References

[1] http://web.media.mit.edu/ push/omcs-research.html.

[2] http://www.openmind.org/lectureslides.html.

[3] Serge Abiteboul, Rakesh Agrawal, Philip A. Bernstein, Mikie Carey, Stefano Ceri, Bruce Croft, David J. DeWitt, Michael J. Franklin, Hector Garcia-Molina, Dieter Gawlick, Jim Gray, Laura Haas, Alon Halevy, Joseph M. Hellerstein, Yannis Ioannidis, Martin Kersten, Michael Pazzani, Michael Lesk, David Maier, Jeffrey F. Naughton, Hans Schek, Timos Sellis, Avi Silberschatz, Michael Stonebraker, Rick Snodgrass, Jeffrey D. Ullman, Gerhard Weikum, Jennifer Widom, and Stan Zdonik. The lowell database research self assessment. *http://research.microsoft.com/ Gray/Lowell/*, 2003.

[4] Rakesh Agrawal, Pedro Domingos, and Matthew Richardson. Trust management for the semantic web. In *Proc. of the Int. Semantic Web Conf.*, 2003.

[5] Rafi Ahmed, Philippe De Smedt, Weimin Du, William Kent, Mohammad A. Ketabchi, Witold Litwin, Abbas Rafii, and Ming-Chien Shan. The pegasus heterogeneous multidatabase system. *IEEE Computer*, 24(12):19–27, 1991.

[6] J. Ambite and C. Knoblock. Flexible and scalable query planning in distributed and heterogeneous environments. In *Proc. of the 4th Int. Conf. on AI Planning Systems (AIPS)*, 1998.

[7] R. Ananthakrishna, S. Chaudhuri, and V. Ganti. Eliminating fuzzy duplicates in data warehouses. In *Proc. of 28th Int. Conf. on Very Large Databases*, 2002.

[8] N. Ashish and C. Knoblock. Wrapper Generation for Semi-structured Information Sources. In *Proc. ACM SIGMOD Workshop on Management of Semi-structured Data*, 1997.

[9] R. Avnur and J. Hellerstein. Continuous query optimization. In *SIGMOD '00*, 2000.

[10] C. Batini, M. Lenzerini, and SB. Navathe. A comparative analysis of methodologies for database schema integration. *ACM Computing Survey*, 18(4):323–364, 1986.

[11] J. Berlin and A. Motro. Autoplex: Automated discovery of content for virtual databases. In *Proceedings of the Conf. on Cooperative Information Systems (CoopIS)*, 2001.

[12] J. Berlin and A. Motro. Database schema matching using machine learning with feature selection. In *Proceedings of the Conf. on Advanced Information Systems Engineering (CAiSE)*, 2002.

[13] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 279, 2001.

[14] P. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu. Data management for peer-to-peer computing: A vision. In *Workshop on the Web and Databases, WebDB*, 2002.

[15] Philip Bernstein, Umeshwar Dayal, David J. DeWitt, Dieter Gawlick, Jim Gray, Matthias Jarke, Bruce G. Lindsay, Peter C. Lockemann, David Maier, Erich J. Neuhold, Andreas Reuter, Lawrence A. Rowe, Hans-Jörg Schek, Joachim W. Schmidt, Michael Schrefl, and Michael Stonebraker. Future directions in dbms research - the laguna beach participants. *SIGMOD Record*, 18(1):17–26, 1989.

[16] Philip A. Bernstein, Michael L. Brodie, Stefano Ceri, David J. DeWitt, Michael J. Franklin, Hector Garcia-Molina, Jim Gray, Gerald Held, Joseph M. Hellerstein, H. V. Jagadish, Michael Lesk, David Maier, Jeffrey F. Naughton, Hamid Pirahesh, Michael Stonebraker, and Jeffrey D. Ullman. The asilomar report on database research. *SIGMOD Record*, 27(4):74–80, 1998.

