrating Linguistic Resources in a Uniform Way for Text Classification Tasks*. In Proceedings of LREC-98, 1st International Conference on Language Resources and Evaluation. Grenada, Spain: 1197–1204.
- Etemad, K., Doermann, D. S., and Chellappa, R. (1997). "Multiscale Segmentation of Unstructured Document Pages Using Soft Decision Integration." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(1): 92–96.

- Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A., Shaked, T., Soderland, S., Weld, D., and Yates, A. (2004). *Web-Scale Information Extraction in KnowItAll*. In Proceedings of WWW-04, 13th International World Wide Web Conference. New York, ACM Press, New York: 100–110.
- Etzioni, O., Cafarella, M., Downey, D., Popescu, A., T. Shaked, Soderland, S., Weld, D., and Yates, A. (2004). *Methods for Domain-Independent Information Extraction from the Web: An Experimental Comparison*. In Proceedings of the 19th National Conference on Artificial Intelligence.
- Ezawa, K., and Norton, S. (1995). *Knowledge Discovery in Telecommunication Services Data Using Bayesian Network Models*. In Proceedings of the First International Conference on Knowledge Discovery (KDD-95). Montreal, AAAI Press, Menlo Park, CA: 100–105.
- Fall, C. J., Torcsvari, A., Benzineb, K., and Karetka, G. (2003). “Automated Categorization in the International Patent Classification.” *SIGIR Forum* 37(1): 10–25.
- Fangmeyer, H., and Lustig, G. (1968). *The EURATOM Automatic Indexing Project*. In Proceedings of the IFIP Congress (Booklet J). Edinburgh, North Holland Publishing Company, Amsterdam: 66–70.
- Fangmeyer, H., and Lustig, G. (1970). *Experiments with the CETIS Automated Indexing System*. In Proceedings of the Symposium on the Handling of Nuclear Information, International Atomic Energy Agency: 557–567.
- Fayyad, U., Grinstein, G., and Wierse, A., Eds. (2001). *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco, Morgan Kaufmann Publishers.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). “From Data Mining to Knowledge Discovery in Databases.” In *Advances in Knowledge Discovery and Data Mining*. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthuruswamy, eds. Cambridge, MA, AAAI/MIT Press: 1–36.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthuruswamy, R., eds. (1996). *Advances in Knowledge Discovery and Data Mining*. Cambridge, MA, AAAI/MIT Press.
- Fayyad, U. M., Reina, C. A., and Bradley, P. S. (1998). *Initialization of Iterative Refinement Clustering Algorithms*. Technical Report MSR-TR-98-38, Jet Propulsion Laboratories.
- Feldman, R. (1993). *Probabilistic Revision of Logical Domain Theories*. Ph.D. thesis, Department of Computer Science, Cornell University.
- Feldman, R. (1996). *The KDT System – Using Prolog for KDD*. In Proceedings of 4th Conference of Practical Applications of Prolog. London: 91–110.
- Feldman, R. (1998). *Practical Text Mining*. In Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery. London: 478.
- Feldman, R. (2002). “Text Mining.” In *Handbook of Data Mining and Knowledge Discovery*. W. Kloeegen and J. Zytkow, eds. New York, Oxford University Press.
- Feldman, R., Amir, A., Aumann, Y., and Zilberstein, A. (1996). *Incremental Algorithms for Association Generation*. In Proceedings of the First Pacific Conference on Knowledge Discovery. Singapore.
- Feldman, R., Aumann, Y., Amir, A., Zilberstein, A., and Kloeegen, W. (1997). *Maximal Association Rules: A New Tool for Keyword Co-occurrences in Document Collections*. In Proceedings of 3rd International Conference on Knowledge Discovery and Data Mining. Newport Beach, CA, AAAI Press, Menlo Park, CA: 167–170.
- Feldman, R., Aumann, Y., Finkelstein-Landau, M., Hurvitz, E., Regev, Y., and Yaroshevich, A. (2002). *A Comparative Study of Information Extraction Strategies*. In Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics. Mexico City, Springer, New York: 349–359.
- Feldman, R., Aumann, Y., Zilberstein, A., and Ben-Yehuda, Y. (1997). *Trend Graphs: Visualizing the Evolution of Concept Relationships in Large Document Collections*. In Proceedings of the 2nd European Symposium of Principles of Data Mining and Knowledge Discovery. Nantes, France, Springer, Berlin: 38–46.

- Feldman, R., and Dagan, I. (1995). *Knowledge Discovery in Textual Databases (KDT)*. In Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining. Montreal, Canada, AAAI Press, Menlo Park, CA: 112–117.
- Feldman, R., Dagan, I., and Hirsh, H. (1998). “Mining Text Using Keyword Distributions.” *Journal of Intelligent Information Systems* **10**(3): 281–300.
- Feldman, R., Dagan, I., and Kloesgen, W. (1996a). *Efficient Algorithms for Mining and Manipulating Associations in Texts*. In Proceedings of the 13th European Meeting on Cybernetics and Systems Research. Vienna, Austria: 949–954.
- Feldman, R., Dagan, I., and Kloesgen, W. (1996b). *KDD Tools for Mining Associations in Textual Databases*. In Proceedings of the 9th International Symposium on Methodologies for Intelligent Systems. Zakopane, Poland: 96–107.
- Feldman, R., Fresko, M., Hirsh, H., Aumann, Y., Liphstat, O., Schler, Y., and Rajman, M. (1998). *Knowledge Management: A Text Mining Approach*. In Proceedings of the 2nd International Conference on Practical Aspects of Knowledge Management (PAKM98). Basel, Switzerland.
- Feldman, R., Fresko, M., Kinar, Y., Lindell, Y., Liphstar, O., Rajman, M., Schler, Y., and Zamir, O. (1998). *Text Mining at the Term Level*. In Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery. Nantes, France, Springer, Berlin: 65–73.
- Feldman, R., and Hirsh, H. (1996a). “Exploiting Background Information in Knowledge Discovery from Text.” *Journal of Intelligent Information Systems* **9**(1): 83–97.
- Feldman, R., and Hirsh, H. (1996b). *Mining Associations in Text in the Presence of Background Knowledge*. In Proceedings of the 2nd International Conference on Knowledge Discovery from Databases. Portland, OR, AAAI Press, Menlo Park, CA: 343–346.
- Feldman, R., and Hirsh, H. (1997). “Finding Associations in Collections of Text.” In *Machine Learning and Data Mining: Methods and Applications*. R. S. Michalski, I. Bratko, and M. Kubat, eds. New York, John Wiley and Sons: 223–240.
- Feldman, R., Kloesgen, W., Ben-Yehuda, Y., Kedar, G., and Reznikov, V. (1997). *Pattern Based Browsing in Document Collections*. In Proceedings of the 1st European Symposium of Principles of Data Mining and Knowledge Discovery. Trondheim, Norway, Springer, Berlin: 112–122.
- Feldman, R., Kloesgen, W., and Zilberstein, A. (1997a). *Document Explorer: Discovering Knowledge in Document Collections*. In Proceedings of the 10th International Symposium on Methodologies for International Systems. Trondheim, Norway, Springer, Berlin: 137–146.
- Feldman, R., Kloesgen, W., and Zilberstein, A. (1997b). *Visualization Techniques to Explore Data Mining Results for Document Collections*. In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. Newport Beach, CA, AAAI Press, Menlo Park, CA: 16–23.
- Feldman, R., Regev, Y., Hurvitz, E., and Landau-Finkelstein, M. (2003). “Mining the Biomedical Literature Using Semantic Analysis and Natural Language Processing Techniques.” *Biosilico* **1** (2): 69–72.
- Fellbaum, C. D., ed. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA, MIT Press.
- Fensel, D., Angele, J., Decker, S., Erdmann, M., Schnurr, H.-P., Staab, S., Studer, R., and Witt, A. (1999). “On2broker: Semantic-Based Access to Information Sources at the WWW.” *WebNet* **1**: 366–371.
- Ferilli, S., Fanizzi, N., and Semeraro, G. (2001). *Learning Logic Models for Automated Text Categorization*. In Proceedings of AI*IA-01, 7th Congress of the Italian Association for Artificial Intelligence. F. Esposito, ed. Bari, Italy, Springer-Verlag, Heidelberg: 81–86.
- Ferrndez, A., Palomar, M., and Moreno, L. (1998). *Anaphor Resolution in Unrestricted Texts with Partial Parsing*. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics. Montreal, Morgan Kaufmann Publishers, San Francisco: 385–391.

- Field, B. J. (1975). "Towards Automatic Indexing: Automatic Assignment of Controlled-Language Indexing and Classification from Free Indexing." *Journal of Documentation* **31**(4): 246–265.
- Finch, S. (1994). *Exploiting Sophisticated Representations for Document Retrieval*. In Proceedings of the 4th Conference on Applied Natural Language Processing. Stuttgart, Germany, Morgan Kaufmann Publishers, San Francisco: 65–71.
- Finn, A., Kushmerick, N., and Smyth, B. (2002). *Genre Classification and Domain Transfer for Information Filtering*. In Proceedings of ECIR-02, 24th European Colloquium on Information Retrieval Research. Glasgow, Springer-Verlag, Heidelberg: 353–362.
- Fisher, D., Soderland, S., McCarthy, J., Feng, F., and Lehnert, W. (1995). *Description of the UMass System as Used for MUC-6*. In Proceedings of the 6th Message Understanding Conference (MUC-6). Columbia, MD, Morgan Kaufmann Publishers, San Francisco: 127–140.
- Fisher, M., and Everson, R. (2003). *When Are Links Useful? Experiments in Text Classification*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Berlin: 41–56.
- Forsyth, R. S. (1999). "New Directions in Text Categorization." In *Causal Models and Intelligent Data Management*. A. Gammerman, ed. Heidelberg, Springer-Verlag: 151–185.
- Frank, E., Chui, C., and Witten, I. H. (2000). *Text Categorization Using Compression Models*. In Proceedings of DCC-00, IEEE Data Compression Conference. Snowbird, UT, IEEE Computer Society Press, Los Alamitos, CA: 200–209.
- Frank, E., Paynter, G. W., Witten, I. H., Gutwin, C., and Neville-Manning, C. G. (1999). *Domain-Specific Keyphrase Extraction*. In Proceedings of the 16th International Joint Conference on Artificial Intelligence. Stockholm, Morgan Kaufmann Publishers, San Francisco: 668–673.
- Frasconi, P., Soda, G., and Vullo, A. (2001). *Text Categorization for Multi-page Documents: A Hybrid Naive Bayes HMM Approach*. In Proceedings of JCDL, 1st ACM-IEEE Joint Conference on Digital Libraries. Roanoke, VA, IEEE Computer Society Press, Los Alamitos, CA: 11–20.
- Frasconi, P., Soda, G., and Vullo, A. (2002). "Text Categorization for Multi-page Documents: A Hybrid Naive Bayes HMM Approach." *Journal of Intelligent Information Systems* **18**(2/3): 195–217.
- Frawley, W. J., Piatetsky-Shapiro, G., and Matheus, C. J. (1991). "Knowledge Discovery in Databases: An Overview." In *Knowledge Discovery in Databases*. G. Piatetsky-Shapiro and W. J. Frawley, eds. Cambridge, MA, MIT Press: 1–27.
- Freeman, L. C. (1977). "A Set of Measures of Centrality Based on Betweenness." *Sociometry* **40**: 35–41.
- Freeman, L. C. (1979). "Centrality in Social Networks: Conceptual Clarification." *Social Networks* **1**: 215–239.
- Freitag, D. (1997). *Using Grammatical Inference to Improve Precision in Information Extraction*. In Proceedings of the Workshop on Grammatical Inference, Automata Induction, and Language Acquisition (ICML '97). Nashville, TN, Morgan Kaufmann Publishers, San Mateo, CA.
- Freitag, D. (1998a). *Information Extraction from HTML: Application of a General Machine Learning Approach*. In Proceedings of the 15th National Conference on Artificial Intelligence. Madison, WI, AAAI Press, Menlo Park, CA: 517–523.
- Freitag, D. (1998b). *Machine Learning for Information Extraction in Informal Domains*. Ph.D. thesis, Computer Science Department, Carnegie Mellon University.
- Freitag, D., and Kushmerick, N. (2000). *Boosted Wrapper Induction*. In Proceedings of AAAI 2000. Austin, TX, AAAI Press, Menlo Park, CA: 577–583.
- Freitag, D., and McCallum, A. (2000). *Information Extraction with HMM Structures Learned by Stochastic Optimization*. In Proceedings of the 17th National Conference on Artificial Intelligence. Austin, TX, AAAI Press, Menlo Park, CA: 584–589.

- Freitag, D., and McCallum, A. L. (1999). *Information Extraction with HMMs and Shrinkage*. In Papers from the AAAI-99 Workshop on Machine Learning for Information Extraction: 31–36.
- Freund, J., and Walpole, R. (1990). *Estadística Matemática con Aplicaciones*. Prentice Hall.
- Frommholz, I. (2001). *Categorizing Web Documents in Hierarchical Catalogues*. In Proceedings of ECIR-01, 23rd European Colloquium on Information Retrieval Research. Darmstadt, Germany: 18–20.
- Fruchterman, T., and Reingold, E. (1991). “Graph Drawing by Force-Directed Placement.” *Software – Practice and Experience* **21**(11): 1129–1164.
- Fuhr, N. (1985). *A Probabilistic Model of Dictionary-Based Automatic Indexing*. In Proceedings of RIAO-85, 1st International Conference “Recherche d’Information Assistée par Ordinateur.” Grenoble, France: 207–216.
- Fuhr, N., Hartmann, S., Knorz, G., Lustig, G., Schwantner, M., and Tzeras, K. (1991). *AIR/X – A Rule-Based Multistage Indexing System for Large Subject Fields*. In Proceedings of RIAO-91, 3rd International Conference “Recherche d’Information Assistée par Ordinateur.” A. Lichnerowicz, ed. Barcelona, Elsevier Science Publishers, Amsterdam: 606–623.
- Fuhr, N., and Knorz, G. (1984). *Retrieval Test Evaluation of a Rule-Based Automated Indexing (AIR/PHYS)*. In Proceedings of SIGIR-84, 7th ACM International Conference on Research and Development in Information Retrieval. C. J. v. Rijsbergen, ed. Cambridge, UK, Cambridge University Press, Cambridge: 391–408.
- Fuhr, N., and Pfeifer, U. (1991). *Combining Model-Oriented and Description-Oriented Approaches for Probabilistic Indexing*. In Proceedings of SIGIR-91, 14th ACM International Conference on Research and Development in Information Retrieval. Chicago, ACM Press, New York: 46–56.
- Fuhr, N., and Pfeifer, U. (1994). “Probabilistic Information Retrieval as Combination of Abstraction Inductive Learning and Probabilistic Assumptions.” *ACM Transactions on Information Systems* **12**(1): 92–115.
- Fukuda, K., Tamura, A., Tsunoda, T., and Takagi, T. (1998). *Toward Information Extraction: Identifying Protein Names*. In Proceedings of the Pacific Symposium on Biocomputing. Maui, Hawaii, World Scientific Publishing Company, Hackensack, NJ: 707–718.
- Fung, G. P. C., Yu, J. X., and Lu, H. (2002). *Discriminative Category Matching: Efficient Text Classification for Huge Document Collections*. In Proceedings of ICDM-02, 2nd IEEE International Conference on Data Mining. Maebashi City, Japan, IEEE Computer Society Press, Los Alamitos, 187–194.
- Furnas, G. (1981). “The FISHEYE View: A New Look at Structured Files.” Bell Laboratories Technical Report, reproduced in *Reading in Information Visualization: Using Vision to Think*. S. K. Card, J. D. Mackinlay, and B. Schneiderman, eds. San Francisco, Morgan Kaufmann Publishers: 312–330.
- Furnas, G. (1986). *Generalized Fisheye Views*. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. ACM Press, New York: 16–23.
- Furnkranz, J. (1999). *Exploiting Structural Information for Text Classification on the WWW*. In Proceedings of IDA-99, 3rd Symposium on Intelligent Data Analysis. Amsterdam, Springer-Verlag, Heidelberg: 487–497.
- Furnkranz, J. (2002). “Hyperlink Ensembles: A Case Study in Hypertext Classification.” *Information Fusion* **3**(4): 299–312.
- Gaizauskas, R., and Humphreys, K. (1997). “Using a Semantic Network for Information Extraction.” *Natural Language Engineering* **3**(2): 147–196.
- Galavotti, L., Sebastiani, F., and Simi, M. (2000). *Experiments on the Use of Feature Selection and Negative Evidence in Automated Text Categorization*. In Proceedings of ECDL-00, 4th European Conference on Research and Advanced Technology for Digital Libraries. Lisbon, Springer-Verlag, Heidelberg: 59–68.
- Gale, W. A., Church, K. W., and Yarowsky, D. (1993). “A Method for Disambiguating Word Senses in a Large Corpus.” *Computers and the Humanities* **26**(5): 415–439.

- Gall, H., Jazayeri, M., and Riva, C. (1999). *Visualizing Software Release Histories: The Use of Color and Third Dimension*. In Proceedings of the International Conference on Software Maintenance (ICSM '99). Oxford, UK, IEEE Computer Society Press, Los Alamitos, CA: 99.
- Gansner, E., Koutsofias, E., North, S., and Vo, K. (1993). "A Technique for Drawing Directed Graphs." *IEEE Transactions on Software Engineering* **19**(3): 214–230.
- Gansner, E., North, S., and Vo, K. (1988). "DAG – A Program that Draws Directed Graphs." *Software Practice and Experience* **18**(11): 1047–1062.
- Ganter, B., and Wille, R. (1999). *Formal Concept Analysis: Mathematical Foundations*. Berlin, Springer-Verlag.
- Gao, S., Wu, W., Lee, C.-H., and Chua, T.-S. (2003). *A Maximal Figure-of-Merit Learning Approach to Text Categorization*. In Proceedings of SIGIR-03, 26th ACM International Conference on Research and Development in Information Retrieval. Toronto, ACM Press, New York: 174–181.
- Gaussier, E., Goutte, C., Popat, K., and Chen, F. (2002). *A Hierarchical Model for Clustering and Categorising Documents*. In Proceedings of ECIR-02, 24th European Colloquium on Information Retrieval Research. Glasgow, Springer-Verlag, Heidelberg: 229–247.
- Gelbukh, A., ed. (2002). *Computational Linguistics and Intelligent Text Processing*. In Proceedings of 3rd International Conference, CICLing 2001. Mexico City, Springer-Verlag, Berlin and New York.
- The Gene Ontology (GO) Consortium. (2000). "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* **25**: 25–29.
- The Gene Ontology (GO) Consortium. (2001). "Creating the Gene Ontology Resource: Design and Implementation." *Genome Research* **11**: 1425–1433.
- Gentili, G. L., Marinilli, M., Micarelli, A., and Sciarone, F. (2001). "Text Categorization in an Intelligent Agent for Filtering Information on the Web." *International Journal of Pattern Recognition and Artificial Intelligence* **15**(3): 527–549.
- Geutner, P., Bodenhausen, U., and Waibel, A. (1993). *Flexibility through Incremental Learning: Neural Networks for Text Categorization*. In Proceedings of WCNN-93, World Congress on Neural Networks. Portland, OR, Lawrence Erlbaum Associates, Hillsdale, NJ: 24–27.
- Ghani, R. (2000). *Using Error-Correcting Codes for Text Classification*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco: 303–310.
- Ghani, R. (2001). *Combining Labeled and Unlabeled Data for Text Classification with a Large Number of Categories*. In Proceedings of the IEEE International Conference on Data Mining. San Jose, CA, IEEE Computer Society Press, Los Alamitos, CA: 597–598.
- Ghani, R. (2002). *Combining Labeled and Unlabeled Data for MultiClass Text Categorization*. In Proceedings of ICML-02, 19th International Conference on Machine Learning. Sydney, Australia, Morgan Kaufmann Publishers, San Francisco: 187–194.
- Ghani, R., Slattery, S., and Yang, Y. (2001). *Hypertext Categorization Using Hyperlink Patterns and Meta Data*. In Proceedings of ICML-01, 18th International Conference on Machine Learning. Williams College, Morgan Kaufmann Publishers, San Francisco: 178–185.
- Giorgetti, D., and Sebastiani, F. (2003a). "Automating Survey Coding by Multiclass Text Categorization Techniques." *Journal of the American Society for Information Science and Technology* **54**(12): 1269–1277.
- Giorgetti, D., and Sebastiani, F. (2003b). *Multiclass Text Categorization for Automated Survey Coding*. In Proceedings of SAC-03, 18th ACM Symposium on Applied Computing. Melbourne, Australia, ACM Press, New York: 798–802.
- Giorgio, M. D. N., and Micarelli, A. (2003). *Does a New Simple Gaussian Weighting Approach Perform Well in Text Categorization?* In Proceedings of IJCAI-03, 18th International Joint Conference on Artificial Intelligence. Acapulco, Morgan Kaufmann Publishers, San Francisco: 581–586.
- Glover, E. J., Tsioutsoulouklis, K., Lawrence, S., Pennock, D. M., and Flake, G. W. (2002). *Using Web Structure for Classifying and Describing Web Pages*. In Proceedings of WWW-02,

