

# Xin (Luna) Dong

180 Park Ave.  
Florham Park, NJ 07932  
<http://www.research.att.com/~lunadong>

Phone: (650) 283-2063  
Fax: (650) 253-8603  
E-mail: lunadong@research.att.com

---

## Research Interests

**Databases, information retrieval and machine learning**, with an emphasis on data integration, data cleaning, probabilistic data management, schema matching, personal information management, web search, web-service discovery and composition, Semantic Web.

## Education

**University of Washington** 2001-2007

**Ph.D.**, Computer Science, August 2007  
Dissertation title: Providing best-effort services in Dataspace systems.  
Advisor: Alon Halevy  
**M.S.**, Computer Science, November 2003

**Peking University** 1998-2001

**M.S.**, Computer Science, June 2001  
Graduated with **Outstanding Graduate Award**

**Nankai University** 1994-1998

**B.S.**, Computer Science, June 1998  
**A.S.**, International Finance, June 1998  
Graduated with **Outstanding Graduate Award**  
*(Granted to eight students in the university in that year)*

## Honors

Best Demo Award in ACM Int. Conf. on Management of Data (SIGMOD), 2005  
*(One of the top 3 demos in that year)*

Kodak Scholarship, Kodak Corporation, 2 times: 2000, 2001  
*(The 1<sup>st</sup> Rank Scholarship in the department)*

Computer World Scholarship, Computer World Newspaper, 1997  
*(Granted to one student in the department every year)*

First-class Nankai Scholarship, Nankai University, 4 times: 1994-1997  
*(Granted to three students in the department every year)*

Baogang Education Scholarship, Baogang Corporation, 1996  
*(Granted to ten students in the university every year)*

Motorola Scholarship, Motorola Corporation, 1996  
*(The 1<sup>st</sup> Rank Scholarship in the university)*

## Research Experience

### **Senior Member of Technical Staff, AT&T Research**

2007/10-present

Manager: Divesh Srivastava

- Researching on detecting dependence between data sources to facilitate finding truth from conflicting information provided by these data sources. Developing the Solomon system that finds truth from conflicting information provided by various data sources.
- Researching on update certificates in database systems.
- Researching on finding experts in a queried area based on Web2.0 enterprise data..
- Researching on generating multi-table mediated schema for a large number of heterogeneous data sources.

### **Research Assistant, University of Washington**

2002-2007

Advisor: Alon Halevy

- Researched on dataspace service platform and developing the SEMEX Personal Information Management System. Worked on automatically extracting instances and associations from personal information, reconciling instance references that refer to the same real-world entity, indexing loosely-coupled heterogeneous data, supporting seamless search and querying on structured and unstructured data, and visualizing heterogeneous data by applying on-the-fly schema matching.
- Designed and developed the Woogle ([www.cs.washington.edu/woogle](http://www.cs.washington.edu/woogle)) Web-Service Search Engine to provide novel mechanisms for searching and composing web services.
- Participated in the PIAZZA peer data management project. Studied the theoretical foundation of XML query containment.

### **Research Intern, Google Inc.**

2006-2007

Supervisor: Alon Halevy

- Studied the theoretical underpinning of probabilistic schema matching.
- Designed and developed a system that can automatically integrate a large set of heterogeneous data sources and provide ranked answers to SQL queries.

### **Research Intern, Microsoft Research**

Summer 2005

Supervisor: Raghav Kaushik and Venkatesh Ganti

- Participated in the Data Cleaning project. Designed and implemented an algorithm that optimizes the computation of query difference.

### **Research Intern, Bell Labs, Lucent Technologies**

Summer 2002

Supervisor: Hank Korth

- Participated in the ROLEX project. Designed and implemented an algorithm that updates XML views over relational databases.

### **Research Assistant, Peking University**

1998-2001

Advisor: Xiao'ou Chen

- Designed and implemented an SVG webpage editor, which is integrated with the FOUNDER ARTWORLD software.