[17] M. Bilenko and R. Mooney. Learning to combine trained distance metrics for duplicate detection in databases. Technical Report Technical Report AI 02-296, Artificial Intelligence Laboratory, University of Texas at Austin, Austin, TX, February 2002.

[18] A. Brown, G. Kar, and A. Keller. An active approach to characterizing dynamic dependencies for problem determination in a distributed environment. In *Proceedings of the Seventh IFIP/IEEE International Symposium on Integrated Network Management*, 2001.

[19] H. Chalupsky. Ontomorph: A translation system for symbolic knowledge. In *Principles of Knowledge Representation and Reasoning*, 2000.

[20] K. Chang, B. He, C. Li, and Z. Zhang. Structured databases on the web: Observations and implications. Technical Report UIUCDCS-R-2003-2321, 2003.

[21] Kevin Chen-Chuan Chang, Bin He, Chengkai Li, and Zhen Zhang. The uiuc deep web repository. Accessible at http://eagle.cs.uiuc.edu/MetaQuerier, December 2002.

[22] Surajit Chaudhuri and Umeshwar Dayal. An overview of data warehousing and olap technology. *SIGMOD Record*, 26(1):65–74, 1997.

[23] Surajit Chaudhuri and Gerhard Weikum. Rethinking database system architecture: Towards a self-tuning risc-style database system. In *VLDB 2000, Proc. of 26th Int. Conference on Very Large Data Bases*, 2000.

[24] J. Chen, D. DeWitt, F. Tian, and Y. Wang. Niagaracq: A scalable continuous query system for internet databases. In *SIGMOD '00*, 2000.

[25] M. Chen, E. Kiciman, E. Fratkin, E. Brewer, and A. Fox. Pinpoint: Problem determination in large, dynamic, internet services. In *Proceedings of the International Conference on Dependable Systems and Networks*, 2002.

[26] W. Cohen. Integration of heterogeneous databases without common domains using queries based on textual similarity. In *Procceedings of SIGMOD-98*, 1998.

[27] W. Cohen. Some practical observations on integration of web information. In *Proc. of the WebDB-99 Workshop*, 1999.

[28] W. Cohen, M. Hurst, and L. Jensen. A flexible learning system for wrapping tables and lists in html documents. In *Proc. of the Int. World-Wide Web Conf. (WWW)*, 2002.

[29] Illinois Fire Services Blue Ribbon Committee. Recommendations for the legistrative needs and funding for the fire services of illinois, 2002.

[30] http://www.cra.org/reports/gc.systems.pdf Computing Research Association. Grand research challenges in information systems.

[31] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery. Learning to construct knowledge bases from the World Wide Web. *Artificial Intelligence*, 118(1-2):69–113, 2000.

[32] V. Crescenzi, G. Mecca, and P. Merialdo. Roadrunner: Towards automatic data extraction from large web sites. In *The VLDB Journal*, pages 109–118, 2001.

[33] Neil Daswani, Hector Garcia-Molina, and Beverly Yang. Problems in data-sharing peer-to-peer systems. In *Int. Conf. on Database Theory*, 2003.

[34] H. Do, S. Melnik, and E. Rahm. Comparison of schema matching evaluations. In *Proceedings of the 2nd Int. Workshop on Web Databases (German Informatics Society)*, 2002.

[35] H. Do and E. Rahm. Coma: A system for flexible combination of schema matching approaches. In *Proceedings of the 28th Conf. on Very Large Databases (VLDB)*, 2002.

[36] A. Doan. Modeling probabilistic actions for practical decision-theoretic planning. In *Proc. of the 3rd Int. Conference on AI Planning Systems (AIPS)*, 1996.

[37] A. Doan. Learning to map between structured representations of data, 2002. PhD thesis, University of Washington, http://anhai.cs.uiuc.edu/home/thesis.html.

[38] A. Doan, P. Domingos, and A. Halevy. Reconciling Schemas of Disparate Data Sources: A Machine Learning Approach. In *Proceedings of the ACM SIGMOD Conference*, 2001.

