Prasenjit Mitra

E-mail: mitra@db.stanford.edu
1714 Heavenly Bamboo Court,
San Jose, CA 95131.

Phone: (408) 254-4904
(650) 723-3605
http://www-db.stanford.edu/~prasen9

Research Interests

Databases, with an emphasis on designing techniques and tools to enable integration and interoperation of data and knowledge over the Internet and across enterprises. In particular: schema matching, query optimization, ontology composition, and information extraction.

Education

Ph.D., Electrical Engineering, Stanford University, CA (expected) 2003
Dissertation: Enabling Interoperation of Information Sources by Resolving Semantic Heterogeneity and Composing Ontologies.
Advisor: Professor Gio Wiederhold
M.S., Computer Science, The University of Texas at Austin, TX Dec.,1994
B.Tech., (Honours), Computer Science and Engineering,
Indian Institute of Technology, Kharagpur, India May, 1993

Dissertation

The primary roadblock preventing the success of data integration and interoperation systems is the absence of **effective tools to establish semantic mappings between large data sources**. My thesis describes a **Semantic Knowledge Articulation Toolkit** that employs a strategy that uses lexical, structural and instance information present in the data sources to find such semantic mappings. Using these semantic mappings, I demonstrated how to compose information from multiple data sources. I utilized ontologies – schema-like structures containing metadata regarding the terms and relationships in the sources – wherever available, to add precision to our data integration efforts. To systematize the **composition of information from multiple sources** I designed an **Ontology-Composition Algebra** and showed how to optimize the composition of information from multiple data sources based on the properties of the algebraic operators.

Work Experience

- **Research Assistant**, Stanford University, CA 1998 – present
  Conducted research on interoperation of information sources. Designed, implemented and experimented with a tool to generate semantic mappings between schemas and ontologies related to data sources. Developed a Ontology-Composition Algebra to enable systematic composition of information. Designed efficient algorithms for answering queries using views in data integration systems.
  Mentored graduate students: Shrish Agarwal, and Alex Carobus.
- **Senior Software Engineer** (part-time), Narus, Inc., CA 2000 – 2001
  As a member of the database group, worked on Narus' billing-mediation system. I designed and implemented an efficient table join algorithm used to generate reports for a database of streaming network traffic data.
- **Senior Member of Technical Staff**, Oracle Corporation, 1995 – 2000
  Designed and implemented the connection load balancer, oracle trace logging utilities on Massively Parallel Processing systems. Was member of the team that designed and maintained the object cache, and implemented inheritance and schema evolution in Oracle 8i.
- **Research Assistant**, The University of Texas at Austin, TX 1994

Developed efficient algorithms for BLACS (Basic Linear Algebra Communication Subroutines) for use in parallel numerical computation.

- **Intern**, NCKBC, Indian Statistical Institute, India Summer, 1992
  At the National Center for Knowledge Based Computing, developed an image-processing algorithm for the automated testing of PCBs.

Teaching Experience

- **Instructor**, Stanford University, CA Winter, 1999
  CS 545: Database Seminar.
- **Teaching Assistant**, The University of Texas at Austin, TX Spring,1994
  CS 352: Computer System Architecture. Helped students understand the material, and graded assignments, mid-term and end-term examinations.
- **Teaching Assistant**, The University of Texas at Austin, TX. Fall,1993
  CS 310: Computer Organization and Programming. Lectured in discussion session, answered questions from students during office hours and via email, graded assignments, mid-term and end-term examinations.
- **Visiting Lecturer**, Mahila Samity Polytechnic, ONGC, India 1992
  Computer Systems: Analysis and Design. As instructor for the class taught to undergraduate students in the polytechnic, lectured on introductory topics in computer system design, data structures, algorithms, and programming, prepared examinations and graded them.

Achievements and Honours

Public Speaking: won several Toastmasters' Humorous Speech Contests and Tall Tale Contests. 1997-1999
Member, Phi Beta Kappa Honor Society Jan. 1995
ACM Programming Contest, UT Austin: member of winning team, represented the University of Texas at Austin. Nov. 1994
Order of Merit for Literary Activities, IIT, India May, 1993
National Talent Search Merit Scholarships, India 1989 –1993

Professional Activities and Service

External Referee for TODS, WWW '02, WAIM '02, and ECAI '02 Workshop on Knowledge Transformation for the Semantic Web.
Ex-Member of MENSA, ACM, ACM-SIGMOD, IEEE
Vice-President and **President** of Oracle Orators 1997-1998

Invited Talks

Enabling Interoperation Among Knowledge Bases via Articulation of Ontologies, **IBM Almaden Research Center**, San Jose, CA, Jan 12, 2001

Publications

***Book Chapters*:**

1. A Scalable Framework for the Interoperation of Information Sources. P. Mitra, G. Wiederhold, and S. Decker. (extended version of SWWS '01 paper). In "The Emerging Semantic Web". IOS Press, Vol. 75 in the Frontiers in Artificial Intelligence and Applications Series, 2002.

