

Bibliography

- Abney S. "Parsing by chunks." In Berwick R, Abney S, and Tenny C, editors, *Principle-based Parsing*. Kluwer Academic Publishers, 1991 36
- Agrawal S, Chaudhuri S, Das G. "DBXplorer: A System for Keyword-Based Search over Relational Databases", *18th International Conference on Data Engineering (ICDE'02)*, pp. 5–16, 2002. DOI: [10.1109/ICDE.2002.994693](https://doi.org/10.1109/ICDE.2002.994693) 49
- Agrawal R. "Search and Data Management". Summary presentation at the Claremont Database Research Self-Assessment Meeting, Berkeley, CA, May 29-30, 2008, <http://db.cs.berkeley.edu/claremont/>
- Agrawal R et al. "The Claremont Report on Database Research". Report issued by the Claremont Database Research Self-Assessment Meeting, Berkeley, CA, May 29-30, 2008, <http://db.cs.berkeley.edu/claremont/> DOI: [10.1145/1462571.1462573](https://doi.org/10.1145/1462571.1462573) 62
- Andrews W. "Gartner MarketScope for Enterprise Search," Gartner, Inc., ID Number: G00206087, November 2010. 3
- Andrews W and Knox RE. "Gartner Magic Quadrant for Information Access Technology," Gartner, Inc., ID Number: G00161178, 2008 61, 62
- Angles R and Gutierrez C. "Survey of graph database models." *ACM Computing Surveys*, vol. 40, no. 1, pp. 1–39, 2008. DOI: [10.1145/1322432.1322433](https://doi.org/10.1145/1322432.1322433) 27
- Baeza-Yates C and Ribeiro-Neto B. *Modern Information Retrieval*. Second Edition. Addison Wesley Longman, 2010. 36
- Basu R. *Implementing quality: A practical guide to tools and techniques*. Thomson, London, 2004. 87
- Bernstein P and Newcomer E. *Principles of transaction processing*, Morgan Kaufmann, Chapter 6, 2009. 30
- Bizer C, Heath T and Berners-Lee T. "Linked Data – The Story So Far". In Heath T, Hepp M, Bizer C. (eds) *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 5, no. 3, pp. Pages 1-22, 2009. 41
- Boag S, Chamberlin D, Fernandez MF, Florescu D, Robie J and Simeon J. "XQuery 1.0: An XML query language". Technical report, World Wide Web Consortium, <http://www.w3.org/TR/xquery/> January 2007. 49

