

# Bibliography

- [1] Harold Abelson, Gerald Jay Sussman, and Julie Sussman. *Structure and Interpretation of Computer Programs*. MIT Press, 1996.
- [2] Serge Abiteboul and Victor Vianu. Queries and computation on the web. In *The Sixth International Conference on Database Theory (ICDT)*, Delphi, Greece, January 1997.
- [3] Jacques André, Richard Furuta, and Vincent Quint, editors. *Structured Documents*. Cambridge University Press, 1989.
- [4] Gustavo O. Arocena, Alberto O. Mendelzon, and George A. Mihaila. Applications of a web query language. In *Proceedings of the Sixth International World Wide Web Conference*, Santa Cruz, CA, April 1997.
- [5] Justin Boyan, Dayne Freitag, and Thorsten Joachims. A machine learning architecture for optimizing web search engines. In Franz [10]. <http://www.cs.cmu.edu/reinf/papers/boyan.laser.ps>.
- [6] E. F. Codd. *The Relational Model for Database Management: Version 2*. Addison-Wesley, Reading, MA, 1990.
- [7] C. J. Date and Hugh Darwen. *A Guide to the SQL Standard, 3rd Edition*. Addison-Wesley, Reading, MA, 1992.
- [8] Arthur Do. *HtmlStreamTokenizer*, 1997. <http://www-cs-students.stanford.edu/~do/html-streamer.html>.
- [9] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and padhr Smyth, editors. *Advances in Knowledge Discovery and Data Mining*. MIT Press, Cambridge, MA, 1996.
- [10] Alexander Franz, editor. *The AAAI-96 Workshop on Internet-based Information Systems*. AAAI, August 1996.
- [11] H. P. Frei and Daniel Stieger. Making use of hypertext links when retrieving information. In *Proceedings of the Fourth ACM Conference on Hypertext*, pages 102–111, Washington, DC, 1992. Association for Computing Machinery.
- [12] Richard Furuta. Defining and using structure in digital documents. In *Digital Libraries '94*, 1994.
- [13] Claire L. Green and Peter Edwards. Using machine learning to enhance software tools for internet information management. In Franz [10].
- [14] J. Hammer, H. Garcia-Molina, J. Cho, R. Aranha, and A. Crespo. Extracting semistructured information from the web. In *Proceedings of the Workshop on Management of Semistructured Data*, Tucson, Arizona, May 1997.
- [15] Eric N. Hanson and Jennifer Widom. An overview of production rules in database systems. *The Knowledge Engineering Review*, 8(2), June 1993.