## Teaching and Mentoring Experience

- Supervisor, AT&T Research** 2007-present
- Songtao Guo, graduate student at University of North Carolina at Charlotte (UNCC).
- Supervising project on record linkage and truth discovery on data provided by multiple sources.
- Su Chen, graduate student at National University of Singapore .
- Supervising project on update certificate in database systems.
- Anish Das Sarma, graduate student at Stanford Univ.
- Supervising project on query answering on uncertain data w. dependencies between data sources.
  - Supervised project on automatically integrating a large number of heterogeneous data sources.
- Daisy Zhe Wang, graduate student at Univ. of California at Berkeley.
- Supervising project on automatically generating a multi-table mediated schema for a large number of heterogeneous data sources.
- Supervisor, University of Washington** 2004-2006
- Jing Liu, graduate student.
- Supervised qualify-exam project on answering structured queries on unstructured data.
- Yuhan Cai, graduate student.
- Supervised project on browsing and querying personal information
  - Mentored in the first-year-student mentoring program.
- Anthony Vigil, undergraduate student.
- Supervised project on ranking objects in an association network.
- Colin Pop, undergraduate student.
- Supervised project on extracting objects and associations from personal information.
- Dinh Lam, undergraduate student.
- Supervised independent-study project on automatically trying web services.
- Teaching Assistant, University of Washington** 2001-2003
- CSE444, Introduction to Database Systems.
- Facilitated in designing a new Database course (see [Halevy, SIGMOD Record 32(3), 2003])
- CSE373, Data Structures and Algorithms
- Mentor, Making Connection Program, University of Washington** 2001-2003
- Mentored two high-school female students
- Co-instructor, Founder Computer Science & Technology Institute** 2001
- Taught XML to employees in the Founder Institute
- Teaching Assistant, Peking University** 2000
- Graduate Course: Digital Image Processing

## **Publications**

## Book Chapters

Xin Dong and Divesh Srivastava: XML Indexing. In “Encyclopedia of Database Systems.” *Springer*, to appear.

Anish Das Sarma, Xin Dong, and Alon Halevy: Uncertainty in Data Integration. In “Managing and Mining Uncertain Data.” *Springer*, to appear.

Tiziana Catarci, Xin Dong, Alon Halevy and Antonella Poggio: Structure everything. In “Personal Information Management.” *University of Washington Press*, Oct 2007.

Xin Dong: Chpt 1. The origin of XML. Chap 2. XML grammar. In “Step by Step to XML.” *Tsinghua Publication House*, March 2001. (In Chinese)

## Publications on Databases

### Refereed Conference Papers

Laure Berti-Equille, Anish Das Sarma, Xin Dong, Amelie Marian, and Divesh Srivastava: Sailing the Information Ocean with Awareness of Currents: Discovery and Application of Source Dependence. In *the 4<sup>th</sup> Conf. on Innovative Data Systems Research (CIDR)*, 2009.

Anish Das Sarma, Xin Dong, and Alon Y. Halevy: Bootstrapping pay-as-you-go data integration systems. In *ACM Int. Conf. on Management of Data (SIGMOD)*, 2008. *Acceptance rate: 17.9%*.

Xin Dong, Alon Y. Halevy and Cong Yu: Probabilistic schema matching. In *the 33<sup>rd</sup> Int. Conf. on Very Large Data Bases (VLDB)*, 2007.

Michael Cammarano, Xin Dong, Bryan Chan, Jeff Klingner, Joel Galenson, Alon Y. Halevy and Pat Haranhan: Visualization of heterogeneous data. In *IEEE InfoVis*, 2007.

Xin Dong and Alon Y. Halevy: Indexing Dataspaces. In *ACM Int. Conf. on Management of Data (SIGMOD)*, 2007. *Acceptance rate: 14.4%*.

Jayant Madhavan, Shawn R. Jeffery, Shirley Cohen, Xin Dong, David Ko, Cong Yu and Alon Y. Halevy: Web-scale data integration: You can only afford to pay as you go. In *the 3<sup>rd</sup> Conf. on Innovative Data Systems Research (CIDR)*, 2007.

Xin Dong, Alon Y. Halevy and Jayant Madhavan: Reference reconciliation in complex information spaces. In *ACM Int. Conf. on Management of Data (SIGMOD)*, pages 85-96, 2005. *Acceptance rate: 15.1%*.

Xin Dong and Alon Y. Halevy: A platform for personal information management and integration. In *the 2<sup>nd</sup> Conf. on Innovative Data Systems Research (CIDR)*, pages 119-130, 2005.

