Bibliography

[ABS99]   S. Abiteboul, P. Buneman, and D. Suciu. *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufmann, San Francisco, California, USA, 1999.

[ACS02]   S. Amer-Yahia, S. Cho, and D. Srivastava. Tree pattern relaxation. In *Proceedings of the 8th Conference on Extending Database Technology (EDBT)*, pages 496–513, Prague, Czech Republic, March 2002.

[ACVW00]   V. Aguilera, S. Cluet, P. Veltri, and F. Wattez. Querying XML documents in Xyleme. In *ACM SIGIR Workshop On XML and Information Retrieval*, pages 79–94, Athens, Greece, July 2000.

[AFL⁺95]   T. Arnold-Moore, M. Fuller, B. Lowe, J. Thom, and R. Wilkinson. The ELF data model and SGQL query language for structured document databases. In *Proceedings of the Australian Database Conference (ADC)*, pages 17–26, Adelaide, Australia, January 1995.

[AG97]   A. Apostolico and Z. Galil, editors. *Pattern Matching Algorithms*, chapter 14: Approximate Tree Pattern Matching, pages 341–371. Oxford University Press, Oxford, UK, June 1997.

[AHU74]   A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, Reading, Massachusetts, USA, 1974.

[AKJK⁺02]   S. Al-Khalifa, H. V. Jagadish, N. Koudas, J. M. Patel, D. Srivastava, and Y. Wu. Structural joins: A primitive for efficient XML query pattern matching. In *Proceedings of the 18th International Conference on Data Engeneering (ICDE)*, San Jose, California, USA, February 2002.

[ANZ01]   A. Aboulnaga, J. F. Naughton, and C. Zhang. Generating synthetic complex-structured XML data. In *Proceedings of the Fourth International*

Workshop on the Web and Databases (WebDB'01), pages 79–85, Santa Barbara, California, USA, May 2001.

[AQM⁺97]   S. Abiteboul, D. Quass, J. McHugh, J. Widom, and J. Wiener. The Lorel query language for semistructured data. *International Journal on Digital Libraries*, 1(1):68–88, 1997.

[ASU86]   A. V. Aho, R. Sethi, and J. D. Ullman. *Compilers: Principles, Techniques, and Tools*. Addison-Wesley, Reading, Massachusetts, USA, 1986.

[BC00]   A. Bonifati and S. Ceri. Comparative analysis of five XML query languages. *SIGMOD Record*, 29(1):68–79, 2000.

[BCF⁺02]   S. Boag, D. Chamberlin, M. F. Fernandez, D. Florescu, J. Robie, J. Siméon, and M. Stefanescu, editors. XQuery 1.0: An XML query language. W3C Working Draft, April 2002. http://www.w3.org/TR/xquery.

[BCML99]   R. Bourret, J. Cowan, I. Macherius, and S. St. Laurent, editors. Document definition markup language (DDML) specification, version 1.0. W3C Note, January 1999. http://www.w3c.org/TR/NOTE-ddml.

[BER02]   The Berkeley Database. Sleepycat Software Inc., Lincoln, Nebraska, USA, 2002. http://www.sleepycat.com/.

[BGP00]   L. Buck, C. F. Goldfarb, and P. Prescod, editors. Datatypes for DTDs (DT4DTD) 1.0. W3C Note, January 2000. http://www.w3.org/TR/dt4dtd.

[BHL99]   T. Bray, D. Hollander, and A. Layman. Namespaces in XML. W3C Recommendation, January 1999. http://www.w3.org/TR/REC-xml-names.

[BM01]   P. V. Biron and A. Malhotra. XML schema part 2: Datatypes. W3C Recommendation, May 2001. http://www.w3.org/TR/xmlschema-2.

[Bos99]   J. Bosak (editor). Four great works of religion. Archive of XML documents, 1999. http://metalab.unc.edu/bosak/xml/eg/rel200.zip.

[BPSM00]   T. Bray, J. Paoli, C. M. Sperberg-McQueen, and E. Maler, editors. Extensible markup language (XML) 1.0 (second edition). W3C Recommendation, October 2000. http://www.w3.org/TR/REC-xml.

