

---

## Curriculum Vitae

### Min Wang

Director, HP Labs China  
HP Distinguished Technologist  
Hewlett Packard Company

Website: [http://www.hpl.hp.com/people/min\\_wang6](http://www.hpl.hp.com/people/min_wang6)

---

### Research Interests

- Query Processing and Optimization for Relational, XML, and Streaming Data
- Information Integration
- Information Retrieval and Analysis
- Electronic Commerce
- Data Mining

### Education

- **Ph.D. Computer Science, Duke University**, Durham, North Carolina, 1999  
*Dissertation: Approximation and Learning Techniques in Database Systems*  
*Advisor: Professor Jeffrey Scott Vitter*
- **M.S. Computer Science, Duke University**, Durham, North Carolina
- **M.S. Computer Science, Tsinghua University**, Beijing, P.R.China
- **B.S. Computer Science, Tsinghua University**, Beijing, P.R.China

### Other Training

- **Leadership Development Program, LEAP** (Leadership Education for Asian Pacifics, Inc.), San Jose, California, May 2008.
- **From Specialist to Strategist: Business Excellence for Women in Science, Technology and Engineering**, Smith Executive Education for Women, Smith College, Northampton, Massachusetts, June 2006.
- **Micro MBA**, IBM Research, Yorktown Heights, New York, February 2005-July 2005.

### Employment History

- **Distinguished Technologist**, Hewlett Packard Company, 7/2011-present.
- **Director**, HP Labs China, HP Labs, Beijing, China, 1/2010-present.
- **Research Staff Member and Manager**, Unified Data Analytics Department, IBM T.J. Watson Research Center, Hawthorne, New York, 9/2008-12/2009.
- **Research Staff Member and Manager**, Event-Based Systems Department, IBM T.J. Watson Research Center, Hawthorne, New York, 11/2006-9/2008.
- **Research Staff Member and Technical Assistant to Dr. Stu Feldman, Vice President of Computer Science**, IBM T.J. Watson Research Center, Hawthorne, New York, 7/2005-10/2006.
- **Research Staff Member**, Database Research Group, IBM T.J. Watson Research Center, Hawthorne, New York, 10/1999-6/2005.
- **Research Assistant**, Department of Computer Science, Duke University, Durham, North Carolina, 8/1994-8/1999.
- **Summer Intern**, DB2/MVS Optimizer Development, IBM Santa Teresa Laboratory (now IBM Silicon Valley Laboratory), San Jose, California, 5/1997-9/1997.
- **Intern**, Database Technology Institute, IBM Santa Teresa Laboratory (now IBM Silicon Valley Laboratory), San Jose, California, 5/1996-4/1997.
- **Summer Intern**, DB2/MVS Optimizer Development, IBM Santa Teresa Laboratory (now IBM Silicon Valley Laboratory), San Jose, California, 5/1995-8/1995.
- **Lecturer**, Department of Computer Science, Tsinghua University, Beijing China, 8/1990

## Teaching and Related Experiences

- **Member of the Advisory Committee**, Department of Computer Science, The University of Hong Kong, Feb. 2010-present.
- **Campus Relationship Manager**, IBM and Duke University, Jan. 2007-Dec. 2009.
- **Research Mentor**, IBM T. J. Watson Research Center. Mentored multiple summer interns from Duke University, UIUC, George Mason University, University of Toronto, and Florida State University to conducted research in database systems and information management. The work with the students resulted in multiple publications in top conferences (e.g., SIGMOD and VLDB). 2000—2009.
- **Ph.D. Committee Member** for Lipyeow Lim, Duke University, August, 2004. The major part of Lipyeow Lim’s dissertation work was done when he was an intern at IBM T. J. Watson Research Center under my supervision during the summer of 2000, 2001, and 2003.
- **Teaching Assistant**, Department of Computer Science, Duke University. Responsible for the lab sessions and grading the homework and exams for a class of 100+ sophomores for the course “Program Design and Analysis II”, 1996.
- **Teaching Assistant**, Department of Computer Science, Duke University. Responsible for the lab sessions and grading the homework and exams for a class of 100+ freshmen for the course “Introduction to Programming Design and Analysis”, 1995.
- **Lecturer**, Department of Computer Science, Tsinghua University, Beijing China, Conducted research in cryptography, supervised several seniors on their senior projects, and taught the course “mathematical logic” to a class of 20+ juniors for two semesters.