- [16] Michiaki Iwazume, Kengo Shirakami, Kazuaki Hatadani, Hideaki Takeda, and Toyoaki Nishida. Iica: An ontology-based internet navigation system. In Franz [10]. [http://ai-www.aist-nara.ac.jp/doc/people/mitiak-i/aaai96/aaai96\\_5\\_12.html](http://ai-www.aist-nara.ac.jp/doc/people/mitiak-i/aaai96/aaai96_5_12.html).
- [17] M. M. Kessler. Bibliographic coupling between scientific papers. *American Documentation*, 14:10–25, 1963.
- [18] W. Eliot Kimber. HyTime and SGML: Understanding the HyTime HyQ query language 1.1. Available at “<http://www.sgmlu.com/documents/HyTime/HyQ-1.1.txt>”, August 1993.
- [19] Thomas Kistler and Hannes Marais. Webl – a programming language for the web. Technical Report 1997-029, Digital Systems Research Center, December 1997.
- [20] David Konopnicki and Oded Shmueli. Information gathering in the world-wide web: The W3QL query language and the W3QS system. Available from “[www.cs.technion.ac.il/~konop/w3qs.html](http://www.cs.technion.ac.il/~konop/w3qs.html)”, 1997.
- [21] Brian A. LaMacchia. *Internet Fish*. PhD thesis, MIT, 1996.
- [22] Michael Mauldin and John R. R. Leavitt. Web agent related research at the center for machine translation. In *Proceedings of the ACM Special Interest Group on Networked Information Discovery and Retrieval*, 1994. <http://fuzine.mt.cs.cmu.edu/mlm/signidr94.html>.
- [23] Alberto Mendelzon, George Mihaila, and Tova Milo. Querying the world wide web. In *Proceedings of the Fourth International Conference on Parallel and Distributed Information Systems*, Miami, FL, 1996.
- [24] Alberto O. Mendelzon and Tova Milo. Formal models of web queries. In *Proceedings of the Sixteenth ACM Symposium on Principles of Database Systems*, Tucson, Arizona, May 1997.
- [25] George A. Mihaila. WebSQL – an SQL-like query language for the world wide web. Master’s thesis, University of Toronto, 1996.
- [26] Marvin Minsky, editor. *Semantic Information Processing*. MIT Press, Cambridge, MA, 1968.
- [27] Tom M. Mitchell. Challenge problems for artificial intelligence. Panel at AAAI-96, moderated by Bart Selman, 1996.
- [28] Daniel E. O’Leary. The relationship between relevance and reliability in internet-based information and retrieval systems. In Franz [10]. <http://www.usc.edu/dept/sba/atisp/AI/AI-Bus/internet/inter-rr.htm>.
- [29] Original Reusable Objects, Inc. *OROMatcher 1.0 User’s Guide*, 1997. <http://www.oro-inc.com/developers/docs/OROMatcher/index.html>.
- [30] Santeri Paavolainen. MD5 in Java JDK Beta-2. <http://www.cs.hut.fi/~santtu/java/>, 1996.
- [31] M. R. Quillian. *Semantic Memory*. PhD thesis, Carnegie Institute of Technology, 1966.
- [32] Dave Raggett. Html 3.2 reference specification. World-Wide Web Consortium technical report, January 1997.
- [33] Paul Resnick and Hal R. Varian. Recommender systems (introduction to special section). *Communications of the ACM*, 40(3):56–58, March 1997.
- [34] Ronald Rivest. The MD5 message-digest algorithm. Network Working Group Request for Comments: 1321, April 1992.
- [35] Ronald Rousseau. Sitations: an exploratory study. *Cybermetrics*, 1(1), 1997. <http://www.cin-doc.csic.es/cybermetrics/articles/v1i1p1.html>.

- [36] Sriram Sankar, Sreenivasa Viswanadha, and Rob Duncan. *Java Compiler Compiler (JavaCC) - The Java Parser Generator*. Located at “<http://www.suntest.com/JavaCC/>”.
- [37] Jacques Savoy. Citation schemes in hypertext information retrieval. In Maristella Agosti and Alan F. Smeaton, editors, *Information Retrieval and Hypertext*, pages 99–120. Kluwer Academic Press, 1996.
- [38] Erik Selberg and Oren Etzioni. Multi-service search and comparison using the metacrawler. In *Proceedings of the 4th International World Wide Web Conference*, 1995.
- [39] Jonathan Shakes, Marc Langheinrich, and Oren Etzioni. Dynamic reference sifting: A case study in the homepage domain. In *The Sixth International World Wide Web Conference*, April 1997. See “<http://ahoy.cs.washington.edu:6060>”.
- [40] Upendra Shardanand and Pattie Maes. Social information filtering: Algorithms for automating “word of mouth”. In *Computer-Human Interaction (CHI)*, 1995.
- [41] H. Small. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24:265–269, 1973.
- [42] Ellen Spertus. Mining links. Thesis proposal, MIT Department of Electrical Engineering and Computer Science, September 1996.
- [43] Randall H. Trigg. *A network-based approach to text handling for the online scientific community*. PhD thesis, University of Maryland, 1983. Also technical report TR-1346.
- [44] S. Weibel, W. Cathro, and R. Iannella. The 4th dublin core metadata workshop report. *D-Lib Magazine*, June 1997. <http://www.dlib.org/dlib/june97/metadata/06weibel.html>.
- [45] Ross Wilkinson and Michael Fuller. Integrated information access via structure. In Maristella Agosti and Alan Smeaton, editors, *Information Retrieval and Hypertext*, chapter 11. Kluwer Academic Publishers, 1996.
- [46] William A. Woods. What’s in a link: Foundations for semantic networks. In *Readings in Knowledge Representation*, chapter 11. Morgan Kaufmann, 1985.