

Xin Luna Dong

9845 Willow Rd NE
Redmond, WA
<http://lunadong.com>

Phone: (201) 650-3494
(650) 788-0228
E-mail: lunadong@meta.com

Highlights

- World renowned expert in knowledge graph and data integration. Ten years of experience in building the world's largest generic and product knowledge graphs at Google and Amazon.
- Science leader for end-to-end science innovation cycles including identifying customer needs, inventing state-of-the-art ML techniques, implementing and deploying production features, and launching new customer experiences.
- High visibility in multiple science communities including data mining, NLP, and databases. Co-authored 2 books on knowledge graph and big data integration; co-chaired 3 top-tier conferences; 100+ keynotes/invited-talks/tutorials; 100+ papers in top academic conferences and journals.

Research Interests

Databases, Data Mining, NLP and Machine Learning, with an emphasis on smart virtual assistant, knowledge graph, data integration, data cleaning, graph mining, and search.

Honors

ACM Fellow, 2024
(For "Significant contributions to knowledge graph construction and data integration")

IEEE Fellow, 2024
(For "Significant contributions to knowledge graph construction and data integration")

VLDB Women in Database Research Award, 2023
(For "Significant contributions to knowledge graph construction and data integration")

ACM Distinguished Member, 2018
(For "Significant contributions to data and knowledge integration")

VLDB Early Career Research Contribution Award, 2016
(For "Advancing the state of the art of knowledge fusion")

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

Research Experience

Principal Scientist, Meta

2021/7-Present

- Leading research and AI production for Meta smart virtual assistant as the chief scientist. Supervising works from AI techniques such as dialog system, search, recommendation, computer vision, and federated learning, to underlying quality framework such as metrics

tracking, benchmarking, and tooling. The techniques have been enabling first-generation smart glasses Ray-ban Stories, second-generation smart glasses Ray-ban Meta, and later devices.

- Designed AI architecture for future-generation smart glasses as the chief architect. Seeing through research and development of key AI components including retrieval-augmented generation, multi-modal dialog, contextual AI, proactive recommendation, and on-device learning, to connect dots from long-term research to production.

Senior Principal Scientist, Amazon

2016/7-2021/6

- Founded the Product Graph project to build an authoritative knowledge graph for retail and digital products, grew the team from 3 to 70+ people, and playing a key role in technical vision, architect, and data deliveries. The product graph has been used in 30+ downstream applications across Retail, Amazon Digital, and Alexa, increased Amazon revenue by hundreds of millions of dollars, and supported an industry-first experience.
- Founded and playing a key role in leading the efforts on web-scale knowledge extraction and integration for Alexa knowledge graph. Improved web extraction accuracy from 60% in state-of-the-art to 90%+ to support production data deliverables.
- Growing a science team to conduct cutting-edge research in knowledge extraction, integration, cleaning, and mining. The team has published 20+ papers and 8 tutorials in top-tier conferences in NLP, Web, data mining, and, and filed 1 patent.
- General chair for Amazon Machine Learning Conference in 2017, and area chair for the conference in 2018-2021. Area chair for Amazon Faculty Research Awards.

Senior Research Scientist, Google

2013/1-2016/6

- Building a scalable system to collect long-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 is serving multiple products, ranging from global and personal knowledge extraction, fact verification, to source recommendation. Knowledge-Based Trust was reported in 200+ news articles. 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.'"*
- 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.

Senior Member of Technical Staff, AT&T Research

2007/10-2012/12

- Initiated 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.
- Researching on source selection and developing the ALEXANDER system where the goal is to select from available sources a subset of sources that together balances quality of integrated data and integration cost.
- Researching on record linkage where tolerance to high value variety is required, including linking records incrementally, linking business listings that belong to the same business chain, linking temporal records that describe the same entity over time, and linking records provided by different sources when errors may exist.

- Researching on update certificates in database systems to prevent erroneous updates.

Research Assistant, Univ. of Washington

2002/10-2017/8

- 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 Woogole 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

Dissertation title: Providing best-effort services in Dataspace systems.

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

Board of Trustees of the VLDB Endowment (2017-2023).

Chair of DBCares (2018-2023); member of DBCares (2023-present)

Member of PVLDB Advisory Committee (2018-present), WSDM Best Paper Award Committee 2024, VLDB Conference Strategy Committee (2018-2023), KDD Best Paper Award Committee 2022, KDD 10-Year-Best-Paper Award Committee 2021, TCDE Award Committee 2017-2019, CIKM Best Paper Award Committee 2017.

PC co-chair: KDD 2022 Applied Data Science (ADS) Track, WSDM 2022, VLDB 2021, Sigmod 2018, WAIM 2015.