- International Conference on the World Wide Web. Honolulu, ACM Press, New York: 562–569.
- Goldberg, J. L. (1995). *CDM: An Approach to Learning in Text Categorization*. In Proceedings of ICTAI-95, 7th International Conference on Tools with Artificial Intelligence. Herndon, VA, IEEE Computer Society Press, Los Alamitos, CA: 258–265.
- Goldberg, J. L. (1996). “CDM: An Approach to Learning in Text Categorization.” *International Journal on Artificial Intelligence Tools* 5(1/2): 229–253.
- Goldstein, J., and Roth, S. (1994). *Using Aggregation and Dynamic Queries for Exploring Large Data Sets*. In Proceedings of Human Factors in Computing Systems CHI '94 Conference. Boston, ACM, New York: 23–29.
- Goldszmidt, M., and Sahami, M. (1998). *Probabilistic Approach to Full-Text Document Clustering*. Technical Report ITAD-433-MS-98-044, SRI International.
- Gomez-Hidalgo, J. M. (2002). *Evaluating Cost-Sensitive Bulk Email Categorization*. In Proceedings of SAC-02, 17th ACM Symposium on Applied Computing. Madrid, ACM Press, New York: 615–620.
- Gomez-Hidalgo, J. M., Rodriguez, J. M. D. B., Lopez, L. A. U., Valdivia, M. T. M., and Vega, M. G. (2002). *Integrating Lexical Knowledge in Learning-Based Text Categorization*. In Proceedings of JADT-02, 6th International Conference on the Statistical Analysis of Textual Data. St.-Malo, France.
- Goodman, M. (1990). *Prism: A Case-Based Telex Classifier*. In Proceedings of IAAI-90, 2nd Conference on Innovative Applications of Artificial Intelligence. Boston, AAAI Press, Menlo Park, CA: 25–37.
- Gotlieb, C. C., and Kumar, S. (1968). “Semantic Clustering of Index Terms.” *Journal of the ACM* 15(4): 493–513.
- Govert, N., Lalmas, M., and Fuhr, N. (1999). *A Probabilistic Description-Oriented Approach for Categorising Web Documents*. In Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management. Kansas City, MO, ACM Press, New York: 475–482.
- Graham, M. (2001). *Visualising Multiple Overlapping Classification Hierarchies*. Ph.D. diss., Napier University.
- Gray, W. A., and Harley, A. J. (1971). “Computer-Assisted Indexing.” *Information Storage and Retrieval* 7(4): 167–174.
- Greene, B. B., and Rubin, G. M. (1971). *Automatic Grammatical Tagging of English*. Technical Report. Providence, RI, Brown University.
- Grieser, G., Jantke, K. P., Lange, S., and Thomas, B. (2000). *A Unifying Approach to HTML Wrapper Representation and Learning. Discovery Science*. In Proceedings of 3rd International Conference, DS 2000. Kyoto, Japan, Springer-Verlag, Berlin: 50–64.
- Grinstein, G. (1996). *Harnessing the Human in Knowledge Discovery*. In Proceedings of the 2nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Portland, OR, AAAI Press, CA: 384–385.
- Grishman, R. (1996). “The Role of Syntax in Information Extraction.” In *Advances in Text Processing: Tipster Program Phase II*. San Francisco, Morgan Kaufmann Publishers.
- Grishman, R. (1997). “Information Extraction: Techniques and Challenges.” In Materials of Information Extraction International Summer School – SCIE '97. Springer, Berlin: 10–27.
- Gruber, T. R. (1993). “A Translation Approach to Portable Ontologies.” *Knowledge Acquisition* 5: 199–220.
- Guthrie, L., Guthrie, J. A., and Leistensnider, J. (1999). “Document Classification and Routing.” In *Natural Language Information Retrieval*. T. Strzalkowski, ed. Dordrecht, Kluwer Academic Publishers: 289–310.
- Guthrie, L., Walker, E., and Guthrie, J. A. (1994). *Document Classification by Machine: Theory and Practice*. In Proceedings of COLING-94, 15th International Conference on Computational Linguistics. Kyoto, Japan, Morgan Kaufmann Publishers, San Francisco: 1059–1063.
- Hadany, R., and Harel, D. (2001). “A Multi-Scale Method for Drawing Graphs Nicely.” *Discrete Applied Mathematics* 113: 3–21.

- Hadjarian, A., Bala, J., and Pachowicz, P. (2001). *Text Categorization through Multistrategy Learning and Visualization*. In Proceedings of CICLING-01, 2nd International Conference on Computational Linguistics and Intelligent Text Processing. A. Gelbukh, ed. Mexico City, Springer-Verlag, Heidelberg: 423–436.
- Hahn, U., and Schnattinger, K. (1997). *Knowledge Mining from Textual Sources*. In Proceedings of the 6th International Conference on Information and Knowledge Management. Las Vegas, ACM, New York: 83–90.
- Hamill, K. A., and Zamora, A. (1978). *An Automatic Document Classification System Using Pattern Recognition Techniques*. In Proceedings of ASIS-78, 41st Annual Meeting of the American Society for Information Science. E. H. Brenner, ed. New York, American Society for Information Science, Washington, DC: 152–155.
- Hamill, K. A., and Zamora, A. (1980). "The Use of Titles for Automatic Document Classification." *Journal of the American Society for Information Science* **33**(6): 396–402.
- Han, E.-H., Karypis, G., and Kumar, V. (2001). *Text Categorization Using Weight-Adjusted k-Nearest Neighbor Classification*. In Proceedings of PAKDD-01, 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Hong Kong, Springer-Verlag, Heidelberg: 53–65.
- Han, J., and Fu, Y. (1995). *Discovery of Multiple-Level Association Rules from Large Databases*. In Proceedings of the 1995 International Conference on Very Large Data Bases (VLDB'95). Zurich, Morgan Kaufmann Publishers, San Francisco: 420–431.
- Hanauer, D. (1996). "Integration of Phonetic and Graphic Features in Poetic Text Categorization Judgements." *Poetics* **23**(5): 363–380.
- Hao, M., Dayal, U., Hsu, M., Sprenger, T., and Gross, M. (2001). *Visualization of Directed Associations in E-commerce Transaction Data*. In Proceedings of Data Visualization (EG and IEEE's VisSym '01). Ascona, Switzerland, Springer, Berlin: 185–192.
- Haralick, R. M. (1994). *Document Image Understanding: Geometric and Logical Layout*. In Proceedings of CVPR94, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Seattle, IEEE Computer Society Press, Los Alamitos, CA: 385–390.
- Hardt, D., and Romero, M. (2002). *Ellipsis and the Structure of Discourse*. In Proceedings of Sinn und Bedeutung VI, Osnabrück, Germany, Institute for Cognitive Science, University of Osnabrück: 85–98.
- Harel, D., and Koren, Y. (2000). *A Fast Multi-Scale Method for Drawing Large Graphs*. In Proceedings of the 8th International Symposium on Graph Drawing. Williamsburg, VA, Springer-Verlag, Heidelberg: 282–285.
- Hatzivassiloglou, V., Duboue, P. A., and Rzhetsky, A. (2001). "Disambiguating Proteins, Genes, and RNA in Text: A Machine Learning Approach." *Bioinformatics* **17**(Suppl 1): S97–106.
- Havre, S., Hetzler, B., and Nowell, L. (1999). *ThemeRiver(TM): In Search of Trends, Patterns and Relationships*. In Proceedings of IEEE Symposium on Information Visualization (InfoVis 1999). San Francisco, IEEE Press, New York: 115–123.
- Hayes, P. (1992). "Intelligent High-Volume Processing Using Shallow, Domain-Specific Techniques." In *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval*. P. S. Jacobs, ed. Hillsdale, NJ, Lawrence Earlbaum: 227–242.
- Hayes, P. J., Andersen, P. M., Nirenburg, I. B., and Schmandt, L. M. (1990). *Tcs: A Shell for Content-Based Text Categorization*. In Proceedings of CAIA-90, 6th IEEE Conference on Artificial Intelligence Applications. Santa Barbara, CA, IEEE Computer Society Press, Los Alamitos, CA: 320–326.
- Hayes, P. J., Knecht, L. E., and Cellio, M. J. (1988). *A News Story Categorization System*. In Proceedings of ANLP-88, 2nd Conference on Applied Natural Language Processing. Austin, TX, Association for Computational Linguistics, Morristown, NJ: 9–17.
- Hayes, P. J., and Weinstein, S. P. (1990). *Construe/Tis: A System for Content-Based Indexing of a Database of News Stories*. In Proceedings of IAAI-90, 2nd Conference on Innovative Applications of Artificial Intelligence. Boston, AAAI Press, Menlo Park, CA: 49–66.

- He, J., Tan, A.-H., and Tan, C.-L. (2003). "On Machine Learning Methods for Chinese Document Categorization." *Applied Intelligence* **18**(3): 311–322.
- Heaps, H. S. (1973). "A Theory of Relevance for Automatic Document Classification." *Information and Control* **22**(3): 268–278.
- Hearst, M. (1992). *Automatic Acquisition of Hyponyms From Large Text Corpora*. In Proceedings of the 14th International Conference on Computational Linguistics. Nantes, France, Association for Computational Linguistics, Morristown, NJ: 539–545.
- Hearst, M. (1995). *TileBars: Visualization of Term Distribution Information in Full-Text Information Access*. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Denver, CO, ACM, New York: 59–66.
- Hearst, M. (1999a). *Untangling Text Mining*. In Proceedings of the 37th Annual Meeting of the Association of Computational Linguistics. College Park, MD, Association of Computational Linguistics, Morristown, NJ: 3–10.
- Hearst, M. (1999b). "User Interfaces and Visualization." In *Modern Information Retrieval*. R. Baeza-Yates and B. Ribeiro-Neto, eds. Boston, Addison-Wesley Longman Publishing Company: 257–323.
- Hearst, M. (2003). *Information Visualization: Principles, Promise and Pragmatics Tutorial Notes*. In Proceedings of CHI 03. Fort Lauderdale, FL.
- Hearst, M., and Hirsh, H. (1996). *Machine Learning in Information Access*. Papers from the 1996 AAAI Spring Symposium. Stanford, CA, AAAI Press, Menlo Park, CA.
- Hearst, M., and Karadi, C. (1997). *Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results Using a Large Category Hierarchy*. In Proceedings of the 20th Annual International ACM/SIGIR Conference. Philadelphia, ACM Press, New York: 246–255.
- Hearst, M. A. (1991). *Noun Homograph Disambiguation Using Local Context in Large Corpora*. In Proceedings of the 7th Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary. Oxford, UK: 1–22.
- Hearst, M. A., Karger, D. R., and Pedersen, J. O. (1995). *Scatter/Gather as a Tool for the Navigation of Retrieval Results*. Working Notes, AAAI Fall Symposium on AI Applications in Knowledge Navigation. Cambridge, MA, AAAI Press, Menlo Park, CA: 65–71.
- Hearst, M. A., and Pedersen, J. O. (1996). *Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results*. In Proceedings of ACM SIGIR '96. Zurich, ACM Press, New York: 76–84.
- Hersh, W., Buckley, C., Leone, T. J., and Hickman, D. (1994). *OHSUMED: An Interactive Retrieval Evaluation and New Large Text Collection for Research*. In Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval. Dublin, Springer-Verlag, Heidelberg: 192–201.
- Hetzler, B., Harris, W. M., Havre, S., and Whitney, P. (1998). *Visualizing the Full Spectrum of Document Relationships*. In Proceedings of the 5th International Society for Knowledge Organization (ISKO) Conference. Lille, France, Ergon-Verlog, Würzburg, Germany: 168–175.
- Hetzler, B., Whitney, P., Martucci, L., and Thomas, J. (1998). *Multi-Faceted Insight through Interoperable Visual Information Analysis Paradigms*. In Proceedings of Information Visualization '98. Research Triangle Park, NC, IEEE Computer Society Press, Los Alamitos, CA: 137–144.
- Hill, D. P., Blake, J. A., Richardson, J. E., and Ringwald, M. (2002). "Extension and Integration of the Gene Ontology (GO): Combining GO Vocabularies with External Vocabularies." *Genome Research* **12**: 1982–1991.
- Hindle, D. (1989). *Acquiring Disambiguation Rules from Text*. In Proceedings of 27th Annual Meeting of the Association for Computational Linguistics. Vancouver, Association for Computational Linguistics, Morristown, NJ: 118–125.
- Hirschman, L., Park, J. C., Tsujii, J., Wong, L., and Wu, C. H. (2002). "Accomplishments and Challenges in Literature Data Mining for Biology." *Bioinformatics Review* **18**(12): 1553–1551.

- Hoashi, K., Matsumoto, K., Inoue, N., and Hashimoto, K. (2000). *Document Filtering Methods Using Non-Relevant Information Profile*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. Athens, ACM Press, New York: 176–183.
- Hobbs, J. (1986). “Resolving Pronoun References.” In *Readings in Natural Language Processing*. B. J. Grosz, K. S. Jones and B. L. Webber, eds. Los Altos, CA, Morgan Kaufmann Publishers: 339–352.
- Hobbs, J., Douglas, R., Appelt, E., Bear, J., Israel, D., Kameyama, M., Stickel, M., and Tyson, M. (1996). “FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text.” In *Finite State Devices for Natural Language Processing*. E. Roche, and Y. Schabes, eds. Cambridge, MA, MIT Press: 383–406.
- Hobbs, J. R. (1993). *FASTUS: A System for Extracting Information from Text*. In Proceedings of DARPA Workshop on Human Language Technology. Princeton, NJ, Morgan Kaufmann Publishers, San Mateo, CA: 133–137.
- Hobbs, J. R., Appelt, D. E., Bear, J., Tyson, M., and Magerman, D. (1991). *The TACITUS System: The MUC-3 Experience*. Menlo Park, CA, SRI.
- Hoch, R. (1994). *Using IR Techniques for Text Classification in Document Analysis*. In Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval. Dublin, Springer-Verlag, Heidelberg: 31–40.
- Honkela, T. (1997). *Self-Organizing Maps in Natural Language Processing*. Neural Networks Research Centre. Helsinki, Helsinki University of Technology.
- Honkela, T., Kaski, S., Kohonen, T., and Lagus, K. (1998). “Self-Organizing Maps of Very Large Document Collections: Justification for the WEBSOM Method.” In *Classification, Data Analysis and Data Highways*. I. Balderjahn, R. Mathar and M. Schader, eds. Berlin, Springer-Verlag: 245–252.
- Honkela, T., Kaski, S., Lagus, K., and Kohonen, T. (1997). *WEBSOM – Self-Organizing Maps of Document Collections*. In Proceedings of WSOM ’97, Workshop on Self-Organizing Maps. Espoo, Finland, Helsinki University of Technology. Helsinki: 310–315.
- Honkela, T., Lagus, K., and Kaski, S. (1998). “Self-Organizing Maps of Large Document Collections.” In *Visual Explorations in Finance with Self-Organizing Maps*. G. Deboeck and T. Kohonen, eds. London, Springer: 168–178.
- Hornbaek, K., Bederson, B., and Plaisant, C. (2002). “Navigation Patterns and Usability of Zoomable User Interfaces With and Without an Overview.” *ACM Transactions on Computer-Human Interaction* 9(4): 362–389.
- Hotho, A., Maedche, A., Staab, S., and Zacharias, V. (2002). “On Knowledgeable Supervised Text Mining.” In *Text Mining: Theoretical Aspects and Applications*. J. Franke, G. Nakhaeizadeh, and I. Renz, eds. Heidelberg, Physica-Verlag (Springer): 131–152.
- Hotho, A., Staab, S., and Maedche, A. (2001). *Ontology-Based Text Clustering*. In Proceedings of the IJCAI-2001 Workshop Text Learning: Beyond Supervision. Seattle.
- Hotho, A., Staab, S., and Stumme, G. (2003). *Text Clustering Based on Background Knowledge*. Institute of Applied Informatics and Formal Descriptive Methods, University of Karlsruhe, Germany: 1–35.
- Hoyle, W. G. (1973). “Automatic Indexing and Generation of Classification by Algorithm.” *Information Storage and Retrieval* 9(4): 233–242.
- Hsu, W.-L., and Lang, S.-D. (1999). *Classification Algorithms for NETNEWS Articles*. In Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management. Kansas City, MO, ACM Press, New York: 114–121.
- Hsu, W.-L., and Lang, S.-D. (1999). *Feature Reduction and Database Maintenance in NET-NEWS Classification*. In Proceedings of IDEAS-99, 1999 International Database Engineering and Applications Symposium. Montreal, IEEE Computer Society Press, Los Alamitos, CA: 137–144.
- Huang, S., Ward, M., and Rudensteiner, E. (2003). *Exploration of Dimensionality Reduction for Text Visualization*. Worcester, MA, Worcester Polytechnic Institute.