[BR99]   R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, Harlow, UK, 1999.

[Bro95]   E. W. Brown. *Execution Performance Issues in Full-Text Information Retrieval*. PhD thesis, University of Massachusetts, USA, October 1995.

[Bur92a]   F. Burkowski. An algebra for hierarchically organized text-dominated databases. *Information Processing and Management*, 28(3):333–348, 1992.

[Bur92b]   F. Burkowski. Retrieval activities in a database consisting of heterogeneous collections of structured texts. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 112–125, Copenhagen, Denmark, June 1992.

[CACS94]   V. Christophides, S. Abiteboul, S. Cluet, and M. Scholl. From structured documents to novel query facilities. In *Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data*, pages 313–324, Minneapolis, Minnessota, USA, June 1994.

[CAM02]   G. Cobena, S. Abiteboul, and A. Marian. Detecting changes in XML documents. In *Proceedings of the 18th International Conference on Data Engeneering (ICDE)*, San Jose, California, USA, February 2002.

[CC92]   K. P. Chan and Y. S. Cheung. Fuzzy attribute graph with application to chinese character recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(1):153–160, 1992.

[CCB95a]   C. Clarke, G. Cormack, and F. Burkowski. An algebra for structured text search and a framework for its implementation. *The Computer Journal*, 38(1):43–56, 1995.

[CCB95b]   C. Clarke, G. Cormack, and F. Burkowski. Schema-independent retrieval from heterogeneous structured text. In *Proceedings of the 4th Annual Symposium on Document Analysis and Information Retrieval*, pages 279–289, Las Vegas, Nevada, USA, April 1995.

[CCD+99]   S. Ceri, S. Comai, E. Damiani, P. Fraternali, S. Paraboschi, and L. Tanca. XML-GL: A graphical language for querying and restructuring XML documents. *WWW8 / Computer Networks*, 31(11-16):1171–1187, 1999.

[CCH92]   J. P. Callan, W. B. Croft, and S. M. Harding. The INQUERY retrieval system. In *Proceedings of the 3rd International Conference on Database and Expert Systems Applications (DEXA)*, pages 78–83, Valencia, Spain, September 1992.

BIBLIOGRAPHY

[CCM96]   V. Christophides, S. Cluet, and G. Moerkotte. Evaluating queries with
            generalized path expressions. In *Proceedings of the 1996 ACM SIGMOD
            International Conference on Management of Data*, pages 413–422, Montreal,
            Canada, June 1996.

[CD99]   J. Clark and S. DeRose, editors. XML path language (XPath) version 1.0.
            W3C Recommendation, November 1999. http://www.w3.org/TR/xpath.

[CDSS98]   S. Cluet, C. Delobel, J. Siméon, and K. Smaga. Your mediators need data
            conversion! In *Proceedings of the 1998 ACM SIGMOD International
            Conference on Management of Data*, pages 177–188, Seattle, Washington,
            USA, June 1998.

[CG99]   S. Chaudhuri and L. Gravano. Evaluating top-k selection queries. In
            *Proceedings of the 25h International Conference on Very Large Databases
            (VLDB)*, pages 397–410, Edinburgh, UK, September 1999.

[Cha90]   S. Chaudhuri. Generalization and a framework for query modification. In *In
            Proceedings of the 6th International Conference on Data Engineering (ICDE)*,
            pages 138–145, Los Angeles, California, USA, February 1990.

[CK96]   Y. Chiaramella and A. Kheirbek. An integrated model for hypermedia and
            information retrieval. In M. Agosti and A. Smeaton, editors, *Information
            Retrieval and Hypertext*, pages 139–176. Kluwer Academic Press, 1996.

[CK97]   M. J. Carey and D. Kossmann. On saying "Enough already!" in SQL. In
            *Proceedings of the 1997 ACM SIGMOD International Conference on
            Management of Data*, pages 219–230, Tucson, Arizona, USA, May 1997.

[CK02]   T. T. Chinenyanga and N. Kushmerick. An expressive and efficient language
            for XML information retrieval. *Journal of the American Society for
            Information Science and Technology (JASIST)*, 53(6):438–453, 2002.