## Honors and Awards

- **2009 SIGMOD Test of Time Award** for the paper from the 1999 SIGMOD Conference that has had the most impact (research, products, methodology) over the intervening decade, ACM SIGMOD/PODS 2009 Conference, June 29-July 2, 2009, Providence, Rhode Island, USA.  
*Paper*: Approximate Computation of Multidimensional Aggregates of Sparse Data Using Wavelets (with Jeff Vitter).
- **ER 2008 Conference Best Paper Award (DKE 25Year Award)** (given by the Elsevier), the 27th International Conference on Conceptual Modeling (ER 2008), Oct. 2008, Barcelona, Spain.  
*Paper*: Modeling and Querying E-Commerce Data in Hybrid Relational-XML DBMSs (with Lipyeow Lim and Haixun Wang).
- **CIKM 2010 Best Paper Runner-up**, the 18th ACM Conference on Information and Knowledge Management (CIKM 2009), Hong Kong, China, November 2-6, 2009.  
*Paper*: Provenance Query Evaluation: What's So Special about It? (A. Kementsietsidis and M. Wang).
- **The Eighth Plateau Invention Achievement Award**, IBM, July 2009. (The previous seven Plateau Invention Achievement Awards were awarded by IBM during 2001-2008.)
- **Outstanding Technical Achievement Award (OTAA)** for contribution to DB2XML, Jan 2007, IBM.
- **Outstanding Ph.D. Dissertation Award**, Department of Computer Science, Duke University, 1999-2000 academic year.  
*Dissertation*: Approximation and Learning Techniques in Database Systems
- **IBM Graduate Fellowship**, 1998-1999 academic year.
- **IBM Graduate Fellowship**, 1997-1998 academic year.
- **Duke University Research Assistantship**, 1995-1997.
- **Duke University Graduate Fellowship**, 1994-1995.