[39] A. Doan, P. Domingos, and A. Halevy. Learning to match the schemas of databases: A multistrategy approach. *Machine Learning Journal*, 50:279–301, 2003.

[40] A. Doan, P. Domingos, and A. Levy. Data integration: A killer app for multi-strategy learning. In *Proc. of the Workshop on Multi-Strategy Learning (MSL-00)*, 2000.

[41] A. Doan, P. Domingos, and A. Levy. Learning source descriptions for data integration. In *Proceedings of the Third International Workshop on Web and Database*, pages 81–86, 2000.

[42] A. Doan and P. Haddawy. Sound abstraction of probabilistic actions in the constraint mass assigment framework. In *Proc. of the 12th Nat. Conference on Uncertainty in AI (UAI)*, 1996.

[43] A. Doan, P. Haddawy, and C. Kahn. Decision-theoretic refinement planning: A new method for clinical decision analysis. In *Proc. of the 19th AMIA Annual Symposium on Computer Applications in Medical Care (SCAMC)*, 1995.

[44] A. Doan, P. Haddawy, and C. Kahn. Decision-theoretic planning for clinical decision analysis. In *Proc. of the Annual AI in Medicine Spring Symposium*, 1996.

[45] A. Doan and A. Halevy. Efficiently ordering query plans for data integration. In *Proc. of the 18th IEEE Int. Conference on Data Engineering (ICDE)*, 2002.

[46] A. Doan and A. Levy. Efficiently ordering query plans for data integration. In *Proc. of the IJCAI-99 Workshop on Intelligent Information Integration*, 1999.

[47] A. Doan, Y. Lu, Y. Lee, and J. Han. Object matching for data integration: A profile-based approach. In *Proc. of the IJCAI-03 Workshop on Information Integration on the Web, www.isi.edu/info-agents/workshops/ijcai03/proceedings.htm*, 2003.

[48] A. Doan, Y. Lu, Y. Lee, and J. Han. Object matching for data integration: A profile-based approach. In *IEEE Intelligent Systems, Special Issue on Information Integration*, 2003. To appear.

[49] A. Doan, J. Madhavan, R. Dhamankar, P. Domingos, and A. Halevy. Learning to match ontologies on the Semantic Web. *VLDB Journal*, 2003. To appear.

[50] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map ontologies on the Semantic Web. In *Proceedings of the World-Wide Web Conference (WWW-02)*, 2002.

[51] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Ontology matching: A machine learning approach. In S. Staab and R. Studer, editors, *Handbook on Ontologies in Information Systems*. Springer-Velag, 2003.

[52] A. Doan and R. McCann. Building data integration systems: a mass collaboration approach. In *Proc. of the IJCAI-03 Workshop on Information Integration on the Web*, 2003.

[53] O. Duschka. Query optimization using local completeness. In *Proc. of the 14th Nat. Conf. on AI*, 1997.

[54] O. Duschka and M. Genesereth. Answering recursive queries using views. In *Proc. of PODS '97*.

[55] Ahmed Elmagarmid (Editor), Marek Rusinkiewicz (Editor), and Amit Sheth (Editor). *Management of Heterogeneous and Autonomous Database Systems*. Morgan Kaufmann, 1999. First Edition.

[56] AK. Elmagarmid and C. Pu. Guest editors' introduction to the special issue on heterogeneous databases. *ACM Computing Survey*, 22(3):175–178, 1990.

[57] D. Embley, D. Jackman, and L. Xu. Multifaceted exploitation of metadata for attribute match discovery in information integration. In *Proc. of the WIIW-01*, 2001.

[58] D. Embley, Y. Jiang, and Y. Ng. Record-boundary discovery in Web documents. In *Proc. of the ACM SIGMOD Conf.*, 1999.

[59] Oren Etzioni, Alon Halevy, AnHai Doan, Zachary Ives, Jayant Madhavan, Luke McDowell, and Igor Tatarinov. Crossing the structure chasm. In *Conf. on Innovative Database Research*, 2003.