PC track co-chair: SigKDD Applied Data Science Invited Talks 2022, VLDB Tutorial track 2019, ICDE Industry track 2019, APWeb Industry track 2016, APWeb Distinguished Lecture Series 2014, CIKM Demo track 2013

Workshop co-chair: Sigmod/PODS Ph.D. Symposium 2012/2013, Sigmod New Researcher Symposium 2012/2013, Workshop of Integrating Big data in the DIMACS workshop series 2012, Int. Workshop on Quality of Databases (QDB) 2012, WebDB 2010, Int. Conf. on Semantics Knowledge and Grid (SKG) 2009.

Conference area chair / PC senior member: VLDB 2025, CIKM 2024, Sigmod 2024, WebConf 2023, AAAI 2023, VLDB 2020, IJCAI 2019, CIKM 2018, Sigmod 2017, Sigmod 2015, ICDE 2013, CIKM 2011.

Conference reviewer: SigKDD 2019, PVLDB 2009-2018, Sigmod 2016, Sigmod 2014, EDBT 2013 Industry track, Sigmod 2012, EDBT 2012 Industry track, WAIM 2011, AMW 2011, Sigmod 2011, VLDB Demo 2010, WWW 2010, NTII 2010, ICDE 2010, CIKM 2009, WebDB 2009, VLDB 2009, WIDM 2008, CIKM 2008, WebDB 2008, VLDB 2008, WWW 2008.

Journal Associate editor: IEEE Data Engineering Bulletin--Special issue on "Retrieval Augmented Generation" 12/2024, IEEE Data Engineering Bulletin--Special issue on "Personal Data Management" 12/2023, SigKDD Exploration 2018-present, VLDB Journal 2017-2020, TKDE 2015-2018, IEEE Data Engineering Bulletin--Special issue on "Towards Quality Data with Fusion and Cleaning" 9/2011.

Journal referee: Nature Journal, 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).

Proposal reviewer: Review for NSF in 2011, for Research Resources (NCRR) and National Institutes of health (NIH) in 2008.

Publications

Book

Gerhard Weikum, Xin Luna Dong, Simon Razniewski and Fabian Suchanek. Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases. *Barnes & Noble*. 2021.

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

Book Chapters

Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava: Data Fusion: Resolving Conflicts from Multiple Sources. In "Managing and Mining Uncertain Data." *Springer*, 2013.

Anish Das Sarma, Xin Luna Dong, and Alon Halevy: Uncertainty in Data Integration and Dataspace Support Platforms. In "Advances in Schema Matching and Mapping." *Springer*, 2011.

Anish Das Sarma, Xin Luna Dong, and Alon Halevy: Data Modeling in Dataspace Support Platforms. In "Conceptual Modeling: Foundations and Applications." *Springer*, 2009.

Xin Luna Dong and Divesh Srivastava: XML Indexing. In "Encyclopedia of Database Systems." *Springer*, 2009.

Anish Das Sarma, Xin Luna Dong, and Alon Halevy: Uncertainty in Data Integration. In "Managing and Mining Uncertain Data." *Springer*, 2008.

Tiziana Catarci, Xin Luna 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)

Invited Blogs

Azza Abouzied, Paolo Papotti, Xin Luna Dong, Alkis Polyzotis, Jens Dittrich, Arun Kumar, Peter Bailis. Courting ML: Witnessing the marriage of relational & Web data systems to machine learning. In *ACM Sigmod Blog*, February 14, 2018.

Julia Stoyanovich, Fabian Suchanek, Serge Abiteboul, Oren Etzioni, Divesh Srivastava, Xin Luna Dong, Gerhard Weikum. The elephant in the room: Getting value from big data. All you need to know to build a product knowledge graph. In *ACM Sigmod Blog*, February 8, 2015.

Tutorials

Xin Luna Dong, Seungwhan Shane Moon, Yifan Ethan Xu, Kshitiz Malik, Zhou Yu. Towards next-generation intelligent assistants leveraging LLM techniques. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2023.

Xin Luna Dong, Seungwhan Shane Moon, Yifan Ethan Xu, Kshitiz Malik, Zhou Yu. Next-Generation Intelligent Assistant for AR/VR Devices. In *Int. World Wide Web Conference (WebConf)*, 2023.

Nasser Zalmout, Chenwei Zhang, Xian Li, Yan Liang, Xin Luna Dong. All you need to know to build a product knowledge graph. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2021.

Colin Lockard, Prashant Shiralkar, Xin Luna Dong, Hannaneh Hajishirzi. Multi-modal information extraction from text, semi-structured, and tabular data on the web. In *Int. Conf. On Web Search and Data Mining (WSDM)*, 2020.