[CLvRC98]   F. Crestani, M. Lalmas, C. J. van Rijsbergen, and I. Campbell. "is this
            document relevant? ... probably": A survey of probabilistic models in
            information retrieval. *ACM Computing Surveys*, 30(4):528–552, 1998.

[CM95]   M. Consens and T. Milo. Algebras for querying text regions. In *Proceedings of
            the 14th Symposium on Principles of Database Systems (PODS)*, pages 11–22,
            San Jose, California, USA, May 1995.

[Coh98]   W. W. Cohen. Integration of heterogeneous databases without common domains using queries based on textual similarity. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, pages 201–212, Seattle, Washington, USA, June 1998.

[Coh00]   W. W. Cohen. WHIRL: A word-based information representation language. *Artificial Intelligence*, 118(1-2):163–196, 2000.

[CRF00]   D. Chamberlin, J. Robie, and D. Florescu. Quilt: An XML query language for heterogeneous data sources. In *Proceedings of 3rd International Workshop on the Web and Databases (WebDB'00)*, pages 53–62, Dallas, Texas, USA, May 2000.

[CRGW96]   S. S. Chawathe, A. Rajaraman, H. Garcia-Molina, and J. Widom. Change detection in hierarchically structured information. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pages 493–504, Montreal, Canada, June 1996.

[Cro83]   W. B. Croft. Experiments with representations in a document retrieval system. *Information Technology: Research and Development*, 2(1):1–21, 1983.

[Cro94]   V. Cross. Fuzzy information retrieval. *Journal of Intelligent Systems*, 3:29–56, 1994.

[CVZ+02]   S.-Y. Chien, Z. Vagena, D. Zhang, V. J. Tsotras, and C. Zaniolo. Efficient structural joins on indexed XML documents. In *Proceedings of the 28th International Conference on Very Large Databases (VLDB)*, pages 263–274, Hong Kong, China, August 2002.

[DD01]   R. Domenig and K. R. Dittrich. Query preprocessing for integrated search in heterogeneous data sources. In *Datenbanksysteme in Büro, Technik und Wissenschaft (BTW)*, pages 154–163, Oldenburg, Germany, March 2001.

[DFF+98]   A. Deutsch, M. Fernandez, D. Florescu, A. Levy, and D. Suciu. XML-QL: A query language for XML. W3C Note, August 1998. http://www.w3.org/TR/NOTE-xml-ql.

[DFH+99]   A. Davidson, M. Fuchs, M. Hedin, M. Jain, J. Koistinen, C. Lloyd, M. Maloney, and K. Schwarzhof. Schema for object-oriented XML 2.0. W3C Note, July 1999. http://www.w3.org/TR/NOTE-SOX.

[DHJ+97]   B. DasGupta, X. He, T. Jiang, M. Li, J. Tromp, and L. Zhang. On distances
             between phylogenetic trees. In *Proceedings of the 8th ACM-SIAM Symposium
             on Discrete Algorithms (SODA)*, pages 427–436, New Orleans, Lousiana, USA,
             January 1997.

[DMO01]   S. DeRose, E. Maler, and D. Orchard, editors. XML Linking Language
             (XLink) Version 1.0. W3C Recommendation, June 2001.
             http://www.w3.org/TR/xlink.

[DT00]   E. Damiani and L. Tanca. Blind queries to XML data. In *Proceedings of the
             11th International Conference on Database and Expert Systems Applications
             (DEXA)*, pages 345–356, London, UK, September 2000.

[FFK+97]   M. Fernandez, D. Florescu, J. Kang, A. Levy, and D. Suciu. STRUDEL: A
             web site management system. In *Proceedings of the 1997 ACM SIGMOD
             International Conference on Management of Data*, pages 549–552, Tucson,
             Arizona, USA, May 1997.

[FG00]   N. Fuhr and K. Großjohann. XIRQL: An extension of XQL for information
             retrieval. In *ACM SIGIR Workshop On XML and Information Retrieval*,
             pages 11–17, Athens, Greece, July 2000.