[60] D. Florescu, D. Koller, and A. Y. Levy. Using probabilistic information in data integration. In *Proc. of VLDB '97*.

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

[62] D. Freitag. Multistrategy learning for information extraction. In *Proc. 15th Int. Conf. on Machine Learning (ICML-98)*, 1998.

[63] M. Friedman and D. Weld. Efficiently executing information-gathering plans. In *Proc. of the Int. Joint Conf. of AI (IJCAI)*, 1997.

[64] H. Galhardas, D. Florescu, D. Shasha, and E. Simon. An extensible framework for data cleaning. In *Proc. of 16th Int. Conf. on Data Engineering*, 2000.

[65] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. Ullman, and J. Widom. The TSIMMIS project: Integration of heterogeneous information sources. *Journal of Intelligent Inf. Systems*, 8(2), 1997.

[66] Michael R. Genesereth, Arthur M. Keller, and Oliver M. Duschka. Infomaster: An information integration system. In *Proceedings of the 1997 ACM SIGMOD Conference*, Tucson, Ariz., 1997. ACM Press, New York.

[67] Andrew R. Golding and Dan Roth. A winnow-based approach to context-sensitive spelling correction. *Machine Learning*, 34(1-3):107–130, 1999.

[68] L. Gravano, P. Ipeirotis, N. Koudas, and D. Srivastava. Text join for data cleansing and integration in an rdbms. In *Proc. of 19th Int. Conf. on Data Engineering*, 2003.

[69] Steven Gribble, Alon Halevy, Zachary Ives, Maya Rodrig, and Dan Suciu. What can databases do for peer-to-peer? In *WebDB Workshop on Databases and the Web*, 2001.

[70] V. Ha, A. Doan, V. Vu, and P. Haddawy. Geometric foundations for interval-based probabilities. *Annals of Mathematics and Artificial Intelligence*, 24, 1998.

[71] Laura M. Haas, Donald Kossmann, Edward L. Wimmers, and Jun Yang. Optimizing queries across diverse data sources. In *Proc. of VLDB '97*, 1997.

[72] P. Haddawy and A. Doan. Abstracting probabilistic actions. In *Proc. of the 10th Conference on Uncertainty in AI (UAI)*, 1994.

[73] P. Haddawy, A. Doan, and R. Goodwin. Efficient decision theoretic planning: Techniques and empirical analysis. In *Proc. of the Nat. Conf. on Uncertainty in AI (UAI)*, 1995.

[74] P. Haddawy, A. Doan, and R. Goodwin. Efficient decision-theoretic planning: Techniques and empirical analysis. In *Proc. of the 11th Nat. Conference on Uncertainty in AI (UAI)*, 1995.

[75] P. Haddawy, A. Doan, and C.E. Kahn. Decision-theoretic refinement planning in medical decision making: Management of acute deep venous thrombosis. *Journal of Medical Decision Making*, 1996.

[76] Alon Halevy. Data integration: A status report. Invited talk at the German Database Conference (BTW), Leipzig, Germany, http://www.cs.washington.edu/homes/alon/files/btw2003.ppt, 2003.

[77] B. He and K. Chang. Statistical schema matching across web query interfaces. In *Proc. of the ACM SIGMOD Conf. (SIGMOD)*, 2003.

[78] M. Hernandez and S. Stolfo. The merge/purge problem for large databases. In *SIGMOD Conference*, pages 127–138, 1995.

[79] http://research.microsoft.com/research/DMX/autoadmin. Microsoft research - autoadmin project.

[80] http://roc.cs.berkeley.edu. The recovery-oriented computing project.

[81] http://www.research.ibm.com/autonomic. Ibm research - autonomic computing.

[82] Z. Ives, D. Florescu, M. Friedman, A. Levy, and D. Weld. An adaptive query execution system for data integration. In *Proc. of SIGMOD*, 1999.