Colin Lockard, Prashant Shiralkar, Xin Luna Dong, Hannaneh Hajishirzi. Multi-modal information extraction from text, semi-structured, and tabular data on the web. In *Meeting of the Association for Computational Linguistics (ACL)*, 2020.

Colin Lockard, Prashant Shiralkar, Xin Luna Dong, Hannaneh Hajishirzi. Multi-modal information extraction from text, semi-structured, and tabular data on the web. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2020.

Xin Luna Dong and Theodoros Rekatsinas: Data integration and machine learning: a natural synergy. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2019.

Xin Luna Dong, Christos Faloustos, Andrey Kan, Jun Ma, Subhabrata Mukherjee. Graph and tensor mining: for fun and for profit. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2018.

Xin Luna Dong, Christos Faloustos, Xian Li, Subhabrata Mukherjee, Prashant Shiralkar. Fact checking: theory and practice. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2018.

Xin Luna Dong and Theodoros Rekatsinas: Data integration and machine learning: a natural synergy. In *Int. Conf. on Very Large Data Bases (VLDB)*, 2018.

Xin Luna Dong and Theodoros Rekatsinas: Data integration and machine learning: a natural synergy. In *Int. Conf. on Management of Data (SIGMOD)*, 2018.

Xin Luna Dong. Knowledge curation and knowledge fusion: challenges, models, and applications. In *Alberto Mendelzon International Workshop on Foundations of Data Management (AMW)*, 2017.

Xin Luna Dong and Wang-Chiew Tan: A Time Machine for Information: Looking Back to Look Forward. In *Int. Conf. on Very Large Data Bases (VLDB)*, 2015.

Xin Luna Dong and Divesh Srivastava. Knowledge curation and knowledge fusion: challenges,

- models, and applications. In *Int. Conf. on Management of Data (SIGMOD)*, 2015.
- Xin Luna Dong. Big data integration. In *Web-Age Information Management (WAIM) Summer school*, 2014.
- Xin Luna Dong. From data fusion to knowledge fusion. In *Asia Pacific Web (APWeb)*, 2014.
- Xin Luna Dong and Divesh Srivastava. Big data integration. In *Int. Conf. on Very Large Data Bases (VLDB)*, 2013.
- Xin Luna Dong and Divesh Srivastava. Big data integration. In *Int. Conf. on Data Engineering (ICDE)*, 2013.
- Xin Luna Dong and Divesh Srivastava. Large-Scale Copy Detection. In *Int. Conf. on Data Engineering (ICDE)*, 2012.
- Xin Luna Dong and Divesh Srivastava. Large-Scale Copy Detection. In *Int. Conf. of Database Systems for Advanced Applications (DASFAA)*, 2012.
- Xin Luna Dong and Divesh Srivastava. Large-Scale Copy Detection. In *Int. Conf. on Management of Data (SIGMOD)*, 2011.
- Xin Luna Dong. Data fusion--Resolving data conflicts for integration. In *China National Database Conference (NDBC)*, 2009.
- Xin Luna Dong and Felix Naumann. Data fusion--Resolving data conflicts for integration. In *Int. Conf. on Very Large Data Bases (VLDB)*, 2009.

Publications

Refereed Conference Papers

- Kai Sun, Yifan Ethan Xu, Hanwen Zha, Yue Liu, Xin Luna Dong: Head-to-Tail: How Knowledgeable are Large Language Models (LLM)? A.K.A. Will LLMs Replace Knowledge Graphs? To Appear, 2024.
- Kewei Cheng, Xian Li, Zhengyang Wang, Chenwei Zhang, Binxuan Huang, Yifan Ethan Xu, Xin Luna Dong, Yizhou Sun. Tab-Cleaner: Weakly Supervised Tabular Data Cleaning via Pre-training for E-commerce Catalog. In *Meeting of the Association for Computational Linguistics (ACL)*, 2023.
- Kewei Cheng, Xian Li, Yifan Xu, Xin Luna Dong, Yizhou Sun. PGE: Robust product graph embedding learning for error detection. In *Proceedings of the VLDB Endowment (PVLDB)*, 2022.
- Di Jin, Bunyamin Sisman, Hao Wei, Xin Luna Dong, Danai Koutra. Deep transfer learning for multi-source entity linkage via domain adaptation. In *Proceedings of the VLDB Endowment (PVLDB)*, 2022.
- Xinyang Zhang, Chenwei Zhang, Xian Li, Xin Luna Dong, Jingbo Shang, Christos Faloutsos, Jiawei Han. OA-Mine: Open-world attribute mining for e-Commerce products with weak supervision. In *Int. World Wide Web Conference (WebConf)*, 2022.
- Liqiang Xiao, Jun Ma, Xin Luna Dong, Pascual Martinez-Gomez, Nasser Zalmout, Wei Chen, Tong Zhao, Hao He, Yaohui Jin. End-to-end conversational search for online shopping with utterance transfer. In *Conf. On Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
- Rongmei Lin, Xiang He, Jie Feng, Nasser Zalmout, Yan Liang, Li Xiong, Xin Luna Dong. PAM: Understanding product images in cross product category attribute extraction. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2021.
- Yikun Xian, Tong Zhao, Jin Li, Jim Chan, Andrey Kan, Jun Ma, Xin Luna Dong, Christos Faloutsos, George Karypis, S. Muthukrishnan, Yongfeng Zhang. EX3: Explainable Attribute-aware Item-set Recommendations. In *ACM Conf. Series on Recommendation Systems (RecSys)*, 2021.