- Hubona, G. S., Shirah, G., and Fout, D. (1997). "The Effects of Motion and Stereopsis on Three-Dimensional Visualization." *International Journal of Human Computer Studies* 47(5): 609–627.
- Huffman, S. (1995). *Acquaintance: Language-Independent Document Categorization by N-Grams*. In Proceedings of TREC-4, 4th Text Retrieval Conference. Gaithersburg, MD, National Institute of Standards and Technology, Gaithersburg, MD: 359–371.
- Huffman, S., and Damashek, M. (1994). *Acquaintance: A Novel Vector-Space N-Gram Technique for Document Categorization*. In Proceedings of TREC-3, 3rd Text Retrieval Conference, D. K. Harman, ed. Gaithersburg, MD, National Institute of Standards and Technology, Gaithersburg, MD: 305–310.
- Huffman, S. B. (1995). "Learning Information Extraction Patterns from Examples." In *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*. S. Wermter, E. Riloff, and G. Scheler, eds. London, Springer-Verlag: 246–260.
- Hull, D. (1996). "Stemming Algorithms – A Case Study for Detailed Evaluation." *Journal of the American Society for Information Science* 47(1): 70–84.
- Hull, D. A. (1994). *Improving Text Retrieval for the Routing Problem Using Latent Semantic Indexing*. In Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval. Dublin, Springer-Verlag, Heidelberg: 282–289.
- Hull, D. A. (1998). *The TREC-7 Filtering Track: Description and Analysis*. In Proceedings of TREC-7, 7th Text Retrieval Conference. Gaithersburg, MD, National Institute of Standards and Technology, Gaithersburg, MD: 33–56.
- Hull, D. A., Pedersen, J. O., and Schutze, H. (1996). *Method Combination for Document Filtering*. In Proceedings of SIGIR-96, 19th ACM International Conference on Research and Development in Information Retrieval. H.-P. Frei, D. Harman, P. Schable, and R. Wilkinson, eds. Zurich, ACM Press, New York: 279–288.
- Hummon, M. P., and Carley, K. (1993). "Social Networks as Normal Science." *Social Networks* 14: 71–106.
- Humphreys, K., Gaizauskas, R., and Azzam, S. (1997). *Event Coreference for Information Extraction*. In Proceedings of the Workshop on Operational Factors in Practical, Robust, Anaphora Resolution for Unrestricted Texts. Madrid, Spain, Association for Computational Linguistics, Morristown, NJ: 75–81.
- Igarashi, T., and Hinckley, K. (2000). *Speed-Dependent Automatic Zooming for Browsing Large Documents*. In Proceedings of the 11th Annual Symposium on User Interface Software and Technology (UIST '00). San Diego, CA, ACM Press, New York: 139–148.
- IntertekGroup (2002). *Leveraging Unstructured Data in Investment Management*. <http://www.taborcommunications.com/dsstar/02/0604/104317.html>.
- Ipeirotis, P. G., Gravano, L., and Sahami, M. (2001). *Probe, Count, and Classify: Categorizing Hidden Web Databases*. In Proceedings of SIGMOD-01, ACM International Conference on Management of Data. W. G. Aref, ed. Santa Barbara, CA, ACM Press, New York: 67–78.
- Ittner, D. J., Lewis, D. D., and Ahn, D. D. (1995). *Text Categorization of Low Quality Images*. In Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval. Las Vegas, NV, ISRI, University of Nevada, Las Vegas, NV: 301–315.
- Iwayama, M., and Tokunaga, T. (1994). *A Probabilistic Model for Text Categorization: Based on a Single Random Variable with Multiple Values*. In Proceedings of ANLP-94, 4th Conference on Applied Natural Language Processing. Stuttgart, Germany, Association for Computational Linguistics, Morristown, NJ: 162–167.
- Iwayama, M., and Tokunaga, T. (1995a). *Cluster-Based Text Categorization: A Comparison of Category Search Strategies*. In Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval. E. A. Fox, P. Ingwersen, and R. Fidel, eds. Seattle, ACM Press, New York: 273–281.
- Iwayama, M., and Tokunaga, T. (1995b). *Hierarchical Bayesian Clustering for Automatic Text Classification*. In Proceedings of IJCAI-95, 14th International Joint Conference on Artificial

- Intelligence. C. E. Mellish, ed. Montreal, Morgan Kaufmann Publishers, San Francisco: 1322–1327.
- Iwazume, M., Takeda, H., and Nishida, T. (1996). *Ontology-Based Information Gathering and Text Categorization from the Internet*. In Proceedings of IEA/AIE-96, 9th International Conference in Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. T. Tanaka, S. Ohsuga, and M. Ali, eds. Fukuoka, Japan: 305–314.
- Iyer, R. D., Lewis, D. D., Schapire, R. E., Singer, Y., and Singhal, A. (2000). *Boosting for Document Routing*. In Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management. A. Agah, J. Callan, and E. Rundensteiner, eds. McLean, VA, ACM Press, New York: 70–77.
- Jacobs, P. S. (1992). *Joining Statistics with NLP for Text Categorization*. In Proceedings of ANLP-92, 3rd Conference on Applied Natural Language Processing. M. Bates and O. Stock, eds. Trento, Italy, Association for Computational Linguistics, Morristown, NJ: 178–185.
- Jacobs, P. S. (1993). “Using Statistical Methods to Improve Knowledge-Based News Categorization.” *IEEE Expert* 8(2): 13–23.
- Jain, A., and Dubes, R. (1988). *Algorithms for Clustering Data*. Englewood Cliffs, NJ, Prentice Hall.
- Jain, A. K., and Chellappa, R., eds. (1993). *Markov Random Fields: Theory and Application*. Boston, Academic Press.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). “Data Clustering: A Review.” *ACM Computing Surveys* 31(3): 264–323.
- Jensen, J. R. (1996). *Introductory Digital Image Processing – A Remote Sensing Perspective*. Englewood Cliffs, NJ, Prentice Hall.
- Jerding, D., and Stasko, J. (1995). *The Information Mural: A Technique for Displaying and Navigating Large Information Spaces*. In Proceedings of Information Visualization '95 Symposium. Atlanta, IEEE Computer Society, Washington, DC: 43.
- Jo, T. C. (1999a). “News Article Classification Based on Categorical Points from Keywords in Backdata.” In *Computational Intelligence for Modelling, Control and Automation*. M. Mohammadian, ed. Amsterdam, IOS Press: 211–214.
- Jo, T. C. (1999b). “News Articles Classification Based on Representative Keywords of Categories.” In *Computational Intelligence for Modelling, Control and Automation*. M. Mohammadian, ed. Amsterdam, IOS Press: 194–198.
- Jo, T. C. (1999c). *Text Categorization with the Concept of Fuzzy Set of Informative Keywords*. In Proceedings of FUZZ-IEEE '99, IEEE International Conference on Fuzzy Systems. Seoul, KR, IEEE Computer Society Press, Los Alamitos, CA: 609–614.
- Joachims, T. (1997). *A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization*. In Proceedings of ICML-97, 14th International Conference on Machine Learning. D. H. Fisher, ed. Nashville, TN, Morgan Kaufmann Publishers, San Francisco: 143–151.
- Joachims, T. (1998). *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*. In Proceedings of ECML-98, 10th European Conference on Machine Learning. C. Nedellec and C. Rouveirol, eds. Chemnitz, Germany, Springer-Verlag, Heidelberg: 137–142.
- Joachims, T. (1999). *Transductive Inference for Text Classification Using Support Vector Machines*. In Proceedings of ICML-99, 16th International Conference on Machine Learning. I. Bratko and S. Dzeroski, eds. Bled, Morgan Kaufmann Publishers, San Francisco: 200–209.
- Joachims, T. (2000). *Estimating the Generalization Performance of a SVM Efficiently*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco: 431–438.
- Joachims, T. (2001). *A Statistical Learning Model of Text Classification with Support Vector Machines*. In Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval. W. B. Croft, D. J. Harper, D. H. Kraft, and J. Zobel, eds. New Orleans, ACM Press, New York: 128–136.

- Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines*. Dordrecht, Kluwer Academic Publishers.
- Joachims, T., Cristianini, N., and Shawe-Taylor, J. (2001). *Composite Kernels for Hypertext Categorisation*. In Proceedings of ICML-01, 18th International Conference on Machine Learning. C. Brodley and A. Danyluk, eds. Williams College, MA, Morgan Kaufmann Publishers, San Francisco: 250–257.
- Joachims, T., Freitag, D., and Mitchell, T. M. (1997). *WebWatcher: A Tour Guide for the World Wide Web*. In Proceedings of IJCAI-97, 15th International Joint Conference on Artificial Intelligence. M. E. Pollack, ed. Nagoya, Japan, Morgan Kaufmann Publishers, San Francisco: 770–775.
- Joachims, T., and Sebastiani, F. (2002). “Guest Editors’ Introduction to the Special Issue on Automated Text Categorization.” *Journal of Intelligent Information Systems* **18**(2/3): 103–105.
- Johnson, B., and Shneiderman, B. (1991). “Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information.” In Proceedings of IEEE Visualization ’91 Conference. G. Nielson and L. Rosenblum, eds. San Diego, CA, IEEE Computer Society Press, Los Alamitos, CA: 284–291.
- Juan, A., and Vidal, E. (2002). “On the Use of Bernoulli Mixture Models for Text Classification.” *Pattern Recognition* **35**(12): 2705–2710.
- Junker, M., and Abecker, A. (1997). *Exploiting Thesaurus Knowledge in Rule Induction for Text Classification*. In Proceedings of RANLP-97, 2nd International Conference on Recent Advances in Natural Language Processing. Tzigras Chark, Bulgaria: 202–207.
- Junker, M., and Dengel, A. (2001). *Preventing Overfitting in Learning Text Patterns for Document Categorization*. In Proceedings of ICAPR-01, 2nd International Conference on Advances in Pattern Recognition. S. Singh, N. A. Murshed, and W. G. Kropatsch, eds. Rio de Janeiro, Springer-Verlag, Heidelberg: 137–146.
- Junker, M., and Hoch, R. (1998). “An Experimental Evaluation of OCR Text Representations for Learning Document Classifiers.” *International Journal on Document Analysis and Recognition* **1**(2): 116–122.
- Junker, M., Sintek, M., and Rinck, M. (2000). *Learning for Text Categorization and Information Extraction with ILP*. In Proceedings of the 1st Workshop on Learning Language in Logic. Bled, Slovenia, Springer-Verlag, Heidelberg: 247–258.
- Kaban, A., and Girolami, M. (2002). “A Dynamic Probabilistic Model to Visualise Topic Evolution in Text Streams.” *Journal of Intelligent Information Systems* **18**(2/3): 107–125.
- Kamada, T., and Kawai, S. (1989). “An Algorithm for Drawing General Undirected Graphs.” *Information Processing Letters* **31**: 7–15.
- Kar, G., and White, L. J. (1978). “A Distance Measure for Automated Document Classification by Sequential Analysis.” *Information Processing and Management* **14**(2): 57–69.
- Karrer, A., and Scacchi, W. (1990). *Requirements for an Extensible Object-Oriented Tree/Graph Editor*. In Proceedings of ACM SIGGRAPH Symposium on User Interface Software and Technology. Snowbird, UT, ACM Press, New York: 84–91.
- Karypis, G., and Han, E.-H. (2000). *Fast Supervised Dimensionality Reduction Algorithm with Applications to Document Categorization and Retrieval*. In Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management. A. Agah, J. Callan, and E. Rundensteiner, eds. McLean, VA, ACM Press, New York: 12–19.
- Kaski, S. (1997). *Data Exploration Using Self-Organizing Maps*. Tech thesis, Helsinki University of Technology.
- Kaski, S., Honkela, T., Lagus, K., and Kohonen, T. (1998). “WEBSOM-Self-Organizing Maps of Document Collections.” *Neurocomputing* **21**: 101–117.
- Kaski, S., Lagus, K., Honkela, T., and Kohonen, T. (1998). “Statistical Aspects of the WEBSOM System in Organizing Document Collections.” *Computing Science and Statistics* **29**: 281–290.

- Kawatani, T. (2002). *Topic Difference Factor Extraction between Two Document Sets and Its Application to Text Categorization*. In Proceedings of SIGIR-02, 25th ACM International Conference on Research and Development in Information Retrieval. K. Jarvelin, M. Beaulieu, R. Baeza-Yates, and S. H. Myaeng, eds. Tampere, Finland, ACM Press, New York: 137–144.
- Kehagias, A., Petridis, V., Kaburlasos, V. G., and Fragkou, P. (2003). “A Comparison of Word- and Sense-Based Text Categorization Using Several Classification Algorithms.” *Journal of Intelligent Information Systems* **21**(3): 227–247.
- Kehler, A. (1997). *Probabilistic Coreference in Information Extraction*. In Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing. C. Cardie and R. Weischedel, eds. Providence, RI, Association for Computational Linguistics, Somerset, NJ: 163–173.
- Keim, D. (2002). “Information Visualization and Visual Data Mining.” *IEEE Transactions on Visualization and Computer Graphics* **8**(1): 1–8.
- Keller, B. (1992). *A Logic for Representing Grammatical Knowledge*. In Proceedings of European Conference on Artificial Intelligence. Vienna, Austria, John Wiley and Sons, New York: 538–542.
- Kennedy, C., and Boguraev, B. (1997). *Anaphora for Everyone: Pronominal Anaphora Resolution Without a Parser*. In Proceedings of the 16th International Conference on Computational Linguistics. J. Tsujii, ed. Copenhagen, Denmark, Association for Computational Linguistics, Morristown, NJ: 113–118.
- Keogh, E., and Smyth, P. (1997). *A Probabilistic Approach to Fast Pattern Matching in Time Series Databases*. In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD’97). D. Heckerman, H. Mannila, D. Pregibon, and R. Uthurusamy, eds. Newport Beach, CA, AAAI Press, Menlo Park, CA: 24–30.
- Kessler, B., Nunberg, G., and Schutze, H. (1997). *Automatic Detection of Text Genre*. In Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics. P. R. Cohen and W. Wahlster, eds. Madrid, Morgan Kaufmann Publishers, San Francisco: 32–38.
- Khmelev, D. V., and Teahan, W. J. (2003). *A Repetition Based Measure for Verification of Text Collections and for Text Categorization*. In Proceedings of SIGIR-03, 26th ACM International Conference on Research and Development in Information Retrieval. C. Clarke, G. Cormack, J. Callan, D. Hawking, and A. Smeaton, eds. Toronto, ACM Press, New York: 104–110.
- Kim, H. (2002). “Predicting How Ontologies for the Semantic Web Will Evolve.” *CACM* **45**(2): 48–54.
- Kim, J.-T., and Moldovan, D. I. (1995). “Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction.” *TKDE* **7**(5): 713–724.
- Kim, Y.-H., Hahn, S.-Y., and Zhang, B.-T. (2000). *Text Filtering by Boosting Naive Bayes Classifiers*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. N. J. Belkin, P. Ingwersen, and M. K. Leong, eds. Athens, ACM Press, New York: 168–75.
- Kindermann, J., Paass, G., and Leopold, E. (2001). *Error Correcting Codes with Optimized Kullback–Leibler Distances for Text Categorization*. In Proceedings of ECML-01, 12th European Conference on Machine Learning. L. de Raedt and A. Siebes, eds. Freiburg, Germany, Springer-Verlag, Heidelberg: 266–275.
- Kindermann, R., and Snell, J. L. (1980). *Markov Random Fields and Their Applications*. Providence, RI, American Mathematical Society.
- Klas, C.-P., and Fuhr, N. (2000). *A New Effective Approach for Categorizing Web Documents*. In Proceedings of BCSIRSG-00, 22nd Annual Colloquium of the British Computer Society Information Retrieval Specialist Group. Cambridge, UK, BCS, Swinden, UK.
- Klebanov, B., and Wiemer-Hastings, P. M. (2002). *Using LSA for Pronominal Anaphora Resolution*. In Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing. A. F. Gelbukh, ed. Mexico City, Springer, Berlin: 197–199.

- Klingbiel, P. H. (1973a). "Machine-Aided Indexing of Technical Literature." *Information Storage and Retrieval* 9(2): 79–84.
- Klingbiel, P. H. (1973b). "A Technique for Machine-Aided Indexing." *Information Storage and Retrieval* 9(9): 477–494.
- Klinkenberg, R., and Joachims, T. (2000). *Detecting Concept Drift with Support Vector Machines*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco: 487–494.
- Kloesgen, W. (1992). "Problems for Knowledge Discovery in Databases and Their Treatment in the Statistics Interpreter EXPLORA." *International Journal for Intelligent Systems* 7(7): 649–673.
- Kloesgen, W. (1995a). "Efficient Discovery of Interesting Statements in Databases." *Journal of Intelligent Information Systems* 4: 53–69.
- Kloesgen, W. (1995b). "EXPLORA: A Multipattern and Multistrategy Discovery Assistant." In *Advances in Knowledge Discovery and Data Mining*. U. Fayyad, G. Piatetsky-Shapiro, and R. Smyth, eds. Cambridge, MA, MIT Press: 249–271.
- Kloesgen, W., and Zytrow, J., eds. (2002). *Handbook of Data Mining and Knowledge Discovery*. Oxford, UK, Oxford University Press.
- Kloptchenko, A., Eklund, T., Back, B., Karlson, J., Vanharanta, H., and Visa, A. (2002). "Combining Data and Text Mining Techniques for Analyzing Financial Reports." *International Journal of Intelligent Systems in Accounting, Finance, and Management* 12(1): 29–41.
- Knorr, E., Ng, R., and Tucakov, V. (2000). "Distance Based Outliers: Algorithms and Applications." *The VLDB Journal* 8(3): 237–253.
- Knorz, G. (1982). *A Decision Theory Approach to Optimal Automated Indexing*. In Proceedings of SIGIR-82, 5th ACM International Conference on Research and Development in Information Retrieval. G. Salton and H.-J. Schneider, eds. Berlin, Springer-Verlag, Heidelberg: 174–193.
- Ko, Y., Park, J., and Seo, J. (2002). *Automatic Text Categorization Using the Importance of Sentences*. In Proceedings of COLING-02, 19th International Conference on Computational Linguistics. Taipei, Taiwan, Association for Computational Linguistics, Morristown NJ/Morgan Kaufmann Publishers, San Francisco, CA: 1–7.
- Ko, Y., and Seo, J. (2000). *Automatic Text Categorization by Unsupervised Learning*. In Proceedings of COLING-00, 18th International Conference on Computational Linguistics. Saarbrücken, Germany, Association for Computational Linguistics, Morristown, NJ: 453–459.
- Ko, Y., and Seo, J. (2002). *Text Categorization Using Feature Projections*. In Proceedings of COLING-02, 19th International Conference on Computational Linguistics. Taipei, Taiwan, Association for Computational Linguistics, Morristown, NJ/Morgan Kauffman Publishers, San Francisco, CA: 453–459.
- Kobsa, A. (2001). *An Empirical Comparison of Three Commercial Information Visualization Systems*. In Proceedings of Infovis 2001, IEEE Symposium on Information Visualization. San Diego, CA, IEEE Computer Society Press, Washington, DC: 123.
- Koehn, P. (2002). *Combining Multiclass Maximum Entropy Text Classifiers with Neural Network Voting*. In Proceedings of PorTAL-02, 3rd International Conference on Advances in Natural Language Processing. Faro, Portugal, Springer, Berlin: 125–132.
- Kohlhase, M. (2000). "Model Generation for Discourse Representation Theory." In Proceedings of the 14th European Conference on Artificial Intelligence. W. Horn, ed. Berlin, IOS Press, Amsterdam: 441–445.
- Kohonen, T. (1981). *Automatic Formation of Topological Maps of Patterns in a Self-Organizing System*. In Proceedings of 2SCIA, 2nd Scandinavian Conference on Image Analysis. E. Uja and O. Simula, eds. Helsinki, Finland, Suomen Hahmontunnistustutkimuksen Seura r.y.: 214–220.
- Kohonen, T. (1982). "Analysis of Simple Self-Organizing Process." *Biological Cybernetics* 44(2): 135–140.

- Kohonen, T. (1995). *Self-Organizing Maps*. Berlin, Springer-Verlag.
- Kohonen, T. (1997). *Exploration of Very Large Databases by Self-Organizing Maps*. In Proceedings of ICNN '97, International Conference on Neural Networks. Houston, TX, IEEE Service Center Press, Piscataway, NJ: 1–6.
- Kohonen, T. (1998). *Self-Organization of Very Large Document Collections: State of the Art*. In Proceedings of ICANN98, 8th International Conference on Artificial Neural Networks. M. Niklasson and T. Zienkka, eds. Skövde, Sweden, Springer-Verlag, London: 65–74.
- Kohonen, T., Kaski, S., Lagus, K., and Honkela, T. (1996). *Very Large Two-Level SOM for the Browsing of Newsgroups*. In Proceedings of ICANN96, International Conference on Artificial Neural Networks. Bochum, Germany, Springer-Verlag, Berlin: 269–274.
- Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, T., Paatero, V., and Saarela, A. (1999). “Self-Organization of a Massive Text Document Collection.” In *Kohonen Maps*. E. Oja and S. Kaski, eds. Amsterdam, Elsevier: 171–182.
- Koike, H. (1993). “The Role of Another Spatial Dimension in Software Visualization.” *ACM Transactions on Information Systems* **11**(3): 266–286.
- Koike, H. (1995). “Fractal Views: A Fractal-Based Method for Controlling Information Display.” *ACM Transactions on Information Systems* **13**(3): 305–323.
- Koike, T., and Rzhetsky, A. (2000). “A Graphic Editor for Analyzing Signal-Transduction Pathways.” *Gene* **259**: 235–244.
- Kolcz, A., Prabakarmurthi, V., and Kalita, J. K. (2001). *String Match and Text Extraction: Summarization as Feature Selection for Text Categorization*. In Proceedings of CIKM-01, 10th ACM International Conference on Information and Knowledge Management. W. Paques, L. Liu, and D. Grossman, eds. Atlanta, ACM Press, New York: 365–370.
- Koller, D., and Sahami, M. (1997). *Hierarchically Classifying Documents Using Very Few Words*. In Proceedings of ICML-97, 14th International Conference on Machine Learning. D. H. Fisher, ed. Nashville, TN, Morgan Kaufmann Publishers, San Francisco: 170–178.
- Kongovi, M., Guzman, J. C., and Dasigi, V. (2002). *Text Categorization: An Experiment Using Phrases*. In Proceedings of ECIR-02, 24th European Colloquium on Information Retrieval Research. F. Cresteni, M. Girotami, and C. J. v. Rijsbergen, eds. Glasgow, Springer-Verlag, Heidelberg: 213–228.
- Kopanis, I., Avouris, N. M., and Daskalaki, S. (2002). *The Role of Knowledge Mining in a Large Scale Data Mining Project*. In Proceedings of Methods and Applications of Artificial Intelligence, 2nd Hellenic Conference on AI. I. P. Vlahavas and C. Spyropoulos, eds. Thessaloniki, Greece, Springer-Verlag, Berlin: 288–299.
- Koppel, M., Argamon, S., and Shimoni, A. R. (2002). “Automatically Categorizing Written Texts by Author Gender.” *Literary and Linguistic Computing* **17**(4): 401–412.
- Kosmynin, A., and Davidson, I. (1996). *Using Background Contextual Knowledge for Document Representation*. In Proceedings of PODP-96, 3rd International Workshop on Principles of Document Processing. C. Nicholas and D. Wood, eds. Palo Alto, CA, Springer-Verlag, Heidelberg: 123–133.
- Koster, C. H., and Seutner, M. (2003). *Taming Wild Phrases*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Heidelberg: 161–176.
- Krauthammer, M., Rzhetsky, A., Morozov, P., and Friedman, C. (2000). “Using BLAST for Identifying Gene and Protein Names in Journal Articles.” *Gene* **259**: 245–252.
- Krier, M., and Zacc, F. (2002). “Automatic Categorization Applications at the European Patent Office.” *World Patent Information* **24**: 187–196.
- Krishnapuram, R., Chitrapura, K., and Joshi, S. (2003). *Classification of Text Documents Based on Minimum System Entropy*. In Proceedings of ICML-03, 20th International Conference on Machine Learning. Washington, DC, Morgan Kaufmann Publishers, San Francisco: 384–391.
- Kupiec, J. (1992). “Robust Part-of-Speech Tagging Using a Hidden Markov model.” *Computer Speech and Language* **6**: 225–243.