[83] F. Jelinek. *Statistical Methods for Speech Recognition*. MIT Press, 1997.

[84] Barry W. Johnson. *Design and Analysis of Fault Tolerant Digital Systems*. Addison Wesley, 1989.

[85] C. Kahn, A. Doan, and P. Haddawy. Management of acute deep venous thrombosis of the lower extremities (abstract). In *American Roentgen Ray Society Meeting*, 1996.

[86] Olga Kapitskaia, Anthony Tomasic, and Patrick Valduriez. Dealing with discrepancies in wrapper functionality. Technical Report RR-3138, INRIA, 1997.

[87] C. Knoblock, S. Minton, J. Ambite, N. Ashish, P. Modi, I. Muslea, A. Philpot, and S. Tejada. Modeling web sources for information integration. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 1998.

[88] Craig Knoblock and Subbarao Kambhampati. Tutorial on information integration on the web. the Eighteenth National Conference on Artificial Intelligence, http://www.isi.edu/info-agents/courses/aaai02/aaai-i3-tut-all.pdf, 2002.

[89] N. Kushmerick. Wrapper verification. *World Wide Web Journal*, 3(2):79–94, 2000.

[90] N. Kushmerick, D. Weld, and R. Doorenbos. Wrapper Induction for Information Extraction. In *Proc. of IJCAI-97*, 1997.

[91] John Lafferty and Chengxiang Zhai. Document language models, query models, and risk minimization for information retrieval. In *Proceedings of SIGIR'2001*, pages 111–119, Sept 2001.

[92] E. Lambrecht, S. Kambhampati, and S. Gnanaprakasam. Optimizing recursive information gathering plans. In *Proc. of the Int. Joint Conf. on AI (IJCAI)*, 1999.

[93] S. Lawrence, K. Bollacker, and C. Lee Giles. Autonomous citation matching. In *Proc. of the 3rd Int. Conf. on Autonomous Agents*, 1999.

[94] Steve Lawrence, Kurt Bollacker, and C. Lee Giles. Distributed error correction. In *Digital Libraries 99 - The Fourth ACM Conference on Digital Libraries*, New York, 1999. ACM Press.

[95] M. Lenzerini. Data integration; a theoretical perspective. In *Proc. of PODS-02*, 2002.

[96] Kristina Lerman, Steven Minton, and Craig A. Knoblock. Wrapper maintenance: a machine learning approach. *Journal of Artificial Intelligence Research*, 2003. To appear.

[97] U. Leser and F. Naumann. Query planning with information quality bounds. In *Proc. of the Int. Conf. on Flexible Query Answering Systems (FQAS)*, 2000.

[98] A. Y. Levy, A. Rajaraman, and J. Ordille. Querying heterogeneous information sources using source descriptions. In *Proc. of VLDB*, 1996.

[99] W. Li and C. Clifton. SEMINT: A tool for identifying attribute correspondence in heterogeneous databases using neural networks. *Data and Knowledge Engineering*, 33:49–84, 2000.

[100] Sam Lightstone, Guy M. Lohman, and Daniel C. Zilio. Toward autonomic computing with db2 universal database. *SIGMOD Record*, 31(3):55–61, 2002.

[101] Guy M. Lohman and Sam Lightstone. Smart: Making db2 (more) autonomic. In *the VLDB Conf.*, 2002.

[102] J. Madhavan, P. Bernstein, K. Chen, A. Halevy, and P. Shenoy. Matching schemas by learning from a schema corpus. In *Proc. of the IJCAI-03 Workshop on Information Integration on the Web*, 2003.

[103] J. Madhavan, P.A. Bernstein, and E. Rahm. Generic schema matching with cupid. In *Proceedings of the International Conference on Very Large Databases (VLDB)*, 2001.

[104] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, US, 1999.

[105] A. McCallum, K. Nigam, and L. Ungar. Efficient clustering of high-dimensional data sets with application to reference matching. In *Proc. 6th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2000.