- Zhengbao Jiang, Jialong Han, Bunyamin Sisman, Xin Luna Dong. CoRI: Collective relation integration with data augmentation for open information extraction. In *Meeting of the Association for Computational Linguistics (ACL)*, 2021.
- Jun Yan, Nasser Zalmout, Yan Liang, Christan Grant, Xiang Ren, Xin Luna Dong. AdaTag: Multi-attribute value extraction from product profiles with adaptive decoding. In *Meeting of the Association for Computational Linguistics (ACL)*, 2021.
- Xinyang Zhang, Chenwei Zhang, Xin Luna Dong, Jingbo Shang, Jiawei Han. Minimally-supervised structure-rich text categorization via learning on text-rich networks. In *Int. World Wide Web Conference (WebConf)*, 2021.
- Daheng Wang, Prashant Shiralkar, Colin Lockard, Binxuan Huang, Xin Luna Dong, Meng Jiang. TCN: Table convolutional network for web table interpretation. In *Int. World Wide Web Conference (WebConf)*, 2021.
- Zhengyang Wang, Bunyamin Sisman, Hao Wei, Xin Luna Dong, Shuiwang Ji. CorDEL: A contrastive deep learning approach for entity linkage. In *Int. Conf. On Data Mining (ICDM)*, 2020.
- Namyong Park, Andrey Kan, Christos Faloutsos, Xin Luna Dong. J-Recs: Principled and scalable recommendation justification. In *Int. Conf. On Data Mining (ICDM)*, 2020.
- Junheng Hao, Tong Zhao, Jin Li, Xin Luna Dong, Christos Faloutsos, Yizhou Sun, Wei Wang. P-Companion: A principled framework for diversified complementary product recommendation. In *Conf. On Information and Knowledge Management (CIKM)*, 2020.
- Xin Luna Dong, Xiang He, Andrey Kan, Xian Li, Yan Liang, Jun Ma, Yifan Ethan Xu, Chenwei Zhang, Tong Zhao, Gabriel Blanco Saldana, Saurabh Deshpande, Alexandre Michetti Manduca, Jay Ren, Surender Pal Singh, Fan Xiao, Haw-Shiuan Chang, Giannis Karamanolakis, Yuning Mao, Yaqing Wang, Christos Faloutsos, Andrew McCallum, Jiawei Han. AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2020.
- Yuning Mao, Tong Zhao, Andrey Kan, Chenwei Zhang, Xin Luna Dong, Christos Faloutsos, Jiawei Han. Octet: Online catalog taxonomy enrichment with self-supervision. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2020.
- Yaqing Wang, Yifan Ethan Xu, Xian Li, Xin Luna Dong, Jing Gao. Automatic validation of textual attribute values in eCommerce Catalog by learning with limited labeled data. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2020.
- Namyong Park, Andrey Kan, Tong Zhao, Christos Faloutsos, Xin Luna Dong. MultiImport: Inferring node importance in a knowledge graph from multiple input signals. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2020.
- Varun R. Embar, Bunyamin Sisman, Hao Wei, Xin Luna Dong, Christos Faloutsos, Lise Getoor. Contrastive Entity Linkage: Mining Variational Attributes from Large Catalogs for Entity Linkage. In *Conf. Of Automated Knowledge Base Construction (AKBC)*, 2020.
- Giannis Karamanolakis, Jun Ma, Xin Luna Dong. TXtract: Taxonomy-aware knowledge extraction for thousands of product categories. In *Meeting of the Association for Computational Linguistics (ACL)*, 2020.
- Colin Lockard, Prashant Shiralkar, Hannaneh Hajishirzi, Xin Luna Dong. ZeroShotCeres: Zero-shot relation extraction from semi-structured webpages. In *Meeting of the Association for Computational Linguistics (ACL)*, 2020.

