

References

- [1] AGRAWAL, R., AND SRIKANT, R. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Databases (VLDB'94)* (Santiago, Chile, September 1994), Morgan Kaufmann, pp. 487–499.
- [2] AITKEN, S., AND REID, S. Evaluation of an ontology-based information retrieval tool. In *Workshop on the Applications of Ontologies and Problem-Solving Methods, European Conference on Artificial Intelligence (ECAI'00)* (Berlin, Germany, September 2000), A. Gómez and V.R. Benjamins and N. Guarino and M. Uschold, pp. 34–43.
- [3] ARC. Arc - a cross archive search service. old dominion university digital library research group. june 30th 2008, 2008.
- [4] BAILEY, J., BRY, F., FURCHE, T., AND SCHAFFERT, S. *Web and Semantic Web Query Languages: A Survey*. Springer, Berlin / Heidelberg, August 2005, ch. 3, pp. 35–133.
- [5] BORST, W. Construction of engineering ontologies. Tech. rep., Centre for Telematica and Information Technology, University of Twente. Enschede, The Netherlands, 1997.
- [6] BRASE, J., AND NEJDL, W. Ontologies and metadata for elearning. In *HandBook on Ontologies*. S. Staab, R. Studer. *International Handbooks on Information Systems* (Berlin, 2004), Springer-Verlag, pp. 555–574.
- [7] BRICKLEY, D., AND GUHA, R. Rdf vocabulary description language 1.0: Rdf schema: W3c recommendation 10 february 2004., 2004.

- [8] CLARK, J., AND DEROSE, S. Xml path language (xpath) version 1.0.: W3c recommendation 16 november 1999, 1999.
- [9] CUI, Z., AND P., O. Domain ontology management environment. In *Proceedings of the 33rd Hawaii International Conference on System Sciences'00 (HICSS'00, Hawaii, January)* (Washington, DC, USA, January 2000), IEEE Conference Proceedings, pp. 8–15.
- [10] DE SOMPEL, H. V., AND LAGOZE, C. Notes from the interoperability front: a progress report on the open archives initiative. In *Proceedings of the 6th European Conference on Research and Advance Technology for Digital Libraries* (September 2002), Joint Conference on Digital Libraries, pp. 144–157.
- [11] DIEDERICH, J., AND BALKE, W. The semantic growbag algorithm: Automatically deriving categorization systems. In *Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries'07 (ECDL'07, Budapest, Hungary, September)* (Berlin, September 2007), vol. 4675 of *Lecture Notes in Computer Science*, Springer, pp. 33–40.
- [12] DOCCLUSTER. Clustering c++ library documentation, 2007.
- [13] DUBLINCORE. Dublin core. 1997. dublin core metadata element set version 1.1: Reference. copyright 1995-2008 dcmi, 1997.
- [14] FERNANDEZ, M., GOMEZ-PEREZ, A., AND JURISTO, N. Methontology: from ontological art towards ontological engineering. In *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering* (Stanford, USA, March 1997), AAAI Press, pp. 33–40.
- [15] FRIDMAN, N., AND MCGUINNESS, D. Ontology development 101: A guide to creating your first ontology. Ksl-01-05 technical report, Stanford Knowledge Systems Laboratory, March 2001.

- [16] FUNG, B., WANG, K., AND ESTER, M. Hierarchical document clustering using frequent itemsets. In *Proceedings of the Third SIAM International Conference on Data Mining, (SDM'03, San Francisco, California, May)*, (San Francisco, CA, USA, May 2003), SIAM, pp. 59–70.
- [17] FUNG, B., WANG, K., AND ESTER, M. *The Encyclopedia of Data Warehousing and Mining*. Idea Group, Montclair State University, USA, 2005, ch. Hierarchical Document Clustering.
- [18] GÓMEZ, A., FERNÁNDEZ, M., AND CORCHO, O. *Ontological Engineering*. Springer-Verlag, London, England, 2004.
- [19] GRIGORIS, A., AND VAN HARMELEN, F. *A Semantic Web Primer*. The MIT Press, London, England, 2004.
- [20] GRUBER, T. A translation approach to portable ontology specification. *Knowledge Acquisition* 5, 2 (1993), 199–220.
- [21] HAN, J., PEI, J., YIN, Y., AND MAO, R. Mining frequent patterns without candidate generation: a frequent-pattern tree approach. In *Data Mining and Knowledge Discovery* 8, (1), 53–87. (Hingham, MA, USA, January 2004), Kluwer Academic Publishers, pp. 53–87.
- [22] HARRISON, T. L., ELANGO, A., BOLLEN, J., AND NELSON, M. Initial experiences re-exporting duplicate and similarity computation with an oai-pmh aggregator. *Computer Research Repository (CoRR) cs.DL/0401001* (January 2004).
- [23] HERMAN, I., SWICH, R., AND BRICKLEY, D. Resource description framework (rdf). <http://www.w3.org/RDF/>, May.
- [24] JANAS, J. An enhanced a priori algorithm for mining multidimensional association rules. In *Proceedings of the 25th International Conference on Information Technology*

