Bibliography

1. *The Description Logic Handbook: Theory, Implementation and Applications, 2nd Edition*. Cambridge University Press, 2007.
2. Karl Aberer, Philippe Cudré-Mauroux, and Manfred Hauswirth. The Chatty Web: Emergent semantics through gossiping. In *12th World Wide Web Conference*, 2003.
3. Serge Abiteboul, Omar Benjelloun, Bogdan Cauytis, Ioana Manolescu, Tova Milo, and Nicoleta Preda. Lazy query evaluation for Active XML. In *SIGMOD*, June 2004.
4. Serge Abiteboul, Omar Benjelloun, and Tova Milo. The Active XML project: An overview. *VLDB J.*, 17(5), 2008.
5. Serge Abiteboul, Angela Bonifati, Gregory Cobena, Ioana Manolescu, and Tova Milo. Dynamic XML documents with distribution and replication. In *SIGMOD*, June 2003.
6. Serge Abiteboul and Oliver Duschka. Complexity of answering queries using materialized views. In *PODS*, Seattle, WA, 1998.
7. Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995.
8. Azza Abouzeid, Kamil Bajda-Pawlikowski, Daniel J. Abadi, Alexander Rasin, and Avi Silberschatz. HadoopDB: An architectural hybrid of MapReduce and DBMS technologies for analytical workloads. *PVLDB*, 2(1), 2009.
9. B. Adelberg. Nodose – A tool for semi-automatically extracting semi-structured data from text documents. In *SIGMOD*, 1998.
10. P. Adjiman, Philippe Chatalic, François Goasdoué, Marie-Christine Rousset, and Laurent Simon. Distributed reasoning in a peer-to-peer setting. In *ECAI*, pages 945–946, 2004.
11. Philippe Adjiman, François Goasdoué, and Marie-Christine Rousset. SomerDFsin the semantic web. *J. Data Semantics*, 8:158–181, 2007.
12. Foto Afrati, Chen Li, and Prasenjit Mitra. Rewriting queries using views in the presence of arithmetic comparisons. *Theoretical Computer Science*, 368(1-2):88–123, 2006.
13. Foto Afrati, Chen Li, and Jeffrey Ullman. Generating efficient plans for queries using views. In *Proceedings of the ACM SIGMOD Conference*, pages 319–330, 2001.
14. Foto N. Afrati and Rada Chirkova. Selecting and using views to compute aggregate queries. *J. Comput. Syst. Sci.*, 77(6), 2011.
15. Foto N. Afrati and Phokion G. Kolaitis. Answering aggregate queries in data exchange. In *PODS*, 2008.
16. Foto N. Afrati, Chen Li, and Prasenjit Mitra. On containment of conjunctive queries with arithmetic comparisons. In *EDBT*, pages 459–476, 2004.
17. Charu Aggrawal, editor. *Managing and Mining Uncertain Data*. Kluwer Academic Publishers, 2009.
18. Sanjay Agrawal, Surajit Chaudhuri, and Gautam Das. DBXplorer: A system for keyword-based search over relational databases. In *ICDE*, 2002.
19. Rafi Ahmed, Phillippe De Smedt, Weimin Du, William Kent, Mohammad A. Ketabchi, Witold A. Litwin, Abbas Rafii, and Ming-Chien Shan. The Pegasus heterogeneous multidatabase system. *IEEE Computer*, pages 19–26, December 1991.

20. Shurug Al-Khalifa, H. V. Jagadish, Nick Koudas, Divesh Srivastava, and Yuqing Wu. Structural joins: A primitive for efficient XML query pattern matching. In *ICDE*, 2002.

21. Bogdan Alexe, Laura Chiticariu, Renée J. Miller, and Wang Chiew Tan. Muse: Mapping understanding and design by example. In *ICDE*, pages 10–19, 2008.

22. Bogdan Alexe, Mauricio Hernández, Lucian Popa, and Wang-Chiew Tan. Mapmerge: Correlating independent schema mappings. *Proc. VLDB Endow.*, 3, September 2010.

23. Bogdan Alexe, Phokion G. Kolaitis, and Wang Chiew Tan. Characterizing schema mappings via data examples. In *PODS*, pages 261–272, 2010.

24. Bogdan Alexe, Balder ten Cate, Phokion G. Kolaitis, and Wang Chiew Tan. Designing and refining schema mappings via data examples. In *SIGMOD Conference*, pages 133–144, 2011.

25. Alsayed Algergawy, Sabine Massmann, and Erhard Rahm. A clustering-based approach for large-scale ontology matching. In *ADBIS*, pages 415–428, 2011.

26. Mehmet Altinel and Michael J. Franklin. Efficient filtering of XML documents for selective dissemination of information. In *VLDB*, 2000.

27. Yael Amsterdamer, Susan B. Davidson, Daniel Deutch, Tova Milo, Julia Stoyanovich, and Val Tannen. Putting lipstick on pig: Enabling database-style workflow provenance. *PVLDB*, 5(4), 2011.

28. Yael Amsterdamer, Daniel Deutch, Tova Milo, and Val Tannen. On provenance minimization. In *PODS*, 2011.

29. Yael Amsterdamer, Daniel Deutch, and Val Tannen. Provenance for aggregate queries. In *PODS*, 2011.

30. Yuan An, Alexander Borgida, Renée J. Miller, and John Mylopoulos. A semantic approach to discovering schema mapping expressions. In *ICDE*, pages 206–215, 2007.

31. Rohit Ananthakrishna, Surajit Chaudhuri, and Venkatesh Ganti. Eliminating fuzzy duplicates in data warehouses. In *VLDB*, 2002.

32. Lyublena Antova, Christoph Koch, and Dan Olteanu. 10^{10^6} worlds and beyond: Efficient representation and processing of incomplete information. In *ICDE*, 2007.

33. A. Arasu and H. Garcia-Molina. Extracting structured data from web pages. In *SIGMOD*, 2003.

34. Arvind Arasu, Venkatesh Ganti, and Raghav Kaushik. Efficient exact set-similarity joins. In *VLDB*, pages 918–929, 2006.

35. Arvind Arasu, Michaela Götz, and Raghav Kaushik. On active learning of record matching packages. In *SIGMOD Conference*, pages 783–794, 2010.

36. Marcelo Arenas, Pablo Barceló, Leonid Libkin, and Filip Murlak. *Relational and XML Data Exchange*. Synthesis Lectures on Data Management. 2011.

37. Yigal Arens, Chin Y. Chee, Chun-Nan Hsu, and Craig A. Knoblock. Retrieving and integrating data from multiple information sources. *International Journal on Intelligent and Cooperative Information Systems*, 1994.

38. Yigal Arens, Craig A. Knoblock, and Wei-Min Shen. Query reformulation for dynamic information integration. *International Journal on Intelligent and Cooperative Information Systems*, (6)2/3, June 1996.

39. Gustavo Arocena and Alberto Mendelzon. WebOQL: Restructuring documents, databases and webs. In *Proceedings of the International Conference on Data Engineering (ICDE)*, Orlando, Florida, 1998.

40. P. Atzeni and R. Torlone. Management of multiple models in an extensible database design tool. In *Proc. EDBT*, pages 79–95, 1996.

41. Paolo Atzeni, Giansalvatore Mecca, and Paolo Merialdo. To weave the web. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, 1997.

42. Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. Dbpedia: A nucleus for a web of open data. In *ISWC/ASWC*, 2007.

43. David Aumueller, Hong Hai Do, Sabine Massmann, and Erhard Rahm. Schema and ontology matching with COMA++. In *SIGMOD Conference*, pages 906–908, 2005.

44. Ron Avnur and Joseph M. Hellerstein. Eddies: Continuously adaptive query processing. In *SIGMOD*, 2000.

45. Brian Babcock, Shivnath Babu, Mayur Datar, and Rajeev Motwani. Chain: Operator scheduling for memory minimization in data stream systems. In *SIGMOD*, 2003.

46. Brian Babcock and Surajit Chaudhuri. Towards a robust query optimizer: A principled and practical approach. In *SIGMOD*, New York, NY, USA, 2005.

47. Brian Babcock, Mayur Datar, and Rajeev Motwani. Load shedding for aggregation queries over data streams. In *ICDE*, 2004.

48. Shivnath Babu and Pedro Bizarro. Adaptive query processing in the looking glass. In *CIDR*, 2005.

49. Shivnath Babu, Rajeev Motwani, Kamesh Munagala, Itaru Nishizawa, and Jennifer Widom. Adaptive ordering of pipelined stream filters. In *SIGMOD*, 2004.

50. Shivnath Babu, Utkarsh Srivastava, and Jennifer Widom. Exploiting k-constraints to reduce memory overhead in continuous queries over streams. Technical Report, Stanford University, 2002.

51. Akanksha Baid, Ian Rae, AnHai Doan, and Jeffrey F. Naughton. Toward industrial-strength keyword search systems over relational data. In *ICDE*, 2010.

52. Akanksha Baid, Ian Rae, Jiexing Li, AnHai Doan, and Jeffrey F. Naughton. Toward scalable keyword search over relational data. *PVLDB*, 3(1):140–149, 2010.

53. François Bancilhon and Nicolas Spyratos. Update semantics of relational views. *TODS*, 6(4), 1981.

54. Michele Banko, Michael J. Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. Open information extraction from the web. In Manuela M. Veloso, editor, *IJCAI*, pages 2670–2676, 2007.

55. Luciano Barbosa and Juliana Freire. Siphoning hidden-web data through keyword-based interfaces. In *SBBD*, 2004.

56. Robert Baumgartner, Sergio Flesca, and Georg Gottlob. Declarative information extraction, Web crawling, and recursive wrapping with Lixto. In *Proc. of the 6th Int Conf. on Logic Programming and Nonmonotonic Reasoning*, 2001.

57. Robert Baumgartner, Sergio Flesca, and Georg Gottlob. Visual Web information extraction with Lixto. In *VLDB*, 2001.

58. Louis Bavoil, Steven P. Callahan, Patricia J. Crossno, Juliana Freire, Carlos E. Scheidegger, Claudio T. Silva, and Huy T. Vo. VisTrails: Enabling interactive multiple-view visualizations. *IEEE Visualization*, 2005.

59. Roberto J. Bayardo, Yiming Ma, and Ramakrishnan Srikant. Scaling up all pairs similarity search. In *WWW*, pages 131–140, 2007.

60. Catriel Beeri, Alon Y. Levy, and Marie-Christine Rousset. Rewriting queries using views in description logics. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 99–108, Tucson, Arizona., 1997.

61. Z. Bellahsene, A. Bonifati, and E. Rahm, editors. *Schema Matching and Mapping*. Springer, 2011.

62. Zohra Bellahsene and Fabien Duchateau. Tuning for schema matching. In *Schema Matching and Mapping*, pages 293–316. 2011.

63. Randall Bello, Karl Dias, Alan Downing, James Feenan, Jim Finnerty, William Norcott, Harry Sun, Andrew Witkowski, and Mohamed Ziauddin. Materialized views in Oracle. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, pages 659–664, 1998.

64. Michael Benedikt and Georg Gottlob. The impact of virtual views on containment. *PVLDB*, 3(1):297–308, 2010.

65. Omar Benjelloun, Hector Garcia-Molina, Heng Gong, Hideki Kawai, Tait Eliott Larson, David Menestrina, and Sutthipong Thavisomboon. D-swoosh: A family of algorithms for generic, distributed entity resolution. In *ICDCS*, page 37, 2007.

66. Omar Benjelloun, Hector Garcia-Molina, David Menestrina, Qi Su, Steven Euijong Whang, and Jennifer Widom. Swoosh: A generic approach to entity resolution. *VLDB J.*, 18(1):255–276, 2009.

67. Omar Benjelloun, Anish Das Sarma, Alon Y. Halevy, and Jennifer Widom. ULDBs: Databases with uncertainty and lineage. In *VLDB*, 2006.

68. Sonia Bergamaschi, Silvana Castano, and Maurizio Vincini. Semantic integration of semistructured and structured data sources. *SIGMOD Record*, 28(1):54–59, 1999.

69. Sonia Bergamaschi, Elton Domnori, Francesco Guerra, Raquel Trillo Lado, and Yannis Velegrakis. Keyword search over relational databases: A metadata approach. In *Proceedings of the 2011 International Conference on Management of Data*, SIGMOD '11, New York, NY, USA, 2011. Available from http://doi.acm.org/10.1145/1989323.1989383.

70. Michael K. Bergman. The deep web: Surfacing hidden value. *Journal of Electronic Publishing*, 2001.

71. Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, May 2001.

72. Philip A. Bernstein. Applying model management to classical meta-data problems. In *Proceedings of the Conference on Innovative Data Systems Research (CIDR)*, 2003.