- Qi Zhu, Hao Wei, Bunyamin Sisman, Xin Luna Dong, Christos Faloutsos, Jiawei Han. Collective multi-type entity alignment between knowledge graphs. In *Int. World Wide Web Conference (WebConf)*, 2020.
- Wei Zhang, Hao Wei, Bunyamin Sisman, Xin Luna Dong, Christos Faloutsos, David Page. AutoBlock: A Hands-off Blocking Framework for Entity Matching. In *Int. Conf. On Web Search and Data Mining (WSDM)*, 2020.
- Junyang Gao, Xian Li, Yifan Ethan Xu, Bunyamin Sisman, Xin Luna Dong, and Jun Yang. Efficient knowledge graph accuracy evaluation. In *Proceedings of the VLDB Endowment (PVLDB)*, 2019.
- Colin Lockard, Prashant Shiralkar, Xin Luna Dong. OpenCeres: When open information extraction meets the semi-structured web. In *Annual Conf. Of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019.
- Dongxu Zhang, Subhabrata Mukherjee, Colin Lockard, Xin Luna Dong, Andrew McCallum. OpenKI: Integrating Open Information Extraction and Knowledge Bases with Relation Inference. In *Annual Conf. Of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019.
- Xiaolan Wang, Xin Luna Dong, Yang Li, Alexandra Meliou. MIDAS: Finding the right web sources to fill knowledge gaps. In *Int. Conf. on Data Engineering (ICDE)*, 2019.
- Namyong Park, Andrey Kan, Xin Luna Dong, Tong Zhao, and Christos Faloutsos. Estimating node importance in knowledge graphs using graph neural networks. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2019.
- Colin Lockard, Xin Luna Dong, Prashant Shiralkar, Arash Einolghozati. CERES: Distantly Supervised Relation Extraction from the Semi-Structured Web. In *Proceedings of the VLDB Endowment (PVLDB)*, 2018.
- Guineng Zheng, Subhabrata Mukherjee, Xin Luna Dong, Feifei Li. OpenTag: Open Attribute Value Extraction from Product Profiles. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2018.
- Rakshit Trivedi, Bunyamin Sisman, Jun Ma, Christos Faloutsos, Hongyuan Zha, Xin Luna Dong. LinkNBed: Multi-Graph Representation Learning with Entity Linkage. In *Meeting of the Association for Computational Linguistics (ACL)*, 2018.
- Furong Li, Xin Luna Dong, Anno Langen, Yang Li. Knowledge Verification for LongTail Verticals. In *Proceedings of the VLDB Endowment (PVLDB)*, 2017.
- Qi Song, Yinghui Wu, Xin Luna Dong. Mining Summaries for Knowledge Graph Search. In *IEEE Int. Conf. on Data Mining (ICDM)*, 2016.
- Disheng Qiu, Luciano Barbosa, Xin Luna Dong, Yanyan Shen, Divesh Srivastava. DEXTER: Large-Scale Discovery and Extraction of Product Specifications on the Web. In *Proceedings of the VLDB Endowment (PVLDB)*, 2015.
- Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, Wei Zhang. Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources. In *Proceedings of the VLDB Endowment (PVLDB)*, 2015.
- Wenfei Fan, Zhe Fan, Chao Tian, Xin Luna Dong. Keys for Graphs. In *Proceedings of the VLDB Endowment (PVLDB)*, 2015.
- Theodoros Rekatsinas, Xin Luna Dong, Lise Getoor, Divesh Srivastava. Finding Quality in Quantity: The Challenge of Discovering Valuable Sources for Integration. In *Conf. on Innovative Data Systems Research (CIDR)*, 2015.