## Publications

### *Refereed Publications*

1. gSketch: On Query Estimation in Graph Streams (P. Zhao, C. Aggarwal, and M. Wang), *Proceedings of the 38th International Conference on Very Large Databases (VLDB 2012)*, Istanbul, Turkey, August 27-31, 2012.
2. Optimizing Statistical Information Extraction Programs Over Evolving Text (X. Feng, F. Chen, C. Ré, and M. Wang), *Proceedings of the 28<sup>th</sup> IEEE International Conference on Data Engineering (ICDE 2012)*, Washington, DC, USA, April 1-5, 2012.
3. Finding Relevant Information of Certain Types from Enterprise Data (X. Liu, H. Fang (C. Yao, and M. Wang), *Proceedings of the 20<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2011)*, Glasgow, Scotland, UK, October 24-28<sup>th</sup>, 2011.
4. AWETO: Efficient Incremental Update and Querying in RDF Storage System (X. Pu, J. Wang, P. Luo, and M. Wang), *Proceedings of the 20<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2011)*, Glasgow, Scotland, UK, October 24-28. Poster paper.
5. Search Result Diversification for Enterprise Data (W. Zhang, H. Fang, C. Yao, and Min Wang), *Proceedings of the 20<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2011)*, Glasgow, Scotland, UK, October 24-28. Poster paper.
6. Prominent Streak Discovery in Sequence Data (X. Jiang, C. Li, P. Luo, M. Wang, and Y. Yu), *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2011)*, San Diego, CA, USA, August 21-24, 2011, pp. 1280-1288.
7. Flexible Aggregate Similarity Search (Y. Li, F. Li, K. Yi, B. Yao, and M. Wang), *Proceedings of 30th ACM SIGMOD International Conference on Management of Data (SIGMOD 2011)*, Athens, Greece, June 12-16, 2011, pp. 1009-1020.
8. Rewriting Queries on SPARQL Views (W. Le, S. Duan, A. Kementsietsidis, F. Li, and M. Wang), *Proceedings of the 20<sup>th</sup> International World Wide Web Conference (WWW 2011)*, Hyderabad, India, March 28-April 1, 2011, pp. 655-664.
9. REACTOR: A Framework for Semantic Relation Extraction and Tagging over Enterprise Data (W. Shen, J. Wang, P. Luo, M. Wang, C. Yao), *Proceedings of the 20<sup>th</sup> International World Wide Web Conference (WWW 2011) Companion Volume*, Hyderabad, India, March 28-April 1, 2011, pp. 121-122. Poster paper.
10. Optimizing Content Freshness of Relations Extracted From the Web Using Keyword Search (M. Yang, H. Wang, L. Lim, and M. Wang), *Proceedings of the 29<sup>th</sup> ACM SIGMOD International Conference on Management of Data (SIGMOD 2010)*, Indianapolis, Indiana, USA, June 6-11, 2010, pp.819-830.
11. Profile-Based Retrieval of Records in Medical Databases (A. Kementsietsidis, L. Lim, and M. Wang), *Proceedings of American Medical Informatics Association Annual Symposium (AMIA 2009)*, San Francisco, CA, USA, November 14-18, 2009.
12. A Framework for Semantic Link Discovery over Relational Data (O. Hassanzadeh, A. Kementsietsidis, L. Lim, R. Miller, and M. Wang), *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, Hong Kong, China, November 2-6, 2009, pp. 1027-1036.
13. Provenance Query Evaluation: What's So Special about It? (A. Kementsietsidis and M. Wang), *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, Hong Kong, China, November 2-6, 2009, pp. 681-690.
14. Semantic Queries in Databases: Problems and Challenges (L. Lim, H. Wang, and M. Wang), *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, Hong Kong, China, November 2-6, 2009, pp. 1505-1508. Short paper.
15. Scalable Matching of Industry Models - A Case Study (B. Byrne, A. Fokoue, A. Kalyanpur, K. Srinivas, and M. Wang), *Proceedings of the Fourth International Workshop on Ontology Matching (OM-2009)*, Washington, DC, October 25, 2009.
16. Linkage Query Writer (O. Hassanzadeh, R. Xin, R. Miller, A. Kementsietsidis, L. Lim, and M. Wang), *Proceedings of the 35th International Conference on Very Large Data Bases (VLDB 2009)*, Lyon, France, August 24-28, 2009, pp.1590-1593.

