
Curriculum Vitae

Wang-Chiew TAN
Associate Professor
Department of Computer Science
University of California, Santa Cruz

EMPLOYMENT

Jul 2008 – present Associate Professor, UC Santa Cruz

Oct-Dec 2008 Visiting Scientist, IBM Almaden Research Center, CA
(On sabbatical leave from UC Santa Cruz)

Jan-Mar 2007 Visiting Scientist, IBM Almaden Research Center, CA
(On leave of absence from UC Santa Cruz)

2002 – 2008 Assistant Professor, UC Santa Cruz

July 2004 Visiting Scientist, Institute for Infocomm Research, Singapore

1999 – 2002 Research Assistant, University of Pennsylvania

Summer 1999 Summer Intern, AT&T Laboratories in New Jersey

1998 – 1999 Teaching Assistant, University of Pennsylvania

1997 – 1998 Research Assistant (Dean's Fellow), University of Pennsylvania

1996 – 1997 Software Engineer, Kent Ridge Digital Laboratories and Port of
Singapore Authority, Singapore

EDUCATION

2002 Ph.D. in Computer Science, University of Pennsylvania
Thesis: Data Annotations, Provenance and Archiving
Advisors: Peter Buneman and Sanjeev Khanna

1999 M.Sc. in Computer Science, University of Pennsylvania

1996 B.Sc. (First Class Honors) in Computer Science, National University of
Singapore, Singapore

RESEARCH INTERESTS

Data Provenance, Information Integration

AWARDS

- 2005 2004 IBM Pat Goldberg Memorial Best Paper Award for the paper
“Composing Schema Mappings: Second-Order Dependencies to the Rescue”
co-authored with R. Fagin, Ph. G. Kolaitis and L. Popa.
- 2002 Best Paper Award for the paper
“Keys for XML”
in Computer Networks vol. 39, no. 5, August 2002
co-authored with P. Buneman, S. B. Davidson, W. Fan, and C. S. Hara

EXTRAMURAL FUNDINGS

- 2009 National Science Foundation (NSF) Grant. IIS-0905276
“Data Interoperability via Schema Mappings”
PI: Ph. G. Kolaitis, Co-PI: W.-C. Tan
Amount: \$1,151,548
- 2004 National Science Foundation (NSF) Grant. IIS-0430994.
“Metadata model management: Schema Mappings and Data Exchange”
PI: Ph. G. Kolaitis, Co-PI: W.-C. Tan
Amount: \$759,722
- 2004 National Science Foundation (NSF) CAREER Award. IIS-0347065
“Tracing the Provenance and Flow of Data”
PI: W.-C. Tan
Amount: \$449,944
- 2003 IBM Shared University Research (SUR) Award.
(Joint award with Ph. G. Kolaitis.)
Amount: \$21,804

INTRAMURAL FUNDINGS

- 2009 UCSC Special Research Grant
PI: W.-C. Tan, Amount: \$6000
- 2009 CAAD Diversity Fund
PI: W.-C. Tan, Amount: \$1980
- 2003 UCSC General Non-Tenured Faculty Development Award.
PI: W.-C. Tan, Amount: \$1,000
- 2002 UCSC Committee on Research: Faculty Research Grant.
PI: W.-C. Tan, Amount: \$1,673
- 2002 UCSC General Non-Tenured Faculty Development Award.
PI: W.-C. Tan, Amount: \$1,500

PATENTS

- “Quasi Inverses of Schema Mappings”
R. Fagin, Ph. G. Kolaitis, L. Popa, and W.-C. Tan
(Patent Filed No. ARC920070055), 2007.
- “Sequential Composition of XML Schema Mappings”
R. Fagin, Ph. G. Kolaitis, L. Popa, W.-C. Tan, and C. Yu
(Patent Filed No: ARC920050087US1), 2005.
- “Method for Converting Relational Data into XML”
M. F. Fernández, A. Morishima, D. Suciú, and W.-C. Tan. (US Patent No.: 6,785,673),
2004.