[FG01]   N. Fuhr and K. Großjohann. XIRQL: A query language for information
             retrieval in XML documents. In *Proceedings of the 24th Annual International
             ACM SIGIR Conference on Research and Development in Information
             Retrieval*, pages 172–180, New Orleans, Lousiana, USA, September 2001.

[FR97]   N. Fuhr and T. Rölleke. A probabilistic relational algebra for the integration
             of information retrieval and database systems. *ACM Transactions on
             Information Systems*, 15(1):32–66, 1997.

[FSW99]   M. Fernandez, J. Siméon, and P. Wadler. XML and query languages:
             Experiences and exemplars. Draft manuscript, September 1999.
             http://www-db.research.bell-labs.com/user/simeon/xquery.ps.

[Fuh92]   N. Fuhr. Probabilistic models in information retrieval. *The Computer Journal*,
             35(3):243–254, 1992.

[GMW99]   R. Goldman, J. McHugh, and J. Widom. From semistructured data to XML:
             Migrating the lore data model and query language. In *Proceedings of the 2nd*

International Workshop on the Web and Databases (WebDB'99), pages 25–30, Philadelphia, Pennsylvania, USA, June 1999.

[GSVGM98] R. Goldman, N. Shivakumar, S. Venkatasubramanian, and H. Garcia-Molina. Proximity search in databases. In *Proceedings of the 24th International Conference on Very Large Databases (VLDB)*, pages 26–37, New York, USA, August 1998.

[GW97] R. Goldman and J. Widom. DataGuides: Enabling query formulation and optimization in semistructured data. In *Proceedings of the 23rd International Conference on Very Large Databases (VLDB)*, pages 436–445, Athens, Greece, August 1997.

[Hay94] P. Hayes. Namefinder: Software that finds names in text. In *Proceedings of the 4th RIAO Conference on Computer Assisted Information Searching on the Internet*, pages 762–774, New York, USA, October 1994.

[HMG97] J. Hammer, J. McHugh, and H. Garcia-Molina. Semistructured data: The TSIMMIS experience. In *Proceedings of the First East-European Symposium on Advances in Databases and Information Systems (ADBIS'97)*, pages 1–8, St. Petersburg, Russia, September 1997.

[HP00] H. Hosoya and B. C. Pierce. XDuce: A typed XML processing language (preliminary report). In *Proceedings of 3rd International Workshop on the Web and Databases (WebDB'00)*, pages 226–244, Dallas, Texas, USA, May 2000.

[HTK00] Y. Hayashi, J. Tomita, and G. Kikui. Searching text-rich XML documents with relevance ranking. In *ACM SIGIR Workshop On XML and Information Retrieval*, Athens, Greece, July 2000.

[HWC+99] M. J. Healey, J. T.-L. Wang, G. Chang, A. Revankar, and G. Samtani. Precise environmental searches: EnviroDaemon with hierarchical information search. *Environmental Quality Management*, 9(1):51–62, 1999.

[ISO86] ISO (International Organization for Standardization). *ISO 8879:1986(E). Information processing – Text and Office Systems – Standard Generalized Markup Language (SGML)*, 1986.

[JWZ94] T. Jiang, L. Wang, and K. Zhang. Alignment of trees - an alternative to tree edit. In *Proceedings of the 5th Annual Symposium on Combinatorial Pattern Matching*, pages 75–86, Asilomar, California, USA, June 1994.

[KBP99]   D. H. Kraft, G. Bordogna, and G. Pasi. Fuzzy set techniques in information retrieval. In J. C. Bezdek, D. Dubois, and H. Prade, editors, *Fuzzy Sets in Approximate Reasoning and Information Systems*. Kluwer Academic Publishers, Boston, Massachusetts, USA, 1999.

[Kil92]   P. Kilpeläinen. *Tree Matching Problems with Applications to Structured Text Databases*. PhD thesis, University of Helsinki, Finland, November 1992.