73. Philip A. Bernstein and Dah-Ming W. Chiu. Using semi-joins to solve relational queries. *J. ACM*, 28, 1981.

74. Philip A. Bernstein, Fausto Giunchiglia, Anastasios Kementsietsidis, John Mylopoulos, Luciano Serafini, and Ilya Zaihrayeu. Data management for peer-to-peer computing: A vision. In *Proceedings of the WebDB Workshop*, 2002.

75. Philip A. Bernstein, Todd J. Green, Sergey Melnik, and Alan Nash. Implementing mapping composition. In *Proc. of VLDB*, pages 55–66, 2006.

76. Philip A. Bernstein, Alon Y. Halevy, and Rachel Pottinger. A vision for management of complex models. *SIGMOD Record*, 29(4):55–63, 2000.

77. Philip A. Bernstein and Sergey Melnik. Model management 2.0: Manipulating richer mappings. In *Proc. of SIGMOD*, pages 1–12, 2007.

78. Philip A. Bernstein, Sergey Melnik, and John E. Churchill. Incremental schema matching. In *VLDB*, pages 1167–1170, 2006.

79. Laure Berti-Equille, Anish Das Sarma, Xin Dong, Amélie Marian, and Divesh Srivastava. Sailing the information ocean with awareness of currents: Discovery and application of source dependence. In *CIDR*, 2009.

80. Gaurav Bhalotia, Arvind Hulgeri, Charuta Nakhe, Soumen Chakrabarti, and S. Sudarshan. Keyword searching and browsing in databases using BANKS. In *ICDE*, 2002.

81. I. Bhattacharya and L. Getoor. A latent Dirichlet model for unsupervised entity resolution. In *Proc. of the SIAM Int. Conf. on Data Mining (SDM)*, 2006.

82. I. Bhattacharya and L. Getoor. Collective entity resolution in relational data. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

83. Indrajit Bhattacharya, Lise Getoor, and Louis Licamele. Query-time entity resolution. In *KDD*, pages 529–534, 2006.

84. M. Bilenko. Learnable similarity functions and their applications to clustering and record linkage. In *AAAI*, pages 981–982, 2004.

85. M. Bilenko and R. J. Mooney. Adaptive duplicate detection using learnable string similarity measures. In *Proc. of the ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 39–48, 2003.

86. M. Bilenko, R. J. Mooney, W. W. Cohen, P. D. Ravikumar, and S. E. Fienberg. Adaptive name matching in information integration. *IEEE Intelligent Systems*, 18(5):16–23, 2003.

87. Olivier Biton, Sarah Cohen Boulakia, Susan B. Davidson, and Carmem S. Hara. Querying and managing provenance through user views in scientific workflows. In *ICDE*, 2008.

88. Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data – The story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.

89. José A. Blakeley, Neil Coburn, and Per-Åke Larson. Updating derived relations: Detecting irrelevant and autonomously computable updates. *TODS*, 14(3), 1989.

90. Burton H. Bloom. Space/time trade-offs in hash coding with allowable errors. *CACM*, 13(7), July 1970.

91. Lukas Blunschi, Jens-Peter Dittrich, Olivier Girard, Shant Krakos Karakashian, and Marcos Antonio Vas Salles. The iMemex personal dataspace management system (demo). In *CIDR*, 2007.

92. Philip Bohannon, Eiman Elnahrawy, Wenfei Fan, and Michael Flaster. Putting context into schema matching. In *VLDB*, pages 307–318, 2006.

93. Alex Borgida. Description logics in data management. *IEEE Trans. on Know. and Data Engineering*, 7(5):671–682, 1995.

94. V. R. Borkar, K. Deshmukh, and S. Sarawagi. Automatic segmentation of text into structured records. In *SIGMOD*, 2001.

95. J. Boulos, N. Dalvi, B. Mandhani, S. Mathur, C. Re, and D. Suciu. MYSTIQ: A system for finding more answers by using probabilities. In *Proc. of ACM SIGMOD*, 2005.

96. Michael Brundage. *XQuery: The XML Query Language*. February 2004.

97. Nicolas Bruno, Nick Koudas, and Divesh Srivastava. Holistic twig joins: Optimal xml pattern matching. In *SIGMOD Conference*, 2002.

98. Peter Buneman, Sanjeev Khanna, and Wang Chiew Tan. Why and where: A characterization of data provenance. In *ICDT*, 2001.

99. Peter Buneman, Anthony Kosky, and Susan Davidson. Theoretical aspects of schema merging. In *Proc. of EDBT*, pages 152–167, 1992.

100. Douglas Burdick, Mauricio A. Hernández, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana Stanoi, Shivakumar Vaithyanathan, and Sanjiv R. Das. Extracting, linking and integrating data from public sources: A financial case study. *IEEE Data Eng. Bull.*, 34(3):60–67, 2011.

101. Michael J. Cafarella, Alon Halevy, Yang Zhang, Daisy Zhe Wang, and Eugene Wu. Uncovering the Relational Web. In *WebDB*, 2008.

102. Michael J. Cafarella, Alon Halevy, Yang Zhang, Daisy Zhe Wang, and Eugene Wu. WebTables: Exploring the power of tables on the web. In *VLDB*, 2008.

103. D. Cai, S. Yu, J. Wen, and W. Ma. Extracting content structure for Web pages based on visual representation. In *Proc. of the 5th Asian-Pacific Web Conference (APWeb)*, 2003.

104. Andrea Cali, Diego Calvanese, Giuseppe DeGiacomo, and Maurizio Lenzerini. Data integration under integrity constraints. In *Proceedings of CAiSE*, pages 262–279, 2002.

105. M. E. Califf and R. J. Mooney. Relational learning of pattern-match rules for information extraction. In *AAAI*, 1999.

106. James P. Callan and Margaret E. Connell. Query-based sampling of text databases. *ACM Transactions on Information Systems*, 19(2):97–130, 2001.

107. D. Calvanese, G. De Giacomo, and M. Lenzerini. Answering queries using views over description logics. In *Proceedings of AAAI*, pages 386–391, 2000.

108. D. Calvanese, G. De Giacomo, M. Lenzerini, and M. Vardi. View-based query processing for regular path queries with inverse. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 58–66, 2000.

109. Michael J. Carey and Donald Kossmann. On saying "enough already!" in SQL. In *SIGMOD*, 1997.

110. Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. Toward an architecture for never-ending language learning. In Maria Fox and David Poole, editors, *AAAI*. AAAI Press, 2010.

111. S. Castano and V. De Antonellis. A discovery-based approach to database ontology design. *Distributed and Parallel Databases – Special Issue on Ontologies and Databases*, 7(1), 1999.

112. T. Catarci and M. Lenzerini. Representing and using interschema knowledge in cooperative information systems. *Journal of Intelligent and Cooperative Information Systems*, pages 55–62, 1993.

113. Xiaoyong Chai, Ba-Quy Vuong, AnHai Doan, and Jeffrey F. Naughton. Efficiently incorporating user feedback into information extraction and integration programs. In *SIGMOD*, New York, NY, USA, 2009.

114. A.K. Chandra and P.M. Merlin. Optimal implementation of conjunctive queries in relational databases. In *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing*, pages 77–90, 1977.

115. C. Chang, M. Kayed, M. R. Girgis, and K. F. Shaalan. A survey of web information extraction systems. *IEEE Trans. Knowl. Data Eng.*, 18(10):1411–1428, 2006.

116. C. Chang and S. Lui. IEPAD: Information extraction based on pattern discovery. In *WWW*, 2001.

117. Adriane Chapman and H. V. Jagadish. Why not? In *SIGMOD Conference*, 2009.

118. Sam Chapman. Sam's string metrics. 2006. Available at http:staffwww.dcs.shef.ac.uk/people/sam.chapman@k-now.co.uk/stringmetrics.html.

119. A. Chatterjee and A. Segev. Data manipulation in heterogeneous databases. *SIGMOD Record*, 20(4):64–68, 1991.

120. S. Chaudhuri, K. Ganjam, V. Ganti, and R. Motwani. Robust and efficient fuzzy match for online data cleaning. In *SIGMOD*, 2003.

121. Surajit Chaudhuri. An overview of query optimization in relational systems. In *PODS*, 1998.

122. Surajit Chaudhuri and Umeshwar Dayal. An overview of data warehousing and olap technology. *SIGMOD Record*, 26(1), 1997.

123. Surajit Chaudhuri, Umeshwar Dayal, and Vivek R. Narasayya. An overview of business intelligence technology. *Commun. ACM*, 54(8):88–98, 2011.

124. Surajit Chaudhuri, Venkatesh Ganti, and Raghav Kaushik. A primitive operator for similarity joins in data cleaning. In *ICDE*, page 5, 2006.

125. Surajit Chaudhuri, Ravi Krishnamurthy, Spyros Potamianos, and Kyuseok Shim. Optimizing queries with materialized views. In *Proceedings of the International Conference on Data Engineering (ICDE)*, pages 190–200, Taipei, Taiwan, 1995.

126. Surajit Chaudhuri, Anish Das Sarma, Venkatesh Ganti, and Raghav Kaushik. Leveraging aggregate constraints for deduplication. In *SIGMOD Conference*, pages 437–448, 2007.

127. Surajit Chaudhuri and Moshe Vardi. Optimizing real conjunctive queries. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 59–70, Washington D.C., 1993.

128. Sudarshan Chawathe, Hector Garcia-Molina, Joachim Hammer, Kelly Ireland, Yannis Papakonstantinou, Jeffrey Ullman, and Jennifer Widom. The TSIMMIS project: Integration of heterogeneous information sources. In *Proceedings of IPSJ*, Tokyo, Japan, October 1994.

129. Sudarshan S. Chawathe and Hector Garcia-Molina. Meaningful change detection in structured data. In *SIGMOD*, pages 26–37, 1997.

130. Chandra Chekuri and Anand Rajaraman. Conjunctive query containment revisited. *Theor. Comput. Sci.*, 239(2):211–229, 2000.

131. Yi Chen, Susan B. Davidson, and Yifeng Zheng. Vitex: A streaming xpath processing system. In *ICDE*, 2005.

132. James Cheney, Umut A. Acar, and Amal Ahmed. Provenance traces. *CoRR*, abs/0812.0564, 2008.

133. James Cheney, Laura Chiticariu, and Wang Chiew Tan. Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*, 1(4), 2009.

134. Rada Chirkova, Alon Y. Halevy, and Dan Suciu. A formal perspective on the view selection problem. *VLDB J.*, 11(3), 2002.

135. L. Chiticariu, P. G. Kolaitis, and L. Popa. Interactive generation of integrated schemas. In *Proc. of SIGMOD*, 2008.

136. Laura Chiticariu, Wang Chiew Tan, and Gaurav Vijayvargiya. Dbnotes: A post-it system for relational databases based on provenance. In *SIGMOD*, 2005.

137. Eric Chu, Akanksha Baid, Xiaoyong Chai, AnHai Doan, and Jeffrey F. Naughton. Combining keyword search and forms for ad hoc querying of databases. In *SIGMOD Conference*, 2009.

138. Francis C. Chu, Joseph Y. Halpern, and Johannes Gehrke. Least expected cost query optimization: What can we expect? In *PODS*, 2002.

139. S. Chuang, K. C. Chang, and C. Zhai. Collaborative wrapping: A turbo framework for Web data extraction. In *ICDE*, 2007.

140. Chris Clifton, E. Housman, and Arnon Rosenthal. Experience with a combined approach to attribute-matching across heterogeneous databases. In *DS-7*, 1997.

141. M. Cochinwala, V. Kurien, G. Lalk, and D. Shasha. Efficient data reconciliation. *Inf. Sci.*, 137(1-4):1–15, 2001.

142. Sara Cohen. Containment of aggregate queries. *SIGMOD Record*, 34(1):77–85, 2005.

143. W. Cohen. A mini-course on record linkage and matching, 2004. http://www.cs.cmu.edu/~wcohen.

144. W. Cohen and J. Richman. Learning to match and cluster large high-dimensional data sets for data integration. In *Proc. of the ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 475–480, 2002.

145. W. W. Cohen. Data integration using similarity joins and a word-based information representation language. *ACM Trans. Inf. Syst.*, 18(3):288–321, 2000.

146. W. W. Cohen. Record linkage tutorial: Distance metrics for text. 2001. PPT slides, available at www.cs.cmu.edu/~wcohen/Matching-2.ppt.

147. W. W. Cohen, M. Hurst, and L. S. Jensen. A flexible learning system for wrapping tables and lists in HTML documents. In *WWW*, 2002.

148. W. W. Cohen, P. D. Ravikumar, and S. E. Fienberg. A comparison of string distance metrics for name-matching tasks. In *IIWeb*, 2003.

149. Latha S. Colby, Timothy Griffin, Leonid Libkin, Inderpal Singh Mumick, and Howard Trickey. Algorithms for deferred view maintenance. In *SIGMOD*, 1996.