- Demonstration.
17. [A Declarative Framework for Semantic Link Discovery over Relational Data](#) ( O. Hassanzadeh, L. Lim, A. Kementsietsidis, M. Wang), *Proceedings of the 18th International Conference on World Wide Web (WWW 2009)*, Madrid, Spain, April 20-24, 2009, pp.1101-1102. Poster paper.
  18. [On the Efficiency of Provenance Queries](#) (A. Kementsietsidis and M. Wang), *Proceedings of 25th International Conference on Data Engineering (ICDE 2009)*, Shanghai, China, March 29 - April 4, 2009, pp.1223-1226. Short paper.
  19. [Supporting Ontology-Based Keyword Search over Medical Databases](#) (A. Kementsietsidis, L. Lim, and M. Wang), *Proceedings of American Medical Informatics Association Annual Symposium (AMIA 2008)*, Washington, DC, November 8-12, 2008.
  20. [Modeling and Querying E-Commerce Data in Hybrid Relational-XML DBMSs](#) (L. Lim, H. Wang, and M. Wang), *Proceedings of 27th International Conference on Conceptual Modeling (ER 2008)*, Barcelona, Spain, October 20-24, 2008, pp.294-310. **Received the Best Paper Award.**
  21. [Advances and Challenges for Scalable Provenance in Stream Processing Systems](#) (A. Misra, M. Blount, A. Kementsietsidis, D. M. Sow, and M. Wang), *Proceedings of 2nd International Provenance and Annotation Workshop (IPAW 2008)*, Salt Lake City, UT, USA, June 17-18, 2008, pp.253-265.
  22. [Unifying Data and Domain Knowledge Using Virtual Views](#) (L. Lim, H. Wang, and M. Wang), *Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB 2007)*, Vienna, Austria, September 23-27, 2007, pp.255-266.
  23. [Century: Automated Aspects of Patient Care](#) (M. Blount, J. S. Davis, M. Ebling, J.H. Kim, K.H. Kim, K.Y. Lee, A.Misra, S.H. Park, D. M. Sow, Y.J. Tak, M. Wang, and K. Witting), *Proceedings of 13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2007)*, Daegu, Korea, August 21-24, 2007, pp.504-509.
  24. [Persisting and Querying Biometric Event Streams with Hybrid Relational-XML DBMS](#) (D. M. Sow, L. Lim, M. Wang, and K. H. Kim), *Proceedings of the 2007 Inaugural International Conference on Distributed Event-Based Systems (DEBS 2007)*, Toronto, Ontario, Canada, June 20-22, 2007, pp.189-197.
  25. [Supporting Ranking and Clustering as Generalized Order-By and Group-By](#) (C. Li, M. Wang, L. Lim, H. Wang, and K. Chang), *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2007)*, Beijing, China, June 12-14, 2007, pp.127-138.
  26. [A Time-and-Value Centric Provenance Model and Architecture for Medical Event Streams](#) (M. Wang, M. Blount, J. S. Davis, A. Misra, and D. M. Sow), *Proceedings of the 1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments (HealthNet 2007)*, San Juan, Puerto Rico, June 11, 2007, pp.95-100.
  27. [Semantic Data Management: Towards Querying Data with their Meaning](#) (L. Lim, H. Wang, and M. Wang), *Proceedings of 23rd International Conference on Data Engineering (ICDE 2007)*, Istanbul, Turkey , April 16-20, 2007, pp.1438-1442.
  28. [Efficient Update of Indexes for Dynamically Changing Web Documents](#) (R. C. Agarwal, L. Lim, M. Wang, S. Padmanabhan, and J. S. Vitter), *Journal of World Wide Web*, Vol. 10, No. 1, March 2007, pp.37-69.
  29. [Cost-Based Optimization in DB2XML](#) (A. Balmin, T. Eliaz, J. Hornibrook, L. Lim, G. Lohman, D. Simmen, M. Wang, and C. Zhang), *IBM Systems Journal*, Vol. 45, No. 2, 2006 pp.299-320.
  30. [Boolean + Ranking: Querying a Database By K-Constrained Optimization](#) (Z. Zhang, S. Hwang, K. Chang, M. Wang, C. Lang, and Y. C. Chang), *Proceedings of 2006 ACM SIGMOD International Conference on Management of Data (SIGMOD 2006)*, Chicago, Illinois, June 26-29, 2006, pp.359-370.
  31. [Finding the Plateau in an Aggregated Time Series](#) (M. Wang and X. S. Wang), *Proceedings of the 7th International Conference on Web-Age Information Management (WAIM 2006)*, Hong Kong, China, June 17-19, 2006, pp.325-336.
  32. [Managing E-Commerce Catalogs in a DBMS with Native XML Support](#) (L. Lim and M. Wang), *Proceedings of IEEE International Conference on e-Business*