[KLR01]   G. Kazai, M. Lalmas, and T. Rölleke. A model for the representation and focussed retrieval of structured documents based on fuzzy aggregation. In *Proceedings of the 8th South American Symposium on String Processing and Information Retrieval (SPIRE'01)*, pages 123–135, Laguna de San Rafael, Chile, November 2001.

[KNS99]   Y. Kanza, W. Nutt, and Y. Sagiv. Queries with incomplete answers over semistructured data. In *Proceedings of the 18th Symposium on Principles of Database Systems (PODS)*, pages 227–236, Philadelphia, Pennsylvania, USA, May–June 1999.

[Knu69]   D. E. Knuth. *The Art of Computer Programming*, volume 1. Addison-Wesley, Reading, Massachusetts, USA, 1969.

[KS01]   Y. Kanza and Y. Sagiv. Flexible queries over semistructured data. In *Proceedings of the 20th Symposium on Principles of Database Systems (PODS)*, pages 40–51, Santa Barbara, California, USA, May 2001.

[Kwo89]   K. L. Kwok. A neural network for probabilistic information retrieval. In *Proceedings of the 12th International Conference on Research and Development in Information Retrieval*, pages 21–30, Cambridge, Massachusetts, USA, June 1989.

[Lal97]   M. Lalmas. Dempster-Shafer's theory of evidence applied to structured documents: Modelling uncertainty. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 110–118, Philadelphia, Pennsylvania, USA, July 1997.

[Lal98]   M. Lalmas. Information retrieval and Dempster-Shafer's theory of evidence. In A. Hunter and S. Parson, editors, *Applications of Uncertainty Formalisms*, Lecture Notes in Computer Science, pages 157–176. Springer Verlag, Heidelberg, Germany, 1998.

[Lal00]   M. Lalmas. Uniform representation of content and structure for structured document retrieval. In *Proceedings of the 20th SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence*, pages 215–228, Cambridge, UK, December 2000.

[Lee94]   J. H. Lee. Properties of extended Boolean models in information retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 182–190, Dublin, Ireland, July 1994.

[Lee02]   D. Lee. *Query Relaxation for XML Model*. PhD thesis, University of California, USA, June 2002.

[Lit80]   W. Litwin. Linear hashing: A new tool for file and table addressing. In *Proceedings of the 6th International Conference on Very Large Databases (VLDB)*, pages 212–223, Montreal, Canada, October 1980.

[Loe94]   A. Loeffen. Text databases: A survey of text models and systems. *SIGMOD RECORD*, 23(1):97–106, 1994.

[LR98]   M. Lalmas and I. Ruthven. Representing and retrieving structured documents using the dempster-shafer theory of evidence: Modelling and evaluation. *Journal of Documentation*, 54(5):529–565, 1998.

[Mac90]   I. A. Macleod. Storage and retrieval of structured documents. *Information Processing and Management*, 26(2):197–208, 1990.

[Mac91]   I. A. MacLeod. A query language for retrieving information from hierarchic text structures. *The Computer Journal*, 34(3):254–264, 1991.

[May01]   W. May. XPathLog: A declarative, native XML data manipulation language. In *Proceedings of the International Database Engineering and Applications Symposium (IDEAS'01)*, pages 123–128, Grenoble, France, July 2001.

[Meg98]   D. Megginson. SAX 1.0: The simple API for XML. Web page, 1998. http://www.megginson.com/SAX/index.html.

[Mer99]   P. Merialdo. ACM SIGMOD record: XML version. Web page, 1999. http://www.acm.org/sigmod/record/xml.

[Meu00]   H. Meuss. *Logical Tree Matching with Complete Answer Aggregates for Retrieving Structured Documents*. PhD thesis, University of Munich, Germany, April 2000.

[MJKZ98]   S. H. Myaeng, D.-H. Jang, M.-S. Kim, and Z.-C. Zhoo. A flexible model for retrieval of SGML documents. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 138–145, Melbourne, Australia, August 1998.

[MM90]   U. Manber and E. W. Myers. Suffix arrays: A new method for on-line string searches. In *Proceedings of the ACM SIAM Symposium on Discrete Algorithms*, pages 319–327, San Francisco, California, USA, January 1990.

