

# Yu Xu

Department of Computer Science and Engineering  
University of California, San Diego  
La Jolla, CA 92093-0114, USA  
Phone: 858-699-8109  
E-mail: [yxu@cs.ucsd.edu](mailto:yxu@cs.ucsd.edu)  
WWW: <http://www.cs.ucsd.edu/~yxu>

## Research Interests

Query processing for XML, semi-structured, and relational data; data integration; XML query languages (XQuery, XPath, XSLT), storage, and indexing; information retrieval in semi-structured data; keyword search in XML; databases and the Web; Web Services.

## Education

*Ph.D.*, Computer Science, University of California, San Diego (expected) Summer 2005  
*M.S.*, Computer Science, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, June 1999  
*B.S.E.*, Computer Science, Northern Jiaotong University, Beijing, June 1996

## Work Experience

**IBM Almaden Research Center** San Jose, CA (June-September 2004)

*Summer intern.* Worked on a proposal to W3C's XQuery working group to support analytical queries in XQuery; worked on defining the syntaxes, semantics and examples of the new features in the proposal; implemented a prototype in IBM DB2 V9.1. The work was published in SIGMOD 2005 and submitted to W3C XML Query Working Group.

**IBM Silicon Valley Lab/Almaden Research Center** San Jose, CA (June-September 2003) *Summer intern.* Worked on optimizing ordering operations in XQuery; designed and implemented XQuery order related rewriting rules in IBM DB2.

**AT&T Labs Research** Florham Park, NJ (June-August 2002)

*Summer intern* worked on the design and implementation of PIX, a system that permits phrase matching in XML documents that contain "mixed content". A key feature of PIX is that users can specify which markup and annotations to ignore when matching a phrase. The work was published in VLDB 2003, and the system was demonstrated in SIGMOD 2003 and ICDE 2003. PIX uses inverted indices and an efficient evaluation algorithm to compute the set of matches and returns answers where phrases, ignored tags and content are highlighted. In addition, query answers are sorted using a ranking function. PIX is implemented as an extension of GALAX, a full-fledged XQuery engine. The functionality of PIX is fully integrated into XQuery and permits a natural combination of XPath-based structure matching with phrase matching.

**Enosys Software** San Diego, CA (June-December 2000; June-September 2001)  
*Summer intern* Enosys was a startup which built the first generally available distributed XQuery processor and was acquired in 2003 by BEA Systems. Worked on the design and implementation of the translation algorithm from XQuery to Enosys's algebra; designed and implemented several algebra rewriting rules; worked on the design and implementation of the query composition algorithm.

## Teaching Experience

- Winter 2002 Teaching Assistant for CSE 130 Programming Languages, University of California, San Diego
- Fall 2001 Teaching Assistant for CSE 21Math, University of California, San Diego
- Spring 2001 Teaching Assistant for CSE 126 Multimedia Systems University of California, San Diego
- Spring 2000 Teaching Assistant for CSE 120 Operating Systems University of California, San Diego
- Winter 2000 Teaching Assistant for CSE 131B Compiler Construction II University of California, San Diego

## Publications

- **Efficient Keyword Search for Smallest LCAs in XML Databases**  
Yu Xu and Yannis Papakonstantinou. ACM SIGMOD 2005
- **Extending XQuery for Analytics**  
Kevin Beyer, Don Chamberlin, Latha Colby, Fatma Ozcan, Hamid Pirahesh, and Yu Xu. ACM SIGMOD 2005
- **The NEXT Logical Framework for XQuery**  
Alin Deutsch, Yannis Papakonstantinou and Yu Xu. VLDB 2004
- **Phrase Matching in XML**  
Sihem Amer-Yahia, Mary F. Fernandez, Divesh Srivastava and Yu Xu. VLDB 2003
- **Minimization and Groupby Detection for Nested XQueries**  
Alin Deutsch, Yannis Papakonstantinou and Yu Xu. ICDE Poster 2004
- **PIX: A System for Phrase Matching in XML Documents**  
Sihem Amer-Yahia, Mary F. Fernandez, Divesh Srivastava and Yu Xu. ICDE Demo 2003
- **PIX: Exact and Approximate Phrase Matching in XML**  
Sihem Amer-Yahia, Mary F. Fernandez, Divesh Srivastava and Yu Xu. ACM SIGMOD Demo 2003

## Patents

Phrase matching in documents having nested-structure arbitrary (document-specific) markup Sihem Amer-Yahia, Mary F. Fernandez, Divesh Srivastava, Yu Xu (US-Patent pending)

## Online Demonstrations

XML Query Algebra: <http://feast.ucsd.edu/People/yu/xmlqademo/index.htm>

XML Keyword Search: <http://feast.ucsd.edu/People/yu/xksearch/index.htm>

Phrase Matching in XML: <http://teriyaki.ucsd.edu:9099/pix/index.htm>

## Skills

Java, C++/C, XQuery, XSL, JavaCC, YACC, JFlex, Java Servlet, ML/OCaml, Prolog, Apache Tomcat, Berkeley DB, XML, HTML, SOAP, Web Service, TCP/IP, Oracle, SQL Server, MySQL, PHP, Bash, Qt, Linux, Matlab

## Community Services

[International Conference on Data Engineering 2006](#) Program Committee member for "XML & Semi-structured Data"

Reviewer for [SIGMOD](#) [VLDB](#)

Volunteer Webmaster for [San Diego Chinese Association](#)

## References

Yannis Papakonstantinou, Associate Professor  
Department of Computer Science and Engineering  
University of California, San Diego  
La Jolla, CA 92093-0114, USA  
Phone : (858) 822-1612  
E-mail: yannis@cs.ucsd.edu

Alin Deutsch, Assistant Professor  
Department of Computer Science and Engineering  
University of California, San Diego  
La Jolla, CA 92093-0114, USA  
Phone : (858) 822-2276  
E-mail: deutsch@cs.ucsd.edu

Mary Fernandez, Principal Technical Staff Member  
AT&T Labs – Research  
180 Park Ave., Bldg 103, E243  
Florham park, NJ 07932-0971  
Phone: (973) 360-8679  
E-mail: mff@research.att.com

Divesh Srivastava, Head of the Database Research Department  
AT&T Labs – Research  
180 Park Ave., Bldg 103, A163  
Florham park, NJ 07932-0971  
Phone: (973) 360-8776  
E-mail: divesh@research.att.com

Don Chamberlin, IBM Fellow, ACM Fellow and a member of the National  
Academy of Engineering.  
Dept. K55/B1  
IBM Almaden Research Center  
650 Harry Road  
San Jose, CA 95120-6099  
E-mail: chamberlin@almaden.ibm.com