- Engineering (ICEBE 2005)*, Beijing, China, October 18-20, 2005, pp.564-571.
33. [Evaluating Trigger Conditions on Streaming Time Series with User-given Quality Requirements](#) (L. Gao, M. Wang, and X. S. Wang ), *Journal of Universal Computer Science, Vol. 11, No. 8, 2005*, pp.1397-1410.
  34. [CXHist: An On-line Classification-Based Histogram for XML String Selectivity Estimation](#) (L. Lim, M. Wang, and J. S. Vitter), *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 2005)*, Trondheim, Norway, August 30-September 2, 2005, pp.1187-1198.
  35. [Quality Driven Evaluation of Trigger Conditions on Streaming Time Series](#)(L. Gao, M. Wang, and X. S. Wang), *Proceedings of the 20th ACM Symposium on Applied Computing -Special Track on Data Streams (SAC 2005)*, New Mexico, USA, March 14-17, 2005, pp.563-567.
  36. [Expressing and Optimizing Similarity-Based Queries in SQL](#) (L. Gao, M. Wang, X. S. Wang, and S. Padmanabhan), *Proceedings of the 3rd International Conference on Conceptual Modeling (ER2004)*, Shanghai, China, November 8-12, 2004, pp.464-478.
  37. [Modeling Autonomous Catalog for Electronic Commerce](#) (Y. C. Chang, V. R. Chillakuru, and M. Wang), *Proceedings of the 3rd International Conference on Conceptual Modeling (ER2004)*, Shanghai, China, November 8-12, 2004, pp.828-830.
  38. [Condition Evaluation for Speculative Systems: a Streaming Time Series Case](#) (X. S. Wang, L. Gao, and M. Wang), *Proceedings of the 2nd International Workshop on Spatio-Temporal Database Management (STDBM 2004)*, Toronto, Canada, August 30, 2004, pp.65-72.
  39. [Scalable Mining for Classification Rules in Relational Databases](#) (M. Wang, B. Iyer, and J. S. Vitter), *Institute of Mathematical Statistics Lecture Notes - Monograph Series, Vol. 45, 2004*, pp. 348-377.
  40. [SASH: A Self-Adaptive Histogram Set for Dynamically Changing Workloads](#) (L. Lim, M. Wang, and J. S. Vitter), *Proceedings of the 29th International Conference on Very Large Databases (VLDB 2003)*, Berlin, Germany, September 9-12, 2003, pp.369-380.
  41. [Efficient Evaluation of Composite Correlations for Streaming Time Series](#) (M. Wang and X. S. Wang), *Proceedings of the 4th International Conference on Web-Age Information Management (WAIM 2003)*, Chengdu, China, August 17-19, 2003, pp.369-380.
  42. [Epi-SPIRE: A Bio-Surveillance System for Environmental and Public Health Activity Monitoring](#) (C. S. Li, C. Aggarwal, M. Campbell, Y. C. Chang, G. Glass, V. S. Iyengar, M. Joshi, C. Y. Lin, M. J. R. Smith, B. L. Tseng, M. Wang, K. L. Wu, and P. S. Yu), *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME 2003)*, Baltimore, Maryland, July 6-9, 2003, pp.713-716.
  43. [A Learning-Based Approach to Estimate Statistics of Operators in Continuous Queries: a Case Study](#) ( L. Gao, M. Wang, X. S. Wang, and S. Padmanabhan), *Proceedings of the 8th ACM SIGMOD Workshop on Reseach Issues in Data Mining and Knowledge Discovery (DMKD 2003)*, San Diego, California, June 13, 2003.
  44. [Dynamic Maintenance of Web Indexes Using Landmarks](#) (L. Lim, M. Wang, S. Padmanabhan, J. S. Vitter, and R. Agarwal), *Proceedings of the 12th International World Wide Web Conference (WWW 2003)*, Budapest, Hungary, May 20-24, 2003, pp.102-111.
  45. [XPathLearner: An On-line Self-Tuning Markov Histogram for XML Path Selectivity Estimation](#) (L. Lim, M. Wang, S. Padmanabhan, J. S. Vitter, and R. Parr), *Proceedings of the 28th International Conference on Very Large Databases (VLDB 2002)*, Hong Kong, China, August 20-23, 2002, pp.442-453.
  46. [Supporting Efficient Parametric Search of E-Commerce Data: a Loosely-Coupled Solution](#) (M. Wang, Y. C. Chang, and S. Padmanabhan), *Proceedings of the 8th Conference on Extending Database Technology (EDBT 2002)*, Prague, the Czech Republic, March 24-28, 2002, pp.409-426.
  47. [Optimizing Relational Store for E-Catalog Queries: a Data Mining Approach](#) (M. Wang and X. S. Wang), *Proceedings of 2002 ACM Symposium on Applied Computing -Special Track on WWW and E-business Applications (SAC 2002)*, Madrid, Spain, March 10-14, 2002, pp.1147-1152.
  48. [Cryptography and Relational Database Management Systems](#) (J. He and M. Wang), *Proceedings of the 5th International Database Engineering and Applications*