[MNF58]   G. A. Miller, E. B. Newman, and E. A. Friedman. Length-frequency statistics for written english. *Information and Control*, 1:370–389, 1958.

[Moe00]   G. Moerkotte. YAXQL: A powerful and web-aware query language supporting query reuse. Technical report, University of Mannheim, Germany, January 2000.

[MS99]   T. Milo and D. Suciu. Index structures for path expressions. In *Proceedings of the 7th International Conference of Database Theory (ICDT)*, pages 277–295, Jerusalem, Israel, January 1999.

[MW99]   J. McHugh and J. Widom. Query optimization for XML. *The International Journal on Very Large Databases*, pages 315–326, 1999.

[MZ92]   E. Mackie and J. Zobel. Retrieval of tree-structured data from disc. In *Proceedings of the 3rd Australian Database Conference*, pages 209–216, Melbourne, Australia, February 1992.

[Nav95]   G. Navarro. A language for queries on structure and contents of textual databases. Master's thesis, Department of Computer Science, University of Chile, April 1995.

[NB96]   G. Navarro and R. Baeza-Yates. Integrating content and structure in text retrieval. *SIGMOD Record*, 25(1):67–79, 1996.

[NB97]   G. Navarro and R. Baeza-Yates. Proximal nodes: A model to query document databases by contents and structure. *ACM Transactions on Information Systems*, 15(4):401–435, 1997.

[NB02]   L. Gravano N. Bruno, S. Chaudhuri. Top-k selection queries over relational databases: Mapping strategies and performance evaluation. *ACM Transactions on Database Systems (TODS)*, 27(2), 2002.

[NBVF98]   G. Navarro, R. Baeza-Yates, J. Vegas, and P. de la Fuente. A model and a
            visual query language for structured text. In *Proceedings of the 5th South
            American Symposium on String Processing and Information Retrieval
            (SPIRE'98)*, pages 7–13, Sta. Cruz de la Sierra, Bolivia, September 1998.

[NJ02]   A. Nierman and H. V. Jagadish. Evaluating structural similarity in XML
            documents. In *Proceedings of the Fifth International Workshop on the Web
            and Databases (WebDB'02)*, Madison, Wisconsin, USA, June 2002.

[NS00]   F. Neven and T. Schwentick. Expressive and efficient pattern languages for
            tree-structured data. In *Proceedings of the 19th Symposium on Principles of
            Database Systems (PODS)*, pages 145–156, Dallas, Texas, USA, May 2000.

[PGF00]   P. Pathak, M. Gordon, and W. Fan. Effective information retrieval using
            genetic algorithms based matching functions adaptation. In *Proceedings of the
            33th Hawaii International Conference on System Sciences (HICSS)*, Maui,
            Hawaii, January 2000.

[PGW95]   Y. Papakonstantinou, H. Garcia-Molina, and J. Widom. Object exchange
            across heterogeneous information sources. In *Proceedings of the 11th
            International Conference on Data Engineering (ICDE)*, pages 251–260, Taipei,
            Taiwan, March 1995.

[Por80]   M. F. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[RH90]   B. N. Rossiter and M. A. Heather. Strengths and weaknesses of database
            models for textual documents. In *Cambridge Series on Electronic Publishing*,
            pages 125–138. Cambridge University Press, Gaithersburg, Maryland, USA,
            1990.

[RJB98]   J. Rumbaugh, I. Jacobson, and G. Booch. *The Unified Modeling Language
            Reference Manual*. Addison-Wesley, Reading, Massachusetts, USA, 1998.

[RLS92]   E. H. Ruspini, J. D. Lowrance, and T. M. Strat. Understanding evidental
            reasoning. *International Journal of Approximate Reasoning*, 6:401–424, 1992.

[RLS98]   J. Robie, J. Lapp, and D. Schach. XML query language (XQL). Web page,
            September 1998. http://www.w3.org/TandS/QL/QL98/pp/xql.html.

[RW97]   Y. Ravin and N. Wacholder. Extracting names from natural language text.
            Technical Report RC 20338, IBM, San Jose, California, USA, October 1997.