2. An algebra for the composition of ontologies, P. Mitra, G. Wiederhold. In "Knowledge Transformation for the Semantic Web". IOS Press, to appear.

***Refereed Journal Article*:**

3. An Algebra for Semantic Interoperability of Information Sources. P. Mitra, G. Wiederhold. (extended version of BIBE paper). In International

Journal of Computational Bioinformatics and Bioengineering, 2002.

Periodical:

4. Framework for the Semantic Web: An RDF Tutorial. S. Decker, P. Mitra, S. Melnik. In IEEE Internet Computing. November/December 2000.

Refereed Conferences & Symposiums:

5. Answering Queries Using Views with Arithmetic Comparisons.
F. Afrati, C. Li, and P. Mitra. In Proc. of the 21st ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems *(PODS)*, 2002.

6. An Algebra for Semantic Interoperability of Information Sources.
P. Mitra, and G. Wiederhold. In Proc. of 2nd. IEEE Symp. on BioInformatics and Bioengineering *(BIBE)*, 2001.

7. A Scalable Framework for Interoperation of Information Sources.
P. Mitra, G. Wiederhold and S. Decker. In Proc. of the 1st International Semantic Web Working Symposium *(SWWS)*, 2001.

8. An Algorithm for Answering Queries Efficiently Using Views. P. Mitra
In Proc. of the Australasian Database Conference *(ADC)*, 2001.
Also in Stanford University Technical Report, 21st September, 1999.

9. An Information Food Chain for Advanced Applications on the WWW.
S. Decker, J. Jannink, S. Melnik, P. Mitra, S. Staab, R. Studer and G. Wiederhold. In Proc. of European Conf. Digital Libraries, *(ECDL)*, 2000.

10. A Graph-Oriented Model for Articulation of Ontology Interdependencies.
P. Mitra, M. Kersten and G. Wiederhold. In Proc. of the 7th International Conference On Extending Database Technology *(EDBT)*, 2000.

11. Semi-automatic Integration of Knowledge Sources.
P. Mitra, G. Wiederhold and J. Jannink.
In Proc. of the 2nd International Conf. on Information Fusion (FUSION), 1999.

12. Multiobjective Search in VLSI Design.
P. Dasgupta, P. Mitra, and P.P. Chakraborty.
In Proc. of the 7th International Symposium on VLSI Design (VLSI), 1994.

Refereed User Group Meeting:

13. Fast Collective Communication Libraries, Please.
P. Mitra, D. Payne, L. Shuler, R. van de Geijn, and J. Watts.
In Proc. of the Intel Supercomputing Users' Group Meeting, 1995.

Refereed Workshops:

14. An Algebra for the Composition of Ontologies.
P. Mitra, and G. Wiederhold.
In Proc. of Workshop on Knowledge Transformation for the Semantic Web at the 15th European Conference on Artificial Intelligence *(ECAI)*, 2002.

15. Resolving Terminological Heterogeneity in Ontologies.
P. Mitra, and G. Wiederhold.
In Proc. of Workshop on Knowledge Transformation for the Semantic Web at the 15th European Conference on Artificial Intelligence *(ECAI)*, 2002.

16. An Algebra for Semantic Interoperation of Seminstructured Data. J. Jannink, P. Mitra, S. Pichai, G. Wiederhold, E. Neuhold, and R. Studer. In IEEE Knowledge and Data Engineering Exchange Workshop *(KDEX)*, 1999.

B-Tech. Thesis:

Multiobjective Heuristic Search.
P. Mitra, B. Tech Thesis, Indian Institute of Technology, 1994.

***Invited Book-Chapter submitted for publication*:**

An Ontology Composition Algebra. P. Mitra, G. Wiederhold.
Handbook on Ontologies in Information Systems.
Editors: S. Staab, and R. Studer.
Springer-Verlag series on International Handbooks on Information Systems.

References:

Professor Gio Wiederhold,
Professor Emeritus,
Computer Science Department, Gates Hall 4A,
Stanford University, Stanford, CA, 94305-9040.
1-650-725-8363
gio@cs.stanford.edu

Professor Foto Afrati,
Professor, Computer Science Division,
National Technical University of Athens,
9, Heroon Politechniou, Zographou,
157 73 Athens, Greece.
30-1-7722498
afrati@cs.ece.ntua.gr

Dr. Stefan Decker,
Computer Scientist,
Information Sciences Institute,
Admiralty Way, Suite 901 ,
Marina del Rey, CA 90292-6695.
1-310-448-8473
stefan@isi.edu

Professor Jeffrey D. Ullman
Stanford W. Ascherman Professor,
Computer Science Department, Gates Hall 4A,
Stanford University, Stanford, CA 94305-9040.
1-650-725-4802
ullman@cs.stanford.edu

Professor Dr. Martin L. Kersten,
INS, CWI,
Kruislaan 413, P.O. Box 94079,
1090 GB Amsterdam, The Netherlands.
31-20-5924066
Martin.Kersten@cwi.nl