- Xian Li, Xin Luna Dong, Kenneth B. Lyons, Weiyi Meng, Divesh Srivastava: Scaling up copy detection. In *Int. Conf. on Data Engineering (ICDE)*, 2015.
- Tim Althoff, Xin Luna Dong, Kevin Murphy, Safa Alai, Van Dang, Wei Zhang: TimeMachine: Timeline Generation for Knowledge-Base Entities. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2015
- Xiaolan Wang, Xin Luna Dong, Alexandra Meliou: Data X-Ray: A Diagnostic Tool for Data Errors. In *Int. Conf. on Management of Data (SIGMOD)*, 2015.
- Pei Li, Xin Luna Dong, Songtao Guo, Andrea Maurino, Divesh Srivastava: Robust Group Linkage. In *Int. World Wide Web Conf. (WWW)*, 2015.
- Anja Gruenheid, Xin Luna Dong, Divesh Srivastava: Incremental Record Linkage. In *Proceedings of the VLDB Endowment (PVLDB)*, 2014.
- Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Kevin Murphy, Shaohua Sun, Wei Zhang: From Data Fusion to Knowledge Fusion. In *Proceedings of the VLDB Endowment (PVLDB)*, 2014.
- Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmman, Shaohua Sun, Wei Zhang: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In *Int. Conf. on Management of Knowledge Discovery and Data Mining (SigKDD)*, 2014.
- Ravali Pochampally, Anish Das Sarma, Xin Luna Dong, Alexandra Meliou, Divesh Srivastava: Fusing data with correlations. In *Int. Conf. on Management of Data (SIGMOD)*, 2014.
- Theodoros Rekatsinas, Xin Luna Dong, Divesh Srivastava: Characterizing and selecting fresh data sources. In *Int. Conf. on Management of Data (SIGMOD)*, 2014.
- Mariam Salloum, Xin Luna Dong, Divesh Srivastava, Vassilis J. Tsotras: Online Ordering of Overlapping Data Sources. In *Proceedings of the VLDB Endowment (PVLDB)*, 2013.
- Xin Luna Dong, Laure Berti-Équille, Divesh Srivastava: Data Fusion: Resolving Conflicts from Multiple Sources. In *Int. Conf. on Web-Age Information Management (WAIM)*, 2013.
- Xin Luna Dong, Divesh Srivastava: Compact explanation of data fusion decisions. In *Int. World Wide Web Conf. (WWW)*, 2013.
- Xin Luna Dong, Barna Saha and Divesh Srivastava: Less is More: Selecting Sources Wisely for Integration. In *Proceedings of the VLDB Endowment (PVLDB)*, 2013.
- Xian Li, Xin Luna Dong, Kenneth Lyons, Weiyi Meng and Divesh Srivastava: Truth Finding on Deep Web: Is the Problem Solved. In *Proceedings of the VLDB Endowment (PVLDB)*, 2013.
- Pei Li, Xin Luna Dong, Andrea Maurino and Divesh Srivastava: Linking Temporal Records. In *Int. Conf. on Very Large Data Bases (VLDB)*, pages 956-967, 2011.
- Xuan Liu, Xin Luna Dong, Beng Chin Ooi and Divesh Srivastava: Online Data Fusion. In *Int. Conf. on Very Large Data Bases (VLDB)*, pages 932-943, 2011.
- Su Chen, Xin Luna Dong, Laks V. S. Lakshmanan and Divesh Srivastava: We challenge you to certify your updates. In *Int. Conf. on Management of Data (SIGMOD)*, pages 481-492, 2011.
- Anish Das Sarma, Xin Luna Dong and Alon Y. Halevy: Data integration with dependent sources. In *Int. Conf. on Extending Database Technology (EDBT)*, pages 401-412, 2011.
- Xin Luna Dong, Laure Berti-Equille, Yifan Hu and Divesh Srivastava: Global Detection of Complex Copying Relationships Between Sources. In *Int. Conf. on Very Large Data Bases (VLDB)*, pages 1358-1369, 2010.
- Songtao Guo, Xin Luna Dong, Divesh Srivastava and Remi Zajac: Record Linkage with Uniqueness Constraints and Erroneous Values. In *Int. Conf. on Very Large Data Bases (VLDB)*,

pages 417-428, 2010.

Xin Luna Dong, Laure Berti-Equille and Divesh Srivastava: Integrating Conflicting Data: The Role of Source Dependence. In *Int. Conf. on Very Large Data Bases (VLDB)*, pages 550-561, 2009.

Xin Luna Dong, Laure Berti-Equille and Divesh Srivastava: Truth Discovery and Copying Detection in a Dynamic World. In *Int. Conf. on Very Large Data Bases (VLDB)*, pages 562-573, 2009.

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

Anish Das Sarma, Xin Luna Dong and Alon Y. Halevy: Bootstrapping pay-as-you-go data integration systems. In *ACM Int. Conf. on Management of Data (SIGMOD)*, pages 861-874, 2008.

Xin Dong, Alon Y. Halevy and Cong Yu: Probabilistic schema matching. In *the 33rd Int. Conf. on Very Large Data Bases (VLDB)*, pages 687-698, 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 3rd Conf. on Innovative Data Systems Research (CIDR)*, pages 342-350, 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.

Xin Dong and Alon Y. Halevy: A platform for personal information management and integration. In *the 2nd 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 30th Int. Conf. on Very Large Data Bases (VLDB)*, pages 372-383, 2004.

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

Invited Journal Articles

Xin Luna Dong. Generations of Knowledge Graphs: The crazy ideas and the business impact. In *Proceedings of the VLDB Endowment (PVLDB)*, 2023.