110 16. FUTURE DIRECTIONS

- Bonifati A, Cattaneo F, Ceri S, Fuggetta A and Paraboschi S. “Designing data marts for data warehouses”, *ACM Transactions on Software Engineering Methodology* vol 4, pp. 452–483, 2001. DOI: [10.1145/384189.384190](https://doi.org/10.1145/384189.384190) 13
- Brin S and Page L. “Anatomy of a large-scale hypertextual web search engine” In *Proceedings of the 7th International World Wide Web Conference* pp. 107–117, Brisbane, Australia, Apr. 14–18, 1998. DOI: [10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X) 12, 43
- Brewer E. “Towards Robust Distributed Systems”, Keynote Address, Annual ACM Symposium on Principles of Distributed Computing (PODC), July 19, 2000. <http://www.cs.berkeley.edu/~brewer/cs262b-2004/PODC-keynote.pdf> DOI: [10.1145/343477.343502](https://doi.org/10.1145/343477.343502) 24
- Büttcher S, Clarke CLA and Cormack GV. *Information Retrieval: Implementing and Evaluating Search Engines*, MIT press, 2010. 18
- Cafarella MJ, Madhavan J and Halevy A. “Web-scale extraction of structured data.” *SIGMOD Rec.*, vol. 37, no. 4, pp.55–61, 2009 DOI: [10.1145/1519103.1519112](https://doi.org/10.1145/1519103.1519112) 41
- Carey MJ, Haas LM, Kleewein J and Reinwald B. “Data access interoperability in the IBM database family.” *IEEE Quarterly Bulletin on Data Engineering; Special Issue on Interoperability*, vol. 21, no. 3, pp. 4–11, 1998. 24
- Caverlee J and Liu L. “Countering web spam with credibility-based link analysis.” In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Principles of Distributed Computing*, Portland, Oregon, August 12 - 15, PODC '07. ACM, New York, NY, pp. 157-166, 2007. DOI: [10.1145/1281100.1281124](https://doi.org/10.1145/1281100.1281124) 37
- Chang F, Dean J, Ghemawat S, Hsieh W, Wallach D, Burrows M, Chandra T, Fikes A, and Gruber R. “Bigtable: A distributed storage system for structured data.” In *Proceedings of the 7th USENIX Symposium on Operating Systems Design and Implementation*. pp. 205–218, 2006. 26
- Cheng J, Ke Y, and Ng W. “Efficient query processing on graph databases.” *ACM Transactions on Database Systems (TODS)* vol. 34, no. 1, April 2009. DOI: [10.1145/1508857.1508859](https://doi.org/10.1145/1508857.1508859) 27
- Chaudhuri S and Dayal U. “An overview of data warehousing and OLAP technology.” *SIGMOD Record*, vol 26, 1997. DOI: [10.1145/248603.248616](https://doi.org/10.1145/248603.248616) 13
- Claybrook B. “On Line Transaction Processing Systems”, John Wiley & Sons, 1992. 13
- Clarke CLA and Cormack GV. “Shortest-substring retrieval and ranking.” *ACM Trans. Inf. Syst.* vol 18, no. 1, pp. 44–78, 2000. DOI: [10.1145/333135.333137](https://doi.org/10.1145/333135.333137) 37
- Cleverdon CW and Mills J. “The testing of index language devices”. *Aslib Proceedings*, vol 15, no. 4, pp. 106–130, 1963. DOI: [10.1108/eb049925](https://doi.org/10.1108/eb049925) 11

- Codd EF. *The Relational Model for Database Management*, Addison-Wesley Publishing Company, 1990 19
- Copestake A and Sparck Jones K. "Natural Language Interfaces to Databases", *Knowledge Engineering Review*, vol. 5, no. 4, pp. 225–249, 2005. DOI: [10.1017/S0269888900005476](https://doi.org/10.1017/S0269888900005476) 10
- Comer D. "The ubiquitous B-tree." *ACM Computing Surveys*, vol 11, no. 2, pp. 121–137, 1979. DOI: [10.1145/356770.356776](https://doi.org/10.1145/356770.356776) 19
- Councill IG, Giles CL, Iorio ED, Gori M, Maggini M and Pucci A. "Towards next generation cite-seer: A flexible architecture for digital library deployment". In *Research and Advanced Technology for Digital Libraries, ECDL2006*, pp. 111–122, 2006. DOI: [10.1007/11863878_10](https://doi.org/10.1007/11863878_10) 19
- DeCandia G, Hastorun D, Jampani M, Kakulapati G, Laksham A, Pilchin A, Sivasubramanian S, Vosshall P and Vogels W. "Dynamo: Amazon's highly available key-value store." In *Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles* ACM Press, New York, pp. 205–220, 2007. DOI: [10.1145/1294261.1294281](https://doi.org/10.1145/1294261.1294281) 25
- Dedrick J, Xu S and Zhu K. "How Does Information Technology Shape Supply-Chain Structure? Evidence on the Number of Suppliers." *Journal of Management Information Systems* vol 25, no. 2, pp. 41–72, 2008. DOI: [10.2753/MIS0742-1222250203](https://doi.org/10.2753/MIS0742-1222250203) 9
- Deutsch P and Emtage A. "Thearchie System: An Internet Electronic Directory Service," *ConneXions*, vol 6, no. 2, February 1992. 11
- Fagin R, Kumar R, McCurley KS, Novak J, Sivakumar D, Tomlin JA and Williamson DP. "Searching the workplace web", in *Proceedings of the 12th international conference on World Wide Web*, Budapest, Hungary, ACM, pp. 366–375, 2003. DOI: [10.1145/775152.775204](https://doi.org/10.1145/775152.775204) 38
- Feldman S and Reynolds H. "Worldwide Search and Discovery 2009 Vendor Shares and Update on Market Trends," IDC #223926, July, 2010. 2, 3
- Filo D and Yang J. *Yahoo! unplugged. Your Discovery Guide to the Web* Foster City: IDG Books Worldwide, 1995. 11
- Gabrilovich E and Markovitch R. "Wikipedia-based semantic interpretation for natural language processing." *Journal of Artificial Intelligence Research*, vol 34, 443–498, 2009. DOI: [10.1613/jair.2669](https://doi.org/10.1613/jair.2669) 38
- Gantz J and Reinsel D. "The Digital Universe Decade – Are You Ready?", IDC, May 2010. Sponsored by EMC Corporation. 9
- Giannadakis N, Rowe A, Ghanem M and Guo Y. "InfoGrid: providing information integration for knowledge discovery," *Information Sciences*, vol 155, no. 3–4, pp. 199–226, 2003. DOI: [10.1016/S0020-0255\(03\)00170-1](https://doi.org/10.1016/S0020-0255(03)00170-1) 27