- Kushmerick, N. (1997). *Wrapper Induction for Information Extraction*. Ph.D. thesis, Department of Computer Science and Engineering, University of Washington.
- Kushmerick, N. (2000). "Wrapper Induction: Efficiency and Expressiveness." *Artificial Intelligence* **118**(1–2): 15–68.
- Kushmerick, N. (2002). *Finite-State Approaches to Web Information Extraction*. In Proceedings of the 3rd Summer Convention on Information Extraction in the Web Era: Natural Language Communication for Knowledge Acquisition and Intelligent Information Agents. M. Pazienza, ed. Rome, Springer-Verlag, Berlin: 77–91.
- Kushmerick, N., Johnston, E., and McGuinness, S. (2001). *Information Extraction by Text Classification*. In Proceedings of IJCAI-01 Workshop on Adaptive Text Extraction and Mining. Seattle, Morgan Kaufmann Publishers, San Francisco.
- Kushmerick, N., Weld, D. S., and Doorenbos, R. B. (1997). *Wrapper Induction for Information Extraction*. In Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI). Nagoya, Japan, Morgan Kaufmann Publishers, San Francisco: 729–735.
- Kwok, J. T. (1998). *Automated Text Categorization Using Support Vector Machine*. In Proceedings of ICONIP '98, 5th International Conference on Neural Information Processing. Kitakyushu, Japan: 347–351.
- Kwon, O.-W., Jung, S.-H., Lee, J.-H., and Lee, G. (1999). *Evaluation of Category Features and Text Structural Information on a Text Categorization Using Memory-Based Reasoning*. In Proceedings of ICCPOL-99, 18th International Conference on Computer Processing of Oriental Languages. Tokushima, Japan: 153–158.
- Kwon, O.-W., and Lee, J.-H. (2003). "Text Categorization Based on k -nearest Neighbor Approach for Web Site Classification." *Information Processing and Management* **39**(1): 25–44.
- Labrou, Y., and Finin, T. (1999). *Yahoo! as an Ontology: Using Yahoo! Categories to Describe documents*. In Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management. Kansas City, MO, ACM Press, New York: 180–187.
- Lafferty, J., McCallum, A., and Pereira, F. (2001). *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*. In Proceedings of 18th International Conference on Machine Learning. Williamstown, MA, Morgan Kaufmann Publisher, San Francisco: 282–289.
- Lager, T. (1998). *Logic for Part-of-Speech Tagging and Shallow Parsing*. In Proceedings of NODALIDA '98. Copenhagen, Denmark, Center for Sprogteknologi, Univio Copenhagen, Copenhagen.
- Lagus, K. (1998). *Generalizability of the WEBSOM Method to Document Collections of Various Types*. In Proceedings of 6th European Congress on Intelligent Techniques and Soft Computing (EUFIT'98). Aachen, Germany, Verlag Mainz, Mainz: 210–215.
- Lagus, K. (2000a). *Text Mining with the WEBSOM*. D. Sc. (Tech) thesis, Department of Computer Science and Engineering, Helsinki University of Technology.
- Lagus, K. (2000b). *Text Retrieval Using Self-Organized Document Maps*. Technical Report A61, Laboratory of Computer and Information Science, Helsinki University of Technology.
- Lagus, K., Honkela, T., Kaski, S., and Kohonen, T. (1999). "WEBSOM for Textual Data Mining." *Artificial Intelligence Review* **13**(5/6): 345–364.
- Lai, K.-Y., and Lam, W. (2001). *Meta-Learning Models for Automatic Textual Document Categorization*. In Proceedings of PAKDD-01, 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining. D. Cheung, Q. Li, and G. Williams, eds. Hong Kong, Springer Verlag, Heidelberg: 78–89.
- Lai, Y.-S., and Wu, C.-H. (2002). "COLUMN: Meaningful Term Extraction and Discriminative Term Selection in Text Categorization via Unknown-Word Methodology." *ACM Transactions on Asian Language Information Processing* **1**(1): 34–64.
- Lam, S. L., and Lee, D. L. (1999). *Feature Reduction for Neural Network Based Text Categorization*. In Proceedings of DASFAA-99, 6th IEEE International Conference on Database

- Advanced Systems for Advanced Application. A. L. Chen and F. H. Lochovsky, eds. Hsinchu, Taiwan, IEEE Computer Society Press, Los Alamitos, CA: 195–202.
- Lam, W., and Ho, C. Y. (1998). *Using a Generalized Instance Set for Automatic Text Categorization*. In Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval. W. B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson, and J. Zobel, eds. Melbourne, Australia, ACM Press, New York: 81–89.
- Lam, W., and Lai, K.-Y. (2001). *A Meta-Learning Approach for Text Categorization*. In Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval. W. B. Croft, D. J. Harper, D. H. Kraft, and J. Zobel, eds. New Orleans, ACM Press, New York: 303–309.
- Lam, W., Low, K. F., and Ho, C. Y. (1997). *Using a Bayesian Network Induction Approach for Text Categorization*. In Proceedings of IJCAI-97, 15th International Joint Conference on Artificial Intelligence. M. E. Pollack, ed. Nagoya, Japan, Morgan Kaufmann Publishers, San Francisco: 745–750.
- Lam, W., Ruiz, M. E., and Srinivasan, P. (1999). “Automatic Text Categorization and Its Applications to Text Retrieval.” *IEEE Transactions on Knowledge and Data Engineering* **11**(6): 865–879.
- Lamping, J., and Rao, R. (1994). *Laying Out and Visualizing Large Trees Using a Hyperbolic Space*. In Proceedings of the ACM UIST (UIST '94). P. Szekely, ed. Marina Del Rey, CA, ACM Press, New York: 13–14.
- Lamping, L., Rao, R., and Piroli, P. (1995). *A Focus-Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies*. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computer Systems. I. Katz, R. Mack, L. Marks, M. B. Rosson, and J. Nielsen, eds. Denver, CO, ACM Press, New York: 401–408.
- Landau, D., Feldman, R., Aumann, Y., Fresko, M., Lindell, Y., Liphstat, O., and Zamir, O. (1998). *TextVis: An Integrated Visual Environment for Text Mining*. In Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD98). Nantes, France, Springer-Verlag, Heidelberg: 56–64.
- Landauer, T. K., Foltz, P. W., and Laham, D. (1998). “Introduction to Latent Semantic Analysis.” *Discourse Processes* **25**: 259–284.
- Lang, K. (1995). *NewsWeeder: Learning to Filter Netnews*. In Proceedings of ICML-95, 12th International Conference on Machine Learning. A. Friedl and S. J. Russell, eds. Lake Tahoe, NV, Morgan Kaufmann Publishers, San Francisco: 331–339.
- Lanquillon, C. (2000). *Learning from Labeled and Unlabeled Documents: A Comparative Study on Semi-Supervised Text Classification*. In Proceedings of PKDD-00, 4th European Conference on Principles of Data Mining and Knowledge Discovery. D. A. Zighed, H. J. Komorowsky and J. M. Zytkow, eds. Lyon, France, Springer-Verlag, Heidelberg: 490–497.
- Lappin, S., and Leass, H. J. (1994). “An Algorithm for Pronominal Anaphora Resolution.” *Computational Linguistics* **20**(4): 535–561.
- Larkey, L. S. (1998). *Automatic Essay Grading Using Text Categorization Techniques*. In Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval. W. B. Croft, A. Moffat, C. J. v. Rijsbergen, R. Wilkinson, and J. Zobel, eds. Melbourne, Australia, ACM Press, New York: 90–95.
- Larkey, L. S. (1999). *A Patent Search and Classification System*. In Proceedings of DL-99, 4th ACM Conference on Digital Libraries. E. A. Fox and N. Roweeds, eds. Berkeley, CA, ACM Press, New York: 179–187.
- Larkey, L. S., and Croft, W. B. (1996). *Combining Classifiers in Text Categorization*. In Proceedings of SIGIR-96, 19th ACM International Conference on Research and Development in Information Retrieval. H. P. Frei, D. Harmon, P. Schaubie, and R. Wilkinson, eds. Zurich, ACM Press, New York: 289–297.
- Lavelli, A., Califf, M. E., Ciravegna, F., Freitag, D., Giuliano, C., Kushmerick, N., and Romano, L. (2004). *A Critical Survey of the Methodology for IE Evaluation*. In Proceedings of the 4th

- International Conference on Language Resources and Evaluation. Lisbon, ELRA, Paris: 1655–1658.
- Lavelli, A., Magnini, B., and Sebastiani, F. (2002). *Building Thematic Lexical Resources by Bootstrapping and Machine Learning*. In Proceedings of the LREC 2002 Workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data. Las Palmas, Canary Islands, ELRA, Paris: 53–62.
- Lee, K. H., Kay, J., Kang, B. H., and Rosebrock, U. (2002). *A Comparative Study on Statistical Machine Learning Algorithms and Thresholding Strategies for Automatic Text Categorization*. In Proceedings of PRICAI-02, 7th Pacific Rim International Conference on Artificial Intelligence. Milshizuka and A. Sattar, eds. Tokyo, Springer-Verlag, Heidelberg: 444–453.
- Lee, M. D. (2002). *Fast Text Classification Using Sequential Sampling Processes*. In Proceedings of the 14th Australian Joint Conference on Artificial Intelligence. M. Stumptner, D. Corbett and M. J. Brooks, eds. Adelaide, Australia, Springer-Verlag, Heidelberg: 309–320.
- Lee, Y.-B., and Myaeng, S. H. (2002). *Text Genre Classification with Genre-Revealing and Subject-Revealing Features*. In Proceedings of SIGIR-02, 25th ACM International Conference on Research and Development in Information Retrieval. M. Beaulieu, E. Beaulieu-Yakes, S. Myaeng, and K. Jarvelin, eds. Tampere, Finland, ACM Press, New York: 145–150.
- Leek, T. R. (1997). *Information Extraction Using Hidden Markov Models*. Master's thesis, Computer Science Department, University of California San Diego.
- Lehnert, W., Soderland, S., Aronow, D., Feng, F., and Shmueli, A. (1994). "Inductive Text Classification for Medical Applications." *Journal of Experimental and Theoretical Artificial Intelligence* 7(1): 49–80.
- Lent, B., Agrawal, R., and Srikant, R. (1997). *Discovering Trends in Text Databases*. In Proceedings of the 3rd Annual Conference on Knowledge Discovery and Data Mining (KDD-97) D. Heckerman, H. Mannila, D. Pregibon, and R. Uthysamy, eds. Newport Beach, CA, AAAI Press, Menlo Park, CA: 227–230.
- Leopold, E., and Kindermann, J. (2002). "Text Categorization with Support Vector Machines: How to Represent Texts in Input Space?" *Machine Learning* 46(1/3): 423–444.
- Lesk, M. (1997). *Practical Digital Libraries: Books, Bytes and Bucks*. San Francisco, Morgan Kaufmann Publishers.
- Leung, C.-H., and Kan, W.-K. (1997). "A Statistical Learning Approach to Automatic Indexing of Controlled Index Terms." *Journal of the American Society for Information Science* 48(1): 55–67.
- Leung, Y. K., and Apperley, M. D. (1994). "A Review and Taxonomy of Distortion-Oriented Presentation Techniques." *ACM Transactions on Computer-Human Interaction* 1(2): 126–160.
- Lewin, I., Becket, R., Boye, J., Carter, D., Rayner, M., and Wir'en, M. (1999). *Language Processing for Spoken Dialogue Systems: Is Shallow Parsing Enough?* Technical Report CRC-074, SRI, Cambridge, MA: 107–110.
- Lewis, D., and Catlett, J. (1994). *Heterogeneous Uncertainty Sampling for Supervised Learning*. In Proceedings of the 11th International Conference on Machine Learning. New Brunswick, NJ, Morgan Kaufmann Publishers, San Francisco: 148–156.
- Lewis, D. D. (1991). *Data Extraction as Text Categorization: An Experiment with the MUC-3 Corpus*. In Proceedings of MUC-3, 3rd Message Understanding Conference. San Diego, CA, Morgan Kaufmann Publishers, San Francisco: 245–255.
- Lewis, D. D. (1992a). *An Evaluation of Phrasal and Clustered Representations on a Text Categorization task*. In Proceedings of SIGIR-92, 15th ACM International Conference on Research and Development in Information Retrieval. N. Belkin, P. Ingwersen, and A. M. Pejtersen, eds. Copenhagen, ACM Press, New York: 37–50.

- Lewis, D. D. (1992b). *Representation and Learning in Information Retrieval*. Ph.D. thesis, Department of Computer Science, University of Massachusetts.
- Lewis, D. D. (1995a). *Evaluating and Optimizing Autonomous Text Classification Systems*. In Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval. E. A. Fox, P. Ingwersen, and R. Fidel, eds. Seattle, ACM Press, New York: 246–254.
- Lewis, D. D. (1995b). “A Sequential Algorithm for Training Text Classifiers: Corrigendum and Additional Data.” *SIGIR Forum* **29**(2): 13–19.
- Lewis, D. D. (1995c). *The TREC-4 Filtering Track: Description and Analysis*. In Proceedings of TREC-4, 4th Text Retrieval Conference. D. K. Warmon, and E. M. Voorhees, eds. Gaithersburg, MD, National Institute of Standards and Technology, Gaithersburg, MD: 165–180.
- Lewis, D. D. (1997). “Reuters-21578 Text Categorization Test Collection. Distribution 1.0.” AT&T Labs-Research, <http://www.research.att.com/lewis>.
- Lewis, D. D. (1998). *Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval*. In Proceedings of ECML-98, 10th European Conference on Machine Learning. C. Nédellec and C. Rouveirol, eds. Chemnitz, Germany, Springer-Verlag, Heidelberg: 4–15.
- Lewis, D. D. (2000). *Machine Learning for Text Categorization: Background and Characteristics*. In Proceedings of the 21st Annual National Online Meeting. M. E. Williams, ed. New York, Information Today, Medford, OR: 221–226.
- Lewis, D. D., and Gale, W. A. (1994). *A Sequential Algorithm for Training Text Classifiers*. In Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval. W. B. Croft and C. J. v. Rijsbergen, eds. Dublin, Springer-Verlag, Heidelberg: 3–12.
- Lewis, D. D., and Hayes, P. J. (1994). “Guest Editors’ Introduction to the Special Issue on Text Categorization.” *ACM Transactions on Information Systems* **12**(3): 231.
- Lewis, D. D., Li, F., Rose, T., and Yang, Y. (2003). “Reuters Corpus Volume I as a Text Categorization Test Collection.” *Journal of Machine Learning Research* **5**: 361–391.
- Lewis, D. D., and Ringuette, M. (1994). *A Comparison of Two Learning Algorithms for Text Categorization*. In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval. Las Vegas, NV, IRSI, University of Nevada, Las Vegas: 81–93.
- Lewis, D. D., Schapire, R. E., Callan, J. P., and Papka, R. (1996). *Training Algorithms for Linear Text Classifiers*. In Proceedings of SIGIR-96, 19th ACM International Conference on Research and Development in Information Retrieval. Zurich, ACM Press, New York: 298–306.
- Lewis, D. D., Stern, D. L., and Singhal, A. (1999). *Attics: A Software Platform for On-line Text Classification*. In Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval. M. A. Wearst, F. Gey, and R. Tong, eds. Berkeley, CA, ACM Press, New York: 267–268.
- Li, C., Wen, J.-R., and Li, H. (2003). *Text Classification Using Stochastic Keyword Generation*. In Proceedings of ICML-03, 20th International Conference on Machine Learning. Washington, DC, Morgan Kaufmann Publishers, San Francisco: 469–471.
- Li, F., and Yang, Y. (2003). *A Loss Function Analysis for Classification Methods in Text Categorization*. In Proceedings of ICML-03, 20th International Conference on Machine Learning. Washington, DC, Morgan Kaufmann Publishers, San Francisco: 472–479.
- Li, H., and Yamanishi, K. (1997). *Document Classification Using a Finite Mixture Model*. In Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics. P. Cohen and W. Wahlster, eds. Madrid, Morgan Kaufmann Publishers, San Francisco: 39–47.
- Li, H., and Yamanishi, K. (1999). *Text Classification Using ESC-Based Stochastic Decision Lists*. In Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management. Kansas City, MO, ACM Press, New York: 122–130.

- Li, H., and Yamanishi, K. (2002). "Text Classification Using ESC-based Stochastic Decision Lists." *Information Processing and Management* **38**(3): 343–361.
- Li, W., Lee, B., Krausz, F., and Sahin, K. (1991). *Text Classification by a Neural Network*. In Proceedings of the 23rd Annual Summer Computer Simulation Conference. D. Pace, ed. Baltimore, Society for Computer Simulation, San Diego, CA: 313–318.
- Li, X., and Roth, D. (2002). *Learning Question Classifiers*. In Proceedings of COLING-02, 19th International Conference on Computational Linguistics. Taipei, Taiwan, Morgan Kaufmann Publishers, San Francisco: 556–562.
- Li, Y. H., and Jain, A. K. (1998). "Classification of Text Documents." *The Computer Journal* **41**(8): 537–546.
- Liang, J., Phillips, I., Ha, J., and Haralick, R. (1996). *Document Zone Classification Using the Sizes of Connected Components*. In Proceedings of Document Recognition III. San Jose, CA, SPIE, Bellingham, WA: 150–157.
- Liang, J., Phillips, I., and Haralick, R. (1997). *Performance Evaluation of Document Layout Analysis on the UW Data Set*. In Proceedings of Document Recognition IV. San Jose, CA, SPIE, Bellingham, WA: 149–160.
- Liao, Y., and Vemuri, V. R. (2002). *Using Text Categorization Techniques for Intrusion Detection*. In Proceedings of the 11th USENIX Security Symposium. D. Boneh, ed. San Francisco: 51–59.
- Liddy, E. D., Paik, W., and Yu, E. S. (1994). "Text Categorization for Multiple Users Based on Semantic Features from a Machine-Readable Dictionary." *ACM Transactions on Information Systems* **12**(3): 278–295.
- Liere, R., and Tadepalli, P. (1997). *Active Learning with Committees for Text Categorization*. In Proceedings of AAAI-97, 14th Conference of the American Association for Artificial Intelligence. Providence, RI, AAAI Press, Menlo Park, CA: 591–596.
- Liere, R., and Tadepalli, P. (1998). *Active Learning with Committees: Preliminary Results in Comparing Winnow and Perceptron in Text Categorization*. In Proceedings of CONALD-98, 1st Conference on Automated Learning and Discovery. Pittsburgh, PA, AAAI Press, Menlo Park, CA.
- Lim, J. H. (1999). *Learnable Visual Keywords for Image Classification*. In Proceedings of DL-99, 4th ACM Conference on Digital Libraries. E. A. Fox and N. Rowe, eds. Berkeley, CA, ACM Press, New York: 139–145.
- Lima, L. R. D., Laender, A. H., and Ribeiro-Neto, B. A. (1998). A Hierarchical Approach to the Automatic Categorization of Medical Documents. In Proceedings of CIKM-98, 7th ACM International Conference on Information and Knowledge Management. G. Gardarin, G. J. French, N. Pissinou, K. Makki, and L. Bouganim, eds. Bethesda, MD, ACM Press, New York: 132–139.
- Lin, D. (1995). "A Dependency-based Method for Evaluating Broad-Coverage Parsers." *Natural Language Engineering* **4**(2): 97–114.
- Lin, X. (1992). *Visualization for the Document Space*. In Proceedings of Visualization '92. Los Alamitos, CA, Center for Computer Legal Research, Pace University/IEEE Computer Society Press, Piscataway, NJ: 274–281.
- Lin, X. (1997). "Map Displays for Information Retrieval." *Journal of the American Society for Information Science* **48**: 40–54.
- Lin, X., Soergel, D., and Marchionini, G. (1991). *A Self-Organizing Semantic Map for Information Retrieval*. In Proceedings of 14th Annual International ACM/SIGIR Conference on Research & Development in Information Retrieval. Chicago, ACM Press, New York: 262–269.
- Litman, D. J., and Passonneau, R. J. (1995). *Combining Multiple Knowledge Sources for Discourse Segmentation*. In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics. Cambridge, MA, Association for Computational Linguistics, Morristown, NJ: 108–115.