Xin Luna Dong, Hamid Palangi: Industry Leaders in Signal Processing and Machine Learning: Luna Dong. *Inside Signal Processing Newsletter*, December 2021.

Vinay K. Chaudhri, Chaitanya K. Baru, Naren Chittar, Xin Luna Dong, Michael R. Genesereth, James A. Hendler, Aditya Kalyanpur, Douglas B. Lenat, Juan Sequeda, Denny Vrandecic, Kuansan Wang. Knowledge Graphs: Introduction, History and, Perspectives. *AI Magazine* 43(1): 17-29, 2022.

Daniel Abadi, Anastasia Ailamaki, David G. Andersen, Peter Bailis, Magdalena Balazinska, Philip A. Bernstein, Peter A. Boncz, Surajit Chaudhuri, Alvin Cheung, AnHai Doan, Luna Dong, Michael J. Franklin, Juliana Freire, Alon Y. Halevy, Joseph M. Hellerstein, Stratos Idreos, Donald Kossmann, Tim Kraska, Sailesh Krishnamurthy, Volker Markl, Sergey Melnik, Tova Milo, C. Mohan, Thomas Neumann, Beng Chin Ooi, Fatma Ozcan, Jignesh M. Patel, Andrew Pavlo, Raluca A. Popa, Raghu Ramakrishnan, Christopher Ré, Michael Stonebraker, Dan Suciu: The Seattle report on database research. *Commun. ACM* 65(8): 72-79, 2022

Philippe Bonnet, Xin Luna Dong, Felix Naumann, Pinar Tözün: VLDB 2021: Designing a Hybrid Conference. *SIGMOD Record* 50(4): 50-53, 2021

Daniel Abadi, Anastasia Ailamaki, David G. Andersen, Peter Bailis, Magdalena Balazinska, Philip A. Bernstein, Peter A. Boncz, Surajit Chaudhuri, Alvin Cheung, AnHai Doan, Luna Dong, Michael J. Franklin, Juliana Freire, Alon Y. Halevy, Joseph M. Hellerstein, Stratos Idreos, Donald Kossmann, Tim Kraska, Sailesh Krishnamurthy, Volker Markl, Sergey Melnik, Tova Milo, C. Mohan, Thomas Neumann, Beng Chin Ooi, Fatma Ozcan, Jignesh M. Patel, Andrew Pavlo, Raluca A. Popa, Raghu Ramakrishnan, Christopher Ré, Michael Stonebraker, Dan Suciu: The Seattle Report on Database Research. *SIGMOD Record* 48(4): 44-53, 2019

Luciano Barbosa, Valter Crescenzi, Xin Luna Dong, Paolo Merialdo, Federico Piai, Disheng Qiu, Yanyan Shen, Divesh Srivastava: *Big Data Integration for Product Specifications*. *IEEE Data Eng. Bull.* 41(2): 71-81, 2018.

Shazia Wasim Sadiq, Tamraparni Dasu, Xin Luna Dong, Juliana Freire, Ihab F. Ilyas, Sebastian Link, Miller J. Miller, Felix Naumann, Xiaofang Zhou, Divesh Srivastava: Data Quality: The Role of Empiricism. *SIGMOD Record* 46(4): 35-43, 2017.

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, Wei Zhang: Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources. *IEEE Data Eng. Bull.* 39(2): 106-117, 2016.

Xin Luna Dong, Anastasios Kementsietsidis, Wang-Chiew Tan: A Time Machine for Information: Looking Back to Look Forward. *SIGMOD Record* 45(2): 23-32, 2016.

Pei Li, Xin Luna Dong, Andrea Maurino and Divesh Srivastava: Linking Temporal Records. *Frontier Computer Science*, 2012.

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, Nilesch N. Dalvi, Xin Dong, Yana Kadiyska, Gerome Miklau and Peter Mork: The Piazza peer data management project. *SIGMOD Record* 32(3): 47-52, 2003.

Conference Demos

Theodoros Rekatsinas, Amol Deshpande, Xin Luna Dong, Lise Getoor, Divesh Srivastava: SourceSight: Enabling Effective Source Selection. In *ACM Int. Conf. on Management of Data (SIGMOD)*, 2016.

Xiaolan Wang, Mary Feng, Yue Wang, Xin Luna Dong, Alexandra Meliou: Error Diagnosis and Data Profiling with Data X-Ray. In *Int. Conf. on Very Large Data Bases (VLDB)*, 2015.

Pei Li, Haidong Wang, Christina Tziviskou, Xin Luna Dong, Xiaoguang Liu, Andrea Maurino and Divesh Srivastava: CHRONOS: Facilitating History Discovery by Linking Temporal Records. In *Int. Conf. on Very Large Data Bases (VLDB)*, 2012.