112 16. FUTURE DIRECTIONS

- Gilbert S and Lynch N. "Brewer's Conjecture and the feasibility of Consistent, Partition Tolerant Web Services," *ACM SIGACT News*, vol 155, no. 3-4, pp. 199-226, 2003. vol 33, no. 2, pp. 51-59, 2002. DOI: [10.1145/564585.564601](https://doi.org/10.1145/564585.564601) 24
- Grefenstette G. "Comparing two language identification schemes". In *Proceedings of the 3rd International Conference on Statistical Analysis of Textual Data (JADT'95)*. pp. 263-268, 1995. 35
- Grefenstette G, Qu Y, Evans DA and Shanahan JG. "Validating the Coverage of Lexical Resources for Affect Analysis and Automatically Classifying New Words along Semantic Axes." In Qu Y., Shanahan JG and Wiede J (eds) *Exploring Attitude and Affect in Text: Theories and Applications*, AAAI-2004 Spring Symposium Series, pp. 1-78, 2004. DOI: [10.1007/1-4020-4102-0_9](https://doi.org/10.1007/1-4020-4102-0_9) 41
- Grover C, Matheson C, Mikheev A and Moens M. "LT TTT - A flexible tokenisation tool." In LREC 2000 *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens, pp 1147-1154, 2000. 36
- Guttman A. "R-Trees: A Dynamic Index Structure for Spatial Searching," *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 47-54, 1984 DOI: [10.1145/971697.602266](https://doi.org/10.1145/971697.602266) 19
- Hasselbring W. "Information system integration," *Communications of the ACM* vol 43, no. 6, pp. 32-38, June, 2000. 15
- Heydon A and Najork M. "Mercator: A scalable, extensible Web crawler." *World Wide Web*, vol 2, no. 43, pp. 219-229, December 1999. DOI: [10.1023/A:1019213109274](https://doi.org/10.1023/A:1019213109274) 29
- Hull R. "Managing Semantic Heterogeneity in Databases: A Theoretical Perspective." In *SIGART Symposium on Principles of Database Systems*, pp. 51-61, 1997. DOI: [10.1145/263661.263668](https://doi.org/10.1145/263661.263668) 24
- Ilyas IF, Aref WG and Elmagarmid AK. "Supporting top-k join queries in relational databases." *VLDB Journal*, vol. 13, no. 3, pp. 207-221, 2004 DOI: [10.1007/s00778-004-0128-2](https://doi.org/10.1007/s00778-004-0128-2) 43
- Ji H and Grishman R. "Refining Event Extraction through Cross-Document Inference." In *Proceedings of ACL-08: HLT*, pp.254-262, Columbus, OH, June, 2008 40
- Kazama J and Torisawa K. "Exploiting Wikipedia as External Knowledge for Named Entity Recognition", in *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2007. 39
- Kofler M. *The Definitive Guide to MySQL 5*, 3rd edn. Apress, Berkeley, 2005. 38
- Lamel L and Gauvain JL. "Speech processing for audio indexing," *GoTAL 2008 - Advances in NLP*, 5221/2008 LNCS pp. 4-15, Springer Verlag, 2008. 41
- Laudon KC and Laudon JP. *Essential of management information systems* (9th ed). Englewood cliffs, NJ: Prentice Hall, 2010.