INVITED TALKS

- 2008 Mapping Understanding and Design by Example, Computer Science Seminar, School of Computing, National University of Singapore.
- 2007 “Debugging Schema Mappings with Routes”, Database Group Seminar, University of Edinburgh.
“Debugging Schema Mappings with Routes”, K55 Lunch Talk, IBM Almaden Research Center.
- 2006 “Debugging Schema Mappings with Routes”, Computer Science Seminar, School of Computing, National University of Singapore.
“Tools and Techniques for *Understanding* Data Exchange Systems”, Workshop on Information Integration, University of Pennsylvania.
“A Tool for Debugging Schema Mappings”, Database Group Seminar, Università Roma Tre.
“A Tool for Debugging Schema Mappings”, Second International Workshop on Exchange and Integration of Data.
- 2005 “DBNotes: A Post-It System for Relational Databases based on Provenance”, SIGMOD Program Committee Workshop.
- 2004 “An Annotation Management System for Relational Databases: Semantics and Implementation”, IBM Silicon Valley Lab.
“An Annotation Management System for Relational Databases: Semantics and Implementation”, Computer Science Seminar, School of Computing, National University of Singapore.
“Tracing the Provenance and Flow of Data”, Stanford Database Seminar, Stanford University.
- 2003 “Tracing the Provenance and Flow of Data”, BEA Systems.
“Tracing the Provenance and Flow of Data”, Computer Science Department, San Jose State University.
- 2002 “Data Annotations and Data Archiving”, Database Group Seminar, University of Washington.
“Archiving Biological Databases” Workshop on Data Derivation and Provenance.
“Computing Provenance and Annotations for Views”, Workshop on Data Derivation and Provenance.

SERVICE TO THE UNIVERSITY

2006-09	Personnel Review Committee Member
2008-09	Ad hoc Committee Service
2003-09	Graduate Committee Member
2004-05	Participant, Faculty Recruitment Search
2003-04	Faculty Recruitment Committee Member
2002-03	Science Library Liaison

SERVICE TO THE PROFESSION

- Program Committee Chair/co-Chair:
 - 2010
 - 2nd USENIX Workshop on the Theory and Practice of Provenance (TaPP '10). Program co-chair.
 - 2008
 - International Conference on Information and Knowledge Management (CIKM) – Track chair for Information Integration, Data Provenance, Probabilistic Databases.
- Program Committee Member:
 - 2010
 - ACM SIGMOD International Conference on Management of Data (SIGMOD).
 - International Conference on Very Large Data Bases (VLDB)
 - 2009
 - ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS).
 - International Workshop on the Web and Databases (WebDB).
 - International Conference on Data Engineering (ICDE).
 - 2008
 - VLDB Workshop on “New Trends in Information Integration”.
 - International Provenance and Annotation Workshop (IPAW).
 - International Conference on Very Large Data Bases (VLDB).
 - 2007
 - International Symposium on Database Programming Languages (DBPL).
 - ACM International Workshop on Web Information and Data Management (WIDM).
 - International Conference on Very Large Data Bases (VLDB).
 - 2006
 - ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS).
 - International Conference on Database Systems for Advanced Applications (DASFAA).
 - International Workshop on the Web and Databases (WebDB).
 - 2005
 - International XML Database Symposium (XSym).
 - International Conference on Very Large Data Bases (VLDB).
 - ACM SIGMOD International Conference on Management of Data (SIGMOD).

2004

- International Workshop on the Web and Databases (WebDB).
- International Conference on Very Large Data Bases (VLDB).
- International Conference on Data Engineering (ICDE).

Postdoctoral Scholars Sponsored

- 2007-08 Dr. Balder ten Cate, University of Amsterdam, The Netherlands; visited UC Santa Cruz and the IBM Almaden Research Center from Jan 2008 to Jun 2008.
- 2007 -08 Dr. Paolo Papotti, Università Roma Tre, visited UC Santa Cruz and the IBM Almaden Research Center from Sep 2007 to Jun 2008.
- 2006-07 Dr. Vassia Pavlaki, National Technical University of Athens, Greece; visited UC Santa Cruz and the IBM Almaden Research Center from Mar 2007 to Jun 2007.