- Symposium (IDEAS 2001)*, Grenoble, France, July 16-18, 2001, pp.273-284.
49. Wavelet-Based Cost Estimation for Spatial Queries (M. Wang, J. S. Vitter, L. Lim, and S. Padmanabhan), *Proceedings of the 7th International Symposium on Spatial and Temporal Databases (SSTD 2001)*, Redondo Beach, CA, USA, July 12-15, 2001, pp. 175-196.
  50. Characterizing Web Document Change (L. Lim, M. Wang, S. Padmanabhan, J. S. Vitter and R. C. Agarwal), *Proceedings of the 2nd International Conference on Web-Age Information Management (WAIM 2001)*, Xi'an, China, July 9-11, 2001, pp.133-144.
  51. Dynamic Maintenance of Wavelet-Based Histograms (Y. Matias, J. S. Vitter, and M. Wang), *Proceedings of the 26th International Conference on Very Large Databases (VLDB 2000)*, Cairo, Egypt, September 10-14, 2000, pp.101-110.
  52. Approximate Computation of Multidimensional Aggregates of Sparse Data Using Wavelets (J. S. Vitter and M. Wang), *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data (SIGMOD 1999)*, Philadelphia, June 1999, pp.193-204. **Received SIGMOD 2009 Test of Time Award.**
  53. Data Cube Approximation and Histograms via Wavelets (J. S. Vitter, M. Wang, and B. Iyer), *Proceedings of Seventh International Conference on Information and Knowledge Management (CIKM 1998)*, Washington D.C., November 1998, pp.96-104.
  54. Scalable Mining for Classification Rules in Relational Databases (M. Wang, B. Iyer, and J. S. Vitter), *Proceedings of International Database Engineering & Applications Symposium (IDEAS 1998)*, Cardiff, Wales, U.K., July 1998, pp. 58-67.
  55. Wavelet-Based Histograms for Selectivity Estimation (Y. Matias, J. S. Vitter, and M. Wang), *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data (SIGMOD 1998)*, Seattle, Washington, June 1998, pp. 448-459.
  56. MIND: A Scalable Classifier in Relational Databases (M. Wang, B. Iyer, and J. S. Vitter), *Proceedings of ACM-SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD 1998)*, Seattle, Washington, June 1998.
  57. Efficient Roll-Up and Drill-Down Analysis in Relational Databases (M. Wang and B. Iyer), *Proceedings of ACM-SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD 1997)*, Tucson, Arizona, May 1997.
  58. Selectivity Estimation in the Presence of Alphanumeric Correlations (M. Wang, J. S. Vitter, and B. Iyer), *Proceedings of the 13th Annual IEEE Conference on Data Engineering (ICDE 1997)*, Birmingham, England, April 1997, pp.169-180.
  59. A New Cryptographic Key Management Scheme (M. Wang), *Chinese Journal of Computers*, Vol. 16, No. 2, 1993, pp.106-112 (In Chinese).

#### **Other Publications**

60. Context-Aware Computing: Applications and Challenges, keynote talk (M. Wang), *Proceedings of the 28<sup>th</sup> National Database Conference of China (NDBC 2011)*, Shanghai, China, October 21-2, 2011.
61. DBMSs with Native XML Support: Towards Faster, Richer, and Smarter Data Management, keynote talk (M. Wang), *Proceedings of International Workshop on Emerging Trends of Web Technologies and Applications (WebETrends, APWeb/WAIM Workshop 2007)*, Huang Shan, China, June 16-18, 2007, pp.253-254.
62. Approximation and Learning Techniques in Database Systems, **Ph.D. Thesis**, Duke University, August 1999.
63. A Method for Estimating Filter Factors for Multi-Column Queries (M. Wang, T. Beavin, B. Iyer, and H. Tie), *Technical Report*, IBM Santa Teresa Laboratory, August 1995.
64. An Experimental Study of Alphanumeric Selectivity Estimation (M. Wang, T. Beavin, B. Iyer, and H. Tie), *Technical Report*, IBM Santa Teresa Laboratory, August 1995.