- Interfaces, 2003. (ITI 2003)* (Cavtat / Dubrovnik, Croatia, June 2003), IEEE Conference Proceedings, pp. 193–198.
- [25] JANIL, A., AND DUBES, R. *Algorithms for clustering data*. Prentice Hall College Div., Upper Saddle River, NJ, USA, 1988.
- [26] JOVANOSKI, V., AND LAVRAČ, N. Classification rule learning with apriori-c. In *Progress in Artificial Intelligence*. Springer Berlin / Heidelberg, 2001, pp. 111–135.
- [27] KAROURI, L., AUFAURE, M., AND BENNACER, N. Context-based hierarchical clustering for the ontology learning. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, (WI'06)* (Hong Kong, Japan, December 2006), IEEE Conference Proceedings, pp. 420–427.
- [28] KAUFMAN, L., AND ROUSSEEUW, P. *Finding Groups in Data: an introduction to cluster analysis*. Wiley, New York, USA, 1990.
- [29] KOSTERS, W., PIJLS, W., AND POPOVA, V. Complexity analysis of depth first and fp-growth implementations of apriori. In *In: Perner P., Rosenfeld A. (Eds) Mining Learning and Dataming in Pattern Recognition. Proceedings MLDM 2003, Springer Lecture Notes in Artificial Intelligence 2734* (2003), Springer Verlag, pp. 284–292.
- [30] LAGOZE, C., AND DE SOMPEL, H. V. The open archives initiative: Building a low-barrier interoperability framework. In *Conference Proceedings of Joint Conference on Digital Libraries (JCDL'01, Roanoke, VA, USA)* (June 2001), Joint Conference on Digital Libraries, pp. 54–62.
- [31] LASSILA, O., AND MCGUINNESS, D. The role of frame-based representation on the semantic web. *Linköping Electronic Articles in Computer and Information Science* 6, 5 (2001).
- [32] LEWIS, D. Reuters-21578 text categorization test collection distribution 1.0, september 26., 1997.

- [33] LIU, B., HSU, W., AND YIMING, M. Integrating classification and association rule mining. In *Knowledge Discovery and Data Mining* (1998), pp. 80–86.
- [34] LJUBIČ, P., LAVRAČ, N., PLISSON, J., MLADENIAE, D., BOLLHALTER, S., AND JERMOL, M. Automated structuring of company competencies in virtual organizations. In *Proceedings of the Conference on Data Mining and Data Warehouses 2005 (SiKDD 2005, Ljubljana, Slovenia, October)* (October 2005), 7th International Multi-conference on Information Society IS'05, pp. 190–193.
- [35] LUCENE. Apache lucene - overview . copyright 2006 the apache software foundation., 2006.
- [36] MADRID, J., AND GAUCH, S. Keyconcept: Un motor de bu'squeda conceptual. *Revista Sistemas y Telema'tica, Universidad Icesi, Cali, Colombia (Enero-Junio)*, ISSN 1692-5238 (2003), 47–62.
- [37] MAEDCHE, A., AND STAAB, S. Semi-automatic engineering of ontologies from text. In *Proceedings of 12th International Conference on Software and Knowledge Engineering (SEKE 2000, Lake Shore Drive, Chicago, IL, USA, July)* (Chicago, IL, 2000), pp. 47–62.
- [38] MAEDCHE, A., STAAB, S., STOJANOVIC, N., STUDER, R., AND SURE, Y. Semantic portal: The seal approach. In *In Fensel, D., Hendler, J.A., Lieberman, H., Wahster, W. (eds): Spinning the Semantic Web. MIT Press, Cambridge London* (2001), pp. 317–359.
- [39] MAYFIELD, J., AND FININ, T. Information retrieval on the semantic web: integrating inference and retrieval. In *Proceedings of the Special Interest Group on Information Retrieval, Workshop on the Semantic Web, (SIGIR'03, Toronto, Canada, August)*, (August 2003).
- [40] MCGUINNESS, D., AND HARMELEN, F. Owl web ontology language overview: W3c recommendation 10 february 2004., 2004.

- [41] MEDINA, M., CHÁVEZ, A., AND CHÁVEZ, R. Construction, implementation and maintenance of ontologies of records. In *Proceedings of the Fourth Latin American Web Congress (LA-WEB'06, Puebla, México, May)* (May 2006), IEEE Computer Society, pp. 67–73.
- [42] MEDINA, M., AND SÁNCHEZ, J. Ontoair: a method to construct lightweight ontologies from document collections. In *Proceedings of the Ninth Mexican International Conference on Computer Science 2008, (ENC 08, Baja Californiaí, México, October)* (October 2006), IEEE Computer Society, p. 12.
- [43] MEDINA, M., AND SÁNCHEZ, J. Rdf-based model for encoding document hierarchies. In *Proceedings of 17th International Conference on Electronics, Communications and Computers* (February 2007), pp. 22–27.
- [44] MEDINA, M., SÁNCHEZ, J., AND CASTELLANOS, N. Ontological agents model based on mas-common kads methodology. In *Proceedings of 14th International Conference on Electronics, Communications and Computers* (February 2004), pp. 260–263.
- [45] MEDINA, M., SÁNCHEZ, J., CHÁVEZ, A., AND BENÍTEZ, A. Designing ontological agents: an alternative to improve information retrieval in federated digital libraries. In *Proceedings of the Atlantic Web Intelligence Conference 2004, (AWIC'04, Cancn, Mxico, May)*. (May 2004), pp. 155–163.
- [46] MEDINA, M., SÁNCHEZ, J., OSTRÓVSKAYA, Y., AND BRISABOA, N. Ontosir: An oai service for multi-collection document retrieval based on ontologies of metadata records. In *LA-WEB '06: Proceedings of the Fourth Latin American Web Congress (LA-WEB'05)* (November 2005), Argentine Society on Computer Science and Operations Research, SADIO, pp. 111–114.

