

# Xin Luna Dong

4030 Wilkie Way  
Palo Alto, CA 94306  
<http://lunadong.com>

Phone: (201) 650-3494  
(408) 898-4402  
E-mail: [lunadong@gmail.com](mailto:lunadong@gmail.com)

---

## Highlights

- Experiences in building both end-to-end **vertical products**, and **horizontal systems** that serve multiple other products.
- Strong research background in **knowledge management, big data integration, data quality, and applied machine learning**. Co-authored book "Big Data Integration"; 65+ papers in top academic conferences and journals; 20+ keynotes/invited-talks/tutorials.
- Proficient in Java and Python, and have coded in C++ and GO.

## Experiences

### Senior Research Scientist, Google Inc.

2013/1-present

- Building a scalable system to collect tail knowledge. The system interacts with ~20 other systems, collecting tail knowledge from raw data and triggering knowledge panels in Google Search. Leading machine-learning efforts in the project and playing a key role in data quality control.
- Invented Knowledge-Based Trust. Tech lead for a system to decide truthfulness of facts and trustworthiness of web sources, and coded the core algorithms. The system serves multiple products, ranging from global and personal knowledge extraction, fact verification, to source recommendation.
- Initiated the Knowledge Fusion research area and contributed to the Knowledge Vault project, which automatically collects knowledge from various types of web contents on billions of web pages, and fuses them into a probabilistic knowledge base containing billions of knowledge facts.
- 300+ news articles regarding our products, systems and papers.

#### **Quotes from Washington Post:**

*"Still, even the possibility of a search engine that evaluates truth is a pretty incredible breakthrough. And it definitely gives new meaning to the phrase 'let me Google that for you.'"*

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

2007/10-2012/12

Manager: Divesh Srivastava

- One of the people who started the Data Fusion research area--resolving conflicts from multiple data sources. Built a data fusion system Solomon, which decides correctness of facts, trustworthiness of data sources, and copying relationship between sources
- Built a Source Selection system that recommends the best sources for data integration
- Researched on various aspects of Entity Reconciliation--linking references to the same real-world entity

### Research Assistant, University of Washington

2002-2007

Advisor: Alon Halevy

- One of the first people researching on Dataspaces--data integration methodology that automatically bootstraps approximate schema-mapping and entity-resolution services, and

evolves the performance over time in a pay-as-you-go fashion. Published a paper in **VLDBJ Special Issue of "Best papers of VLDB 2007"**.

- Built Semex Personal Information Management system, which generates and manages personal knowledge graph from personal data. The system won **Best Demo Award in Sigmod'2005**.
- Built the Woogle Web-Service Search Engine to provide novel mechanisms for searching and composing web services. The paper has 750+ citations.

## Education

**University of Washington** 2001-2007

**Ph.D.**, Computer Science, August 2007

**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)*

## Professional Activities and Services

- PC co-chair: APWeb'16 Industry track, WAIM'15, CIKE'13 Demo Track, WebDB'10, SKG'09.
- Conference PC area chair: Sigmod 2017, Sigmod 2015, ICDE 2013, CIKM 2011.
- Journal Associate editor: TKDE, JDIQ, IEEE Data Engineering Bulletin Special issue.
- Keynotes at WEBDB'16, ICWE'16, NorCal DB Day 2015, WISA'14, WACCK'14, DEOS'14, DESWEB'13, QDB'10.
- Invited talks at Quora Inc. 2015, Stanford Computer Systems Colloquium (EE380) 2015, UC Santa Cruz CSE 2013, QCRI 2012, HKUST CSE 2012, BEWEB 2011, US Census Bureau 2011, etc.

## Publications

**Full list:** [lunadong.com/publications.htm](http://lunadong.com/publications.htm)

### **Selected list**

Xin Luna Dong, and Divesh Srivastava: Big Data Integration. (Synthesis Lectures on Data Management). *Morgan Claypool Publishers*. 2015.

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. Knowledge-based trust: estimating the trustworthiness of web sources. In *VLDB*, 2015.

Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmman, Shaohua Sun, and Wei Zhang. Knowledge Vault: A Web-scale approach to probabilistic knowledge fusion. In *SIGKDD*, 2014.

Xian Li, Xin Luna Dong, Kenneth Lyons, Weiyi Meng, and Divesh Srivastava. Truth finding on the Deep Web: Is the problem solved? In *VLDB*, 2013.

Xin Dong, Alon Y. Halevy and Cong Yu: Data Integration with Uncertainties. In *VLDB*, 2007.

Xin Dong, Alon Y. Halevy and Jayant Madhavan: Reference Reconciliation in Complex Information Spaces. In *SIGMOD* 2005.

Xin Dong, Alon Y. Halevy, Jayant Madhavan, Ema Nemes and Jun Zhang: Similarity Search for Web Services. In *VLDB* 2004.

## **Patents**

“Providing User-Interactive Graphical Timelines,” Xin Dong, Tim Althoff, Kevin Murphy, Safa Alai, Van Dang and Wei Zhang, United States Patent, filed 9/2015, to be issued.

“Method and Apparatus for Exploring and Selecting Data Sources,” Xin Dong and Divesh Srivastava, United States Patent 20130138480, issued 5/30/2013.

“Online Data Fusion,” Xuan Liu, Xin Dong, Ben Chin Ooi and Divesh Srivastava, United States Patent 20130144843, issued 12/5/2011.

“Update Certificates,” Su Chen, Xin Dong, Laks Lakshmana and Divesh Srivastava, United States Patent, filed 9/2010, to be issued.

“Dependency Between Sources in Truth Discovery,” Xin Dong, Laure Berti-Equille and Divesh Srivastava. United States Patent 8190546, issued 5/29/2012.

“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 and Suryanarayan Perinkulam. United States Patent 20050165866, issued 7/28/2005.