- Leavitt N. "Will NoSQL Databases Live Up to Their Promise?" *Computer*, vol 43, pp. 12–14, 2010. DOI: [10.1109/MC.2010.58](https://doi.org/10.1109/MC.2010.58) 24, 32
- Lew MS, Sebe N, Djeraba C and Jain R. "Content-based multimedia information retrieval: State of the art and challenges." *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, pp. 1–19, 2006 DOI: [10.1145/1126004.1126005](https://doi.org/10.1145/1126004.1126005) 41
- Mani I and Maybury MT. (Eds.) *Advances in Automatic Text Summarization*, The MIT Press, 1999. 39
- Manning C and Schütze H. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999. 36, 38
- Masterman M, Needham RM and Sparck Jones K. "The analogy between mechanical translation and library retrieval", *Proceedings of the International Conference on Scientific Information (1958)*, National Academy of Sciences - National Research Council, Washington, D.C., vol. 2, pp. 917–935, 1959 11
- McGuinness DL. "Ontologies come of age." In Fensel D, HendlerJ, Lieberman H, Wahlster W, editors. *The semantic web: why, what, and how*. Cambridge: MIT Press, 2001
- Mitkov R. "Outstanding issues in anaphora resolution." In Gelbukh, A., ed.: *Proceedings of the 2nd International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2001)*. Volume 2004 of Lecture Notes in Computer Science., Berlin, Springer, pp. 110–125, 2001 40
- Negash S and Gray P. *Business Intelligence*, Springer, Berlin, Heidelberg, 2008. 13
- Nijssen G and Halpin T. *Conceptual Schema and Relational Database Design: a fact oriented approach* Prentice-Hall, Sydney, Australia, 1989 13
- Page L, Brin S, Motwani R and Winograd T "The PageRank citation ranking: bringing order to the Web." Technical report, Stanford Digital Library Technologies Project, 1998. 37
- Pinkerton B. "Finding What People Want: Experiences with the WebCrawler." The Second International WWW Conference Chicago, USA, October 17-20, 1994. 11
- Qi X and Davison BD. "Web page classification: Features and algorithms." *ACM Comput. Surv.* vol. 41, no. 2, 2009. DOI: [10.1145/1459352.1459357](https://doi.org/10.1145/1459352.1459357) 41
- Reese G. *Database programming with JDBC and Java*. Sebastopol CA: O'Reilly & Associates, 2000. 38
- Russell SJ and Norvig P, *Artificial Intelligence: A Modern Approach*, 3rd edition, Prentice Hall, 2009. 41

114 16. FUTURE DIRECTIONS

- Salton G. *The SMART Retrieval System's Experiments in Automatic Document Processing*, Prentice-Hall, Inc. 1971. 11
- Salton G. *Automatic Text Processing*. Addison-Wesley, Reading, Mass, 1989. 36
- Thomson, A. and Abadi, D. "The Case for Determinism in Database Systems," *Proceedings of the VLDB Endowment*, Vol. 3, No. 1, 2010. <http://db.cs.yale.edu/determinism-vldb10.pdf> 33
- Tunkelang D. *Faceted Search Synthesis Lectures on Information Concepts, Retrieval, and Services*. Morgan & Claypool Publishers, 2009. 14
- Vicknair C, Macias M, Zhao Z, Nan X, Chen Y, Wilkins D. "A Comparison of a Graph Database and a Relational Database: A Data Provenance Perspective," *Proc. of the ACM Southeast Conference (ACMSE)*, Oxford, Mississippi, April 2010. 27
- Wolf JL, Squillante MS, Yu PS, Sethuraman J, Ozsen L. "Optimal crawling strategies for web search engines." In *Proceedings of the 11th International World Wide Web conference*, pp. 136–147, 2002. DOI: 10.1145/511446.511465 29
- Yoshikawa M, Amagasa T, Shimura T, Uemura S. "XRel: A path-based approach to storage and retrieval of XML documents using relational databases" *ACM Transactions on Internet Technology*, vol 1, no. 1, pp. 110–141, August 2001 DOI: 10.1145/383034.383038