Ph.D. students currently supervised

Bogdan Alexe (advanced to candidacy, Fall 2008), Enela Pema

Ph.D students graduated

2008 Laura Chiticariu Enabling Data Interoperability with Provenance and Schema Integration (first employment: IBM Almaden Research Center)

Advancement Committee Member

2009	Jessica Gronski	Semantic Web for Search
2008	Karl Schnaitter	Online Index Selection for Physical Database Tuning

Master Projects supervised

2008	Jordan Johnson	A Java Implementation of the Quasi-Inverse Algorithm.
2007	Bogdan Alexe	A Survey of View Maintenance for Relational Databases.
2005	Laura Chiticariu	Computing the Core: Algorithmic Issues.
2005	Gaurav Vijayvargiya	Tracing the Why-Provenance of Data over Schema Mappings.
2003	Xinyu Hua	A Relational Database Annotation Management System.

Member of Ph.D. Thesis Committee

2006	Jonathan Panttaja	Complexity in Databases, Games and Logics
------	-------------------	---

Member of M.Sc. Project/Thesis Committee

2008	Gloria Chatzopoulou	Collaborative Filtering for Interactive Database Exploration
2008	Jonathan Finger	Assessment of Rank Joins with Optimal Costs
2004	Yongqin Xiao	Scale-Invariant Geometric Features and Descriptors for Height Data Registration
2004	Randolph Goldman	A Survey of Logic in Games and Games in Logic
2003	Guozheng Ge	Bamboo: Automatic Generation of Rule-based Software Configuration Management (SCM) Systems
2003	Kai Pan	An Algorithm for Translating DASL Queries into SQL

PUBLICATIONS

Scholarly Journals

1. “Provenance”, W.-C. Tan, *Encyclopedia of Database Systems*, pp. 2202, 2009.
2. “Provenance in Scientific Databases”, S. C. Boulakia, W.-C. Tan, *Encyclopedia of Database Systems*, pp. 2202-2207, 2009.
3. “Schema Mapping Composition”, *Encyclopedia of Database Systems*, W.-C. Tan, pp. 2488-2493, 2009.
4. “Provenance in Databases: Why, How, and Where” J. Cheney, L. Chiticariu, and W.-C Tan, In *Foundations and Trends in Databases*, vol. 1, no. 4, pp 379-474, 2009.
5. “Quasi Inverses of Schema Mappings”, R. Fagin, Ph. G. Kolaitis, L. Popa, and W.-C. Tan, In *ACM Transactions on Database Systems (TODS)*, vol. 33, no. 2, pp. 123-132, 2008. (A preliminary version of this paper appeared in the *2005 Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*. See [22].)
6. “Peer Data Exchange”, A. Fuxman, Ph. G. Kolaitis, R. J. Miller, and W.-C. Tan. In *ACM Transactions on Database Systems (TODS)*, vol. 31, no. 4, pp. 1454-1498, 2006. (A preliminary version of this paper appeared in the *2005 Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*. See [26].)
7. “An Annotation Management System for Relational Databases”, D. Bhagwat, L. Chiticariu, W.-C. Tan, and G. Vijayvardiya, In *Very Large Data Bases (VLDB) Journal*, vol 14, no. 4, pp. 373-396, 2006. (A preliminary version of this paper appeared in *2004 Proceedings of the International Conference of Very Large Data Bases (VLDB)*. See [29].)
8. “Composing Schema Mappings: Second-order dependencies to the rescue”, R. Fagin, Ph. G. Kolaitis, L. Popa, and W.-C. Tan, In *ACM Transactions on Database Systems (TODS)*, vol 30, no. 4, pp. 994-1055, 2005. (A preliminary version of this paper appeared in *2004 Proceedings of ACM Symposium on Principles of Database Systems (PODS)*. See [28].)
9. “Research Problems in Data Provenance”, W.-C. Tan, In *IEEE Data Engineering Bulletin*, vol. 27, no. 4, pp. 45-52, 2004.