150. Richard L. Cole and Goetz Graefe. Optimization of dynamic query evaluation plans. In *SIGMOD*, 1994.

151. Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, and Sean Slattery. Learning to extract symbolic knowledge from the world-wide web. In *Proceedings of the AAAI Fifteenth National Conference on Artificial Intelligence*, 1998.

152. V. Crescenzi and G. Mecca. Grammars have exceptions. *Inf. Syst.*, 23(8):539–565, 1998.

153. V. Crescenzi and G. Mecca. Automatic information extraction from large websites. *J. ACM*, 51(5):731–779, 2004.

154. V. Crescenzi, G. Mecca, and P. Merialdo. Roadrunner: Towards automatic data extraction from large web sites. In *VLDB*, 2001.

155. Yingwei Cui. *Lineage Tracing in Data Warehouses*. PhD thesis, Stanford Univ., 2001.

156. A. Culotta and A. McCallum. Joint deduplication of multiple record types in relational data. In *Proc. of the ACM Int. Conf. on Information and Knowledge Management (CIKM)*, pages 257–258, 2005.

157. Carlo Curino, Hyun Jin Moon, Alin Deutsch, and Carlo Zaniolo. Update rewriting and integrity constraint maintenance in a schema evolution support system: Prism++. *PVLDB*, 4(2), 2010.

158. N. Dalvi, R. Kumar, and M. A. Soliman. Automatic wrappers for large scale web extraction. *PVLDB*, 4(4):219–230, 2011.

159. N. N. Dalvi, P. Bohannon, and F. Sha. Robust Web extraction: An approach based on a probabilistic tree-edit model. In *SIGMOD*, 2009.

160. Nilesh Dalvi and Dan Suciu. Efficient query evaluation on probabilistic databases. In *VLDB*, 2004.

161. Dean Daniels. Query compilation in a distributed database system. Technical Report RJ 3423, IBM, 1982.

162. A. Das Sarma, L. Dong, and A. Halevy. Bootstrapping pay-as-you-go data integration systems. In *Proc. of SIGMOD*, 2008.

163. Arjun Dasgupta, Xin Jin, Bradley Jewell, Nan Zhang, and Gautam Das. Unbiased estimation of size and other aggregates over hidden web databases. In *SIGMOD Conference*, pages 855–866, 2010.

164. Susan B. Davidson, Sanjeev Khanna, Tova Milo, Debmalya Panigrahi, and Sudeepa Roy. Provenance views for module privacy. In *PODS*, 2011.

165. U. Dayal. Processing queries over generalized hierarchies in a multidatabase systems. In *Proc. of the VLDB Conf.*, pages 342–353, 1983.

166. Umeshwar Dayal and Philip A. Bernstein. On the correct translation of update operations on relational views. *TODS*, 7(3), 1982.

167. Filipe de S Mesquita, Altigran Soares da Silva, Edleno Silva de Moura, Pvel Calado, and Alberto H. F. Laender. Labrador: Efficiently publishing relational databases on the web by using keyword-based query interfaces. *Inf. Process. Manage.*, 43(4):983–1004, 2007.

168. L. G. DeMichiel. Resolving database incompatibility: An approach to performing relational operations over mismatched domains. *IEEE Transactions on Knowledge and Data Engineering*, 1989.

169. Pedro DeRose, Xiaoyong Chai, Byron J. Gao, Warren Shen, AnHai Doan, Philip Bohannon, and Xiaojin Zhu. Building community wikipedias: A machine-human partnership approach. In *ICDE*, pages 646–655, 2008.

170. Pedro DeRose, Warren Shen, Fei Chen, AnHai Doan, and Raghu Ramakrishnan. Building structured web community portals: A top-down, compositional, and incremental approach. In *Proc. of VLDB*, 2007.

171. Amol Deshpande and Joseph M. Hellerstein. Lifting the burden of history from adaptive query processing. In *VLDB*, 2004.

172. Amol Deshpande, Zachary Ives, and Vijayshankar Raman. Adaptive query processing. *Foundations and Trends in Databases*, 2007.

173. Alin Deutsch, Mary F. Fernández, Daniela Florescu, Alon Y. Levy, and Dan Suciu. XML-QL. In *QL*, 1998.

174. Alin Deutsch and Val Tannen. MARS: A system for publishing XML from mixed and redundant storage. In *VLDB*, 2003.

175. D. Dey. Entity matching in heterogeneous databases: A logistic regression approach. *Decision Support Systems*, 44(3):740–747, 2008.

176. D. Dey, S. Sarkar, and P. De. A distance-based approach to entity reconciliation in heterogeneous databases. *IEEE Trans. Knowl. Data Eng.*, 14(3):567–582, 2002.

177. Yanlei Diao, Peter M. Fischer, Michael J. Franklin, and Raymond To. YFilter: Efficient and scalable filtering of XML documents. In *ICDE*, 2002.

178. Hong Hai Do and Erhard Rahm. COMA – A system for flexible combination of schema matching approaches. In *VLDB*, 2002.

179. A. Doan and A. Y. Halevy. Semantic integration research in the database community: A brief survey. *AI Magazine*, 26(1):83–94, 2005.

180. A. Doan, N. F. Noy, and A. Y. Halevy. Introduction to the special issue on semantic integration. *SIGMOD Record*, 33(4):11–13, 2004.

181. AnHai Doan, Pedro Domingos, and Alon Y. Halevy. Reconciling schemas of disparate data sources: A machine learning approach. In *Proceedings of the ACM SIGMOD Conference*, 2001.

182. AnHai Doan, Ying Lu, Yoonkyong Lee, and Jiawei Han. Profile-based object matching for information integration. *IEEE Intelligent Systems*, 18(5):54–59, 2003.

183. Anhai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy. Learning to map between ontologies on the semantic web. In *11th World Wide Web Conference*, 2002.

184. AnHai Doan, Jeffrey F. Naughton, Raghu Ramakrishnan, Akanksha Baid, Xiaoyong Chai, Fei Chen, Ting Chen, Eric Chu, Pedro DeRose, Byron Gao, Chaitanya Gokhale, Jiansheng Huang, Warren Shen, and Ba-Quy Vuong. Information extraction challenges in managing unstructured data. *SIGMOD Record*, December 2008.

185. AnHai Doan, Raghu Ramakrishnan, Fei Chen, Pedro DeRose, Yoonkyong Lee, Robert McCann, Mayssam Sayyadian, and Warren Shen. Community information management. *IEEE Data Eng. Bull.*, 29(1):64–72, 2006.

186. AnHai Doan, Raghu Ramakrishnan, and Alon Y. Halevy. Crowdsourcing systems on the world-wide web. *Commun. ACM*, 54(4):86–96, 2011.

187. Pedro Domingos and Micheal Pazzani. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning*, 29:103–130, 1997.

188. X. Dong, A. Y. Halevy, and J. Madhavan. Reference reconciliation in complex information spaces. In *Proc. of the SIGMOD Conf.*, pages 85–96, 2005.

189. X. Dong, A. Y. Halevy, and C. Yu. Data integration with uncertainty. In *Proc. of VLDB*, 2007.

190. Xin Dong, Laure Berti-Equille, Yifan Hu, and Divesh Srivastava. Global detection of complex copying relationships between sources. *PVLDB*, 3(1):1358–1369, 2010.

191. Xin Dong and Alon Halevy. A platform for personal information management and integration. In *Proc. of CIDR*, 2005.

192. Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Werner Nutt. The complexity of concept languages. In *Proceedings of KR-91*, 1991.

193. Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Andrea Schaerf. A hybrid system with datalog and concept languages. In E. Ardizzone, S. Gaglio, and F. Sorbello, editors, *Trends in Artificial Intelligence*, volume LNAI 549, pages 88–97. Springer Verlag, 1991.

194. Mira Dontcheva, Steven M. Drucker, David Salesin, and Michael F. Cohen. Relations, cards, and search templates: User-guided web data integration and layout. In *UIST*, pages 61–70, 2007.

195. R. B. Doorenbos, O. Etzioni, and D. S. Weld. A scalable comparison-shopping agent for the World-Wide Web. In *Agents*, 1997.

196. R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1999.

197. Oliver Duschka, Michael Genesereth, and Alon Levy. Recursive query plans for data integration. *Journal of Logic Programming, special issue on Logic Based Heterogeneous Information Systems*, 43(1), 2000.

198. Oliver M. Duschka and Michael R. Genesereth. Answering recursive queries using views. In *PODS*, 1997.

199. Oliver M. Duschka and Michael R. Genesereth. Query planning in infomaster. In *Proceedings of the ACM Symposium on Applied Computing*, pages 109–111, San Jose, CA, 1997.

200. Oliver M. Duschka and Alon Y. Levy. Recursive plans for information gathering. In *Proc. of the 15th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 778–784, 1997.

201. Marc Ehrig, Steffen Staab, and York Sure. Bootstrapping ontology alignment methods with APFEL. In *International Semantic Web Conference*, pages 186–200, 2005.

202. Charles Elkan. A decision procedure for conjunctive query disjointness. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, Portland, Oregon, 1989.

203. Charles Elkan. Independence of logic database queries and updates. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 154–160, 1990.

204. A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios. Duplicate record detection: A survey. *IEEE Trans. Knowl. Data Eng.*, 19(1):1–16, 2007.

205. A.K. Elmagarmid, P.G. Ipeirotis, and V.S. Verykios. Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):1–16, 2007.

206. Hazem Elmeleegy, Ahmed K. Elmagarmid, and Jaewoo Lee. Leveraging query logs for schema mapping generation in u-map. In *SIGMOD Conference*, pages 121–132, 2011.

207. Hazem Elmeleegy, Jayant Madhavan, and Alon Y. Halevy. Harvesting relational tables from lists on the web. *PVLDB*, 2(1):1078–1089, 2009.

208. Hazem Elmeleegy, Mourad Ouzzani, and Ahmed K. Elmagarmid. Usage-based schema matching. In *ICDE*, pages 20–29, 2008.

209. Mohamed Y. Eltabakh, Walid G. Aref, Ahmed K. Elmagarmid, Mourad Ouzzani, and Yasin N. Silva. Supporting annotations on relations. In *EDBT*, 2009.

210. D. W. Embley, D. M. Campbell, Y. S. Jiang, S. W. Liddle, Y. Ng, D. Quass, and R. D. Smith. Conceptual-model-based data extraction from multiple-record web pages. *Data Knowl. Eng.*, 31(3):227–251, 1999.

211. D. W. Embley, Y. S. Jiang, and Y. Ng. Record-boundary discovery in web documents. In *SIGMOD*, 1999.

212. David W. Embley, David Jackman, and Li Xu. Multifaceted exploitation of metadata for attribute match discovery in information integration. In *Workshop on Information Integration on the Web*, pages 110–117, 2001.

213. O. Etzioni, K. Golden, and D. Weld. Sound and efficient closed-world reasoning for planning. *Artificial Intelligence*, 89(1–2):113–148, January 1997.

214. Jérôme Euzenat and Pavel Shvaiko. *Ontology Matching*. Springer, 2007.

215. Ronald Fagin. Inverting schema mappings. In *Proc. of PODS*, pages 50–59, 2006.

216. Ronald Fagin, Phokion Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. *TCS*, 336:89–124, 2005.

217. Ronald Fagin, Phokion Kolaitis, and Lucian Popa. Data exchange: Getting to the core. *ACM Transactions on Database Systems*, 30(1):174–210, 2005.

218. Ronald Fagin, Phokion G. Kolaitis, and Lucian Popa. Composing schema mappings: Second-order dependencies to the rescue. *ACM Transactions on Database Systems*, 30(4):994–1055, 2005.

219. Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang-Chiew Tan. Quasi-inverses of schema mappings. In *Proc. of PODS*, pages 123–132, 2007.

220. Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang Chiew Tan. Schema mapping evolution through composition and inversion. In *Schema Matching and Mapping*, pages 191–222. 2011.

221. Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. *Journal of Computer and System Sciences*, 66(4), June 2003.

222. Sean M. Falconer and Margaret-Anne D. Storey. A cognitive support framework for ontology mapping. In *ISWC/ASWC*, pages 114–127, 2007.

223. Wenfei Fan and Floris Geerts. Capturing missing tuples and missing values. In *PODS*, pages 169–178, 2010.

224. I. P. Fellegi and A. B. Sunter. A theory for record linkage. *Journal of the American Statistical Society*, 64(328):1183–1210, 1969.

225. Mary Fernandez, Daniela Florescu, Jaewoo Kang, Alon Levy, and Dan Suciu. Catching the boat with Strudel: Experiences with a web-site management system. In *Proceedings of the ACM SIGMOD Conference*, Seattle, WA, 1998.