- Liu, H., Selker, T., and Lieberman, H. (2003). *Visualizing the Affective Structure of a Text Document*. In Proceedings of the Conference on Human Factors in Computing Systems (CHI 2003). Fort Lauderdale, FL, ACM Press, New York: 740–741.
- Liu, X., and Croft, W. B. (2003). “Statistical Language Modeling for Information Retrieval.” *Annual Review of Information Science and Technology* **39**.
- Liu, Y., Carbonell, J., and Jin, R. (2003). *A New Pairwise Ensemble Approach for Text Classification*. In Proceedings of ECML-03, 14th European Conference on Machine Learning. N. Lavrac, D. Gamberger, L. Todorovski, and H. Blockeel, eds. Cavtat-Dubrovnik, Croatia, Springer-Verlag, Heidelberg: 277–288.
- Liu, Y., Yang, Y., and Carbonell, J. (2002). *Boosting to Correct the Inductive Bias for Text Classification*. In Proceedings of CIKM-02, 11th ACM International Conference on Information and Knowledge Management. McLean, VA, ACM Press, New York: 348–355.
- Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., and Watkins, C. (2002). “Text Classification Using String Kernels.” *Journal of Machine Learning Research* **2**: 419–444.
- Lodhi, H., Shawe-Taylor, J., Cristianini, N., and Watkins, C. J. (2001). “Discrete Kernels for Text Categorisation.” In *Advances in Neural Information Processing Systems*. T. K. Leen, T. Ditterich, and V. Tresp, eds. Cambridge, MA, MIT Press: 563–569.
- Lombardo, V. (1991). *Parsing Dependency Grammars*. In Proceedings of the 2nd Congress of the Italian Association for Artificial Intelligence on Trends in Artificial Intelligence. E. Ardiszone, S. Gaglio, and F. Sorbello, eds. Springer-Verlag, London: 291–300.
- Lorrain, F., and White, H. C. (1971). “Structural Equivalence of Individuals in Social Networks.” *Journal of Mathematical Sociology* **1**: 49–80.
- Lu, S. Y., and Fu, K. S. (1978). “A Sentence-to-Sentence Clustering Procedure for Pattern Analysis.” *IEEE Transactions on Systems, Man and Cybernetics*. **8**: 381–389.
- Di., Nunzio, G. M., and Micarelli, A. (2003). *Does a New Simple Gaussian Weighting Approach Perform Well in Text Categorization?* In Proceedings of IJCAI-03, 18th International Joint Conference on Artificial Intelligence. Acapulco, Morgan Kaufmann Publishers, San Francisco: 581–586.
- Macskassy, S. A., Hirsh, H., Banerjee, A., and Dayanik, A. A. (2001). *Using Text Classifiers for Numerical Classification*. In Proceedings of IJCAI-01, 17th International Joint Conference on Artificial Intelligence. B. Nebel, ed. Seattle, Morgan Kaufmann Publishers, San Francisco: 885–890.
- Macskassy, S. A., Hirsh, H., Banerjee, A., and Dayanik, A. A. (2003). “Converting Numerical Classification into Text Classification.” *Artificial Intelligence* **143**(1): 51–77.
- Maderlechner, G., Suda, P., and Bruckner, T. (1997). “Classification of Documents by Form and Content.” *Pattern Recognition Letters* **18**(11/13): 1225–1231.
- Maedche, A., and Staab, S. (2001). “Learning Ontologies for the Semantic Web.” *IEEE Intelligent Systems* **16**(2), Special Issue on the Semantic Web.
- Maltese, G., and Mancini, F. (1991). *A Technique to Automatically Assign Parts-of-Speech to Words Taking into Account Word-Ending Information through a Probabilistic Model*. In Proceedings of Eurospeech 1991. Genoa, Italy, Genovale Institute fuer Kommunikations Forschung und Phonetick, Bonn, Germany: 753–756.
- Manevitz, L. M., and Yousef, M. (2001). “One-Class SVMs for Document Classification.” *Journal of Machine Learning Research* **2**: 139–154.
- Mannila, H., and Toivonen, H. (1996). *On an Algorithm for Finding All Interesting Sentences*. In Proceedings of the 13th European Meeting on Cybernetics and Systems Research. R. Trappl, ed. Vienna, Austria, University of Helsinki, Department of Computer Science: 973–978.
- Mannila, H., Toivonen, H., and Verkamo, A. (1994). *Efficient Algorithms for Discovering Association Rules*. In Proceedings of Knowledge Discovery in Databases, AAAI Workshop (KDD’94). U. M. Eayyad and R. Uthrusamy, eds. Seattle, AAAI Press, Menlo Park, CA: 181–192.

- Mannila, H., Toivonen, H., and Verkamo, A. (1995). *Discovering Frequent Episodes in Sequences*. In Proceedings of the 1st International Conference of Knowledge Discovery and Data Mining. Montreal, AAAI Press, Menlo Park, CA: 210–215.
- Mannila, H., Toivonen, H., and Verkamo, A. (1997). “Discovery of Frequent Episodes in Event Sequences.” *Data Mining and Knowledge Discovery* 1(3): 259–289.
- Manning, C., and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA, MIT Press.
- Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge, UK, Cambridge University Press.
- Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1994). “Building a Large Annotated Corpus of English: The Penn Treebank.” *Computational Linguistics* 19(2): 313–330.
- Maron, M. E. (1961). “Automatic Indexing: An Experimental Inquiry.” *Journal of the Association for Computing Machinery* 8(3): 404–417.
- Martin, P. (1995). *Using the WordNet Concept Catalog and a Relation Hierarchy for Knowledge Acquisition*. In Proceedings of Peirce’95, 4th International Workshop on Peirce. E. Ellis and R. Levinson, eds. Santa Cruz, CA, University of Maryland, MD: 36–47.
- Masand, B. (1994). Optimising Confidence of Text Classification by Evolution of Symbolic Expressions. In *Advances in Genetic Programming*. K. E. Kinnear, ed. Cambridge, MA, MIT Press: 459–476.
- Masand, B., Linoff, G., and Waltz, D. (1992). *Classifying News Stories Using Memory-Based Reasoning*. In Proceedings of SIGIR-92, 15th ACM International Conference on Research and Development in Information Retrieval. N. Belkin, P. Ingwersen, and A. M. Pejtersen, eds. Copenhagen, Denmark, ACM Press, New York: 59–65.
- Masui, T., Minakuchi, M., Borden, G., and Kashiwagi, K. (1995). *Multiple-View Approach for Smooth Information Retrieval*. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST’95). G. Robertson, ed. Pittsburgh, ACM Press, New York: 199–206.
- Matsuda, K., and Fukushima, T. (1999). *Task-Oriented World Wide Web Retrieval by Document-Type Classification*. In Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management. S. Gruch, ed. Kansas City, MO, ACM Press, New York: 109–113.
- McCallum, A., Freitag, D., and Pereira, F. (2000). *Maximum Entropy Markov Models for Information Extraction and Segmentation*. In Proceedings of the 17th International Conference on Machine Learning. Stanford University, Palo Alto, CA, Morgan Kaufmann Publishers, San Francisco: 591–598.
- McCallum, A., and Jensen, D. (2003). *A Note on the Unification of Information Extraction and Data Mining Using Conditional-Probability, Relational Models*. In Proceedings of IJCAI03 Workshop on Learning Statistical Models from Relational Data. D. Jensen and L. Getoo, eds. Acapulco, Mexico, published electronically by IJCAI and AAAI: 79–87.
- McCallum, A. K., and Nigam, K. (1998). *Employing EM in Pool-Based Active Learning for Text Classification*. In Proceedings of ICML-98, 15th International Conference on Machine Learning. J. W. Shavlik, ed. Madison, WI, Morgan Kaufmann Publishers, San Francisco: 350–358.
- McCallum, A. K., Rosenfeld, R., Mitchell, T. M., and Ng, A. Y. (1998). *Improving Text Classification by Shrinkage in a Hierarchy of Classes*. In Proceedings of ICML-98, 15th International Conference on Machine Learning. J. W. Shavlik, ed. Madison, WI, Morgan Kaufmann Publishers, San Francisco: 359–367.
- McCarthy, J. F., and Lehnert, W. G. (1995). *Using Decision Trees for Coreference Resolution*. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95). C. Mellish, ed. Montreal, Morgan Kaufmann Publishers, San Francisco: 1050–1055.

- Melancon, G., and Herman, I. (2000). *DAG Drawing from an Information Visualization Perspective*. In Proceedings of Data Visualization '00, Amsterdam, Springer-Verlag, Heidelberg: 3–12.
- Meretakis, D., Fragoudis, D., Lu, H., and Likothanassis, S. (2000). *Scalable Association-Based Text Classification*. In Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management. A. Agoh, J. Callan, S. Gauch, and E. Rundensteiner, eds. McLean, VA, ACM Press, New York: 373–374.
- Merialdo, B. (1994). “Tagging English text with a Probabilistic Model.” *Computational Linguistics* **20**(2): 155–172.
- Merkel, D. (1998). “Text Classification with Self-Organizing Maps: Some Lessons Learned.” *Neurocomputing* **21**(1/3): 61–77.
- Miller, D., Schwartz, R., Weischedel, R., and Stone, R. (1999). *Named Entity Extraction from Broadcast News*. In Proceedings of DARPA Broadcast News Workshop. Herndon, VA, Morgan Kaufmann Publishers, San Francisco: 37–40.
- Miller, N., Wong, P. C., Brewster, M., and Foote, H. (1998). *TOPIC ISLANDS(TM): A Wavelet-Based Text Visualization System*. In Proceedings of IEEE Visualization '98. Research Triangle Park, NC, ACM Press, New York: 189–196.
- Mitkov, R. (1998). *Robust Pronoun Resolution with Limited Knowledge*. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. Montreal, Canada, Association for Computational Linguistics, Morristown, NJ: 869–875.
- Mladenic, D. (1998a). *Feature Subset Selection in Text Learning*. In Proceedings of ECML-98, 10th European Conference on Machine Learning. C. Nedellec and C. Rouveirol, eds. Chemnitz, Germany, Springer-Verlag, London: 95–100.
- Mladenic, D. (1998b). *Machine Learning on Non-homogeneous, Distributed Text Data*. Ph.D. thesis, J. Stefan Institute, University of Ljubljana.
- Mladenic, D. (1998c). *Turning Yahoo! into an Automatic Web Page Classifier*. In Proceedings of ECAI-98, 13th European Conference on Artificial Intelligence. H. Prade, ed. Brighton, UK, John Wiley and Sons, Chichester, UK: 473–474.
- Mladenic, D. (1999). “Text Learning and Related Intelligent Agents: A Survey.” *IEEE Intelligent Systems* **14**(4): 44–54.
- Mladenic, D., and Grobelnik, M. (1998). *Word Sequences as Features in Text-Learning*. In Proceedings of ERK-98, 7th Electrotechnical and Computer Science Conference. Ljubljana, Slovenia: 145–148.
- Mladenic, D., and Grobelnik, M. (1999). *Feature Selection for Unbalanced Class Distribution and Naive Bayes*. In Proceedings of ICML-99, 16th International Conference on Machine Learning. I. Bratko and S. Dzeroski, eds. Bled, Slovenia, Morgan Kaufmann Publishers, San Francisco: 258–267.
- Mladenic, D., and Grobelnik, M. (2003). “Feature Selection on Hierarchy of Web Documents.” *Decision Support Systems* **35**(1): 45–87.
- Mock, K. (1998). *A Comparison of Three Document Clustering Algorithms: TreeCluster, Word Intersection GQF, and Word Intersection Hierarchical Agglomerative Clustering*. Technical Report, Intel Architecture Labs.
- Moens, M.-F., and Dumortier, J. (2000). “Text Categorization: The Assignment of Subject Descriptors to Magazine Articles.” *Information Processing and Management* **36**(6): 841–861.
- Montes-y-Gomez, M., Gelbukh, A., and Lopez-Lopez, A. (2001a). *Discovering Association Rules in Semi-Structured Data Sets*. In Proceedings of the Workshop on Knowledge Discovery from Distributed, Dynamic, Heterogeneous, Autonomous Data and Knowledge Source at 17th International Joint Conference on Artificial Intelligence (IJCAI'2001). Seattle, AAAI Press, Menlo Park, CA: 26–31.
- Montes-y-Gomez, M., Gelbukh, A., and Lopez-Lopez, A. (2001b). “Mining the News: Trends, Associations and Deviations.” *Computación y Sistemas* **5**(1): 14–25.

- Mooney, R. J., and Roy, L. (2000). *Content-Based Book Recommending Using Learning for Text Categorization*. Proceedings of DL-00, 5th ACM Conference on Digital Libraries. San Antonio, TX, ACM Press, New York: 195–204.
- Moschitti, A. (2003). *A Study on Optimal Parameter Tuning for Rocchio Text Classifier*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Heidelberg: 420–435.
- Mostafa, J., and Lam, W. (2000). “Automatic Classification Using Supervised Learning in a Medical Document Filtering Application.” *Information Processing and Management* 36(3): 415–444.
- Moulinier, I. (1997). *Feature Selection: A Useful Preprocessing Step*. In Proceedings of BCSIRSG-97, 19th Annual Colloquium of the British Computer Society Information Retrieval Specialist Group. J. Furner and D. Harper, eds. Aberdeen, UK, Springer-Verlag, Heidelberg, Germany: 1–11.
- Moulinier, I., and Ganascia, J.-G. (1996). “Applying an Existing Machine Learning Algorithm to Text Categorization.” In *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*. S. Wermter, E. Riloff, and G. Scheler, eds. Heidelberg, Springer-Verlag: 343–354.
- Moulinier, I., Raskinis, G., and Ganascia, J.-G. (1996). *Text Categorization: A Symbolic Approach*. In Proceedings of SDAIR-96, 5th Annual Symposium on Document Analysis and Information Retrieval. Las Vegas, NV, ISRI, University of Nevada, Las Vegas: 87–99.
- Munoz, M., Punyakanok, V., Roth, D., and Zimak, D. (1999). *A Learning Approach to Shallow Parsing*. Technical Report 2087, University of Illinois at Urbana-Champaign: 18.
- Munzner, T., and Burchard, P. (1995). *Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space*. In Proceedings of VRML '95. San Diego, CA, ACM Press, New York: 33–38.
- Mutton, P. (2004). “Inferring and Visualizing Social Networks on Internet Relay Chat.” *Journal of WSCG* 12(1–3).
- Mutton, P., and Golbeck, J. (2003). *Visualization of Semantic Metadata and Ontologies*. In Proceedings of Information Visualization 2003 (IV03). London, UK, IEEE Computer Society Press, Washington, DC: 300.
- Mutton, P., and Rodgers, P. (2002). *Spring Embedder Preprocessing for WWW Visualization*. In Proceedings of 6th International Conference on Information Visualization. London, IEEE Computer Society Press, Washington, DC: 744–749.
- Myers, K., Kearns, M., Singh, S., and Walker, M. A. (2000). *A Boosting Approach to Topic Spotting on Subdialogues*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco: 655–662.
- Nahm, U., and Mooney, R. (2000). *A Mutually Beneficial Integration of Data Mining and Information Extraction*. In Proceedings of the 17th Conference of Artificial Intelligence, AAAI-2000. Austin, TX, AAAI Press, Menlo Park, CA: 627–632.
- Nahm, U., and Mooney, R. (2001). *Mining Soft Matching Rules from Text Data*. In Proceedings of the 7th International Joint Conference on Artificial Intelligence. Seattle, WA, Morgan Kaufmann Publishers, San Francisco: 978–992.
- Nahm, U. Y., and Mooney, R. J. (2002). *Text Mining with Information Extraction*. In Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases. S. Harabagio and V. Chaudhri, eds. Palo Alto, CA, AAAI Press, Menlo Park, CA: 60–68.
- Nardiello, P., Sebastiani F., and Sperduti, A. (2003). *Discretizing Continuous Attributes in AdaBoost for Text Categorization*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Heidelberg: 320–334.

- Nasukawa, T., and Nagano, T. (2001). "Text Analysis and Knowledge Mining System." *IBM Systems Journal* 40(4): 967–984.
- Neuhaus, P., and Broker, N. (1997). *The Complexity of Recognition of Linguistically Adequate Dependency Grammars*. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics. P. R. Cohen and W. Wahlster, eds. Somerset, NJ, Association for Computational Linguistics: 337–343.
- Ng, G. K.-C. (2000). *Interactive Visualisation Techniques for Ontology Development*. Ph.D. thesis, Department of Computer Science, University of Manchester.
- Ng, H. T., Goh, W. B., and Low, K. L. (1997). *Feature Selection, Perceptron Learning, and a Usability Case Study for Text Categorization*. In Proceedings of SIGIR-97, 20th ACM International Conference on Research and Development in Information Retrieval. N. J. Belkin, A. Narasimhalu, W. Hersch, and P. Willett, eds. Philadelphia, ACM Press, New York: 67–73.
- Ng, V., and Cardie, C. (2002). *Improving Machine Learning Approaches to Coreference Resolution*. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, Association for Computational Linguistics, Morristown, NJ: 104–111.
- Ng, V., and Cardie, C. (2003). *Bootstrapping Coreference Classifiers with Multiple Machine Learning Algorithms*. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-2003), Sappora, Japan, Association for Computational Linguistics, Morristown, NJ: 113–120.
- Nigam, K. (2001). *Using Unlabeled Data to Improve Text Classification*. Ph.D. thesis, Computer Science Department, Carnegie Mellon University.
- Nigam, K., and Ghani, R. (2000). *Analyzing the Applicability and Effectiveness of Co-training*. In Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management. A. Agah, J. Callan, S. Gauch, and E. Rundensteiner, eds. McLean, VA, ACM Press, New York: 86–93.
- Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T. M. (1998). *Learning to Classify Text from Labeled and Unlabeled Documents*. In Proceedings of AAAI-98, 15th Conference of the American Association for Artificial Intelligence. Madison, WI, AAAI Press, Menlo Park, CA: 792–799.
- Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T. M. (2000). "Text Classification from Labeled and Unlabeled Documents Using EM." *Machine Learning* 39(2/3): 103–134.
- Niyogi, D. (1995). *A Knowledge-Based Approach to Deriving Logical Structure from Document Images*. Doctoral dissertation, State University of New York, Buffalo.
- Niyogi, D., and Srihari, S. (1996). *Using Domain Knowledge to Derive the Logical Structure of Documents*. In Proceedings of Document Recognition III. SPIE, Bellingham, WA: 114–125.
- Noik, E. (1996). *Dynamic Fisheye Views: Combining Dynamic Queries and Mapping with Database View Definition*. Ph.D. thesis, Graduate Department of Computer Science, University of Toronto.
- Nong, Y., ed. (2003). *The Handbook of Data Mining*. Boston, Lawrence Erlbaum Associates.
- Oh, H.-J., Myaeng, S. H., and Lee, M.-H. (2000). *A Practical Hypertext Categorization Method Using Links and Incrementally Available Class Information*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. N. Belkin, P. Ingwersen, and M.-K. Leong, eds. Athens, ACM Press, New York: 264–271.
- Ontrup, J., and Ritter, H. (2001a). *Hyperbolic Self-Organizing Maps for Semantic Navigation*. In Proceedings of NIPS 2001. T. Dietterich, S. Becker, and Z. Chahramani, eds. Vancouver, MIT Press, Cambridge, MA: 1417–1424.
- Ontrup, J., and Ritter, H. (2001b). *Text Categorization and Semantic Browsing with Self-Organizing Maps on Non-Euclidean Spaces*. In Proceedings of PKDD-01, 5th European