10. “Archiving Scientific Data”, P. Buneman, S. Khanna, K. Tajima, and W.-C. Tan, In *ACM Transactions on Database Systems (TODS)*, vol 29, no. 1, pp. 2-42, 2004. (A preliminary version of this paper appeared in the *2002 Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*). See [32].)
11. “Reasoning about Keys for XML”, P. Buneman, S. B. Davidson, W. Fan, C. Hara, and W.-C. Tan, In *Information Systems*, vol. 28, no.8, pp. 1037-1063, 2003. (A preliminary version of this paper appeared in the *2001 Proceedings of the International Symposium on Data Base Programming Languages (DBPL)*). See [34].)
12. “SilkRoute: A Framework for Publishing Relational Data in XML”, M. F. Fernández, Y. Kadiyska, A. Morishima, D. Suciu, and W.-C. Tan, In *ACM Transactions on Database Systems (TODS)*, vol.27, no.4, pp. 438-493, 2002.
13. “Keys for XML”, P. Buneman, S. B. Davidson, W. Fan, C. S. Hara, and W.-C. Tan. In *Computer Networks*, vol.39, no.5, pp.473-487, 2002. (A preliminary version of this paper appeared in the *2001 Proceedings of International Conference on World Wide Web (WWW10)*). See [35].)
14. “Publishing Relational Data in XML: The SilkRoute Approach”, M. F. Fernández, A. Morishima, D. Suciu, and W.-C. Tan, In *IEEE Data Engineering Bulletin*, vol.24, no. 2, pp. 12-19, 2001.
15. “SilkRoute: trading between relations and XML”, M. F. Fernández, D. Suciu and W.-C. Tan. In *Computer Networks*, vol. 33, no. 1-6: pp. 723-745, 2000. (A preliminary version of this paper appeared in the *2000 Proceedings of the International Conference on World Wide Web (WWW9)*). See [39].)
16. “A Graphical Interface to Genome Multidatabases”, W.-C. Tan, K. Wang and L. Wong. In *Journal of Database Management*, vol. 9, no. 1, pp. 24-32, 1998. (A preliminary version of this paper appeared in *1996 International Workshop on Database and Expert Systems Applications (DEXA Workshop)*). See [42].)

Peer Reviewed Conference and Workshop Articles

17. “Laconic Schema Mappings: Computing the Core with SQL Queries”, B. ten Cate, L. Chiticariu, Ph. G. Kolaitis, W.-C. Tan. In *Proceedings of the Very Large Data Bases Endowment (PVLDB)*, vol. 2, no. 1, pp. 1006-1017, 2009.
18. “Artemis: A System for Analyzing Missing Answers”, M. Herschel, M. A. Hernández, W.-C. Tan, . In *Proceedings of the Very Large Data Bases Endowment (PVLDB) (Demonstration Track)*, vol. 2, no. 2, pp. 1550-1553, 2009.
19. “Reverse Data Exchange”, R. Fagin, Ph. G. Kolaitis, W.-C. Tan, L. Popa, In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 22-32, 2009.
20. “Comparing and Evaluating Mapping Systems with STBenchMark”, B. Alexe, W.-C. Tan, Y. Velegrakis, In *Proceedings of the Very Large Data Bases Endowment (PVLDB) (Demonstration Track)*, pp. 1468-1471, 2008.

21. “STBenchMark: Towards a Benchmark for Mapping Systems”, B. Alexe, W.-C. Tan, Y. Velegarakis. In *Proceedings of the Very Large Data Bases Endowment (PVLDB)*, pp. 230-244, 2008.
22. “Data Exchange with Data-Metadata Translations”, M. A. Hernández, P. Papotti, W.-C. Tan. In *Proceedings of the Very Large Data Bases Endowment (PVLDB)*, pp. 260-273, 2008.
23. “Muse: Mapping Understanding and Design by Example”, B. Alexe, L. Chiticariu, R. J. Miller, W.-C. Tan, In *Proceedings of International Conference of Data Engineering (ICDE)*, pp. 10-19, 2008.
24. Muse: A System for Understanding and Designing Mappings, B. Alexe, L. Chiticariu, R. J. Miller, D. Pepper, W.-C. Tan, In *Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD) (Demonstration Track)*, pp. 1281-1284, 2008.
25. “Curated Databases”, P. Buneman, J. Cheney, W.-C. Tan, S. Vansummeren, In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 1-12, 2008.
26. “Provenance in Databases”, P. Buneman and W.-C. Tan, In *Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD) (Tutorial Track)*, pp. 1171-1173, 2007.
27. “Quasi-inverses of Schema Mappings”, R. Fagin, Ph. G. Kolaitis, L. Popa, and W.-C. Tan. In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 123-132, 2007.
28. “Debugging Schema Mappings with Routes”, L. Chiticariu and W.-C. Tan, In *Proceedings of International Conference on Very Large Data Bases (VLDB)*, pp. 79-90, 2006.
29. “SPIDER: a Schema mapPIng DEbugger”, B. Alexe, L. Chiticariu and W.-C. Tan, In *Proceedings of International Conference on Very Large Data Bases (VLDB) (Demonstration Track)*, pp. 1179-1182, 2006.
30. “The Complexity of Data Exchange”, Ph. G. Kolaitis, J. Panttaja, and W.-C. Tan. In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 30-39, 2006.
31. “Peer Data Exchange”, A. Fuxman, Ph. G. Kolaitis, R. J. Miller, and W.-C. Tan, In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 160-171, 2005.
32. “DBNotes: a Post-It System for Relational Databases based on Provenance”, L. Chiticariu, G. Vijayvargiya and W.-C. Tan, In *Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD) (Demonstration Track)*, pp. 942-944, 2005.
33. “Composing Schema Mappings: Second-Order Dependencies to the Rescue”, R. Fagin, Ph. G. Kolaitis, L. Popa, and W.-C. Tan, In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 83-94, 2004.