226. Jonathan Finger and Neoklis Polyzotis. Robust and efficient algorithms for rank join evaluation. In *SIGMOD*, New York, NY, USA, 2009.

227. Daniela Florescu, Daphne Koller, and Alon Levy. Using probabilistic information in data integration. In *VLDB*, 1997.

228. Daniela Florescu, Alon Levy, Ioana Manolesu, and Dan Suciu. Query optimization in the presence of limited access patterns. In *Proceedings of the ACM SIGMOD Conference*, 1999.

229. Daniela Florescu, Alon Levy, and Alberto Mendelzon. Database techniques for the world-wide web: A survey. *SIGMOD Record*, 27(3):59–74, September 1998.

230. Daniela Florescu, Louiqa Raschid, and Patrick Valduriez. Using heterogeneous equivalences for query rewriting in multidatabase systems. In *Proceedings of the Int. Conf. on Cooperative Information Systems (COOPIS)*, 1995.

231. J. Nathan Foster, Todd J. Green, and Val Tannen. Annotated XML: Queries and provenance. In *PODS*, 2008.

232. Paul Francis, Sugih Jamin, Cheng Jin, Yixin Jin, Danny Raz, Yuval Shavitt, and Lixia Zhang. Idmaps: A global internet host distance estimation service. *IEEE/ACM Trans. Netw.*, 9(5), 2001.

233. Michael Franklin, Alon Halevy, and David Maier. From databases to dataspaces: A new abstraction for information management. *SIGMOD Rec.*, 34(4), 2005.

234. M. Friedman and D. Weld. Efficient execution of information gathering plans. In *Proc. of the 15th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 1997.

235. Marc Friedman, Alon Levy, and Todd Millstein. Navigational Plans for Data Integration. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 1999.

236. N. Fuhr and T. Rölleke. A probabilistic relational algebra for the integration of information retrieval and database systems. *ACM Transactions on Information Systems*, 14(1), 1997.

237. Ariel Fuxman, Mauricio A. Hernández, C. T. Howard Ho, Renée J. Miller, Paolo Papotti, and Lucian Popa. Nested mappings: Schema mapping reloaded. In *VLDB*, 2006.

238. A. Gal. Why is schema matching tough and what can we do about it? *SIGMOD Record*, 35(4):2–5, 2007.

239. A. Gal, G. Modica, H. Jamil, and A. Eyal. Automatic ontology matching using application semantics. *AI Magazine*, 26(1):21–31, 2005.

240. Avigdor Gal. Managing uncertainty in schema matching with top-k schema mappings. *Journal of Data Semantics*, VI:90–114, 2006.

241. Avigdor Gal. *Uncertain Schema Matching*. Synthesis Lectures on Data Management. 2011.

242. Avigdor Gal, Ateret Anaby-Tavor, Alberto Trombetta, and Danilo Montesi. A framework for modeling and evaluating automatic semantic reconciliation. In *VLDB J.*, pages 50–67, 2005.

243. M. Ganesh, J. Srivastava, and T. Richardson. Mining entity-identification rules for database integration. In *Proc. of the ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 291–294, 1996.

244. Hector Garcia-Molina, Yannis Papakonstantinou, Dallan Quass, Anand Rajaraman, Yehoshua Sagiv, Jeffrey Ullman, and Jennifer Widom. The TSIMMIS project: Integration of heterogeneous information sources. *Journal of Intelligent Information Systems*, 8(2), March 1997.

245. Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book*. Prentice Hall, 2002.

246. Wolfgang Gatterbauer, Magdalena Balazinska, Nodira Khoussainova, and Dan Suciu. Believe it or not: Adding belief annotations to databases. *PVLDB*, 2(1), 2009.

247. Wolfgang Gatterbauer, Paul Bohunsky, Marcus Herzog, Bernhard Krüpl, and Bernhard Pollak. Towards domain-independent information extraction from web tables. In *WWW*, pages 71–80, 2007.

248. L. Getoor and R. Miller. Data and metadata alignment, 2007. Tutorial, the Alberto Mendelzon Workshop on the Foundations of Databases and the Web.

249. C. L. Giles, K. D. Bollacker, and S. Lawrence. CiteSeer: An automatic citation indexing system. In *Proc. of the ACM Int. Conf. on Digital Libraries*, pages 89–98, 1998.

250. L. E. Gill. OX-LINK: The Oxford medical record linkage system. In *Proc. of the Int Record Linkage Workshop and Exposition*, 1997.

251. Fausto Giunchiglia, Pavel Shvaiko, and Mikalai Yatskevich. S-match: An algorithm and an implementation of semantic matching. In *ESWS*, pages 61–75, 2004.

252. Boris Glavic and Gustavo Alonso. Perm: Processing provenance and data on the same data model through query rewriting. In *ICDE*, 2009.

253. Boris Glavic and Gustavo Alonso. Provenance for nested subqueries. In *EDBT*, 2009.

254. Boris Glavic, Gustavo Alonso, Renée J. Miller, and Laura M. Haas. Tramp: Understanding the behavior of schema mappings through provenance. *PVLDB*, 3(1), 2010.

255. François Goasdoué and Marie-Christine Rousset. Querying distributed data through distributed ontologies: A simple but scalable approach. *IEEE Intelligent Systems*, 18(5):60–65, 2003.

256. E. M. Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

257. E. M. Gold. Complexity of automaton identification from given data. *Information and Control*, 37(3):302–320, 1978.

258. Roy Goldman, Jason McHugh, and Jennifer Widom. From semistructured data to XML: Migrating the Lore data model and query language. In *WebDB '99*, 1999.

259. Jonathan Goldstein and Per-Ake Larson. Optimizing queries using materialized views: A practical, scalable solution. In *Proceedings of the ACM SIGMOD Conference*, pages 331–342, 2001.

260. Hector Gonzalez, Alon Y. Halevy, Christian S. Jensen, Anno Langen, Jayant Madhavan, Rebecca Shapley, and Warren Shen. Google fusion tables: Data management, integration and collaboration in the cloud. In *SoCC*, 2010.

261. Hector Gonzalez, Alon Y. Halevy, Christian S. Jensen, Anno Langen, Jayant Madhavan, Rebecca Shapley, Warren Shen, and Jonathan Goldberg-Kidon. Google fusion tables: Web-centered data management and collaboration. In *SIGMOD*, 2010.

262. Georg Gottlob, Christoph Koch, Robert Baumgartner, Marcus Herzog, and Sergio Flesca. The Lixto data extraction project — Back and forth between theory and practice. In *PODS*, 2004.

263. Goetz Graefe. Query evaluation techniques for large databases. *ACM Computing Surveys*, 25(2), June 1993.

264. L. Gravano, P. G. Ipeirotis, N. Koudas, and D. Srivastava. Text joins in an RDBMS for web data integration. In *WWW*, 2003.

265. Luis Gravano, Panagiotis G. Ipeirotis, H. V. Jagadish, Nick Koudas, S. Muthukrishnan, and Divesh Srivastava. Approximate string joins in a database (almost) for free. In *VLDB*, pages 491–500, 2001.

266. Luis Gravano, Panagiotis G. Ipeirotis, Nick Koudas, and Divesh Srivastava. Text joins in an RDBMS for web data integration. In *WWW*, 2003.

267. Todd J. Green. Containment of conjunctive queries on annotated relations. In *ICDT*, 2009.

268. Todd J. Green, Grigoris Karvounarakis, Zachary G. Ives, and Val Tannen. Update exchange with mappings and provenance. In *VLDB*, 2007. Amended version available as Univ. of Pennsylvania report MS-CIS-07-26.

269. Todd J. Green, Grigoris Karvounarakis, and Val Tannen. Provenance semirings. In *PODS*, 2007.

270. Todd J. Green, Gerome Miklau, Makoto Onizuka, and Dan Suciu. Processing XML streams with deterministic automata and stream indexes. Available from http://www.cs.washington.edu/homes/suciu/files/paper.ps, February 2002.

271. Todd J. Green and Val Tannen. Models for incomplete and probabilistic information. In *International Workshop on Incompleteness and Inconsistency in databases*, March 2006.

272. Nils Grimsmo, Truls A. Bjørklund, and Magnus Lie Hetland. Fast optimal twig joins. *PVLDBJ*, 3, September 2010.

273. S. Grumbach and G. Mecca. In search of the lost schema. In *ICDT*, 1999.

274. Pankaj Gulhane, Amit Madaan, Rupesh R. Mehta, Jeyashankher Ramamirtham, Rajeev Rastogi, Sandeepkumar Satpal, Srinivasan H. Sengamedu, Ashwin Tengli, and Charu Tiwari. Web-scale information extraction with vertex. In *ICDE*, pages 1209–1220, 2011.

275. Ashish Gupta, Inderpal Singh Mumick, and V. S. Subrahmanian. Maintaining views incrementally. In *SIGMOD*, 1993.

276. Himanshu Gupta. Selection of views to materialize in a data warehouse. In *Database Theory ICDT '97*, volume 1186 of *Lecture Notes in Computer Science*. 1997. Available from http://dx.doi.org/10.1007/3-540-62222-5_39.

277. Himanshu Gupta and Inderpal Singh Mumick. Selection of views to materialize under a maintenance cost constraint. In *ICDT*, 1999.

278. Nitin Gupta, Lucja Kot, Sudip Roy, Gabriel Bender, Johannes Gehrke, and Christoph Koch. Entangled queries: Enabling declarative data-driven coordination. In *SIGMOD Conference*, 2011.

279. Nitin Gupta, Milos Nikolic, Sudip Roy, Gabriel Bender, Lucja Kot, Johannes Gehrke, and Christoph Koch. Entangled transactions. *PVLDB*, 4(11), 2011.

280. Dan Gusfield. *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press, 1999.

281. Laura M. Haas, Donald Kossmann, Edward L. Wimmers, and Jun Yang. Optimizing queries across diverse data sources. In *VLDB*, 1997.

282. Jan Hajic, Sandra Carberry, and Stephen Clark, editors. *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11–16, 2010, Uppsala, Sweden*. The Association for Computer Linguistics, 2010.

283. Alon Halevy, Zachary Ives, Jayant Madhavan, Peter Mork, Dan Suciu, and Igor Tatarinov. The Piazza peer data management system. *TKDE*, 16(7), July 2004.

284. Alon Y. Halevy. Answering queries using views: A survey. *VLDB J.*, 10(4), 2001.

285. Alon Y. Halevy, Naveen Ashish, Dina Bitton, Michael J. Carey, Denise Draper, Jeff Pollock, Arnon Rosenthal, and Vishal Sikka. Enterprise information integration: Successes, challenges and controversies. In *SIGMOD Conference*, pages 778–787, 2005.

286. Alon Y. Halevy, Michael J. Franklin, and David Maier. Principles of dataspace systems. In *PODS*, 2006.

287. Alon Y. Halevy, Zachary G. Ives, Peter Mork, and Igor Tatarinov. Piazza: Data management infrastructure for semantic web applications. In *12th World Wide Web Conference*, May 2003.

288. Alon Y. Halevy, Zachary G. Ives, Dan Suciu, and Igor Tatarinov. Schema mediation in peer data management systems. In *ICDE*, March 2003.

289. Fayçal Hamdi, Brigitte Safar, Chantal Reynaud, and Haïfa Zargayouna. Alignment-based partitioning of large-scale ontologies. In *EGC (best of volume)*, pages 251–269, 2009.

290. J. Hammer, H. Garcia-Molina, S. Nestorov, R. Yerneni, M. M. Breunig, and V. Vassalos. Template-based wrappers in the tsimmis system. In *SIGMOD*, 1997.

291. J. Hammer, J. McHugh, and H. Garcia-Molina. Semistructured data: The tsimmis experience. In *Proc. of the First East-European Symposium on Advances in Databases and Information Systems (ADBIS)*, 1997.

292. Joachim Hammer, Hector Garcia-Molina, Svetlozar Nestorov, Ramana Yerneni, Markus M. Breunig, and Vasilis Vassalos. Template-based wrappers in the TSIMMIS system (system demonstration). In *Proceedings of the ACM SIGMOD Conference*, Tucson, Arizona, 1998.

293. B. He and K. C. Chang. Statistical schema matching across web query interfaces. In *Proc. of SIGMOD*, 2003.

294. Bin He and Kevin Chang. Automatic complex schema matching across web query interfaces: A correlation mining approach. *TODS*, 31(1), 2006.

295. Bin He and Kevin Chen-Chuan Chang. Statistical schema integration across the deep web. In *Proc. of SIGMOD*, 2003.

296. Bin He, Kevin Chen-Chuan Chang, and Jiawei Han. Discovering complex matchings across web query interfaces: A correlation mining approach. In *KDD*, pages 148–157, 2004.

297. Bin He, Mitesh Patel, Zeng Zhang, and Kevin Chen-Chuan Chang. Accessing the deep Web: A survey. *Communications of the ACM*, 50(5):95–101, 2007.