[Sal68]   G. Salton. *Automatic information organization and retrieval.* Mc Graw Hill,
            New York, USA, 1968.

[SC98]   J. Siméon and S. Cluet. Using YAT to build a web server. In *Proceedings of
            the 1st International Workshop on the Web and Databases (WebDB'98)*, pages
            118–135, Valencia, Spain, March 1998.

[Sch95]   U. Schöning. *Logik für Informatiker.* Spektrum Akademischer Verlag,
            Heidelberg, Germany, 1995.

[Sch00]   T. Schlieder. Strukturelle Ähnlichkeitssuche in XML-Dokumenten. In
            Tagungsband des 12. Workshop "Grundlagen von Datenbanken", pages 91–95,
            Plön, Germany, April 2000.

[Sch01a]   T. Schlieder. ApproXQL: Design and implementation of an approximate
            pattern matching language for XML. Technical Report B 01-02, Freie
            Universität Berlin, Germany, May 2001.

[Sch01b]   T. Schlieder. Similarity search in XML data using cost-based query
            transformations. In *Proceedings of the Fourth International Workshop on the
            Web and Databases (WebDB'01)*, pages 19–24, Santa Barbara, California,
            USA, May 2001.

[Sch02a]   T. Schlieder. Schema-driven evaluation of approximate tree-pattern queries. In
            *Proceedings of the 8th International Conference on Extending Database
            Technology (EDBT)*, pages 514–532, Prague, Czech Republic, March 2002.

[Sch02b]   T. Schlieder. Schema-driven evaluation of ApproXQL queries. Technical
            Report B 02-01, Freie Universität Berlin, Germany, January 2002.

[Sel77]   S. M. Selkow. The tree-to-tree editing problem. *Information Processing
            Letters*, 6(6):184–186, 1977.

[SFW83]   G. Salton, E. A. Fox, and H. Wu. Extended Boolean information retrieval.
            Communications of the ACM, 26(11):1022–1036, 1983.

[Sha76]   G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press,
            Princeton, New Jersey, USA, 1976.

[SKW01]   A. Schmidt, M. L. Kersten, and M. Windhouwer. Querying XML documents
            made easy: Nearest concept queries. In *Proceedings of the 17th International*

Conference on Data Engineering (ICDE), pages 321–329, Heidelberg, Germany, April 2001.

[SM83]   G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill, Tokio, Japan, 1983.

[SM00]   T. Schlieder and H. Meuss. Result ranking for structured queries against XML documents. In *DELOS Workshop on Information Seeking, Searching and Querying in Digital Libraries*, Zurich, Switzerland, December 2000.

[SM02]   T. Schlieder and H. Meuss. Querying and ranking XML documents. *Journal of the American Society for Information Science and Technology (JASIST)*, 53(6):489–503, 2002.

[SN00]   T. Schlieder and F. Naumann. Approximate tree embedding for querying XML data. In *ACM SIGIR Workshop On XML and Information Retrieval*, pages 53–67, Athens, Greece, July 2000.

[ST93]   A. Salminen and F. W. Tompa. PAT expressions: An algebra for text search. *Acta Linguistica Hungarica*, 41(1-4):277–306, 1993.

[STW01]   S. Sizov, A. Theobald, and G. Weikum. Ähnlichkeitssuche auf XML-Daten. In *Datenbanksysteme in Büro, Technik und Wissenschaft (BTW)*, pages 364–383, Oldenburg, Germany, March 2001.

[SWG02]   D. Shasha, J. T.-L. Wang, and R. Giugno. Algorithmics and applications of tree and graph searching. In *Proceedings of the 21st Symposium on Principles of Database Systems (PODS)*, pages 39–52, Madison, Wisconsin, USA, June 2002.

[SZ90]   B. A. Shapiro and K. Zhang. Comparing multiple RNA secondary structures using tree comparisons. *Computer Applications in Biosciences*, 6(4):309–318, 1990.

[SZW01]   D. Shasha, K. Zhang, and J. Wang. TreeSearch: Searching among unordered trees. Web page, February 2001. http://cs.nyu.edu/cs/faculty/shasha/papers/treesearch.html.

