# Fan Yang

Phone: 408-349-6648 (office) 607-342-4541 (cell)

Email: yangf@cs.cornell.edu

#### **Research Focus**

- Next generation platform for developing web applications and services
- Scalability and performance of large-scale database systems for hosted web applications
- Query processing and optimization in Database Systems with large datasets

## Education

• Cornell University, Ithaca, NY GPA 3.8/4.0 09/2003-present Ph.D. candidate in Computer Science (expected in 05/2008)
Advisor: Prof Johannes Gehrke

• **Peking University**, Beijing, China. GPA 3.8/4.0 08/1999-06/2003 **B.S. in Computer Science**. Ranked top 1 out 120.

# **Professional Experience**

- Research Intern, Community System Group, Yahoo Research 10/2007 present Advisor: Dr. Jayavel Shanmugasundaram
  - o Designing and developing a scalable database solution for hosting a large number of user-created social network applications.
- Research Assistant, Computer Science Department, Cornell University

Advisor: Prof. Johannes Gehrke

06/2005 - present

- o The main researcher on the HILDA project (National Science Foundation Grant No. 534404).
- o Designed a next generation declarative language for developing data driven web applications
- o Designed and implemented methods for automatically partitioning application functionalities across client and server based on performance and scalability criterion.
- o Designed and implemented systems to allow users to build and share their personalization for web applications.
- Reseach Intern, Community System Group, Yahoo Research. 05/2007 08/2007 Advisor: Dr. Jayavel Shanmugasundaram
  - Developed a WYSIWYG development platform for data driven web applications based on Web 2.0 architectures.
  - o Co-devised algorithms for automatic database schema generation from user-created web pages. (Patent Pending, Submitted for publication)
- Research Intern, Information Management Group, IBM Almaden Research Center. 05/2006 08/2006

Advisor: Dr. Quanzhong Li

- o Co-developed a system for processing queries over XML data in RDMS with optimization techniques using full-text search index.
- Research Assistant, Computer Science Department, Cornell University

Advisor: Prof. Jayavel Shanmugasundaram

05/2004 - 06/2005

- o The main developer of the storage component of the QUARK system (<a href="http://www.cs.cornell.edu/database/quark/quark\_main.html">http://www.cs.cornell.edu/database/quark/quark\_main.html</a>), which is an open source native XML database system with full text search support.
- Software Developer, Bell Labs Research, Beijing, China.

01/2003 - 06/2003

Advisor: Dr. Ruibing Hao

- o Developed a system for verifying network routing protocols with randomized algorithms.
- Research Assistant, Network and Database Group, Peking University, China. Advisor: Prof. Ming Zhang 06/2002 - 12/2002
  - o Co-developed a text classification system with various machine learning approaches.

#### **Professional Skills**

- Proficient in C/C++/STL, Java, Javascript/AJAX and SQL. Programming experience with Perl, Python and scripts.
- Proficient in algorithm design and standard data structures.
- Advanced knowledge on database system, server architectures, programming language design and distributed systems.
- Experience on designing and building large-scale systems.
- Experience on data analysis on large scale datasets

### **Related Courses**

- The Architecture of Large-Scale Information Systems
- Advanced Database System
- Advanced topics in Machine Learning
- Advanced Design & Analysis Algorithms
- Theory of Computing
- Research in Computer Networks
- Advanced Programming Languages

#### **Patents**

- *Method and/or System for Schema and View Generation*. Co-inventor with Chavdar Botev, Nitin Gupta, George Levchenko and Jayavel Shanmugasundaram, US Patent pending, 2007
- Offline Operations for Online Applications. Co-inventor with Nitin Gupta and Jayavel Shanmugasundaram, US Patent pending, 2007

#### **Publications**

- A Unified Platform for Data Driven Web Applications with Automatic Client-Server Partitioning. Fan Yang, Nitin Gupta, Nicholas Gerner, Xin Qi, Alan J. Demers, Johannes Gehrke, Jayavel Shanmugasundaram: WWW Conference 2007
- *Hilda: High Level Language for Data Driven Applications*. Fan Yang, J. Shanmugasundaram, Mirek Riedewald, Johannes Gehrke, and Alan Demers, ICDE Conference 2006
- Efficient Keyword Search over Virtual XML Views. Feng Shao, Lin Guo, Chavdar Botev, Anand Bhaskar, Muthiah M. Muthiah Chettiar, Fan Yang, Jayavel Shanmugasundaram: VLDB Conference 2007
- User-centric personalized extensibility for data-driven web applications. Nitin Gupta, Fan Yang, Alan J. Demers, Johannes Gehrke, Jayavel Shanmugasundaram: SIGMOD Conference 2007
- Automatic Client-Server Partitioning of Data-Driven Web Applications. Nicholas Gerner, Fan Yang, Alan Demers, Johannes Gehrke, Mirek Riedewald and Jayavel Shanmugasundaram SIGMOD Conference 2006
- Quark: An Efficient XQuery Full-Text Implementation. Anand Bhaskar, Chavdar Botev, Muthiah M Muthaia Chettiar, Lin Guo, Jayavel Shanmugasundaram, Feng Shao and Fan Yang. SIGMOD Conference 2006