- Conference on Principles and Practice of Knowledge Discovery in Databases. Freiburg, Germany, Springer-Verlag, Heidelberg: 338–349.
- Paijmans, H. (1999). “Text Categorization as an Information Retrieval Task.” *The South African Computer Journal*. 31: 4–15.
- Paliouras, G., Karkaletsis, V., and Spyropoulos, C. D. (1999). *Learning Rules for Large Vocabulary Word Sense Disambiguation*. In Proceedings of IJCAI-99, 16th International Joint Conference on Artificial Intelligence. T. Dean, ed. Stockholm, Morgan Kaufmann Publishers, San Francisco: 674–679.
- Pang, B., Lee, L., and Vaithyanathan, S. (2002). *Thumbs Up? Sentiment Classification Using Machine Learning Techniques*. In Proceedings of EMNLP-02, 7th Conference on Empirical Methods in Natural Language Processing. Philadelphia, Association for Computational Linguistics, Morristown, NJ: 79–86.
- Patel-Schneider, P., and Simeon, J. (2002). *Building the Semantic Web on XML*. In Proceedings of the 1st International Semantic Web Conference (ISWC). I. Horrocks and J. Hendler, eds. Sardinia, Italy, Springer-Verlag, Heidelberg, Germany: 147–161.
- Pattison, T., Vernik, R., Goodburn, D., and Phillips, M. (2001). *Rapid Assembly and Deployment of Domain Visualisation Solutions*. In Proceedings of Australian Symposium on Information Visualization, ACM International Conference. Sydney, Australian Computer Society, Darlinghurst, Australia: 19–26.
- Pedersen, T., and Bruce, R. (1997). *Unsupervised Text Mining*. Dallas, TX, Department of Computer Science and Engineering, Southern Methodist University.
- Peng, F., and Schuurmans, D. (2003). *Combining Naive Bayes n-gram and Language Models for Text Classification*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Heidelberg: 335–350.
- Peng, F., Schuurmans, D., and Wang, S. (2003). *Language and Task Independent Text Categorization with Simple Language Models*. In Proceedings of HLT-03, 3rd Human Language Technology Conference. Edmonton, CA, ACL Press, Morgan Kaufmann Publishers, San Francisco: 110–117.
- Petasis, G., Cucchiarelli, A., Velardi, P., Paliouras, G., Karkaletsis, V., and Spyropoulos, C. D. (2000). *Automatic Adaptation of Proper Noun Dictionaries through Cooperation of Machine Learning and Probabilistic Methods*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. N. Belkin, Peter Ingwersen, and M.-K. Leong, eds. Athens, ACM Press, New York: 128–135.
- Peters, C., and Koster, C. H. (2002). *Uncertainty-Based Noise Reduction and Term Selection in Text Categorization*. In Proceedings of ECIR-02, 24th European Colloquium on Information Retrieval Research. F. Crestani, M. Girolomi, and C. J. v. Rijsbergen, eds. Glasgow, Springer-Verlag, London: 248–267.
- Phillips, W., and Riloff, E. (2002). *Exploiting Strong Syntactic Heuristics and Co-Training to Learn Semantic Lexicons*. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002). Philadelphia, Association for Computational Linguistics: 125–132.
- Piatetsky-Shapiro, G., and Frawley, W. J., eds. (1991). *Knowledge Discovery in Databases*. Cambridge, MA, MIT Press.
- Pierre, J. M. (2002). *Mining Knowledge from Text Collections Using Automatically Generated Metadata*. In Proceedings of the 4th International Conference on Practical Aspects of Knowledge Management (PAKM-02). D. Karagiannis and Reimer, eds. Vienna, Austria, Springer-Verlag, London: 537–548.
- Pollard, C., and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, University of Chicago Press and CSLI Publications.
- Porter, A. (2002). *Text Mining*. Technology Policy and Assessment Center, Georgia Institute of Technology.
- Pottenger, W., and Yang, T.-h. (2001). *Detecting Emerging Concepts in Textual Data Mining*. Philadelphia, SIAM.

- Punyakanok, V., and Roth, D. (2000). *Shallow Parsing by Inferencing with Classifiers*. In Proceedings of the 4th Conference on Computational Natural Language Learning and of the 2nd Learning Language in Logic Workshop. Lisbon, Association for Computational Linguistics, Somerset, NJ: 107–110.
- Pustejovsky, J., Castano, J., Zhang, J., Kotecki, M., and Cochran, B. (2002). *Robust Relational Parsing over Biomedical Literature: Extracting Inhibit Relations*. In Proceedings of the 2002 Pacific Symposium on Biocomputing (PSB-2002). Lihue, Hawaii, World Scientific Press, Hackensack, NJ: 362–373.
- Rabiner, L. R. (1986). “An Introduction to Hidden Markov Models.” *IEEE ASSP Magazine* 3(1): 4–16.
- Rabiner, L. R. (1990). “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.” In *Readings in Speech Recognition*. A. Waibel and K.-F. Lee, eds. Los Altos, CA, Morgan Kaufmann Publishers: 267–296.
- Ragas, H., and Koster, C. H. (1998). *Four Text Classification Algorithms Compared on a Dutch Corpus*. In Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval. W. B. Croft, A. Moffat, C. J. v. Rijsbergen, R. Wilkinson and J. Zobel, eds. Melbourne, Australia, ACM Press, New York: 369–370.
- Rainsford, C., and Roddick, J. (2000). *Visualization of Temporal Interval Association Rules*. In Proceedings of the 2nd International Conference on Intelligent Data Engineering and Automated Learning. Hong Kong, Springer-Verlag, London: 91–96.
- Rajman, M., and Besancon, R. (1997a). *A Lattice Based Algorithm for Text Mining*. Technical Report TR-LIA-LN1/97, Swiss Federal Institute of Technology.
- Rajman, M., and Besancon, R. (1997b). *Text Mining: Natural Language Techniques and Text Mining Applications*. In Proceedings of the 7th IFIP 2.6 Working Conference on Database Semantics (DS-7). Leysin, Switzerland, Norwell, MA.
- Rajman, M., and Besancon, R. (1998). *Text Mining – Knowledge Extraction from Unstructured Textual Data*. In Proceedings of the 6th Conference of the International Federation of Classification Societies. Rome: 473–480.
- Rambow, O., and Joshi, A. K. (1994). “A Formal Look at Dependency Grammars and Phrase-Structure Grammars, with Special Consideration of Word-Order Phenomena.” In *Current Issues in Meaning-Text Theory*. L. Wanner, ed. London, Pinter.
- Rao, R., and Card, S. (1994). *The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus + Context Visualization for Tabular Information*. In Proceedings of the International Conference on Computer-Human Interaction '94. Boston, MA, ACM Press, New York: 318–322.
- Rao, R., Card, S., Jelinek, H., Mackinlay, J., and Robertson, G. (1992). *The Information Grid: A Framework for Information Retrieval and Retrieval-Centered Applications*. In Proceedings of the 5th Annual Symposium on User Interface Software and Technology (UIST) '92. Monterey, CA, ACM Press, New York: 23–32.
- Raskutti, B., Ferra, H., and Kowalczyk, A. (2001). *Second Order Features for Maximising Text Classification Performance*. In Proceedings of ECML-01, 12th European Conference on Machine Learning. L. D. Raedt and P. A. Flach, eds. Freiburg, Germany, Springer-Verlag, London: 419–430.
- Rau, L. F., and Jacobs, P. S. (1991). *Creating Segmented Databases from Free Text for Text Retrieval*. In Proceedings of SIGIR-91, 14th ACM International Conference on Research and Development in Information Retrieval. Chicago, ACM Press, New York: 337–346.
- Reape, M. (1989). *A Logical Treatment of Semi-free Word Order and Bounded Discontinuous Constituency*. In Proceedings of the 4th Meeting of the European ACL. Monchester, UK, Association for Computational Linguistics, Morristown, NJ: 103–110.
- Rennie, J., and McCallum, A. K. (1999). *Using Reinforcement Learning to Spider the Web Efficiently*. In Proceedings of ICML-99, 16th International Conference on Machine Learning. I. Bratko and S. Dzeroski, eds. Bled, Slovenia, Morgan Kaufmann Publishers, San Francisco: 335–343.

- Rennie, J., Shih, L., Teevan, J., and Karger, D. (2003). *Tackling the Poor Assumptions of Naive Bayes Text Classifiers*. In Proceedings of ICML-03, 20th International Conference on Machine Learning. Washington, DC, Morgan Kaufmann Publishers, San Francisco: 616–623.
- Reynar, J., and Ratnaparkhi, A. (1997). *A Maximum Entropy Approach to Identifying Sentence Boundaries*. In Proceedings of the 5th Conference on Applied Natural Language Processing. Washington, DC, Morgan Kaufmann Publishers, San Francisco: 16–19.
- Ribeiro-Neto, B., Laender, A. H. F., and Lima, L. R. D. (2001). “An Experimental Study in Automatically Categorizing Medical Documents.” *Journal of the American Society for Information Science and Technology* **52**(5): 391–401.
- Rich, E., and LuperFoy, S. (1988). *An Architecture for Anaphora Resolution*. In ACL Proceedings of the 2nd Conference on Applied Natural Language Processing. Austin, TX, Association for Computational Linguistics, Morristown, NJ: 18–24.
- Rijsbergen, C. J. v. (1979). *Information Retrieval*, 2nd ed. London, Butterworths.
- Riloff, E. (1993a). *Automatically Constructing a Dictionary for Information Extraction Tasks*. In Proceedings of the 11th National Congress on Artificial Intelligence. Washington, DC, AAAI/MIT Press, Menlo Park, CA: 811–816.
- Riloff, E. (1993b). *Using Cases to Represent Context for Text Classification*. In Proceedings of CIKM-93, 2nd International Conference on Information and Knowledge Management. Washington, DC, ACM Press, New York: 105–113.
- Riloff, E. (1994). *Information Extraction as a Basis for Portable Text Classification Systems*. Amherst, MA, Department of Computer Science, University of Massachusetts.
- Riloff, E. (1995). *Little Words Can Make a Big Difference for Text Classification*. In Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval. E. A. Fox, P. Ingwersen, and R. Fidel, eds. Seattle, ACM Press, New York: 130–136.
- Riloff, E. (1996a). *Automatically Generating Extraction Patterns from Untagged Text*. In Proceedings of the 13th National Conference on Artificial Intelligence. AAAI/MIT Press, Menlo Park, CA: 1044–1049.
- Riloff, E. (1996b). “Using Learned Extraction Patterns for Text Classification.” In *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*. S. Wermter, E. Riloff, and G. Scheler, eds. Springer-Verlag, London: 275–289.
- Riloff, E., and Jones, R. (1999). *Learning Dictionaries for Information Extraction by Multi-level Boot-Strapping*. In Proceedings of the 16th National Conference on Artificial Intelligence. Orlando, AAAI Press/MIT Press, Menlo Park, CA: 1044–1049.
- Riloff, E., and Lehnert, W. (1994). “Information Extraction as a Basis for High-Precision Text Classification.” *ACM Transactions on Information Systems*, **12**(3): 296–333.
- Riloff, E., and Lehnert, W. (1998). *Classifying Texts Using Relevancy Signatures*. In Proceedings of AAAI-92, 10th Conference of the American Association for Artificial Intelligence. San Jose, CA, AAAI Press, Menlo Park, CA: 329–334.
- Riloff, E., and Lorenzen, J. (1999). “Extraction-Based Text Categorization: Generating Domain-Specific Role Relationships.” In *Natural Language Information Retrieval*. T. Strzalkowski, ed. Dordrecht, Kluwer Academic Publishers: 167–196.
- Riloff, E., and Schmelzenbach, M. (1998). *An Empirical Approach to Conceptual Case Frame Acquisition*. In Proceedings of the 6th Workshop on Very Large Corpora. E. Chmiak, ed. Montreal, Quebec, Association for Computational Linguistics, Morgan Kaufmann Publishers, San Francisco: 49–56.
- Riloff, E., and Shoen, J. (1995). *Automatically Acquiring Conceptual Patterns Without an Automated Corpus*. In Proceedings of the 3rd Workshop on Very Large Corpora. Boston, MA, Association for Computational Linguistics, Somerset, NJ: 148–161.
- Rindflesch, T. C., Hunter, L., and Aronson, A. R. (1999). *Mining Molecular Binding Terminology from Biomedical Text*. In Proceedings of the '99 AMIA Symposium. Washington, DC, AMIA, Bethesda, MD: 127–131.

- Rindflesch, T. C., Tanabe, L., Weinstein, J. N., and Hunter, L. (2000). *EDGAR: Extraction of Drugs, Genes and Relations from the Biomedical Literature*. In Proceedings of the 2000 Pacific Symposium on Biocomputing. Waikiki Beach, Hawaii, World Scientific Press, Hackensack, NJ: 517–528.
- Roark, B., and Johnson, M. (1999). *Efficient Probabilistic Top-Down and Left-Corner Parsing*. In Proceedings of the 37th Annual Meeting of the ACL. College Park, MD, Association for Computational Linguistics, Morristown, NJ: 421–428.
- Robertson, G., Mackinlay, J., and Card, S. (1991). *Cone Trees: Animated 3D Visualizations of Hierarchical Information*. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. New Orleans, ACM Press, New York: 189–194.
- Robertson, S. E., and Harding, P. (1984). “Probabilistic Automatic Indexing by Learning from Human Indexers.” *Journal of Documentation* **40**(4): 264–270.
- Rodriguez, M. D. B., Gomez-Hidalgo, J. M., and Diaz-Agudo, B. (1997). *Using WordNet to Complement Training Information in Text Categorization*. In Proceedings of RANLP-97, 2nd International Conference on Recent Advances in Natural Language Processing. R. Mitkov and N. Nikolov, eds. Tzigrav Chark, Bulgaria, John Benjamins, Philadelphia: 353–364.
- Rokita, P. (1996). “Generating Depth-of-Field Effects in Virtual Reality Applications.” *IEEE Computer Graphics and Applications* **16**(2): 18–21.
- Rose, T., Stevenson, M., and Whitehead, M. (2002). *The Reuters Corpus Volume 1 – From Yesterday’s News to Tomorrow’s Language Resources*. In Proceedings of LREC-02, 3rd International Conference on Language Resources and Evaluation. Las Palmas, Spain, ELRA, Paris: 827–832.
- Rosenfeld, B., Feldman, R., Fresko, M., Schler, J., and Aumann, Y. (2004). *TEG: A Hybrid Approach to Information Extraction*. In Proceedings of CIKM 2004. Arlington, VA, ACM Press, New York: 589–596.
- Roth, D. (1998). *Learning to Resolve Natural Language Ambiguities: A Unified Approach*. In Proceedings of AAAI-98, 15th Conference of the American Association for Artificial Intelligence. Madison, WI, AAAI Press, Menlo Park, CA: 806–813.
- Ruiz, M., and Srinivasan, P. (2002). “Hierarchical Text Classification Using Neural Networks.” *Information Retrieval* **5**(1): 87–118.
- Ruiz, M. E., and Srinivasan, P. (1997). *Automatic Text Categorization Using Neural Networks*. In Proceedings of the 8th ASIS/SIGCR Workshop on Classification Research. E. Efthimiadis, ed. Washington, DC, American Society for Information Science, Washington, DC: 59–72.
- Ruiz, M. E., and Srinivasan, P. (1999a). *Combining Machine Learning and Hierarchical Indexing Structures for Text Categorization*. In Proceedings of the 10th ASIS/SIGCR Workshop on Classification Research. Washington, DC, American Society for Information Science, Washington, DC.
- Ruiz, M. E., and Srinivasan, P. (1999b). *Hierarchical Neural Networks for Text Categorization*. In Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval. M. A. Hearst, F. Gey, and R. Tong, eds. Berkeley, CA, ACM Press, New York: 281–282.
- Rzhetsky, A., Iossifov, I., Koike, T., Krauthammer, M., Kra, P., Morris, M., Yu, H., Duboue, P. A., Weng, W., Wilbur, J. W., Hatzivassiloglou, V., and Friedman, C. (2004). “GeneWays: A System for Extracting, Analyzing, Visualizing, and Integrating Molecular Pathway Data.” *Journal of Biomedical Informatics* **37**: 43–53.
- Rzhetsky, A., Koike, T., Kalachikov, S., Gomez, S. M., Krauthammer, M., Kaplan, S. H., Kra, P., Russo, J. J., and Friedman, C. (2000). “A Knowledge Model for Analysis and Simulation of Regulatory Networks.” *Bionformatics* **16**: 1120–1128.
- Sabidussi, G. (1966). “The Centrality Index of a Graph.” *Psychometrika* **31**: 581–603.
- Sable, C., and Church, K. (2001). *Using Bins to Empirically Estimate Term Weights for Text Categorization*. In Proceedings of EMNLP-01, 6th Conference on Empirical Methods in Natural Language Processing. Pittsburgh, Association for Computational Linguistics, Morristown, NJ: 58–66.

- Sable, C. L., and Hatzivassiloglou, V. (1999). *Text-Based Approaches for the Categorization of Images*. In Proceedings of ECDL-99, 3rd European Conference on Research and Advanced Technology for Digital Libraries. S. Abitebout and A.-M. Vercoustre, eds. Paris, Springer-Verlag, Heidelberg: 19–38.
- Sable, C. L., and Hatzivassiloglou, V. (2000). “Text-Based Approaches for Non-topical Image Categorization.” *International Journal of Digital Libraries* 3(3): 261–275.
- Sahami, M., ed. (1998). *Learning for Text Categorization*. Papers from the 1998 AAAI Workshop. Madison, WI, AAAI Press, Menlo Park, CA.
- Sahami, M., Hearst, M. A., and Saund, E. (1996). *Applying the Multiple Cause Mixture Model to Text Categorization*. In Proceedings of ICML-96, 13th International Conference on Machine Learning. L. Saitta, ed. Bari, Italy, Morgan Kaufmann Publishers, San Francisco: 435–443.
- Sahami, M., Yusufali, S., and Baldonado, M. Q. (1998). *SONIA: A Service for Organizing Networked Information Autonomously*. In Proceedings of DL-98, 3rd ACM Conference on Digital Libraries. I. Witten, R. Aksyn, and F. M. Shipman, eds. Pittsburgh, ACM Press, New York: 200–209.
- Sakakibara, Y., Misue, K., and Koshiba, T. (1996). “A Machine Learning Approach to Knowledge Acquisitions from Text Databases.” *International Journal of Human Computer Interaction* 8(3): 309–324.
- Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C. D., and Stamatopoulos, P. (2001). *Stacking Classifiers for Anti-Spam Filtering of E-Mail*. In Proceedings of EMNLP-01, 6th Conference on Empirical Methods in Natural Language Processing. Pittsburgh, Association for Computational Linguistics, Morristown, NJ: 44–50.
- Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C. D., and Stamatopoulos, P. (2003). “A Memory-Based Approach to Anti-Spam Filtering for Mailing Lists.” *Information Retrieval* 6(1): 49–73.
- Salamonsen, W., Mok, K., Kolatkar, P., and Subbiah, S. (1999). *BioJAKE: A Tool for the Creation, Visualization and Manipulation of Metabolic Pathways*. In Proceedings of the Pacific Symposium on Biocomputing. Hawaii, World Scientific Press, Hackensack NJ: 392–400.
- Salton, G. (1989). *Automatic Text Processing*. Reading, MA, Addison-Wesley.
- Sanchez, S. N., Triantaphyllou, E., and Kraft, D. (2002). “A Feature Mining Based Approach for the Classification of Text Documents into Disjoint Classes.” *Information Processing and Management* 38(4): 583–604.
- Sarkar, M., and Brown, M. (1992). *Graphical Fisheye Views of Graphs*. In Proceedings of the ACM SIGCHI '92 Conference on Human Factors in Computing Systems. Monterey, CA, ACM Press, New York: 83–91.
- Sasaki, M., and Kita, K. (1998). *Automatic Text Categorization Based on Hierarchical Rules*. In Proceedings of the 5th International Conference on Soft Computing and Information. Iizuka, Japan, World Scientific, Singapore: 935–938.
- Sasaki, M., and Kita, K. (1998). *Rule-Based Text Categorization Using Hierarchical Categories*. In Proceedings of SMC-98, IEEE International Conference on Systems, Man, and Cybernetics. La Jolla, CA, IEEE Computer Society Press, Los Alamitos, CA: 2827–2830.
- Schapire, R. E., and Singer, Y. (2000). “BoosTexter: A Boosting-Based System for Text Categorization.” *Machine Learning* 39(2/3): 135–168.
- Schapire, R. E., Singer, Y., and Singhal, A. (1998). *Boosting and Rocchio Applied to Text Filtering*. In Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval. W. S. Croft, A. Moffat, C. J. v. Rijsbergen, R. Wilkinson, and J. Zobel, eds. Melbourne, Australia, ACM Press, New York: 215–223.
- Scheffer, T., and Joachims, T. (1999). *Expected Error Analysis for Model Selection*. In Proceedings of ICML-99, 16th International Conference on Machine Learning. I. Bratko and S. Dzeroski, eds. Bled, Slovenia, Morgan Kaufmann Publishers, San Francisco: 361–370.
- Schneider, K.-M. (2003). *A Comparison of Event Models for Naive Bayes Anti-Spam E-Mail Filtering*. In Proceedings of EACL-03, 11th Conference of the European Chapter of the