**Patents (12 patents issued plus 24 pending patent applications)**

### ***Issued Patents***

1. Method and Apparatus for Functional Model-Based Data Provenance in Stream Processing Environments. Filed as *Docket YOR920070310US1* in US on 06/08/2007. **Issued** as *Patent 7539753* in US on 05/26/2009.
2. System and Method for Searching Deep Web Service. Filed as *Docket YOR920060245US1* in US on 08/14/2006. Filed as *Docket YOR920060245US2* in US on 07/15/2008. **Issued** as *Patent 7533085* in US on 05/12/2009.
3. Method and Apparatus for Organizing Data Sources. Filed as *Docket YOR920060232US1* in US on 08/14/2006. Filed as *Docket YOR920060232US2* in US on 06/27/2008. **Issued** as *Patent 7529740* in US on 05/05/2009.
4. Methods and Apparatus for Incremental Approximate Nearest Neighbor Searching. Filed as *Docket YOR920050240US1* in US on 08/31/2005. Filed as *Docket YOR920050240US2* in US on 03/31/2008. **Issued** as *Patent 7512282* in US on 03/31/2009.
5. Statistics Collection Using Path-Value Pairs for Relational Databases. Filed as *Docket YOR920060207US1* in US 05/16/2006. **Issued** as *Patent 7472108* in US on 12/30/2008.
6. Classification-Based Method and Apparatus for String Selectivity Estimation. Filed as *Docket YOR920060138US1* in US on 06/26/2006. Filed as *Docket YOR920060138US2* in US on 03/28/2008. **Issued** as *Patent 7395270* in US on 07/01/2008.
7. Dynamic Maintenance of Web Indexes Using Landmarks. Filed as *Docket YOR920030009US1* in US on 05/06/2003. **Issued** as *Patent 7299404* in US on 11/20/2007.
8. Relational Database Management Encryption System. Filed as *Docket YOR920010553US1* in US on 12/28/2001. **Issued** as *Patent 7269729* in US on 09/11/2007.
9. Query Processing Method of Name-Value Pairs in Relational Databases. Filed as *Docket YOR920010783US1* in US on 03/26/2002. **Issued** as *Patent 6691102* in US on 02/10/2004.
10. Query Optimization Through the Use of Multi-Column Statistics to Avoid the Problems of Non-Indexed Column Correlation. Filed as *Docket STL919960067US2* in US on 03/26/1999. **Issued** as *Patent 6272487* in US on 08/07/2001.
11. Query Optimization Through the Use of Multi-Column Statistics to Avoid the Problems of Column Correlation. Filed as *Docket STL919960067US1* in US on 02/28/1997. **Issued** as *Patent 5995957* in US on 11/30/1999.
12. Scalable Set-Oriented Classifier. Filed as *Docket STL919960061US1* in US on 02/14/1997. **Issued** as *Patent 5899992* in US on 05/04/1999.

### ***Filed Patents***

13. An Incremental Image Clustering Method Using Hierarchical Model. Filed as 82720380 (700205446WO01) under PCT on 08/01/2011.
14. Mapping of Relationship Entities between Ontologies. Filed as *Docket SVL920100007US1/IBMP062* in US on 05/14/2010.
15. Semantic-Aware Record Matching. Filed as *Docket YOR920090276US1* in US on 10/30/2009.
16. Semantic Link Discovery. Filed as *Docket YOR920090275US1* in US on 10/30/2009.
17. Semantic Query By Example. Filed as *Docket YOR920090133US1* in US on 9/25/2009.
18. Query Optimization over Graph Data Streams. Filed as *Docket YOR920080706US1* in US on 07/29/2009.
19. Ontology-Based Searching in Database Systems. Filed as *Docket YOR920080721US1* in US on 06/09/2009.
20. Indexing Provenance Data and Evaluation Provenance Data Queries in Data Processing Systems. Files as *Docket YOR920080470US1* in US on 05/20/2009.
21. User-Defined Granular Determination of Data Quality. Filed as *Docket YOR920080222US1* in US on 12/31/2008.
22. A Method of Collaborative Evaluation Infrastructure to Assess the Quality of Healthcare Clinical Decision Actors. Filed as *Docket KR920070004US1* in US on 06/04/2008.
23. Method and Apparatus for Determining and Validating Provenance Data in Data Stream Processing System. Filed as *Docket YOR920070681US1* in US on 05/22/2008.
24. Method and Apparatus for Maintaining and Processing Provenance Data in Data Stream Processing System. Filed as *Docket YOR920070680US1* in US on 05/22/2008.
25. System and Method for Provenance Function Window Optimization. Filed as *Docket*