34. “An Annotation Management System for Relational Databases”, D. Bhagwat, L. Chiticariu, G. Vijayvargiya, and W.-C. Tan, In *Proceedings of International Conference on Very Large Data Bases (VLDB)*, pp. 900-911, 2004.
35. “Containment of Relational Queries with Annotation Propagation”, W.-C. Tan, In *Proceedings of International Symposium on Data Base Programming Languages (DBPL)*, pp. 37-53, 2003.
36. “On Propagation of Deletions and Annotations through Views”, P. Buneman, S. Khanna and W.-C. Tan, In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 150-158, 2002.
37. “Archiving Scientific Data”, P. Buneman, S. Khanna, K. Tajima, and W.-C. Tan, In *Proceedings of ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pp. 1-12, 2002.
38. “On Computing Functions with Uncertainty”, S. Khanna and W.-C. Tan, In *Proceedings of ACM Symposium on Principles of Database Systems (PODS)*, pp. 171-182, 2001.
39. “Reasoning about Keys for XML”, P. Buneman, S. B. Davidson, W. Fan, C. S. Hara, and W.-C. Tan. In *Proceedings of International Workshop on Data Base Programming Languages (DBPL)*, pp. 133-148, 2001.
40. “Keys for XML”, P. Buneman, S. B. Davidson, W. Fan, C. S. Hara, and W.-C. Tan, In *Proceedings of International Conference on World Wide Web (WWW10)*, pp. 201-210, 2001.
41. “Why and Where: A Characterization of Data Provenance”, P. Buneman, S. Khanna and W.-C. Tan, In *Proceedings of International Conference on Database Theory (ICDT)*, pp. 316-330, 2001.
42. “Towards a Query Language for Annotation Graphs”, S. Bird, P. Buneman and W.-C. Tan, In *Proceedings of International Conference on Language Resources and Evaluation (LREC)*, pp. 807-814, 2000.
43. “Data Provenance: Some Basic Issues”, P. Buneman, S. Khanna and W.-C. Tan, In *Proceedings of Conference on the Foundations of Software Techniques and Theoretical Computer Science (FSTTCS)*, pp. 87-93, 2000.
44. “SilkRoute: Trading between relations and XML”, M. F. Fernández, D. Suciu and W.-C. Tan. In *Proceedings of International Conference on World Wide Web (WWW9)*, pp. 723-745, 2000.
45. “Beyond XML Query Languages”, P. Buneman, A. Deutsch, W. Fan, H. Liefke, A. Sahuguet, and W.-C. Tan, In *Query Language Workshop (QL'98)*, 1998.
46. “A Deterministic Model for Semistructured Data”, P. Buneman, A. Deutsch and W.-C. Tan, In *Workshop on Query Processing for Semistructured Data and Non-Standard Data Formats*, 1998.

47. "QUICK: Graphical User Interface to Multiple Databases", W.-C. Tan, K. Wang and L. Wong, In *International Workshop on Database and Expert Systems Applications (DEXA Workshop)*, pp. 404-409, 1996.