- Association for Computational Linguistics. Budapest, Hungary, Association for Computational Linguistics, Morristown, NJ: 307–314.
- Schutze, H. (1993). *Part-of-Speech Induction from Scratch*. In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics. Columbus, OH, Association for Computational Linguistics, Morristown, NJ: 251–258.
- Schutze, H. (1998). “Automatic Word Sense Discrimination.” *Computational Linguistics* **24**(1): 97–124.
- Schutze, H., Hull, D. A., and Pedersen, J. O. (1995). *A Comparison of Classifiers and Document Representations for the Routing Problem*. In Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval. E. A. Fox, P. Ingwersen, and R. Fidel, eds. Seattle, ACM Press, New York: 229–237.
- Scott, J. (2000). *Social Network Analysis: A Handbook*. London, Sage Publications.
- Scott, S. (1998). *Feature Engineering for a Symbolic Approach to Text Classification*. Master’s thesis, Computer Science Department, University of Ottawa.
- Scott, S., and Matwin, S. (1999). *Feature Engineering for Text Classification*. In Proceedings of ICML-99, 16th International Conference on Machine Learning. I. Bratko and S. Dzeroski, eds. Bled, Slovenia, Morgan Kaufmann Publishers, San Francisco: 379–388.
- Sebastiani, F. (1999). *A Tutorial on Automated Text Categorisation*. In Proceedings of ASAI-99, 1st Argentinian Symposium on Artificial Intelligence. A. Anandi and R. Zunino, eds. Buenos Aires: 7–35.
- Sebastiani, F. (2002). “Machine Learning in Automated Text Categorization.” *ACM Computing Surveys* **34**(1): 1–47.
- Sebastiani, F., Sperduti, A., and Valdambrini, N. (2000). *An Improved Boosting Algorithm and Its Application to Automated Text Categorization*. In Proceedings of CIKM-00, 9th ACM International Conference on Information and Knowledge Management. A. Ayah, J. Callan, and E. Rundensteiner, eds. McLean, VA, ACM Press, New York: 78–85.
- Seidman, S. B. (1983). “Network Structure and Minimum Degree.” *Social Networks* **5**: 269–287.
- Seymore, K., McCallum, A., and Rosenfeld, R. (1999). *Learning Hidden Markov Model Structure for Information Extraction*. In AAAI 99 Workshop on Machine Learning for Information Extraction. Orlando, FL, AAAI Press, Menlo Park, CA: 37–42.
- Sha, F., and Pereira, F. (2003). *Shallow Parsing with Conditional Random Fields*. In Technical Report C15 TR MS-C15-02-35, University of Pennsylvania.
- Shin, C., Doermann, D., and Rosenfeld, A. (2001). “Classification of Document Pages Using Structure-Based Features.” *International Journal on Document Analysis and Recognition* **3**(4): 232–247.
- Shneiderman, B. (1996). *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In Proceedings of the 1996 IEEE Conference on Visual Languages. Boulder, CO, IEEE Computer Society Press, Washington, DC: 336–343.
- Shneiderman, B. (1997). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Reading, MA, Addison-Wesley.
- Shneiderman, B., Byrd, D., and Croft, W. B. (1998). “Sorting Out Searching: A User Interface Framework for Text Searches.” *Communications of the ACM* **41**(4): 95–98.
- Sigletos, G., Paliouras, G., and Karkaletsis, V. (2002). *Role Identification from Free Text Using Hidden Markov Models*. In Proceedings of the 2nd Hellenic Conference on AI: Methods and Applications of Artificial Intelligence. I. P. Vlahavas and C. D. Spyropoulos, eds. Thessaloniki, Greece, Springer-Verlag, London: 167–178.
- Silberschatz, A., and Tuzhilin, A. (1996). “What Makes Patterns Interesting in Knowledge Discovery Systems.” *IEEE Transactions on Knowledge and Data Engineering* **8**(6): 970–974.
- Silverstein, C., Brin, S., and Motwani, R. (1999). “Beyond Market Baskets: Generalizing Association Rules to Dependence Rules.” *Data Mining and Knowledge Discovery* **2**(1): 39–68.

- Siolas, G., and d'Alche-Buc, F. (2000). *Support Vector Machines Based on a Semantic Kernel for Text Categorization*. In Proceedings of IJCNN-00, 11th International Joint Conference on Neural Networks. Como, Italy, IEEE Computer Society Press, Los Alamitos, CA: 205–209.
- Skarmeta, A. G., Bensaid, A., and Tazi, N. (2000). “Data Mining for Text Categorization with Semi-supervised Agglomerative Hierarchical Clustering.” *International Journal of Intelligent Systems* **15**(7): 633–646.
- Slattery, S., and Craven, M. (1998). *Combining Statistical and Relational Methods for Learning in Hypertext Domains*. In Proceedings of ILP-98, 8th International Conference on Inductive Logic Programming. D. Page, ed. Madison, WI, Springer-Verlag, Heidelberg: 38–52.
- Slattery, S., and Craven, M. (2000). *Discovering Test Set Regularities in Relational Domains*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco: 895–902.
- Slonim, N., and Tishby, N. (2001). *The Power of Word Clusters for Text Classification*. In Proceedings of ECIR-01, 23rd European Colloquium on Information Retrieval Research. Darmstadt, Germany Academic Press, British Computer Society, London.
- Smith, D. (2002). *Detecting and Browsing Events in Unstructured Text*. In Proceedings of the 25th Annual ACM SIGIR Conference. Tampere, Finland, ACM Press, New York: 73–80.
- Soderland, S. (1999). “Learning Information Extraction Rules for Semi-Structured and Free Text.” *Machine Learning* **34**(1–3): 233–272.
- Soderland, S., Etzioni, O., Shaked, T., and Weld, D. S. (2004). *The Use of Web-based Statistics to Validate Information Extraction*. In Proceedings of the AAAI-2004 Workshop on Adaptive Text Extraction and Mining (ATEM-2004). San Jose, CA, AAAI Press, Menlo Park, CA: 21–27.
- Soderland, S., Fisher, D., Aseltine, J., and Lehnert, W. (1995). *CRYSTAL: Inducing a Conceptual Dictionary*. In Proceedings of the 14th International Joint Conference on Artificial Intelligence. C. Mellish, ed. Montreal, Canada, Morgan Kaufmann Publishers, San Francisco: 1314–1319.
- Soh, J. (1998). *A Theory of Document Object Locator Combination*. Doctoral Dissertation, State University of New York of Buffalo.
- Sondag, P.-P. (2001). *The Semantic Web Paving the Way to the Knowledge Society*. In Proceedings of the 27th International Conference on Very Large Databases, (VLDB). Rome, Morgan Kaufmann Publishers, San Francisco: 16.
- Soon, W. M., Ng, H. T., and Lim, D. C. Y. (2001). “A Machine Learning Approach to Coreference Resolution in Noun Phrases.” *Computational Linguistics* **27**(4): 521–544.
- Soucy, P., and Mineau, G. W. (2001a). *A Simple Feature Selection Method for Text Classification*. In Proceedings of IJCAI-01, 17th International Joint Conference on Artificial Intelligence. B. Nebel, ed. Seattle, AAAI Press, Menlo Park, CA: 897–902.
- Soucy, P., and Mineau, G. W. (2001b). *A Simple KNN Algorithm for Text Categorization*. In Proceedings of ICDM-01, IEEE International Conference on Data Mining. N. Cerone, T. Y. Lin, and X. Wu, eds. San Jose, CA, IEEE Computer Society Press, Los Alamitos, CA: 647–648.
- Soucy, P., and Mineau, G. W. (2003). *Feature Selection Strategies for Text Categorization*. In Proceedings of CSCSI-03, 16th Conference of the Canadian Society for Computational Studies of Intelligence. Y. Xiang and B. Chaib-Draa, eds. Halifax: 505–509.
- Spence, B. (2001). *Information Visualization*. Harlow, UK, Addison-Wesley.
- Spence, M., and Beilken, C. (1999). *Visual, Interactive Data Mining with InfoZoom – The Financial Data Set*. In Proceedings of the 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases. Prague, Springer Verlag, Berlin.
- Spitz, L., and Maghbouleh, A. (2000). *Text Categorization Using Character Shape Codes*. In Proceedings of the 7th SPIE Conference on Document Recognition and Retrieval. San Jose, CA, SPIE, The International Society for Optical Engineering, Bellingham, WA: 174–181.

- Spoerri, A. (1999). "InfoCrystal: A Visual Tool for Information Retrieval." In *Readings in Information Visualization: Using Vision to Think*. S. Card, J. Mackinlay, and B. Shneiderman, eds. San Francisco, Morgan Kaufmann Publishers: 140–147.
- Srikant, R., and Agrawal, R. (1995). *Mining Generalized Association Rules*. In Proceedings of the 21st International Conference on Very Large Databases. U. Dayal, P. Gray, and S. Nishio, eds. Zurich, Switzerland, Morgan Kaufmann Publishers, San Francisco, CA: 407–419.
- Srikant, R., and Agrawal, R. (1996). *Mining Sequential Patterns: Generalizations and Performance Improvements*. In Proceedings of the 5th Annual Conference on Extending Database Technology. P. Apers, M. Boozeghoub, and G. Gardarin, eds. Avignon, France, Springer-Verlag, Berlin: 3–17.
- Stamatatos, E., Fakotakis, N., and Kokkinakis, G. (2000). "Automatic Text Categorization in Terms of Genre and Author." *Computational Linguistics* 26(4): 471–495.
- Stapley, B. J., and Benoit, G. (2000). *Biobibliometrics: Information Retrieval and Visualization from Co-occurrences of Gene Names in Medline Abstracts*. In Proceedings of the Pacific Symposium on Biocomputing. Honolulu, Hawaii, World Scientific Press, Hackensack, NJ: 526–537.
- Steinbach, M., Karypis, G., and Kumar, V. (2000). *A Comparison of Document Clustering Techniques*. In Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Boston, ACM Press, New York.
- Sun, A., and Lim, E.-P. (2001). *Hierarchical Text Classification and Evaluation*. In Proceedings of ICDM-01, IEEE International Conference on Data Mining. N. Cercone, T. Lin, and X. Wu, eds. San Jose, CA, IEEE Computer Society Press, Los Alamitos, CA: 521–528.
- Sun, A., Lim, E.-P., and Ng, W.-K. (2003a). "Hierarchical Text Classification Methods and Their Specification." In *Cooperative Internet Computing*. A. T. Chan, S. Chan, H. Y. Leong, and V. T. Y. Ng., eds. Dordrecht, Kluwer Academic Publishers: 236–256.
- Sun, A., Lim, E.-P., and Ng, W.-K. (2003b). "Performance Measurement Framework for Hierarchical Text Classification." *Journal of the American Society for Information Science and Technology* 54(11): 1014–1028.
- Sun, A., Naing, M., Lim, E., and Lam, W. (2003). *Using Support Vector Machine for Terrorism Information Extraction*. In Proceedings of the Intelligence and Security Informatics: 1st NSF/NIJ Symposium on Intelligence and Security Informatics. H. Chen, R. Miranda, D. Zeng, C. Demchek, J. Schroeder, and T. Madhusudan, eds. Tucson, AZ, Springer-Verlag, Berlin: 1–12.
- Taghva, K., Nartker, T. A., Borsack, J., Lumos, S., Condit, A., and Young, R. (2000). *Evaluating Text Categorization in the Presence of OCR Errors*. In Proceedings of the 8th SPIE Conference on Document Recognition and Retrieval. San Jose, CA, SPIE, The International Society for Optical Engineering, Washington, DC: 68–74.
- Taira, H., and Haruno, M. (1999). *Feature Selection in SVM Text Categorization*. In Proceedings of AAAI-99, 16th Conference of the American Association for Artificial Intelligence. Orlando, FL, AAAI Press, Menlo Park, CA: 480–486.
- Taira, H., and Haruno, M. (2001). *Text Categorization Using Transductive Boosting*. In Proceedings of ECML-01, 12th European Conference on Machine Learning. L. D. Raedt and P. A. Flach, eds. Freiburg, Germany, Springer-Verlag, Heidelberg: 454–465.
- Takamura, H., and Matsumoto, Y. (2001). *Feature Space Restructuring for SVMs with Application to Text Categorization*. In Proceedings of EMNLP-01, 6th Conference on Empirical Methods in Natural Language Processing. Pittsburgh, Association for Computational Linguistics, Morristown, NJ: 51–57.
- Tan, A.-H. (2001). *Predictive Self-Organizing Networks for Text Categorization*. In Proceedings of PAKDD-01, 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Hong Kong, Springer-Verlag, Heidelberg: 66–77.
- Tan, A. (1999). *Text Mining: The State of the Art and the Challenges*. In Proceedings of the PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases (KDAD'99). Beijing: 71–76.

- Tan, C.-M., Wang, Y.-F., and Lee, C.-D. (2002). "The Use of Bigrams to Enhance Text Categorization." *Information Processing and Management* 38(4): 529–546.
- Taskar, B., Abbeel, P., and Koller, D. (2002). *Discriminative Probabilistic Models of Relational Data*. In Proceedings of UAI-02, 18th Conference on Uncertainty in Artificial Intelligence. Edmonton, Canada, Morgan Kaufmann Publishers, San Francisco: 485–492.
- Taskar, B., Segal, E., and Koller, D. (2001). *Probabilistic Classification and Clustering in Relational Data*. In Proceedings of IJCAI-01, 17th International Joint Conference on Artificial Intelligence. B. Nebel, ed. Seattle, Morgan Kaufmann Publishers, San Francisco: 870–878.
- Tauritz, D. R., Kok, J. N., and Sprinkhuizen-Kuyper, I. G. (2000). "Adaptive Information Filtering Using Evolutionary Computation." *Information Sciences* 122(2/4): 121–140.
- Tauritz, D. R., and Sprinkhuizen-Kuyper, I. G. (1999). *Adaptive Information Filtering Algorithms*. In Proceedings of IDA-99, 3rd Symposium on Intelligent Data Analysis. D. J. Wand, J. N. Kok, and M. R. Berthold, eds. Amsterdam, Springer-Verlag, Heidelberg: 513–524.
- Teahan, W. J. (2000). *Text Classification and Segmentation Using Minimum Cross-entropy*. In Proceedings of RIAO-00, 6th International Conference "Recherche d'Information Assistée par Ordinateur." Paris: 943–961.
- Teytaud, O., and Jalam, R. (2001). *Kernel Based Text Categorization*. In Proceedings of IJCNN-01, 12th International Joint Conference on Neural Networks. Washington, DC, IEEE Computer Society Press, Los Alamitos, CA: 1892–1897.
- Theeramunkong, T., and Lertnatee, V. (2002). *Multi-Dimensional Text Classification*. In Proceedings of COLING-02, 19th International Conference on Computational Linguistics. Taipei, Taiwan Association for Computational Linguistics, Morristown, NJ.
- Thelen, M., and Riloff, E. (2002). *A Bootstrapping Method for Learning Semantic Lexicons Using Extraction Pattern Contexts*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002). Philadelphia, Association for Computational Linguistics, Morristown, NJ: 214–221.
- Thomas, J., Cook, K., Crow, V., Hetzler, B., May, R., McQuerry, D., McVeety, R., Miller, N., Nakamura, G., Nowell, L., Whitney, P., and Wong, P. C. (1999). *Human Computer Interaction with Global Information Spaces: Beyond Data Mining*. In Proceedings of the British Computer Society Conference. Bradford, UK, Springer-Verlag, London.
- Thompson, P. (2001). *Automatic Categorization of Case Law*. In Proceedings of ICAIL-01, 8th International Conference on Artificial Intelligence and Law. St. Louis, MO, ACM Press, New York: 70–77.
- Toivonen, H., Klemettinen, M., Ronkainen, P., Hatonen, K., and Mannila, H. (1995). *Pruning and Grouping Discovered Association Rules*. In Workshop Notes: Statistics, Machine Learning and Knowledge Discovery in Databases, ECML-95. N. Lavrac and S. Wrobel, eds. Heraklion, Greece, Springer-Verlag, Berlin: 47–52.
- Tombros, A., Villa, R., and Rijsbergen, C. J. (2002). "The Effectiveness of Query-Specific Hierarchic Clustering in Information Retrieval." *Information Processing & Management* 38(4): 559–582.
- Tong, R., Winkler, A., and Gage, P. (1992). *Classification Trees for Document Routing: A Report on the TREC Experiment*. In Proceedings of TREC-1, 1st Text Retrieval Conference. D. K. Harman, ed. Gaithersburg, MD, National Institute of Standards and Technology, Gaithersburg, MD: 209–228.
- Tong, S., and Koller, D. (2000). *Support Vector Machine Active Learning with Applications to Text Classification*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco, CA: 999–1006.
- Tong, S., and Koller, D. (2001). "Support Vector Machine Active Learning with Applications to Text Classification." *Journal of Machine Learning Research* 2: 45–66.
- Toutanova, K., Chen, F., Popat, K., and Hofmann, T. (2001). *Text Classification in a Hierarchical Mixture Model for Small Training Sets*. In Proceedings of CIKM-01, 10th ACM

- International Conference on Information and Knowledge Management. H. Paques, L. Liu, and D. Grossman, eds. Atlanta, ACM Press, New York: 105–113.
- Trastour, D., Bartolini, C., and Preist, C. (2003). "Semantic Web Support for the Business-to-Business E-Commerce Pre-Contractual Lifecycle." *Computer Networks* **42**(5): 661–673.
- Tufte, E. (1983). *The Visual Display of Quantitative Information*. Chelshire, CT, Graphics Press.
- Tufte, E. (1990). *Envisioning Information*. Chelshire, CT, Graphics Press.
- Tufte, E. (1997). *Visual Explanations*. Cheshire, CT, Graphics Press.
- Turney, P. (1997). *Extraction of Keyphrases from Text: Evaluation of Four Algorithms*. Technical Report ERB 1051, National Research Council of Canada, Institute for Information Technology: 1–27.
- Turney, P. D. (2000). "Learning Algorithms for Keyphrase Extraction." *Information Retrieval* **2**(4): 303–336.
- Tzeras, K., and Hartmann, S. (1993). *Automatic Indexing Based on Bayesian Inference Networks*. In Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval. R. Korfhage, E. M. Rasmussen, and P. Willett, eds. Pittsburgh, ACM Press, New York: 22–34.
- Tzoukermann, E., Klavans, J., and Jacquemin, C. (1997). *Effective Use of Natural Language Processing Techniques for Automatic Conflation of Multi-Word Terms: The Role of Derivational Morphology, Part of Speech Tagging, and Shallow Parsing*. In Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Philadelphia, ACM Press, New York: 148–155.
- Ureña-Lopez, L. A., Buenaga, M., and Gomez, J. M. (2001). "Integrating linguistic resources in TC through WSD." *Computers and the Humanities* **35**(2): 215–230.
- Uren, V. S., and Addis, T. R. (2002). "How Weak Categorizers Based upon Different Principles Strengthen Performance." *The Computer Journal* **45**(5): 511–524.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Berlin, Springer-Verlag.
- Varadarajan, S., Kasravi, K., and Feldman, R. (2002). *Text-Mining: Application Development Challenges*. In Proceedings of the 22nd SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligence. Cambridge, UK, Springer-Verlag, Berlin.
- Vel, O. Y. D., Anderson, A., Corney, M., and Mohay, G. M. (2001). "Mining Email Content for Author Identification Forensics." *SIGMOD Record* **30**(4): 55–64.
- Vert, J.-P. (2001). *Text Categorization Using Adaptive Context Trees*. In Proceedings of CICLING-01, 2nd International Conference on Computational Linguistics and Intelligent Text Processing. A. Gelbukh, ed. Mexico City, Springer-Verlag, Heidelberg: 423–436.
- Viechnicki, P. (1998). *A Performance Evaluation of Automatic Survey Classifiers*. In Proceedings of ICGI-98, 4th International Colloquium on Grammatical Inference. V. Honavar and G. Slutzki, eds. Ames, IA, Springer-Verlag, Heidelberg: 244–256.
- Vinokourov, A., and Girolami, M. (2001). *Document Classification Employing the Fisher Kernel Derived from Probabilistic Hierarchic Corpus Representations*. In Proceedings of ECIR-01, 23rd European Colloquium on Information Retrieval Research. Darmstadt, Germany, Springer-Verlag, Berlin: 24–40.
- Vinokourov, A., and Girolami, M. (2002). "A Probabilistic Framework for the Hierarchic Organisation and Classification of Document Collections." *Journal of Intelligent Information Systems* **18**(2/3): 153–172.
- Wang, H., and Son, N. H. (1999). *Text Classification Using Lattice Machine*. In Proceedings of ISMIS-99, 11th International Symposium on Methodologies for Intelligent Systems. A. Skowron and Z. W. Ras, eds. Warsaw, Springer-Verlag, Heidelberg: 235–243.
- Wang, J. T. L., Zhang, K., Chang, G., and Shasha, D. (2002). "Finding Approximate Patterns in Undirected Acyclic Graphs." *Pattern Recognition* **35**(2): 473–483.
- Wang, K., Zhou, S., and He, Y. (2001). *Hierarchical Classification of Real Life Documents*. In Proceedings of the 1st SIAM International Conference on Data Mining. Chicago, SIAM Press, Philadelphia.