298. Hao He, Haixun Wang, Jun Yang, and Philip S. Yu. Blinks: Ranked keyword searches on graphs. In *SIGMOD*, 2007.

299. Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *COLING*, pages 539–545, 1992.

300. M. A. Hernandez and S. J. Stolfo. The merge/purge problem for large databases. In *Proc. of SIGMOD*, 1995.

301. M. A. Hernández and S. J. Stolfo. Real-world data is dirty: Data cleansing and the merge/purge problem. *Data Mining and Knowledge Discovery*, 2:9–37, 1998.

302. Mauricio A. Hernandez, Renée J. Miller, and Laura M. Haas. Clio: A semi-automatic tool for schema mapping. In *SIGMOD*, 2001.

303. Raphael Hoffmann, Congle Zhang, and Daniel S. Weld. Learning 5000 relational extractors. In Hajic et al. [282], pages 286–295.

304. Vagelis Hristidis, Yannis Papakonstantinou, and Andrey Balmin. Keyword proximity search on XML graphs. In *ICDE*, 2003.

305. C. Hsu and M. Dung. Generating finite-state transducers for semi-structured data extraction from the web. *Inf. Syst.*, 23(8):521–538, 1998.

306. Wei Hu, Yuzhong Qu, and Gong Cheng. Matching large ontologies: A divide-and-conquer approach. *Data Knowl. Eng.*, 67(1):140–160, 2008.

307. Jiansheng Huang, Ting Chen, AnHai Doan, and Jeffrey F. Naughton. On the provenance of non-answers to queries over extracted data. *PVLDB*, 1(1), 2008.

308. G. Huck, P. Fankhauser, K. Aberer, and E. J. Neuhold. Jedi: Extracting and synthesizing information from the Web. In *CoopIS*, 1998.

309. Ryan Huebsch, Brent N. Chun, Joseph M. Hellerstein, Boon Thau Loo, Petros Maniatis, Timothy Roscoe, Scott Shenker, Ion Stoica, and Aydan R. Yumerefendi. The architecture of PIER: An Internet-scale query processor. In *CIDR*, 2005.

310. Ihab F. Ilyas, Walid G. Aref, and Ahmed K. Elmagarmid. Supporting top-k join queries in relational databases. In *VLDB*, 2003.

311. Ihab F. Ilyas, Walid G. Aref, Ahmed K. Elmagarmid, Hicham G. Elmongui, Rahul Shah, and Jeffrey Scott Vitter. Adaptive rank-aware query optimization in relational databases. *ACM Trans. Database Syst.*, 31(4), 2006.

312. Ihab F. Ilyas and Mohamed Soliman. *Probabilistic Ranking Techniques in Relational Databases*. Synthesis Lectures on Data Management. 2011.

313. Tomasz Imielinski and Witold Lipski. Incomplete information in relational databases. *JACM*, 31(4), 1984.

314. IBM Inc. IBM AlphaWorks QED Wiki. http://services.alphaworks.ibm.com/qedwiki/.

315. Microsoft Inc. Popfly. http://www.popfly.com/, 2008.

316. Yahoo Inc. Pipes. http://pipes.yahoo.com/pipes/.

317. Infochimps: Smart data for apps & analytics. http://www.infochimps.com/, 2011.

318. Yannis E. Ioannidis. Query optimization. *ACM Comput. Surv.*, 28(1), 1996.

319. Yannis E. Ioannidis, Raymond T. Ng, Kyusheok Shim, and Timos K. Sellis. Parametric query optimization. *VLDB J.*, 6(2), 1997.

320. Yannis E. Ioannidis and Raghu Ramakrishnan. Containment of conjunctive queries: Beyond relations as sets. *ACM Transactions on Database Systems*, 20(3):288–324, 1995.

321. Panagiotis G. Ipeirotis and Luis Gravano. Distributed search over the hidden web: Hierarchical database sampling and selection. In *VLDB*, pages 394–405, 2002.

322. U. Irmak and T. Suel. Interactive wrapper generation with minimal user effort. In *WWW*, 2006.

323. Zachary Ives, Daniela Florescu, Marc Friedman, Alon Levy, and Dan Weld. An adaptive query execution engine for data integration. In *Proceedings of the ACM SIGMOD Conference*, pages 299–310, 1999.

324. Zachary G. Ives, Todd J. Green, Grigoris Karvounarakis, Nicholas E. Taylor, Val Tannen, Partha Pratim Talukdar, Marie Jacob, and Fernando Pereira. The ORCHESTRA collaborative data sharing system. *SIGMOD Rec.*, 2008.

325. Zachary G. Ives, Alon Y. Halevy, and Daniel S. Weld. An XML query engine for network-bound data. *VLDB J.*, 11(4), December 2002.

326. Zachary G. Ives, Alon Y. Halevy, and Daniel S. Weld. Adapting to source properties in processing data integration queries. In *SIGMOD*, June 2004.

327. Zachary G. Ives, Craig A. Knoblock, Steven Minton, Mari Jacob, Partha Pratim Talukdar, Rattapoom Tuchindra, Jose Luis Ambite, Maria Muslea, and Cenk Gazen. Interactive data integration though smart copy & paste. In *Proceedings of the Conference on Innovative Data Systems Research (CIDR)*, 2009.

328. Zachary G. Ives and Nicholas E. Taylor. Sideways information passing for push query processing. In *ICDE*, 2008.

329. P. Jaccard. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Socit Vaudoise des Sciences Naturelles*, 37:547–579, 1901.

330. Ravi Jampani, Fei Xu, Mingxi Wu, Luis Leopoldo Perez, Chris Jermaine, and Peter J. Haas. The monte carlo database system: Stochastic analysis close to the data. *ACM Trans. Database Syst.*, 36(3), 2011.

331. M. A. Jaro. Unimatch: A record linkage system: User's manual. 1976. Technical Report, U.S. Bureau of the Census, Washington D.C.

332. T. S. Jayram, Phokion Kolaitis, and Erik Vee. The containment problem for real conjunctive queries with inequalities. In *Proc. of PODS*, pages 80–89, 2006.

333. S. Jeffery, M. Franklin, and A. Halevy. Pay-as-you-go user feedback for dataspace systems. In *Proc. of SIGMOD*, 2008.

334. Wen Jin and Jignesh M. Patel. Efficient and generic evaluation of ranked queries. In *SIGMOD Conference*, 2011.

335. Vanja Josifovski, Marcus Fontoura, and Attila Barta. Querying XML streams. *The VLDB Journal*, 14(2), 2005.

336. Navin Kabra and David J. DeWitt. Efficient mid-query re-optimization of sub-optimal query execution plans. In *SIGMOD*, 1998.

337. Varun Kacholia, Shashank Pandit, Soumen Chakrabarti, S. Sudarshan, Rushi Desai, and Hrishikesh Karambelkar. Bidirectional expansion for keyword search on graph databases. In *VLDB*, 2005.

338. D. V. Kalashnikov, S. Mehrotra, and Z. Chen. Exploiting relationships for domain-independent data cleaning. In *Proc. of the SDM Conf.*, 2005.

339. Jaewoo Kang and Jeffrey F. Naughton. On schema matching with opaque column names and data values. In *SIGMOD Conference*, pages 205–216, 2003.

340. Carl-Christian Kanne and Guido Moerkotte. Efficient storage of XML data. In *ICDE*, 2000.

341. Verena Kantere, Maher Manoubi, Iluju Kiringa, Timos K. Sellis, and John Mylopoulos. Peer coordination through distributed triggers. *PVLDB*, 3(2), 2010.

342. Grigoris Karvounarakis and Zachary G. Ives. Bidirectional mappings for data and update exchange. In *WebDB*, 2008.

343. Grigoris Karvounarakis and Zachary G. Ives. Querying data provenance. In *SIGMOD*, 2010.

344. Gjergji Kasneci, Maya Ramanath, Mauro Sozio, Fabian M. Suchanek, and Gerhard Weikum. Star: Steiner-tree approximation in relationship graphs. In *ICDE*, 2009.

345. Arthur M. Keller. Algorithms for translating view updates to database updates for views involving selections, projections, and joins. In *SIGMOD*, 1985.

346. Anastasios Kementsietsidis, Marcelo Arenas, and Renée J. Miller. Mapping data in peer-to-peer systems: Semantics and algorithmic issues. In *SIGMOD*, June 2003.

347. A. Klug. On conjunctive queries containing inequalities. *Journal of the ACM*, 35(1): pages 146–160, 1988.

348. D. Koller and N. Friedman. *Probabilistic Graphical Models*. The MIT Press, 2009.

349. George Konstantinidis and José Luis Ambite. Scalable query rewriting: A graph-based approach. In *SIGMOD Conference*, pages 97–108, 2011.

350. Donald Kossmann. The state of the artin distributed query procesing. *ACM Computing Surveys*, 32(4), 2000.

351. N. Koudas. Special issue on data quality. *IEEE Data Engineering Bulletin*, 29(2), 2006.

352. N. Koudas, A. Marathe, and D. Srivastava. Flexible string matching against large databases in practice. In *VLDB*, 2004.

353. N. Koudas, S. Sarawagi, and D. Srivastava. Record linkage: Similarity measures and algorithms. Tutorial, the ACM SIGMOD Conference, 2006.

354. Nick Koudas, Amit Marathe, and Divesh Srivastava. Flexible string matching against large databases in practice. In *VLDB*, pages 1078–1086, 2004.

355. Nick Koudas and Divesh Srivastava. Approximate joins: Concepts and techniques. In *VLDB*, page 1363, 2005.

356. Andries Kruger, C. Lee Giles, Frans Coetzee, Eric J. Glover, Gary William Flake, Steve Lawrence, and Christian W. Omlin. Deadliner: Building a new niche search engine. In *CIKM*, pages 272–281, 2000.

357. N. Kushmerick. Wrapper induction for information extraction, 1997. PhD thesis, University of Washington.

358. N. Kushmerick. Wrapper induction: Efficiency and expressiveness. *Artif. Intell.*, 118(1-2):15–68, 2000.

359. N. Kushmerick. Wrapper verification. *World Wide Web*, 3(2):79–94, 2000.

360. Nick Kushmerick, Robert Doorenbos, and Daniel Weld. Wrapper induction for information extraction. In *IJCAI*, 1997.

361. Chung T. Kwok and Daniel S. Weld. Planning to gather information. In *Proc. of the 13th National Conf. on Artificial Intelligence (AAAI)*, pages 32–39, 1996.

362. A. H. F. Laender, B. A. Ribeiro-Neto, A. S. da Silva, and J. S. Teixeira. A brief survey of web data extraction tools. *SIGMOD Record*, 31(2):84–93, 2002.

363. J. D. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of the Int. Conf. on Machine Learning (ICML)*, pages 282–289, 2001.

364. Laks V. S. Lakshmanan, Nicola Leone, Robert Ross, and V. S. Subrahmanian. Probview: A flexible probabilistic database system. *ACM Trans. Database Syst.*, 22(3), 1997.

365. Eric Lambrecht, Subbarao Kambhampati, and Senthil Gnanaprakasam. Optimizing recursive information gathering plans. In *Proc. of the 16th Int. Joint Conf on Artificial Intelligence (IJCAI)*, pages 1204–1211, 1999.

366. T. Landers and R. Rosenberg. An overview of multibase. In *Proceedings of the Second International Symoposium on Distributed Databases*, pages 153–183. North Holland, Amsterdam, 1982.

367. Veronique Lattes and Marie-Christine Rousset. The use of the CARIN language and algorithms for information integration: The PICSEL project. In *Proceedings of the ECAI-98 Workshop on Intelligent Information Integration*, 1998.

368. Michael K. Lawrence, Rachel Pottinger, and Sheryl Staub-French. Data coordination: Supporting contingent updates. *PVLDB*, 4(11), 2011.

369. Amy J. Lee, Andreas Koeller, Anisoara Nica, and Elke A. Rundensteiner. Data warehouse evolution: Trade-offs between quality and cost of query rewritings. In *ICDE*, 1999.

370. Dongwon Lee. Weighted exact set similarity join. Tutorial Presentation. Available from http://pike.psu.edu/p2/wisc09-tech.ppt, 2009.

371. Y. Lee, M. Sayyadian, A. Doan, and A. Rosenthal. eTuner: Tuning schema matching software using synthetic scenarios. *VLDB J.*, 16(1):97–122, 2007.

372. Yoonkyong Lee, AnHai Doan, Robin Dhamankar, Alon Y. Halevy, and Pedro Domingos. imap: Discovering complex mappings between database schemas. In *Proc. of SIGMOD*, pages 383–394, 2004.

373. Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, 2002.