[Tai78]   K.-C. Tai. Syntactic error correction in programming languages. *IEEE Transactions on Software Engineering*, 4(5):414–425, 1978.

BIBLIOGRAPHY

[Tai79]   K.-C. Tai. The tree-to-tree correction problem. *Journal of the ACM*,
             26(3):422–433, 1979.

[TBMM01]   H. S. Thompson, D. Beech, M. Maloney, and N. Mendelsohn. XML schema
             part 1: Structures. W3C Recommendation, May 2001.
             http://www.w3.org/TR/xmlschema-1.

[TT88]   E. Tanaka and K. Tanaka. The tree-to-tree editing problem. *International
             Journal of Pattern Recognition and Artificial Intelligence*, 2(2):221–240, 1988.

[TW00]   A. Theobald and G. Weikum. Adding relevance to XML. In *Proceedings of 3rd
             International Workshop on the Web and Databases (WebDB'00)*, pages 35–40,
             Dallas, Texas, USA, May 2000.

[TW02]   A. Theobald and G. Weikum. The index-based XXL search engine for
             querying XML data with relevance ranking. In *Proceedings of the 8th
             Conference on Extending Database Technology (EDBT)*, pages 477–495,
             Prague, Czech Republic, March 2002.

[WFC99]   J. E. Wolff, H. Flörke, and A. B. Cremers. XPRES: A ranking approach to
             retrieval on structured documents. Technical Report IAI-TR-99-12, University
             of Bonn, Germany, July 1999.

[WFC00]   J. E. Wolff, H. Flörke, and A. B. Cremers. Searching and browsing collections
             of structural information. In *Proceedings of IEEE Advances in Digital
             Libraries (ADL 2000)*, pages 141–150, Washington DC, USA, May 2000.

[WH91]   R. Wilkinson and P. Hingston. Using the cosine measure in a neural network
             for document. In *Proceedings of the 14th Annual International ACM SIGIR
             Conference on Research and Development in Information Retrieval*, pages
             202–210, Chicago, Illinois, USA, October 1991.

[Wir77]   N. Wirth. What can we do about the unnecessary diversity of notation for
             syntatic definitions? *Communications of the ACM*, 20(11):822–823, 1977.

[XER02]   Xerces XML parser for C++. Web page. The Apache Software Foundation,
             2002. http://xml.apache.org/xerces-c/.

[Yag96]   R. R. Yager. Quantifier guided aggregation using OWA operators.
             International Journal of Intelligent Systems, 11:49–73, 1996.

[YKR93]   J.-J. Yang, R. Korfhage, and E. M. Rasmussen. Query improvement in information retrieval using genetic algorithm: a report on the experiments of the TREC project. In *Proceedings of the Text Retrieval Conference (TREC-1)*, pages 31–58, Gaithesburg, Maryland, USA, November 1993.

[Zha93]   K. Zhang. A new editing based distance between unordered labeled trees. In *Proceedings of the 4th Annual Symposium on Combinatorial Pattern Matching*, pages 254–265, Padova, Italy, June 1993.

[Zip49]   G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, Massachusetts, USA, 1949.

[ZND⁺01]   C. Zhang, J. F. Naughton, D. J. DeWitt, Q. Luo, and G. M. Lohman. On supporting containment queries in relational database management systems. In *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, pages 425–436, Santa Barbara, California, USA, May 2001.

[ZSS92]   K. Zhang, R. Stratman, and D. Shasha. On the editing distance between unordered labeled trees. *Information Processing Letters*, 42:133–139, 1992.

[ZSW94]   K. Zhang, D. Shasha, and J. T.-L. Wang. Approximate tree matching in the presence of variable length don't cares. *Journal of Algorithms*, 16(1):33–66, 1994.

[ZWS95]   K. Zhang, J. T.-L. Wang, and D. Shasha. On the editing distance between undirected acyclic graphs and related problems. In *Proceedings of the 6th Annual Symposium on Combinatorial Pattern Matching*, pages 395–407, Espoo, Finland, July 1995.