- Wang, K., Zhou, S., and Liew, S. C. (1999). *Building Hierarchical Classifiers Using Class Proximity*. In Proceedings of VLDB-99, 25th International Conference on Very Large Data Bases. M. P. Atkinson, M. E. Orlowska, P. Valduriez, S. B. Zdonik, and M. L. Brodie, eds. Edinburgh, Morgan Kaufmann Publishers, San Francisco: 363–374.
- Wang, W., Meng, W., and Yu, C. (2000). *Concept Hierarchy Based Text Database Categorization in a Metasearch Engine Environment*. In Proceedings of WISE-00, 1st International Conference on Web Information Systems Engineering. Hong Kong, IEEE Computer Society Press, Los Alamitos, CA: 283–290.
- Wang, Y., and Hu, J. (2002). *A Machine Learning Based Approach for Table Detection on the Web*. In Proceedings of the 11th International World Web Conference. Honolulu, HI, ACM Press, New York: 242–250.
- Ware, C. (2000). *Information Visualization: Perception for Design*, San Francisco, Morgan Kaufmann Publishers.
- Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, UK, Cambridge University Press.
- Wei, C.-P., and Dong, Y.-X. (2001). *A Mining-based Category Evolution Approach to Managing Online Document Categories*. In Proceedings of HICSS-01, 34th Annual Hawaii International Conference on System Sciences. R. H. Sprague, ed. Maui, HI, IEEE Computer Society Press, Los Alamitos, CA: 7061–7062.
- Weigend, A. S., Wiener, E. D., and Pedersen, J. O. (1999). “Exploiting Hierarchy in Text Categorization.” *Information Retrieval* **1**(3): 193–216.
- Weischedel, R., Meteor, M., Schwartz, R., Ramshaw, L., and Palmucci, J. (1993). “Coping with Ambiguity and Unknown Words through Probabilistic Methods.” *Computational Linguistics* **19**(2): 361–382.
- Weiss, S. M., Apte, C., Damerau, F. J., Johnson, D. E., Oles, F. J., Goetz, T., and Hampp, T. (1999). “Maximizing Text-Mining Performance.” *IEEE Intelligent Systems* **14**(4): 63–69.
- Wermter, S. (2000). “Neural Network Agents for Learning Semantic Text Classification.” *Information Retrieval* **3**(2): 87–103.
- Wermter, S., Arevian, G., and Panchev, C. (1999). *Recurrent Neural Network Learning for Text Routing*. In Proceedings of ICANN-99, 9th International Conference on Artificial Neural Networks. Edinburgh, Institution of Electrical Engineers, London, UK: 898–903.
- Wermter, S., and Hung, C. (2002). *Self-Organizing Classification on the Reuters News Corpus*. In Proceedings of COLING-02, the 19th International Conference on Computational Linguistics. Taipei, Morgan Kaufmann Publishers, San Francisco.
- Wermter, S., Panchev, C., and Arevian, G. (1999). *Hybrid Neural Plausibility Networks for News Agents*. In Proceedings of AAAI-99, 16th Conference of the American Association for Artificial Intelligence. Orlando, FL, AAAI Press, Menlo Park, CA: 93–98.
- Westphal, C., and Bergeron, R. D. (1998). *Data Mining Solutions: Methods and Tools for Solving Real-World Problems*. New York, John Wiley and Sons.
- White, D. R., and Reitz, K. P. (1983). “Graph and Semigroup Homomorphisms on Networks of Relations.” *Social Networks* **5**: 193–234.
- Wibowo, W., and Williams, H. E. (2002). *Simple and Accurate Feature Selection for Hierarchical Categorisation*. In Proceedings of the 2002 ACM Symposium on Document Engineering. McLean, VA, ACM Press, New York: 111–118.
- Wiener, E. D. (1995). *A Neural Network Approach to Topic Spotting in Text*. Boulder, CO, Department of Computer Science, University of Colorado at Boulder.
- Wiener, E. D., Pedersen, J. O., and Weigend, A. S. (1995). *A Neural Network Approach to Topic Spotting*. In Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval. Las Vegas, ISRI, University of Nevada, Las Vegas: 317–332.
- Wilks, Y. (1997). “Information Extraction as a Core Language Technology.” In M. T. Pazienza, ed. *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*. *Lecture Notes in Computer Science* **1229**: 1–9.

- Williamson, C., and Schneiderman, B. (1992). *The Dynamic HomeFinder: Evaluating Dynamic Queries in a Real-Estate Information Exploration System*. In Proceedings of the 15th Annual, ACM-SIGIR. N. Belkin, P. Ingwersen, A. Pejtersen, eds. Copenhagen, ACM Press, New York: 338–346.
- Wills, G. (1999). “NicheWorks’ Interactive Visualization of Very Large Graphs.” *Journal of Computational and Graphical Statistics* 8(2): 190–212.
- Wise, J., Thomas, J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., and Crow, V. (1995). *Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents*. In Proceedings of IEEE Information Visualization ’95. Atlanta, GA, IEEE Computer Society Press, Los Alamitos, CA: 51–58.
- Witten, I. H., Bray, Z., Mahoui, M., and Teahan, W. J. (1999). *Text Mining: A New Frontier for Lossless Compression*. In Proceedings of IEEE Data Compression Conference. J. Ai. Storer and M. Cohn, eds. Snowbird, UT, IEEE Computer Society Press, Los Alamitos, CA: 198–207.
- Wong, J. W., Kan, W.-K., and Young, G. H. (1996). “Action: Automatic Classification for Full-Text Documents.” *SIGIR Forum* 30(1): 26–41.
- Wong, P. C. (1999). “Visual Data Mining – Guest Editor’s Introduction.” *IEEE Computer Graphics and Applications* 19(5): 2–12.
- Wong, P. C., Cowley, W., Foote, H., Jurrus, E., and Thomas, J. (2000). *Visualizing Sequential Patterns for Text Mining*. In Proceedings of the IEEE Information Visualization Conference (INFOVIS 2000). Salt Lake City, UT, ACM Press, New York: 105–115.
- Wong, P. C., Whitney, P., and Thomas, J. (1999). *Visualizing Association Rules for Text Mining*. In Proceedings of IEEE Information Visualization (InfoVis ’99). San Francisco, IEEE Computer Society Press, Washington, DC: 120–124.
- Xu, Z., Yu, K., Tresp, V., Xu, X., and Wang, J. (2003). *Representative Sampling for Text Classification Using Support Vector Machines*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Berlin: 393–407.
- Xue, D., and Sun, M. (2003). *Chinese Text Categorization Based on the Binary Weighting Model with Non-binary Smoothing*. In Proceedings of ECIR-03, 25th European Conference on Information Retrieval. F. Sebastiani, ed. Pisa, Italy, Springer-Verlag, Berlin: 408–419.
- Yamazaki, T., and Dagan, I. (1997). *Mistake-Driven Learning with Thesaurus for Text Categorization*. In Proceedings of NLP97, the Natural Language Processing Pacific Rim Symposium. Phuket, Thailand: 369–374.
- Yang, C. C., Chen, H., and Hong, K. (2003). “Visualization of Large Category Map for Internet Browsing.” *Decision Support Systems* 35: 89–102.
- Yang, H.-C., and Lee, C.-H. (2000a). *Automatic Category Generation for Text Documents by Self-Organizing Maps*. In Proceedings of IJCNN-00, 11th International Joint Conference on Neural Networks, Volume 3. Como, Italy, IEEE Computer Society Press, Los Alamitos, CA, 3581–3586.
- Yang, H.-C., and Lee, C.-H. (2000b). *Automatic Category Structure Generation and Categorization of Chinese Text Documents*. In Proceedings of PKDD-00, 4th European Conference on Principles of Data Mining and Knowledge Discovery. D. Zighed, A. Komorowski, and D. Zytrowski, eds. Lyon, France, Springer-Verlag, Heidelberg, Germany: 673–678.
- Yang, T. (2000). *Detecting Emerging Contextual Concepts in Textual Collections*. M.Sc. thesis, Department of Computer Science, University of Illinois at Urbana-Champaign.
- Yang, Y. (1994). *Expert Network: Effective and Efficient Learning from Human Decisions in Text Categorisation and Retrieval*. In Proceedings of SIGIR-94, 17th ACM International Conference on Research and Development in Information Retrieval. W. B. Croft and C. J. v. Rijsbergen, eds. Dublin, Springer-Verlag, Heidelberg: 13–22.
- Yang, Y. (1995). *Noise Reduction in a Statistical Approach to Text Categorization*. In Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in

- Information Retrieval. E. A. Fox, P. Ingwersen, and R. Fidel, eds. Seattle, ACM Press, New York: 256–263.
- Yang, Y. (1996). *An Evaluation of Statistical Approaches to MEDLINE Indexing*. In Proceedings of AMIA-96, Fall Symposium of the American Medical Informatics Association. J. J. Cimino, ed. Washington, DC, Hanley and Belfus, Philadelphia: 358–362.
- Yang, Y. (1999). “An Evaluation of Statistical Approaches to Text Categorization.” *Information Retrieval* **1**(1/2): 69–90.
- Yang, Y. (2001). *A Study on Thresholding Strategies for Text Categorization*. In Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval. W. B. Croft, D. J. Harper, D. H. Kroft, and J. Zobel, eds. New Orleans, ACM Press, New York: 137–145.
- Yang, Y., Ault, T., and Pierce, T. (2000). *Combining Multiple Learning Strategies for Effective Cross-Validation*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco: 1167–1182.
- Yang, Y., Ault, T., Pierce, T., and Lattimer, C. W. (2000). *Improving Text Categorization Methods for Event Tracking*. In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval. N. J. Belkin, P. Ingwersen, and M.-K. Leong, eds. Athens, Greece, ACM Press, New York: 65–72.
- Yang, Y., and Chute, C. G. (1993). *An Application of Least Squares Fit Mapping to Text Information Retrieval*. In Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval. R. Korthage, E. Rasmussen, and P. Willett, eds. Pittsburgh, ACM Press, New York: 281–290.
- Yang, Y., and Chute, C. G. (1994). “An Example-Based Mapping Method for Text Categorization and Retrieval.” *ACM Transactions on Information Systems* **12**(3): 252–277.
- Yang, Y., and Liu, X. (1999). *A Re-examination of Text Categorization Methods*. In Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval. M. Hearst, F. Gey, and R. Tong, eds. Berkeley, CA, ACM Press, New York: 42–49.
- Yang, Y., and Pedersen, J. O. (1997). *A Comparative Study on Feature Selection in Text Categorization*. In Proceedings of ICML-97, 14th International Conference on Machine Learning. D. H. Fisher. Nashville, TN, Morgan Kaufmann Publishers, San Francisco: 412–420.
- Yang, Y., Slattery, S., and Ghani, R. (2002). “A Study of Approaches to Hypertext Categorization.” *Journal of Intelligent Information Systems* **18**(2/3): 219–241.
- Yang, Y., and Wilbur, J. W. (1996a). “An Analysis of Statistical Term Strength and Its Use in the Indexing and Retrieval of Molecular Biology Texts.” *Computers in Biology and Medicine* **26**(3): 209–222.
- Yang, Y., and Wilbur, J. W. (1996b). “Using Corpus Statistics to Remove Redundant Words in Text Categorization.” *Journal of the American Society for Information Science* **47**(5): 357–369.
- Yang, Y., Zhang, J., and Kisiel, B. (2003). *A Scalability Analysis of Classifiers in Text Categorization*. In Proceedings of SIGIR-03, 26th ACM International Conference on Research and Development in Information Retrieval. J. Callan, G. Cormack, C. Clarke, D. Hawking, and A. Smeaton, eds. Toronto, ACM Press, New York: 96–103.
- Yao, D., Wang, J., Lu, Y., Noble, N., Sun, H., Zhu, X., Lin, N., Payan, D., Li, M., and Qu, K. (2004). *Pathway Finder: Paving the Way Towards Automatic Pathway Extraction*. In Proceedings of the 2nd Asian Bioinformatics Conference. Dunedin, New Zealand, Australian Computer Society, Darlinghurst, Australia: 53–62.
- Yavuz, T., and Guvenir, H. A. (1998). *Application of k-nearest Neighbor on Feature Projections Classifier to Text Categorization*. In Proceedings of ISCIS-98, 13th International Symposium on Computer and Information Sciences. U. Gudukbay, T. Dayar, A. Gorsoy, and E. Gelenbe, eds. Ankara, Turkey, IOS Press, Amsterdam: 135–142.

- Ye, N. (2003). *The Handbook of Data Mining*. Mahwah, NJ, Lawrence Erlbaum Associates.
- Yee, K.-P., Fisher, D., Dhamija, R., and Hearst, M. (2001). *Animated Exploration of Dynamic Graphs with Radial Layout*. In Proceedings of IEEE Symposium on Information Visualization (InfoVis 2001). San Diego, CA, IEEE Computer Society Press, Washington, DC: 43–50.
- Yeh, A., and Hirschman, L. (2002). “Background and Overview for KDD Cup 2002 Task 1: Information Extraction from Biomedical Articles.” *KDD Explorations* 4(2): 87–89.
- Yi, J., and Sundaresan, N. (2000). *A Classifier for Semi-Structured Documents*. In Proceedings of KDD-00, 6th ACM International Conference on Knowledge Discovery and Data Mining. Boston, ACM Press, New York: 340–344.
- Yoon, S., Henschen, L. J., Park, E., and Makki, S. (1999). *Using Domain Knowledge in Knowledge Discovery*. In Proceedings of the ACM Conference CIKM '99. Kansas City, MO, ACM Press, New York: 243–250.
- Yu, E. S., and Liddy, E. D. (1999). *Feature Selection in Text Categorization Using the Baldwin Effect Networks*. In Proceedings of IJCNN-99, 10th International Joint Conference on Neural Networks. Washington, DC, IEEE Computer Society Press, Los Alamitos, CA: 2924–2927.
- Yu, K. L., and Lam, W. (1998). *A New On-Line Learning Algorithm for Adaptive Text Filtering*. In Proceedings of CIKM-98, 7th ACM International Conference on Information and Knowledge Management. G. Gardarin, J. French, N. Pissinou, K. Makki, and L. Bouganim, eds. Bethesda, MD, ACM Press, New York: 156–160.
- Yumi, J. (2000). *Graphical User Interface and Visualization Techniques for Detection of Emerging Concepts*. M.S. thesis, Department of Computer Science, University of Illinois at Urbana-Champaign.
- Zaiane, O. R., and Antonie, M.-L. (2002). *Classifying Text Documents by Associating Terms with Text Categories*. In Proceedings of the 13th Australasian Conference on Database Technologies. Melbourne, Australia, ACM Press, New York: 215–222.
- Zamir, O., and Etzioni, O. (1999). “Grouper: A Dynamic Clustering Interface to Web Search Results.” *Computer Networks*. 31(11–16): 1361–1374.
- Zaragoza, H., Massih-Reza, A., and Gallinari, P. (1999). *A Dynamic Probability Model for Closed-Query Text Mining Tasks*. Draft submission to KDD '99.
- Zelikovitz, S., and Hirsh, H. (2000). *Improving Short Text Classification Using Unlabeled Background Knowledge*. In Proceedings of ICML-00, 17th International Conference on Machine Learning. P. Langley, ed. Stanford, CA, Morgan Kaufmann Publishers, San Francisco: 1183–1190.
- Zelikovitz, S., and Hirsh, H. (2001). *Using LSI for Text Classification in the Presence of Background Text*. In Proceedings of CIKM-01, 10th ACM International Conference on Information and Knowledge Management. H. Paques, L. Liu, and D. Grossman, eds. Atlanta, ACM Press, New York: 113–118.
- Zhang, D., and Lee, W. S. (2003). *Question Classification Using Support Vector Machines*. In Proceedings of SIGIR-03, 26th ACM International Conference on Research and Development in Information Retrieval. J. Callan, G. Cormack, C. Clarke, D. Hawking, and A. Smeaton, eds. Toronto, ACM Press, New York: 26–32.
- Zhang, J., Jin, R., Yang, Y., and Hauptmann, A. (2003). *Modified Logistic Regression: An Approximation to SVM and Its Applications in Large-Scale Text Categorization*. In Proceedings of ICML-03, 20th International Conference on Machine Learning. Washington, DC, Morgan Kaufmann Publishers, San Francisco: 888–895.
- Zhang, J., and Yang, Y. (2003). *Robustness of Regularized Linear Classification Methods in Text Categorization*. In Proceedings of SIGIR-03, 26th ACM International Conference on Research and Development in Information Retrieval. J. Collan, G. Cormack, C. Clarke, D. Hawking, and A. Smeaton, eds. Toronto, ACM Press, New York: 190–197.
- Zhang, K., Wang, J. T. L., and Shasha, D. (1995). “On the Editing Distance Between Undirected Acyclic Graphs.” *International Journal of Foundations of Computer Science* 7(1): 43–57.

- Zhang, T., and Oles, F. J. (2001). "Text Categorization Based on Regularized Linear Classification Methods." *Information Retrieval* **4**(1): 5–31.
- Zhao, Y., and Karypis, G. (2002). *Criterion Functions for Document Clustering: Experiments and Analysis*. Technical Report, TR 01–40. Minneapolis, Department of Computer Science, University of Minnesota.
- Zhdanova, A. V., and Shishkin, D. V. (2002). *Classification of Email Queries by Topic: Approach Based on Hierarchically Structured Subject Domain*. In Proceedings of IDEAL-02, 3rd International Conference on Intelligent Data Engineering and Automated Learning. H. Yin, N. Allinson, R. Freeman, J. Keane, and S. Hubbard, eds. Manchester, UK, Springer-Verlag, Heidelberg: 99–104.
- Zhong, S., and Ghosh, J. (2003). "A Comparative Study of Generative Models for Document Clustering." *Knowledge and Information Systems: An International Journal* **8**: 374–384.
- Zhou, M., and Cui, Y. (2004). "GeneInfoViz: Constructing and Visualizing Gene Relation Networks." *In Silico Biology* **4**(3): 323–333.
- Zhou, S., Fan, Y., Hua, J., Yu, F., and Hu, Y. (2000). *Hierachically Classifying Chinese Web Documents without Dictionary Support and Segmentation Procedure*. In Proceedings of WAIM-00, 1st International Conference on Web-Age Information Management. Shanghai, China, Springer-Verlag, Heidelberg: 215–226.
- Zhou, S., and Guan, J. (2002a). *An Approach to Improve Text Classification Efficiency*. In Proceedings of ADBIS-02, 6th East-European Conference on Advances in Databases and Information Systems. Y. M., and P. Navrat, eds. Bratislava, Slovakia, Springer-Verlag, Heidelberg: 65–79.
- Zhou, S., and Guan, J. (2002b). *Chinese Documents Classification Based on N-Grams*. In Proceedings of CICLING-02, 3rd International Conference on Computational Linguistics and Intelligent Text Processing. A. F. Gelbukh, ed. Mexico City, Springer-Verlag, Heidelberg: 405–414.
- Zhou, S., Ling, T. W., Guan, J., Hu, J., and Zhou, A. (2003). *Fast Text Classification: A Training-Corpus Pruning Based Approach*. In Proceedings of DASFAA-03, 8th IEEE International Conference on Database Advanced Systems for Advanced Application. Kyoto, Japan, IEEE Computer Society Press, Los Alamitos, CA: 127–136.