374. K. Lerman, L. Getoor, S. Minton, and C. A. Knoblock. Using the structure of Web sites for automatic segmentation of tables. In *SIGMOD*, 2004.

375. K. Lerman, S. Minton, and C. A. Knoblock. Wrapper maintenance: A machine learning approach. *J. Artif. Intell. Res. (JAIR)*, 18:149–181, 2003.

376. V. Levenshtein. Binay code capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, 163(4):845–848, 1965. Original in Russian–translation in *Soviet Physics Doklady*, 10(8): 707–710, 1966.

377. Alon Levy and Marie-Christine Rousset. Combining Horn rules and description logics in CARIN. *Artificial Intelligence*, 104:165–209, September 1998.

378. Alon Y. Levy. Obtaining complete answers from incomplete databases. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, pages 402–412, Bombay, India, 1996.

379. Alon Y. Levy. Logic-based techniques in data integration. In Jack Minker, editor, *Logic-Based Artificial Intelligence*, pages 575–595. Kluwer Academic Publishers, Dordrecht, 2000.

380. Alon Y. Levy, Anand Rajaraman, and Joann J. Ordille. Query answering algorithms for information agents. In *Proc. of the 13th National Conf. on Artificial Intelligence (AAAI)*, 1996.

381. Alon Y. Levy, Anand Rajaraman, and Joann J. Ordille. Querying heterogeneous information sources using source descriptions. In *VLDB*, 1996.

382. Alon Y. Levy and Yehoshua Sagiv. Queries independent of updates. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, pages 171–181, Dublin, Ireland, 1993.

383. Chengkai Li, Kevin Chen-Chuan Chang, Ihab F. Ilyas, and Sumin Song. RankSQL: Query algebra and optimization for relational top-k queries. In *SIGMOD*, 2005.

384. W. Li and C. Clifton. Semantic integration in heterogeneous databases using neural networks. In *VLDB*, pages 1–12, 1994.

385. X. Li, P. Morie, and D. Roth. Robust reading: Identification and tracing of ambiguous names. In *Proc. of the HLT-NAACL Conf.*, pages 17–24, 2004.

386. X. Li, P. Morie, and D. Roth. Semantic integration in text: From ambiguous names to identifiable entities. *AI Magazine*, 26(1):45–58, 2005. A. Doan and N. Noy and A. Halevy (editors).

387. Xian Li, Timothy Lebo, and Deborah L. McGuinness. Provenance-based strategies to develop trust in semantic web applications. In *IPAW*, 2010.

388. Y. Li, A. Terrell, and J. M. Patel. WHAM: A high-throughput sequence alignment method. In *SIGMOD Conference*, pages 445–456, 2011.

389. Leonid Libkin. Incomplete information and certain answers in general data models. In *PODS*, pages 59–70, 2011.

390. E. P. Lim, J. Srivastava, S. Prabhakar, and J. Richardson. Entity identification in database integration. In *Proc. of the 5th Int. Conf. on Data Engineering (ICDE-93)*, pages 294–301, 1993.

391. Girija Limaye, Sunita Sarawagi, and Soumen Chakrabarti. Annotating and searching web tables using entities, types and relationships. *PVLDB* 3(1), pages 1338–1347, 2010.

392. B. Liu. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Data-Centric Systems and Applications*. Springer, 2007.

393. B. Liu, R. L. Grossman, and Y. Zhai. Mining data records in Web pages. In *KDD*, 2003.

394. L. Liu, C. Pu, and W. Han. XWRAP: An XML-enabled wrapper construction system for Web information sources. In *Proc. of the IEEE Intl Conf. on Data Engineering (ICDE)*, 2000.

395. Mengmeng Liu, Nicholas E. Taylor, Wenchao Zhou, Zachary G. Ives, and Boon Thau Loo. Recursive computation of regions and connectivity in networks. In *ICDE*, 2009.

396. Xiufeng Liu, Christian Thomsen, and Torben Bach Pedersen. Etlmr: A highly scalable dimensional etl framework based on mapreduce. In *Proceedings of the 13th International Conference on Data Warehousing and Knowledge Discovery*, DaWaK'11, Berlin, Heidelberg, 2011.

397. Boon Thau Loo, Joseph M. Hellerstein, Ion Stoica, and Raghu Ramakrishnan. Declarative routing: Extensible routing with declarative queries. In *SIGCOMM*, 2005.

398. James J. Lu, Guido Moerkotte, Joachim Schue, and V.S. Subrahmanian. Efficient maintenance of materialized mediated views. In *SIGMOD*, 1995.

399. Meiyu Lu, Divyakant Agrawal, Bing Tian Dai, and Anthony K. H. Tung. Schema-as-you-go: On probabilistic tagging and querying of wide tables. In *SIGMOD Conference*, pages 181–192, 2011.

400. Bertram Ludäscher, Ilkay Altintas, Chad Berkley, Dan Higgins, Efrat Jaeger, Matthew Jones, Edward A. Lee, Jing Tao, and Yang Zhao. Scientific workflow management and the kepler system. *Concurrency and Computation: Practice and Experience*, 2006.

401. Bertram Ludäscher, Rainer Himmeröder, Georg Lausen, Wolfgang May, and Christian Schlepphorst. Managing semistructured data with *FLORID*: A deductive object-oriented perspective. *Information Systems*, 23(8), 1998.

402. Qiong Luo, Sailesh Krishnamurthy, C. Mohan, Hamid Pirahesh, Honguk Woo, Bruce G. Lindsay, and Jeffrey F. Naughton. Middle-tier database caching for e-business. In *SIGMOD*, 2002.

403. Yi Luo, Wei Wang, and Xuemin Lin. Spark: A keyword search engine on relational databases. In *ICDE*, 2008.

404. Lothar F. Mackert and Guy M. Lohman. R* optimizer validation and performance evaluation for distributed queries. In *VLDB*, 1986.

405. Lothar F. Mackert and Guy M. Lohman. R* optimizer validation and performance evaluation for local queries. In *SIGMOD*, 1986.

406. Jayant Madhavan, Philip A. Bernstein, AnHai Doan, and Alon Y. Halevy. Corpus-based schema matching. In *Proc. of ICDE*, pages 57–68, 2005.

407. Jayant Madhavan, Philip A. Bernstein, and Erhard Rahm. Generic schema matching with Cupid. In *VLDB*, 2001.

408. Jayant Madhavan and Alon Halevy. Composing mappings among data sources. In *Proc. of VLDB*, 2003.

409. Jayant Madhavan, Shawn Jeffery, Shirley Cohen, Xin Dong, David Ko, Cong Yu, and Alon Halevy. Web-scale data integration: You can only afford to pay as you go. In *CIDR*, 2007.

410. Jayant Madhavan, David Ko, Lucja Kot, Vignesh Ganapathy, Alex Rasmussen, and Alon Halevy. Google's deep-web crawl. In *Proc. of VLDB*, pages 1241–1252, 2008.

411. M. Magnani and D. Montesi. Uncertainty in data integration: Current approaches and open problems. In *VLDB Workshop on Management of Uncertain Data*, pages 18–32, 2007.

412. M. Magnani, N. Rizopoulos, P. Brien, and D. Montesi. Schema integration based on uncertain semantic mappings. *Lecture Notes in Computer Science*, pages 31–46, 2005.

413. Hatem A. Mahmoud and Ashraf Aboulnaga. Schema clustering and retrieval for multi-domain pay-as-you-go data integration systems. In *SIGMOD Conference*, pages 411–422, 2010.

414. C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

415. Amélie Marian, Nicolas Bruno, and Luis Gravano. Evaluating top-k queries over web-accessible databases. *ACM Trans. Database Syst.*, 29(2), 2004.

416. A. McCallum and B. Wellner. Conditional models of identity uncertainty with application to noun coreference. In *Proc. of the Conf. on Advances in Neural Information Processing Systems (NIPS)*, 2004.

417. Andrew McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. A machine learning approach to building domain-specific search engines. In *IJCAI*, pages 662–667, 1999.

418. Andrew K. McCallum, Kamal Nigam, and Lyle H. Ungar. Efficient clustering of high-dimensional data sets with application to reference matching. In *KDD*, 2000.

419. R. McCann, B. K. AlShebli, Q. Le, H. Nguyen, L. Vu, and A. Doan. Mapping maintenance for data integration systems. In *VLDB*, 2005.

420. Robert McCann, AnHai Doan, Vanitha Varadarajan, Alexander Kramnik, and ChengXiang Zhai. Building data integration systems: A mass collaboration approach. In *WebDB*, pages 25–30, 2003.

421. Robert McCann, Warren Shen, and AnHai Doan. Matching schemas in online communities: A web 2.0 approach. In *ICDE*, pages 110–119, 2008.

422. Luke McDowell, Oren Etzioni, Alon Halevy, Henry Levy, Steven Gribble, William Pentney, Deepak Verma, and Stani Vlasseva. Enticing ordinary people onto the semantic web via instant gratification. In *Proceedings of the Second International Conference on the Semantic Web*, October 2003.

423. C. Meek, J. M. Patel, and S. Kasetty. OASIS: An online and accurate technique for local-alignment searches on biological sequences. In *VLDB*, pages 910–921, 2003.

424. Alexandra Meliou, Wolfgang Gatterbauer, Katherine F. Moore, and Dan Suciu. The complexity of causality and responsibility for query answers and non-answers. *PVLDB*, 4(1), 2010.

425. Alexandra Meliou, Wolfgang Gatterbauer, Suman Nath, and Dan Suciu. Tracing data errors with view-conditioned causality. In *SIGMOD*, 2011.

426. Sergey Melnik, Philip A. Bernstein, Alon Y. Halevy, and Erhard Rahm. Supporting executable mappings in model management. In *Proc. of SIGMOD*, pages 167–178, 2005.

427. Sergey Melnik, Hector Garcia-Molina, and Erhard Rahm. Similarity flooding: A versatile graph matching algorithm. In *Proceedings of the 18th International Conference on Data Engineering (ICDE)*, 2002.

428. Sergey Melnik, Erhard Rahm, and Phil Bernstein. Rondo: A programming platform for generic model management. In *Proc. of SIGMOD*, 2003.

429. X. Meng, D. Hu, and C. Li. Schema-guided wrapper maintenance for Web-data extraction. In *WIDM*, 2003.

430. Gerome Miklau and Dan Suciu. Containment and equivalence for a fragment of XPath. *J. ACM*, 51(1), 2004.

431. George A. Miller. Wordnet: A lexical database for English. In *HLT*, 1994.

432. Renée J. Miller, Laura M. Haas, and Mauricio Hernandez. Schema matching as query discovery. In *VLDB*, 2000.

433. Tova Milo, Serge Abiteboul, Bernd Amann, Omar Benjelloun, and Frederic Dang Ngoc. Exchanging intensional XML data. In *Proc. of SIGMOD*, pages 289–300, 2003.

434. Tova Milo and Sagit Zohar. Using schema matching to simplify heterogeneous data translation. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, 1998.

435. Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. Distant supervision for relation extraction without labeled data. In Keh-Yih Su, Jian Su, and Janyce Wiebe, editors, *ACL/AFNLP*, pages 1003–1011. The Association for Computer Linguistics, 2009.

436. Paolo Missier, Satya Sanket Sahoo, Jun Zhao, Carole A. Goble, and Amit P. Sheth. *Janus*: From workflows to semantic provenance and linked open data. In *IPAW*, 2010.

437. Hoshi Mistry, Prasan Roy, S. Sudarshan, and Krithi Ramamritham. Materialized view selection and maintenance using multi-query optimization. In *SIGMOD*, 2001.

438. Tom M. Mitchell. *Machine Learning*. McGraw Hill, 1997.

439. Prasenjit Mitra. An algorithm for answering queries efficiently using views. In *ADC*, pages 99–106, 2001.

440. Prasenjit Mitra, Natasha F. Noy, and Anuj R. Jaiswal. Omen: A probabilistic ontology mapping tool. In *International Semantic Web Conference*, pages 537–547, 2005.

441. R. Mohapatra, K. Rajaraman, and S. Y. Sung. Efficient wrapper reinduction from dynamic Web sources. In *Web Intelligence*, 2004.

442. A. E. Monge and C. Elkan. The field matching problem: Algorithms and applications. In *KDD*, 1996.

443. A. E. Monge and C. P. Elkan. An efficient domain-independent algorithm for detecting approximately duplicate database records. In *Proc. of the Second ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD-97)*, pages 23–29, 1997.

444. Hyun Jin Moon, Carlo Curino, Alin Deutsch, Chien-Yi Hou, and Carlo Zaniolo. Managing and querying transaction-time databases under schema evolution. *PVLDB*, 1(1), 2008.

445. Peter Mork, Philip A. Bernstein, and Sergey Melnik. Teaching a schema translator to produce o/r views. In *Proceedings of Entity Relationship Conference*, pages 102–119, 2007.