Xin Dong, Alon Y. Halevy, Jayant Madhavan, Ema Nemes and Jun Zhang: Similarity search for web services. In *the 30<sup>th</sup> Int. Conf. on Very Large Data Bases (VLDB)*, pages 372-383, 2004. *Acceptance rate: 16%*.

Xin Dong, Alon Y. Halevy and Igor Tatarinov: Containment of nested XML queries. In *the 30<sup>th</sup> Int. Conf. on Very Large Data Bases (VLDB)*, pages 132-143, 2004. *Acceptance rate: 16%*.

### Invited Journal Articles

Xin Dong, Alon Y. Halevy and Cong Yu: Probabilistic schema matching. *The International Journal on Very Large Data Bases (VLDBJ)*, the **Special Issue of VLDB Journal “Best Papers of VLDB 2007”**, 5(10), 2008.

Jayant Madhavan, Shirley Cohen, Xin Dong, Alon Y. Halevy, Shawn Jeffery, David Ko and Cong Yu: Structured data meets the Web: A few observations. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 29(4): 19-26, 2006.

Xin Dong, Jayant Madhavan and Alon Y. Halevy: Mining structures for semantics. *SIGKDD Explorations*, 6(2):53-60, 2004.

Igor Tatarinov, Zachary G. Ives, Jayant Madhavan, Alon Y. Halevy, Dan Suciu, Nilesh N. Dalvi, Xin Dong, Yana Kadiyska, Gerome Miklau and Peter Mork: The Piazza peer data management project. *SIGMOD Record* 32(3): 47-52, 2003.

### Refereed Workshop Papers

Jing Liu, Xin Dong and Alon Y. Halevy: Answering structured queries on unstructured data. In *the Int. Workshop on the Web and Databases (WebDB)*, pages 25-30, 2006.

Xin Dong: A Platform for personal information management and integration. In *the VLDB 2005 PhD Workshop*, pages 26-30, 2005.

Xin Dong and Alon Y. Halevy: Malleable schemas: A preliminary report. In *the Int. Workshop on the Web and Databases (WebDB)*, pages 139-144, 2006.

Xin Dong, Alon Y. Halevy, Ema Nemes, Stephan B. Sigurdsson and Pedro Domingos: SEMEX: Mining for personal information integration. In *Workshop on Mining for and from the Semantic Web (MSW)*, pages 58-63, 2004.

Xin Dong, Alon Y. Halevy, Ema Nemes, Stephan B. Sigurdsson and Pedro Domingos: SEMEX: Toward on-the-fly personal information integration. In *Workshop on Information Integration on the Web (IIWEB)*, 2004.

### Conference Demos

Yuhan Cai, Xin Dong, Alon Y. Halevy, Jing Liu and Jayant Madhavan: Personal information management with SEMEX. In *ACM Int. Conf. on Management of Data (SIGMOD)*, pages 921-923, 2005. (**BEST DEMO**, one of the top 3 demos)

Philip Bohannon, Xin Dong, Sumit Ganguly, Henry F. Korth, Chengkai Li, P. P. S. Narayan and Pradeep Shenoy: ROLEX: Relational on-line exchange with XML. In *ACM Int. Conf. on Management of Data (SIGMOD)*, page 673, 2002.

### Technical Reports

Liu, Xin Dong and Alon Y. Halevy: Answering structured queries on unstructured data. Technical Report UW-CSE-06-06-03, University of Washington, 2006.

Xin Dong, Alon Y. Halevy and Igor Tatarinov: Containment of nested XML queries. Technical Report UW-CSE-03-12-05, University of Washington, 2003.

### Publications on Graphics and Image Processing

Xin Dong and Xiao'ou Chen. Automatic improvement of color-image quality oriented to webpage browsing. *Journal of Computer Assisting Design and Graphics*, 2002. (In Chinese)

Xin Dong, Xiao'ou Chen and Yuqian Wu. The research and implementation of an SVG-based webpage design system. In *National Conference on Multi-Media Technology*, China, October 2001. (In Chinese)

Xin Dong and Xiao'ou Chen. Choosing the best markup languages for homepages. *Journal of Computer Applications*, 2001. (In Chinese)

Zhong Wang, Xin Dong and Xiao'ou Chen. SVG—A new language specification supporting scalable vector graphics for web browsing. *Journal of Image and Graphics*, December 2000. (In Chinese)

Xin Dong and Xiao'ou Chen. The display and browsing of XML files. *Journal of Computer Applications*, August 2000. (In Chinese)

Xin Dong. The kernel technique in building a WWW webpage system. In *Int. Computer Network and Application Forum*, 1998. (In Chinese)

Xin Dong. Building a WWW server with IIS. In *Int. Computer Network and Application Forum*, 1998. (In Chinese)

## **Invited Talks**

“Data Integration with Uncertainty,” University of Pennsylvania, Database Group, Philadelphia, PA, Nov 2008.

