

Marie Jacob

575 Levine Hall North, 3330 Walnut Street
Philadelphia, PA 19104

215-573-2582

majacob@cis.upenn.edu

<http://www.cis.upenn.edu/~majacob>

RESEARCH INTERESTS

My interests lie broadly in database, data integration and web systems. Specifically, I look to work on integration systems that take users and context into account when discovering and querying for information over web-scale settings.

EDUCATION

- **University of Pennsylvania** Philadelphia, PA
Ph.D., Computer and Information Science Sep. 2007 - Present
 - Advisor: Zack Ives
 - Tentative thesis topic: interactively learning data integration mappings and queries.
- **York University** Toronto, ON
B.A. Honours, Computer Science 2003-2007
 - Graduated Summa Cum Laude.
 - Senior undergraduate project: Building rewrite rules for special classes of SQL queries.

RESEARCH EXPERIENCE

Q System January 2008 - present University Of Pennsylvania
Currently developing a query authoring system that tackles data integration for the naive user. We learn user preferences of sources and associations between sources through feedback on query answers. The system incorporates keyword search as an initial means of finding potentially relevant relations and associations, subsequently choosing the best queries and repeatedly updating an internal model through user feedback. My current focus is on efficiency and optimization in retrieving and updating ranked answers for the user.

CopyCat January 2008 - January 2009 University Of Pennsylvania
Developed a system for learning and authoring queries for one-time data integration tasks. This project involves monitoring a user's actions by copying and pasting data from different sources, and then trying to learn and generalize these actions by considering given examples and user feedback. We make use of three levels of learners: one for learning source document structures, one for learning semantic types of data values, and one for learning the integration queries. We use provenance information, in a user friendly form, as part of this process to inform users about where query answers originate from.

ASPEN, SmartCIS September 2007 - July 2009 University Of Pennsylvania
Projects involving data integration in distributed streaming environments. My focus has been on studying and implementing new techniques for in-network joins in mesh-type networks.

2D Shape reconstruction May 2005 - August 2005 Centre For Vision Research, York University
Implemented reconstructions of two-dimensional closed shapes by using extracted fourier descriptors of common shapes in natural images. Conducted a study on the effects of random sampling of fourier descriptors to be used in such reconstructions.

EMPLOYMENT

- **University of Pennsylvania** Philadelphia, PA
Research Assistant & Ph.D. Candidate September 2007 - Present
- **Environment Canada** Toronto, ON
Programmer & Data Analyst January 2006 - August 2007

TEACHING

- **CIS 455/555, Internet and Web Systems, Spring 2009** - Taught occasional lectures, held office hours and handled grading.
- **CIS 550/CSE 330, Database and Information Management Systems, Fall 2008** - Designed course project, assisted in preparing homework/midterm questions solutions and held office hours.

AWARDS & HONOURS

Graduate Fellowship, CIS Department, University of Pennsylvania	2007
York University Computer Science Academic Achievement Award	2007
ACM-SIGMOD Undergraduate Award (declined)	2007
Canada-EU Mobility Award	2006
NSERC Undergraduate Summer Research Award	2005
York University Entrance and Continuing Scholarships	2003-2007
Ontario Scholar	2003

PUBLICATIONS

Submitted:

- Marie Jacob, Zachary G. Ives. **Answering Ranked Queries with Context-Specific Scoring Functions**, 2010.

Peer-reviewed:

- Svilen Mihaylov, Marie Jacob, Zachary Ives, Sudipto Guha. **Dynamic Join Optimization for Multi-Hop Wireless Sensor Networks**. PVLDB Vol. 3/ VLDB 2010, Singapore.
- Mengmeng Liu, Svilen Mihaylov, Zhuowei Bao, Marie Jacob, Zachary G. Ives, Boon Thau Loo, Sudipto Guha. **SmartCIS: Integrating Digital and Physical Environments**. SIGMOD 2009, Demonstration Description. [*Honourable mention for best demo*].
- Zachary G. Ives, Craig Knoblock, Steve Minton, Marie Jacob, Partha Pratim Talukdar, Rattapoom Tuchinda, Jose Luis Ambite, Maria Muslea, Cenk Gazen. **Interactive data integration through Smart Copy & Paste**. CIDR 2009, Asilomar, CA.
- Zachary G. Ives, Todd J Green, Grigoris Karvounarakis, Nicholas E. Taylor, Val Tannen, Partha Pratim Talukdar, Marie Jacob, Fernando Pereira. **The Orchestra Collaborative Data Sharing System**. SIGMOD Record, September 2008.

- Partha Pratim Talukdar, Marie Jacob, M. Salman Mehmood, Koby Crammer, Zachary G. Ives, Fernando Pereira, and Sudipto Guha. **Learning to Create Data-Integrating Queries**. VLDB 2008, Auckland, New Zealand. [*Best Poster Award at DB/IR day, Fall 2008*].
- Svilen R. Mihaylov, Marie Jacob, Zachary G. Ives, Sudipto Guha. **A Substrate for In-Network Sensor Data Integration**. DMSN 2008, Auckland, New Zealand.

SKILLS

Languages: Java (5+ years), C (1 year), Python (1.5 years), Prolog, Lisp (< 1 year)

Operating Systems: UNIX, Windows

Database Management & Languages: SQL, XML; DB2, MYSQL, ORACLE

Other: Matlab (1.5 years)

MEMBERSHIPS & ACTIVITIES

- ACM - Student Member
- GWWE 2009 - Google's workshop for Women Engineers
- CISTers (Female graduate students and faculty at Penn) - Participant and (occasional) organizer.