446. Amihai Motro. Integrity = validity + completeness. *ACM Transactions on Database Systems*, 14(4):480–502, December 1989.

447. Inderpal Singh Mumick, Dallan Quass, and Barinderpal Singh Mumick. Maintenance of data cubes and summary tables in a warehouse. In *SIGMOD*, 1997.

448. Kiran-Kumar Muniswamy-Reddy, David A. Holland, Uri Braun, and Margo I. Seltzer. Provenance-aware storage systems. In *USENIX Annual Technical Conference, General Track*, 2006.

449. I. Muslea, S. Minton, and C. A. Knoblock. A hierarchical approach to wrapper induction. In *Agents*, 1999.

450. I. Muslea, S. Minton, and C. A. Knoblock. Hierarchical wrapper induction for semistructured information sources. *Autonomous Agents and Multi-Agent Systems*, 4(1/2):93–114, 2001.

451. I. Muslea, S. Minton, and C. A. Knoblock. Active learning with strong and weak views: A case study on wrapper induction. In *IJCAI*, 2003.

452. A. Nash, P. Bernstein, and S. Melnik. Composition of mappings given by embedded dependencies. *ACM Transactions on Database Systems*, 32(1), 2007.

453. F. Naumann and M. Herschel. *An Introduction to Duplicate Detection (Synthesis Lectures on Data Management)*. Morgan & Claypool, 2010. M. Tamer Ozsu (editor).

454. Felix Naumann, Johann Christoph Freytag, and Ulf Leser. Completeness of integrated information sources. *Inf. Syst.*, 29(7):583–615, 2004.

455. G. Navarro. A guided tour to approximate string matching. *ACM Comput. Surv.*, 33(1):31–88, 2001.

456. S. Needleman and C. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453, 1970.

457. Frank Neven and Thomas Schwentick. XPath containment in the presence of disjunction, DTDs, and variables. In *ICDT*, 2003.

458. H. B. Newcombe, J. M. Kennedy, S. Axford, and A. James. Automatic linkage of vital records. *Science*, 130(3381):954–959, 1959.

459. W. S. Ng, B. C. Ooi, K.-L. Tan, and A. Zhou. Peerdb: A p2p-based system for distributed data sharing. In *ICDE*, Bangalore, India, 2003.

460. Hoa Nguyen, Ariel Fuxman, Stelios Paparizos, Juliana Freire, and Rakesh Agrawal. Synthesizing products for online catalogs. *PVLDB*, 4(7):409–418, 2011.

461. Zaiqing Nie, Ji-Rong Wen, and Wei-Ying Ma. Object-level vertical search. In *CIDR*, pages 235–246, 2007.

462. H. Nottelmann and U. Straccia. Information retrieval and machine learning for probabilistic schema matching. *Information Processing and Management*, 43(3):552–576, 2007.

463. N. F. Noy, A. Doan, and A. Y. Halevy. Semantic integration. *AI Magazine*, 26(1):7–10, 2005.

464. Natalya F. Noy and Mark A. Musen. PROMPT: Algorithm and tool for automated ontology merging and alignment. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2000.

465. Natalya Freidman Noy and Mark A. Musen. Smart: Automated support for ontology merging and alignment. In *Proceedings of the Knowledge Acquisition Workshop, Banff, Canada*, 1999.

466. Natalya Fridman Noy. Semantic integration: A survey of ontology-based approaches. *SIGMOD Record*, 33(4):65–70, 2004.

467. Alexandros Ntoulas, Petros Zerfos, and Junghoo Cho. Downloading textual hidden web content through keyword queries. In *JCDL*, pages 100–109, 2005.

468. T. Oinn, M. Greenwood, M. Addis, N. Alpdemir, J. Ferris, K. Glover, C. Goble, A. Goderis, D. Hull, D. Marvin, P. Li, P. Lord, M. Pocock, M. Senger, R. Stevens, A. Wipat, and C. Wroe. Taverna: Lessons in creating a workflow environment for the life sciences. *Concurrency and Computation: Practice and Experience*, 18(10), 2006.

469. B. On, N. Koudas, D. Lee, and D. Srivastava. Group linkage. In *ICDE*, 2007.

470. Open provenance model. http://twiki.ipaw.info/bin/view/Challenge/OPM, 2008.

471. M. Tamer Ozsu and Patrick Valduriez. *Principles of Distributed Database Systems*. Springer, 2011.

472. Luigi Palopoli, Domenico Sacc, G. Terracina, and Domenico Ursino. A unified graph-based framework for deriving nominal interscheme properties, type conflicts and object cluster similarities. In *Proceedings of CoopIS*, 1999.

473. Paolo Papotti, Valter Crescenzi, Paolo Merialdo, Mirko Bronzi, and Lorenzo Blanco. Redundancy-driven web data extraction and integration. In *WebDB*, 2010.

474. A. Parameswaran, N. Dalvi, H. Garcia-Molina, and R. Rastogi. Optimal schemes for robust web extraction. In *VLDB*, 2011.

475. H. Pasula, B. Marthi, B. Milch, S. Russell, and I. Shpitser. Identity uncertainty and citation matching. In *Proc. of the NIPS Conf.*, pages 1401–1408, 2002.

476. Feng Peng and Sudarshan S. Chawathe. Xsq: A streaming xpath engine. *ACM Trans. Database Syst.*, 30(2), 2005.

477. L. Philips. Hanging on the metaphone. *Computer Language Magazine*, 7(12):39–44, 1990.

478. L. Philips. The double metaphone search algorithm. *C/C++ Users Journal*, 18(5), 2000.

479. J. C. Pinheiro and D. X. Sun. Methods for linking and mining massive heterogeneous databases. In *Proc. of the ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 309–313, 1998.

480. Lucian Popa and Val Tannen. An equational chase for path conjunctive queries, constraints and views. In *Proceedings of the International Conference on Database Theory (ICDT)*, 1999.

481. Rachel Pottinger and Philip A. Bernstein. Merging models based on given correspondences. In *Proc. of VLDB*, pages 826–873, 2003.

482. Rachel Pottinger and Alon Halevy. Minicon: A scalable algorithm for answering queries using views. *VLDB Journal*, 2001.

483. Jeffrey Pound, Ihab F. Ilyas, and Grant E. Weddell. Expressive and flexible access to web-extracted data: A keyword-based structured query language. In *SIGMOD Conference*, pages 423–434, 2010.

484. C. Pu. Key equivalence in heterogeneous databases. In *Proc. of the 1st Int. Workshop on Interoperability in Multidatabase Systems*, 1991.

485. Sven Puhlmann, Melanie Weis, and Felix Naumann. Xml duplicate detection using sorted neighborhoods. In *EDBT*, pages 773–791, 2006.

486. Dallan Quass and Jennifer Widom. On-line warehouse view maintenance. In *SIGMOD*.

487. Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. *VLDB Journal*, 10(4):334–350, 2001.

488. Anand Rajaraman, Yehoshua Sagiv, and Jeffrey D. Ullman. Answering queries using templates with binding patterns. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 105–112, San Jose, CA, 1995.

489. Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems*. McGraw Hill, 2000.

490. Vijayshankar Raman, Amol Deshpande, and Joseph M. Hellerstein. Using state modules for adaptive query processing. In *ICDE*, 2003.

491. Vijayshankar Raman and Joseph M. Hellerstein. Potter's wheel: An interactive data cleaning system. In *VLDB*, pages 381–390, 2001.

492. Aditya Ramesh, S. Sudarshan, and Purva Joshi. Keyword search on form results. *PVLDB*, 4(11), 2011.

493. J. Raposo, A. Pan, M. Álvarez, and J. Hidalgo. Automatically maintaining wrappers for semi-structured Web sources. *Data Knowl. Eng.*, 61(2):331–358, 2007.

494. P. D. Ravikumar and W. Cohen. A hierarchical graphical model for record linkage. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence (UAI)*, pages 454–461, 2004.

495. Simon Razniewski and Werner Nutt. Completeness of queries over incomplete databases. *PVLDB*, 4(11):749–760, 2011.

496. Christopher Re, Nilesh N. Dalvi, and Dan Suciu. Efficient top-k query evaluation on probabilistic data. In *ICDE*, 2007.

497. E. S. Ristad and P. N. Yianilos. Learning string-edit distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(5):522–532, 1998.

498. Mary Tork Roth, Fatma Ozcan, and Laura M. Haas. Cost models do matter: Providing cost information for diverse data sources in a federated system. In *VLDB*, 1999.

499. Elke A. Rundensteiner, Luping Ding, Timothy M. Sutherland, Yali Zhu, Bradford Pielech, and Nishant Mehta. Cape: Continuous query engine with heterogeneous-grained adaptivity. In *VLDB*, 2004.

500. R. C. Russell. 1918. U.S. Patent 1,261,167.

501. R. C. Russell. 1922. U.S. Patent 1,435,663.

502. Y. Sagiv and M. Yannakakis. Equivalence among relational expressions with the union and difference operators. *Journal of the ACM*, 27(4):633–655, 1981.

503. A. Sahuguet and F. Azavant. Web ecology: Recycling HTML pages as XML documents using W4F. In *WebDB (Informal Proceedings)*, 1999.

504. Marcos Antonio Vaz Salles, Jens-Peter Dittrich, Shant Kirakos Karakashian, Olivier René Girard, and Lukas Blunschi. iTrails: Pay-as-you-go information integration in dataspaces. In *VLDB*, 2007.

505. Yatin Saraiya. *Subtree-elimination algorithms in deductive databases*. PhD thesis, Stanford University, Stanford, California, 1991.

506. S. Sarawagi. Information extraction. *Foundations and Trends in Databases*, 1(3):261–377, 2008.

507. S. Sarawagi and A. Bhamidipaty. Interactive deduplication using active learning. In *Proc. of the ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 269–278, 2002.

508. Sunita Sarawagi and Alok Kirpal. Efficient set joins on similarity predicates. In *SIGMOD Conference*, pages 743–754, 2004.

509. Nikos Sarkas, Stelios Paparizos, and Panayiotis Tsaparas. Structured annotations of web queries. In *SIGMOD Conference*, 2010.

510. Anish Das Sarma, Xin Dong, and Alon Y. Halevy. Bootstrapping pay-as-you-go data integration systems. In *SIGMOD Conference*, pages 861–874, 2008.

511. M. Sayyadian, Y. Lee, A. Doan, and A. Rosenthal. Tuning schema matching software using synthetic scenarios. In *VLDB*, pages 994–1005, 2005.

512. Mayssam Sayyadian, Hieu LeKhac, AnHai Doan, and Luis Gravano. Efficient keyword search across heterogeneous relational databases. In *ICDE*, 2007.

513. Carlos Eduardo Scheidegger, Huy T. Vo, David Koop, Juliana Freire, and Cláudio T. Silva. Querying and re-using workflows with vstrails. In *SIGMOD Conference*, 2008.

514. Karl Schnaitter and Neoklis Polyzotis. Evaluating rank joins with optimal cost. In *PODS*, 2008.

515. Karl Schnaitter, Joshua Spiegel, and Neoklis Polyzotis. Depth estimation for ranking query optimization. In *VLDB*, 2007.

516. Thomas Schwentick. XPath query containment. *SIGMOD Record*, 33(1), 2004.

517. Luc Segoufin and Victor Vianu. Validating streaming XML documents. In *PODS*, 2002.

518. Prithviraj Sen and Amol Deshpande. Representing and querying correlated tuples in probabilistic databases. In *ICDE*, 2007.

519. Warren Shen, Pedro DeRose, Robert McCann, AnHai Doan, and Raghu Ramakrishnan. Toward best-effort information extraction. In *SIGMOD*, 2008.

520. Warren Shen, Pedro DeRose, Long Vu, AnHai Doan, and Raghu Ramakrishnan. Source-aware entity matching: A compositional approach. In *ICDE*, pages 196–205, 2007.

521. Warren Shen, Xin Li, and AnHai Doan. Constraint-based entity matching. In *AAAI*, pages 862–867, 2005.

522. Cláudio T. Silva, Erik W. Anderson, Emanuele Santos, and Juliana Freire. Using vistrails and provenance for teaching scientific visualization. *Comput. Graph. Forum*, 30(1), 2011.

523. Alkis Simitsis, Georgia Koutrika, and Yannis Ioannidis. Précis: From unstructured keywords as queries to structured databases as answers. *The VLDB Journal*, 17, 2008. Available from http://dx.doi.org/10.1007/s00778-007-0075-9.

524. P. Singla and P. Domingos. Object identification with attribute-mediated dependences. In *Proc. of the PKDD Conf.*, pages 297–308, 2005.

525. John Miles Smith, Philip A. Bernstein, Umeshwar Dayal, Nathan Goodman, Terry Landers, Ken W.T. Lin, and Eugene Wong. MULTIBASE – Integrating heterogeneous distributed database systems. In *Proceedings of 1981 National Computer Conference*, 1981.