- YOR920070707US1* in US on 03/12/2008.
26. Method and System for Combining Ranking and Clustering in a Data Management System. Filed as *Docket YOR920070101US1* in US on 04/25/2007.
  27. Method and System for Indexing and Serializing Data. Filed as *Docket YOR920060843US1* in US on 03/02/2007. Filed as *Docket YOR920060843PCT1* under PCT on 02/07/2008.
  28. Method for Supporting Ontology-Related Semantic Queries in DBMSs with XML Support. Filed as *Docket YOR920060225US1* in US on 03/02/2007.
  29. Querying Data and an Associated Ontology in a Database Management System. Filed as *Docket YOR920060649US1* in US on 01/17/2007. Filed as *Docket YOR920060651AU1* in AU on 01/10/2008. Filed as *Docket YOR920060649PCT1* under PCT on 01/10/2008.
  30. Identifying and Annotating Shared Hierarchical Markup Document Trees. Filed as *Docket YOR920060543US1* in US on 10/11/2006.
  31. Processing Queries on Hierarchical Markup Data Using Shared Hierarchical Markup Trees. Filed as *Docket YOR920060541US1* in US on 10/11/2006.
  32. Statistics Collection Using Path-Identifiers for Relational Databases. Filed as *Docket YOR920060098US1* in US on 05/16/2006.
  33. Method for Merging Multiple Ranked Lists with Bounded Memory. Filed as *Docket YOR920040249US1* in US on 02/24/2005.
  34. A Learning-Based Method for Estimating Cost and Statistics of Complex Operators in Continuous Queries. Filed as *Docket YOR920040258US1* in US on 11/08/2004. Filed as *Docket YOR920040258US2* in US on 02/03/2009.
  35. Techniques for Estimating Item Frequencies in Large Data Sets. Filed as *Docket YOR920040290US1* in US on 09/27/2004.
  36. Enabling Interoperability between Participants in a Value Network. Filed as *Docket YOR920030622US1* in US on 04/29/2004. Filed as *Docket YOR920030622US2* in US on 07/24/2008. Filed as *Docket YOR920030622CN1* in CN on 11/30/2004. Filed as *Docket YOR920030622EP1* in EP on 11/30/2004. Filed as *Docket YOR920030622JP1* in JP on 11/30/2004. Filed as *Docket YOR920030622KR1* in KR on 11/30/2004. Filed as *Docket YOR920030622PCT1* under PCT on 11/30/2004. Filed as *Docket YOR920030622TW1* in TW on 04/13/2005.

## Professional Activities

- **Chair**, Industrial and Applications Track, *VLDB 2012*.
- **Program Committee Member and Group Leader**, *SIGMOD 2011*.
- **Review Board Member**, *VLDB Journal Track (PVLDB)*, Sept. 2008-Jan. 2010.
- **Chair**, IBM Research Professional Interest Community on Data Management (DM PIC), Jan. 2008-Dec. 2009.
- **Program Committee Member** for *SIGMOD 2010*, *VLDB 2009*, *WAIM&APWeb 2009*, *SIGMOD 2008 IDAR Workshop*, *ICDE 2008 DBRank Workshop*, *APWeb 2008*, *SAC 2008 (Data Streams Track)*, *VLDB 2007*, *SAC 2007 (Data Streams Track)*, *WAIM&APWeb 2007*, *SAC 2006 (Data Streams Track)*, *WAIM 2006*, *SAC 2005 (Data Streams Track)*, *VLDB 2005*, *SAC 2004 (Data Streams Track)*, *APWeb 2004*, *WAIM 2004*, *WAIM 2003*, *DASFFA 2003*, *WAIM 2002*, *DaMeB 2001*, *WAIM 2001*.
- **Referee** for conferences: *SIGMOD*, *VLDB*, *PODS*, *ICDE*, *EDBT*, *KDD*, *CIKM*, *WAIM*, *SSTD*, *EDBT*, *ICDCS*, *CVIU*, 1996--present.
- **Referee** for journals: *Journal of Information Sciences*, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, and *ACM Transactions on Database Systems (TODS)*, 1996--present.
- **Publication Chair** for *ICEBE 2005*.
- **Tutorial Chair** for *WAIM 2001*.
- **Panelist** at *WAIM 2000*.
- **Member** of *ACM*.

*last updated: October 7<sup>th</sup>, 2011*