[106] R. McCann, A. Doan, A. Kramnik, and V. Varadarajan. Building data integration systems via mass collaboration. In *Proc. of the SIGMOD-03 Workshop on the Web and Databases (WebDB-03)*, 2003.

[107] D. McGuinness, R. Fikes, J. Rice, and S. Wilder. The Chimaera Ontology Environment. In *Proceedings of the 17th National Conference on Artificial Intelligence*, 2000.

[108] S. Melnik, H. Molina-Garcia, and E. Rahm. Similarity Flooding: A Versatile Graph Matching Algorithm. In *Proceedings of the International Conference on Data Engineering (ICDE)*, 2002.

[109] G. Miller. WordNet: an on-line lexical database. *International Journal of Lexicography*, 3(4):235–312, 1991.

[110] T. Milo and S. Zohar. Using schema matching to simplify heterogeneous data translation. In *Proc. of VLDB*, 1998.

[111] P. Mitra, G. Wiederhold, and J. Jannink. Semi-automatic Integration of Knowledge Sources. In *Proceedings of Fusion'99*, 1999.

[112] U. Nahm and R. Mooney. A mutually beneficial integration of data mining and information extraction. In *Proc. of the National AI Conf. (AAAI)*, 2000.

[113] Srini Narayanan and Sheila A. McIlraith. Simulation, verification and automated composition of web services. In *the Eleventh Int. World Wide Web Conference (WWW-11)*, 2002.

[114] Felix Naumann, Ulf Leser, and Johann C. Freytag. Quality-driven integration of heterogeneous information systems. In *Proc. of VLDB '99*, 1999.

[115] F. Neumann, CT. Ho, X. Tian, L. Haas, and N. Meggido. Attribute classification using feature analysis. In *Proceedings of the Int. Conf. on Data Engineering (ICDE)*, 2002.

[116] Zaiqing Nie and Subbarao Kambhampati. Joint optimization of cost and coverage of information gathering plans. Technical report, Dept. of CSE, Arizona State Univ., 2001. http://rakaposhi.eas.asu.edu/ParPlan.pdf.

[117] K. Nigam. Using unlabeled data to improve text classification. Ph.d. thesis, Carnegie-Mellon University, School of Computer Science, 2001.

[118] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Learning to classify text from labeled and unlabeled documents. In *Proc. of the Nat. Conf. on AI (AAAI)*, 1998.

[119] N.F. Noy and M.A. Musen. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2000.

[120] NF. Noy and MA. Musen. PromptDiff: A fixed-point algorithm for comparing ontology versions. In *Proceedings of the Nat. Conf. on Artificial Intelligence (AAAI)*, 2002.

[121] L. Palopoli, D. Sacca, and D. Ursino. Semi-automatic, semantic discovery of properties from database schemes. In *Proc. of the Int. Database Engineering and Applications Symposium (IDEAS-98)*, pages 244–253, 1998.

[122] C. Parent and S. Spaccapietra. Issues and approaches of database integration. *Communications of the ACM*, 41(5):166–178, 1998.

[123] Qujiang Peng, Takeshi Ito, and Teiji Furugori. Word sense disambiguation with a corpus-based semantic network. In *Proc. of the Sixth Natural Language Processing Pacific Rim Symposium*, 2001.

[124] J. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *Proceedings of the ACM SIGIR*, pages 275–281, 1998.

[125] Rachel Pottinger and Alon Y. Levy. A scalable algorithm for answering queries using views. In *Proc. of VLDB '00*, 2000.

[126] E. Rahm and P.A. Bernstein. On matching schemas automatically. *VLDB Journal*, 10(4), 2001.

[127] R. Ramakrishnan. Mass collaboration and data mining, 2001. Keynote address, KDD-01, www.cs.wisc.edu/ raghu/Kddrev.ppt.

[128] V. Raman and J. Hellerstein. Potter's wheel: An interactive data cleaning system. In *The VLDB Journal*, pages 381–390, 2001.