526. T. Smith and M. Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197, 1981.

527. Socrata: The social data cloud company. http://www.socrata.com/, 2011.

528. S. Soderland. Learning information extraction rules for semi-structured and free text. *Machine Learning*, 34(1-3):233–272, 1999.

529. S. Soderland, D. Fisher, J. Aseltine, and W. G. Lehnert. Crystal: Inducing a conceptual dictionary. In *IJCAI*, 1995.

530. Michael Stonebraker. *The Design and Implementation of Distributed INGRES*. Boston, MA, USA, 1986.

531. Michael Stonebraker, Daniel J. Abadi, David J. DeWitt, Samuel Madden, Erik Paulson, Andrew Pavlo, and Alexander Rasin. MapReduce and parallel DBMSs: Friends or foes? *Commun. ACM*, 53(1), 2010.

532. Michael Stonebraker, Paul M. Aoki, Witold Litwin, Avi Pfeffer, Adam Sah, Jeff Sidell, Carl Staelin, and Andrew Yu. Mariposa: A wide-area distributed database system. *VLDB J.*, 5(1), 1996.

533. V.S. Subrahmanian, S. Adali, A. Brink, R. Emery, J. Lu, A. Rajput, T. Rogers, R. Ross, and C. Ward. HERMES: A heterogeneous reasoning and mediator system. Technical Report, University of Maryland, 1995.

534. Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A large ontology from wikipedia and wordnet. *J. Web Sem.*, 6(3), 2008.

535. Dan Suciu, Dan Olteanu, Christopher Ré, and Christoph Koch. *Probabilistic Databases*. Synthesis Lectures on Data Management. 2011.

536. Tableau software. http://www.tableausoftware.com/, 2011.

537. R. L. Taft. Name search techniques. 1970. Technical Report, special report no. 1, New York State Identification and Intelligence System, Albany, N.Y.

538. Partha Pratim Talukdar, Zachary G. Ives, and Fernando Pereira. Automatically incorporating new sources in keyword search-based data integration. In *SIGMOD*, 2010.

539. Partha Pratim Talukdar, Marie Jacob, Muhammad Salman Mehmood, Koby Crammer, Zachary G. Ives, Fernando Pereira, and Sudipto Guha. Learning to create data-integrating queries. In *VLDB*, 2008.

540. Sandeep Tata and Guy M. Lohman. Sqak: Doing more with keywords. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, New York, NY, USA, 2008. Available from http://doi.acm.org/10.1145/1376616.1376705.

541. Igor Tatarinov and Alon Halevy. Efficient query reformulation in peer data management systems. In *Proc. of SIGMOD*, 2004.

542. Igor Tatarinov, Stratis Viglas, Kevin S. Beyer, Jayavel Shanmugasundaram, Eugene J. Shekita, and Chun Zhang. Storing and querying ordered XML using a relational database system. In *SIGMOD*, 2002.

543. Nesime Tatbul, Ugur Cetintemel, Stanley B. Zdonik, Mitch Cherniack, and Michael Stonebraker. Load shedding in a data stream manager. In *VLDB*, 2003.

544. Nicholas E. Taylor and Zachary G. Ives. Reconciling while tolerating disagreement in collaborative data sharing. In *SIGMOD*, 2006.

545. S. Tejada, C. A. Knoblock, and S. Minton. Learning object identification rules for information integration. *Inf. Syst.*, 26(8):607–633, 2001.

546. Andreas Thor and Erhard Rahm. MOMA – A mapping-based object matching system. In *CIDR*, pages 247–258, 2007.

547. Feng Tian and David J. DeWitt. Tuple routing strategies for distributed eddies. In *VLDB*, 2003.

548. Y. Tian, S. Tata, R. A. Hankins, and J. M. Patel. Practical methods for constructing suffix trees. *VLDB J.*, 14(3):281–299, 2005.

549. Kai-Ming Ting and Ian H. Witten. Issues in stacked generalization. *Journal of Artificial Intelligence Research*, 10:271–289, 1999.

550. Guilherme A. Toda, Eli Cortez, Altigran S. da Silva, and Edleno de Moura. A probabilistic approach for automatically filling form-based web interfaces. *PVLDB*, 4(3):151–160, 2011.

551. Odysseas G. Tsatalos, Marvin H. Solomon, and Yannis E. Ioannidis. The GMAP: A versatile tool for physical data independence. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, pages 367–378, Santiago, Chile, 1994.

552. Yi-Cheng Tu, Song Liu, Sunil Prabhakar, Bin Yao, and William Schroeder. Using control theory for load shedding in data stream management. In *ICDE*, 2007.

553. Rattapoom Tuchindra, Pedro Szekely, and Craig Knoblock. Building mashups by example. In *Proceedings of CHI*, pages 139–148, 2008.

554. Jeffrey D. Ullman. *Principles of Database and Knowledge-Base Systems, Volumes I, II*. Computer Science Press, Rockville MD, 1989.

555. Jeffrey D. Ullman. Information Integration using Logical Views. In *Proceedings of the International Conference on Database Theory (ICDT)*, 1997.

556. Tolga Urhan, Michael J. Franklin, and Laurent Amsaleg. Cost based query scrambling for initial delays. In *SIGMOD*, 1998.

557. Ron van der Meyden. The complexity of querying indefinite data about linearly ordered domains. In *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, pages 331–345, San Diego, CA, 1992.

558. Petros Venetis, Alon Y. Halevy, Jayant Madhavan, Marius Pasca, Warren Shen, Fei Wu, Gengxin Miao, and Chung Wu. Recovering semantics of tables on the web. *PVLDB*, 4(9):528–538, 2011.

559. Rares Vernica, Michael J. Carey, and Chen Li. Efficient parallel set-similarity joins using MapReduce. In *SIGMOD Conference*, pages 495–506, 2010.

560. Daisy Zhe Wang, Michael J. Franklin, Minos N. Garofalakis, Joseph M. Hellerstein, and Michael L. Wick. Hybrid in-database inference for declarative information extraction. In *SIGMOD Conference*, 2011.

561. Daisy Zhe Wang, Eirinaios Michelakis, Minos N. Garofalakis, and Joseph M. Hellerstein. BayesStore: Managing large, uncertain data repositories with probabilistic graphical models. *PVLDB*, 1(1), 2008.

562. J. Wang and F. H. Lochovsky. Data extraction and label assignment for Web databases. In *WWW*, 2003.

563. Wei Wang. Similarity join algorithms: An introduction. Tutorial Presentation. Available from http://www.cse.unsw.edu.au/ weiw/project/tutorial-simjoin-SEBD08.pdf, 2008.

564. Y. R. Wang and S. E. Madnick. The inter-database instance identification problem in integrating autonomous systems. In *Proc. of the 5th Int. Conf. on Data Engineering (ICDE-89)*, pages 46–55, 1989.

565. Yalin Wang and Jianying Hu. A machine learning based approach for table detection on the web. In *WWW*, pages 242–250, 2002.

566. M. Waterman, T. Smith, and W. Beyer. Some biological sequence metrics. *Advances in Math*, 20(4):367–387, 1976.

567. Melanie Weis and Felix Naumann. Detecting duplicates in complex XML data. In *ICDE*, page 109, 2006.

568. Steven Euijong Whang and Hector Garcia-Molina. Developments in generic entity resolution. *IEEE Data Eng. Bull.*, 34(3):51–59, 2011.

569. Michael L. Wick, Andrew McCallum, and Gerome Miklau. Scalable probabilistic databases with factor graphs and MCMC. *PVLDB*, 3(1), 2010.

570. Michael L. Wick, Khashayar Rohanimanesh, Karl Schultz, and Andrew McCallum. A unified approach for schema matching, coreference and canonicalization. In *KDD*, pages 722–730, 2008.

571. Gio Wiederhold. Mediators in the architecture of future information systems. *IEEE Computer*, pages 38–49, March 1992.

572. W. E. Winkler. Improved decision rules in the Fellegi-Sunter model of record linkage, 1993. Technical Report, Statistical Research Report Series RR93/12, U.S. Bureau of the Census.

573. W. E. Winkler. The state of record linkage and current research problems, 1999. Technical Report, Statistical Research Report Series RR99/04, U.S. Bureau of Census.

574. W. E. Winkler. Methods for record linkage and Bayesian networks, 2002. Technical Report, Statistical Research Report Series RRS2002/05, U.S. Bureau of the Census.

575. W. E. Winkler and Y. Thibaudeau. An application of the Fellegi-Sunter model of record linkage to the 1990 U.S. decennial census. 1991. Technical Report, Statistical Research Report Series RR91/09, U.S. Bureau of the Census, Washington, D.C.

576. David Wolpert. Stacked generalization. *Neural Networks*, 5:241–259, 1992.

577. Jeffrey Wong and Jason I. Hong. Making mashups with marmite: Towards end-user programming for the web. In *CHI*, pages 1435–1444, 2007.

578. Fei Wu and Daniel S. Weld. Autonomously semantifying wikipedia. In Mário J. Silva, Alberto H. F. Laender, Ricardo A. Baeza-Yates, Deborah L. McGuinness, Bjørn Olstad, Øystein Haug Olsen, and André O. Falcão, editors, *CIKM*, pages 41–50. ACM, 2007.

579. Fei Wu and Daniel S. Weld. Automatically refining the wikipedia infobox ontology. In Jinpeng Huai, Robin Chen, Hsiao-Wuen Hon, Yunhao Liu, Wei-Ying Ma, Andrew Tomkins, and Xiaodong Zhang, editors, *WWW*, pages 635–644. ACM, 2008.

580. Fei Wu and Daniel S. Weld. Open information extraction using wikipedia. In Hajic et al. [282], pages 118–127.

581. Wensheng Wu, Clement T. Yu, AnHai Doan, and Weiyi Meng. An interactive clustering-based approach to integrating source query interfaces on the deep web. In *SIGMOD Conference*, pages 95–106, 2004.

582. Chuan Xiao, Wei Wang, Xuemin Lin, and Jeffrey Xu Yu. Efficient similarity joins for near duplicate detection. In *WWW*, pages 131–140, 2008.

583. Dong Xin, Yeye He, and Venkatesh Ganti. Keyword++: A framework to improve keyword search over entity databases. *PVLDB*, 3(1), 2010.

584. Khaled Yagoub, Daniela Florescu, Valerie Issarny, and Patrick Valduriez. Caching strategies for data-intensive web sites. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, pages 188–199, Cairo, Egypt, 2000.

585. Beverly Yang and Hector Garcia-Molina. Improving search in peer-to-peer networks. In *ICDCS*, pages 5–14, 2002.

586. H. Z. Yang and P. A. Larson. Query transformation for PSJ-queries. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, pages 245–254, Brighton, England, 1987.

587. Jeffrey Xu Yu, Qin Lu, and Lijun Chang. *Keyword Search in Databases*. Synthesis Lectures on Data Management. 2010.

588. Markos Zaharioudakis, Roberta Cochrane, George Lapis, Hamid Pirahesh, and Monica Urata. Answering complex SQL queries using automatic summary tables. In *Proceedings of the ACM SIGMOD Conference*, pages 105–116, 2000.

589. Y. Zhai and B. Liu. Web data extraction based on partial tree alignment. In *WWW*, 2005.

590. Y. Zhang, N. Tang, and P. A. Boncz. Efficient distribution of full-fledged XQuery. In *Engineering*, pages 565–576, April 2009.

591. Jun Zhao, Satya Sanket Sahoo, Paolo Missier, Amit P. Sheth, and Carole A. Goble. Extending semantic provenance into the web of data. *IEEE Internet Computing*, 15(1), 2011.

592. Gang Zhou, Richard Hull, Roger King, and Jean-Claude Franchitti. Data integration and warehousing using h2o. *IEEE Data Eng. Bull.*, 18(2), 1995.

593. Wenchao Zhou, Qiong Fei, Arjun Narayan, Andreas Haeberlen, Boon Thau Loo, and Micah Sherr. Secure network provenance. In *SOSP*, 2011.

594. Wenchao Zhou, Qiong Fei, Shengzhi Sun, Tao Tao, Andreas Haeberlen, Zachary G. Ives, Boon Thau Loo, and Micah Sherr. NetTrails: A declarative platform for maintaining and querying provenance in distributed systems. In *SIGMOD Conference*, 2011.

595. Wenchao Zhou, Micah Sherr, Tao Tao, Xiaozhou Li, Boon Thau Loo, and Yun Mao. Efficient querying and maintenance of network provenance at internet-scale. In *SIGMOD*, 2010.

596. Yue Zhuge, Hector Garcia-Molina, and Janet L. Wiener. Multiple view consistency for data warehousing. In *ICDE*, 1997.