- [47] MEDINA, M., SÁNCHEZ, J., AND PAZ, J. Document retrieval from multiple collections by using lightweight ontologies. In *Proceedings of the Fifteenth International Conference on Computing (CIC-2006)* (November 2006), IEEE Computer Society, pp. 141–146.
- [48] MEDINA, M., SÁNCHEZ, J., AND RAMÍREZ, A. Describing document hierarchies by using markup languages. In *Taller de tecnologías del lenguaje humano. Proceedings of the Seventh Mexican International Conference on Computer Science 2006, (ENC 06, San Luis Potosí, México, September)* (September 2006), IEEE Computer Society, pp. 31–37.
- [49] MIZOGUCHI, R. Tutorial on ontological engineering - part 1: Introduction to ontological engineering. *New Generation Computing, OhmSha&Springer* 21, 4 (2003), 365–384.
- [50] MIZOGUCHI, R., IKEDA, M., AND SETA, K. Ontology for modeling the world from problem solving perspectives. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 1995. Workshop on Basic Ontological Issues in Knowledge Sharing, (IJCAI'95)* (August 1995), pp. 1–12.
- [51] NAVIGLI, R., AND VELARDI, P. Learning domain ontologies from document warehouses and dedicated web sites. In *Computational Linguistics*, vol. 30. MIT Press, 2004, pp. 151–179.
- [52] OAI-HARVESTER. Oai-harvester. copyright (c) 2003 ed summers., 2008.
- [53] OAI-PACK. Oai-pmh facilities for python, zope, silva, and railroad., 2008.
- [54] OPENDIRECTORYPROYECT. Open directory proyect, 2008.
- [55] PLISSON, J., MLADENIAE, D., LJUBIĆ, P., LAVRAČ, N., AND GROBELNIK, M. Using machine learning to structure the expertise of companies: Analysis of the yahoo! business data. In *Conference on Data Mining and Data Warehouses (SiKDD 2005) Proceedings* (2005), 7th International Multi-conference on Information Society IS'05, pp. 186–189.

- [56] RASMUSSEN, E. Clustering algorithms. In *Information Retrieval: Data Structures and Algorithms*. Prentice Hall, 1992, pp. 419–442.
- [57] RASMUSSEN, M., AND KARYPIS, G. gcluto - an interactive clustering, visualization and analysis system. CSE/UMN Technical Report TR04-021, University of Minnesota, Department of Computer Science and Engineering, 2004.
- [58] SALTON, G., AND MCGILL, J. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [59] STEINBACH, M., KARYPIS, G., AND KUMAR, V. A comparison of document clustering techniques. *Proceedings of Workshop on Text Mining, 6th ACM SIGKDD International Conference on Data Mining (KDD'00)* (August 20–23 2000), 109–110.
- [60] VAN HEIJST, G., SCHREIBER, A., AND WIELINGA, B. Using explicit ontologies in kbs development. *International Journal of Human-Computer Studies* 46, Issue 2-3. (February-March) 46, 2-3 (1997), 183–292.
- [61] WARNER, S. Exposing and harvesting metadata using the oai metadata harvesting protocol: A tutorial, 2001.
- [62] WELLMAN, M., BIRMINGHAM, W., AND DURFEE, E. The digital library as a community of information agents. In *IEEE Expert* 11, Issue 3, (June) (1996), pp. 10–11.
- [63] WELTY, C., AND GUARINO, N. Supporting ontological analysis of taxonomic relationships. *Data and Knowledge Engineering* 39, 1 (2001), 51–74.
- [64] WOOLDRIDGE, M., JENNINGS, N., AND KINNY, D. The gaia methodology for agent-oriented analysis and design. In *Autonomous Agents and Multi-Agent Systems*. Wooldridge M., Rosenschein J. Springer Netherlands (2000), pp. 285–312.
- [65] YAHOO. Yahoo! directory. yahoo! inc. june 16th 2008, 2008.
- [66] YATES, R. B., AND NETO, B. R. *Modern Information Retrieval*. Addison Wesley, 1999.

- [67] ZHAO, Y., AND KARYPIS, G. Evaluation of hierarchical clustering algorithms for document datasets. In *Proceedings of the Fourth SIAM International Conference Information and Knowledge Management (CIKM 2002, McLean, Virginia, USA, November (2002))*, pp. 515–524.