“Data Integration with Uncertainty,” Stevens Institute of Technology, Computer Science Dept, Hoboken, NJ, Sept 2008.

“Data Integration with Uncertainty,” AT&T Labs-Research, Florham Park, NJ, May 2008.

“Managing a Space of Heterogeneous Data,” Renmin University, Database Group, Beijing, China, June 2007.

“Managing a Space of Heterogeneous Data,” University of Wisconsin, Madison WI, February 2007.

“Semex: A platform for personal information management and integration,” Microsoft Research, Adaptive System & Interaction Core Group, Seattle, WA, November 2005.

“Semex: A platform for personal information management and integration,” Microsoft Research at Asia, Beijing, China, June 2005.

## **Patents**

“Minimal difference query and view matching,” Raghav Kaushik, Venkatesh Ganti and Xin Dong. United States Patent 7251646, issued 7/31/2007.

“Method and apparatus for updating XML views of relational data,” Philip L. Bohannon, Xin Dong, Henry F. Korth, Suryanarayan Perinkulam. United States Patent 20050165866, filed Jan 28, 2004, to be issued.

## **Professional Activities and Services**

Proposal reviewer: Review proposals for establishing a resource for integration and sharing of data and information about animal models requested by the National Center for Research Resources (NCRR) and National Institutes of health (NIH) in 2008.

Journal referee: Journal of Data Management Research (Proceedings of VLDB Endowment), The International Journal on Very Large Data Bases (VLDB Journal), ACM Transactions on Database Systems (TODS), Theoretical Computer Science (TCS), ACM Transactions on Internet Technology Journal (TOIT), Transactions on Information Systems (TOIS), Data Mining and Knowledge Discovery Journal, Transactions on Knowledge and Data Engineering (TKDE), Information System (IS)

Conference reviewer: VLDB2009, WIDM2008, CIKM 2008, WebDB 2008, VLDB 2008, WWW 2008.

Conference external reviewer: CIDR 2007, AAAI 2005, WWW 2005, SIGMOD 2005, CIDR 2005, VLDB 2004

Scribe, NSF Personal Information Management Workshop, 2005

Member, Graduate Student Admission Committee, CSE Department, Univ. of Washington, 2004

Volunteer, International Youth Summit, 2000

Vice president, Departmental Student Union, CSE Department, Nankai University, 1995-1997

## Contest Awards

### **Prize in Programming and Algorithm Contests**

One of the three team members representing Nankai University at *ACM 21<sup>st</sup> Int. Collegiate Programming Contest*, ranked the 12<sup>th</sup> in the Asian Region, 1996

**Second prize** in *Tianjin Programming Contest for Undergraduate students*, 2 times: 1995, 1996

**Bronze medal** in *National Informatics Olympic Games for High School Students (NOI)*, 1992

**First prize** in *Tianjin Programming Contest for High School Students*, 5 times: 1990-1994

### **Prize in Mathematics Contests**

**Second prize** in *National Mathematics Competition for Senior High-School Students*, 1993

**First prize** in *National Mathematics Competition for Junior High-School Students*, 1991

**Full score** in *the 5<sup>th</sup>/6<sup>th</sup> American Mathematics Invitational Contests*, 2 times: 1989, 1990

Title of **Excellent Female Youth** in *American Test of Mathematics Ability*, 1989

**Third prize** in *National LuogengHua Golden-Cup Mathematics Invitational Contests*, 1989

**Second prize** in *Tianjin "I Love Math" Mathematics Competition for Elementary-School Students*, 1987

## Personal Information

Citizen: People's Republic of China

Visa status: H1B

Language: Fluent in English and Chinese Mandarin, know Japanese