[129] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: Collaborative filtering of netnews. In *Proc. of CSCW-94*, 1994.

[130] the Illinois Fire Service Institute Richard Jaehne, Director. Personal communication, 2003.

[131] M. Richardson and P. Domingos. Building large knowledge bases by mass collaboration. Technical Report UW-TR-03-02-04, Dept. of CSE, Univ. of Washington, 2003.

[132] S. Sarawagi and A. Bhamidipaty. Interactive deduplication using active learning. In *Proc. of 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2002.

[133] R. Schvaneveldt. *PathFinder Associative Networks: Studies in Knowledge Organization*. Albex, 1990.

[134] AP. Seth and JA. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Survey*, 22(3):183–236, 1990.

[135] Abraham Silberschatz and Stanley B. Zdonik et. al. Strategic directions in database systems - breaking out of the box. *ACM Computing Surveys*, 28(4):764–778, 1996.

[136] Abraham Silberschatz, Michael Stonebraker, and Jeffrey D. Ullman. Database systems: Achievements and opportunities. *CACM*, 34(10):110–120, 1991.

[137] Abraham Silberschatz, Michael Stonebraker, and Jeffrey D. Ullman. Database research: Achievements and opportunities into the 21st century. *SIGMOD Record*, 25(1):52–63, 1996.

[138] John Miles Smith, Philip A. Bernstein, Umeshwar Dayal, Nathan Goodman, Terry Landers, Ken W.T. Lin, and Eugene Wong. Multibase – integrating heterogeneous distributed database system. In *AFIPS National Computer Conf.*, 1981.

[139] 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 Journal: Very Large Data Bases*, 5(1):48–63, 1996.

[140] S. Tejada, C. Knoblock, and S. Minton. Learning domain-independent string transformation weights for high accuracy object identification. In *Proc. of the 8th SIGKDD Int. Conf. (KDD-2002)*, 2002.

[141] Jeffrey D. Ullman. Information integration using logical views. In *Proceedings of the 6th International Conference on Database Theory*, Delphi, Greece, January 1997. Springer, Berlin.

[142] Gerhard Weikum, Axel Mnkeberg, Christof Hasse, and Peter Zabback. Self-tuning database technology and information services: from wishful thinking to viable engineering. In *the VLDB Conf.*, 2002.

[143] G Wiederhold. Mediators in the architecture of future information systems. *IEEE Computer*, pages 38–49, March 1992.

[144] www.ncbi.nlm.nih.gov/Entrez. Entrez search and retrieval system.

[145] Extensible markup language (XML) 1.0. www.w3.org/TR/1998/REC-xml-19980210, 1998. W3C Recommendation.

[146] R. Yerneni, F. Naumann, and H. Garcia-Molina. Maximizing coverage of mediated web queries. Technical Report www-db.stanford.edu/ yerneni/pubs/mcmwq.ps, Stanford University, 2000.

[147] Ramana Yerneni, Yannis Papakonstantinou, and Hector Garcia-Molina. Fusion queries over internet databases. In *Proc. of the 6th Int. Conf. on Extending Database Technology (EDBT)*, 1998.

[148] W. Yih and D. Roth. Probabilistic reasoning for entity and relation recognition. In *Proc. of COLING'02*, 2002.

[149] H. Yu, A. Doan, and J. Han. Mining for information discovery on the web: Overview and illustrative research. In N. Zhong, editor, *Intelligent Technologies - Advances in Agents, Data Mining, and Learning*. Springer-Velag, 2003. To appear.

[150] H. Yu, J. Han, and K. Chang. Pebl: Positive example based learning for web page classification using svm. In *Proc. of the Conf. on Knowledge Discovery and Data Mining (KDD)*, 2002.

[151] Chengxiang Zhai and John Lafferty. Two-stage language models for information retrieval. In *Proceedings of SIGIR'2002*, pages 49–56, Aug 2002.