Xin Luna Dong, Laure Berti-Equille, Yifan Hu and Divesh Srivastava: SOLOMON: Seeking the Truth Via Copying Detection. In *Int. Conf. on Very Large Data Bases (VLDB)*, pages 1617-1620,

2010.

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.

Refereed Workshop Papers

Daisy Zhe Wang, Xin Luna Dong, Anish Das Sarma, Michael J. Franklin and Alon Y. Halevy: Functional Dependency Generation and Applications in Pay-As-You-Go Data Integration System. In *the Int. Workshop on the Web and Databases (WebDB)*, 2009

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. Sigundsson 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. Sigundsson and Pedro Domingos: SEMEX: Toward on-the-fly personal information integration. In *Workshop on Information Integration on the Web (IIWEB)*, 2004.

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)

Talks

Interview: Xin Luna Dong's Interview at SEMANTiCS 2023.

Podcast: “Building the Product Knowledge Graph at Amazon”. This Week in Machine Learning & AI (TWIML). 2021.

“The Journey to a Knowledgeable Assistant with Retrieval-Augmented Generation (RAG)”

- Keynote at International Conference on Web Information Systems Engineering (WISE). Doha, Qatar, December 2024.
- Keynote at ACL Workshop on Knowledge Graphs and LLMs (KaLLM). Bangkok, Thailand, August 2024.
- Invited talk at WSDM Industry Day. Merida, Mexico, March 2024.
- Keynote at KDD Workshop on Knowledge Augmented Methods for NLP, Long Beach, CA, August 2023.

“Generations of Knowledge Graphs: The Crazy Ideas and the Business”

- Keynote at Semantics Conference, Leipzig, Germany, September 2023.
- VLDB Women in Database Research Award talk at VLDB, Vancouver, CA, August 2023.
- Keynote at North East Database Day, Boston, MA, March 2023.
- Keynote at The Extraction and Knowledge Management (EGC) Conference, Lyon, France, January 2023.

“Next-Generation Intelligent Assistants for AR/VR Devices”

- Keynote at KDD Workshop on Multimodal Learning, Long Beach, CA, August 2023.
- Invited talk at Stanford Data Science Affiliates Program--Graph Learning Workshop, Palo Alto, CA, September 2022.
- Keynote at RecSys Knowledge-Aware Recommender System (KaRS) Workshop, Seattle, WA, September 2022.
- Invited talk at ML Lab of Danish Business Authority, Virtual, September 2022.
- Invited talk at NAACL Workshop on Structured and Unstructured Knowledge Integrating (SUKI), Seattle, WA, July 2022.
- Keynote at SIAM International Conference on Data Mining (SDM), Online, April 2022.
- Invited talk at Trustworthy Data Science and AI Seminar Series, Simon Fraser University, April 2022.

“Zero to One Billion: The Path to a Rich Product Knowledge Graph”

- Keynote at Northwest Database Society Annual Meeting, Seattle, WA, May 2022.
- Keynote at AAAI Workshop on Deep Learning on Graphs: Method and Applications (DGL), Online, February 2022.
- Keynote at WSDM Workshop on Machine Learning on Graphs (MLOG), Online, February 2022.
- Keynote at AKBC Workshop on Unstructured/Structured KBs, Online, October 2021.
- Keynote at SigKDD Workshop on Data Quality Assessment for Machine Learning (DQAML), Online, August 2021.
- Keynote at SigIR Industry Track, Online, July 2021.
- Keynote at European Semantic Web Conference (ESWC), Online, June 2021.
- Lecture at Stanford CS520 Knowledge Graphs--Data Models, Knowledge Acquisition, Inference and Applications, May, 2021

“Ceres: Harvesting Knowledge from Semi-Structured Web”

- Invited talk at SigKDD Workshop on Deriving Insights from User-Generated Text (WIT), Online, August 2021.
- Keynote at Text Analysis Conference--Knowledge Base Population (TAC-KBP), Online, February 2021.
- Invited talk at UCSB NLP Seminar, Santa Barbara, January 2021.
- Keynote at Conference on Information and Knowledge Management (CIKM), Dublin, Iceland, October 2020.
- Invited talk at Automated Knowledge Base Construction Conference (AKBC), Irvine, CA, June 2020.
- Invited talk at Northwest Database Society Annual Meeting, Seattle, WA, January 2020.
- Keynote at SigKDD Workshop on Truth Discovery and Fact Checking: Theory and Practice, Anchorage, AL, August 2019.
- Keynote at SigIR Workshop on Explainable Recommendation and Search (EARS), Paris, France, July 2019.
- Invited talk at Amsterdam Data Science at Sigmod/PODS'2019, Amsterdam, Netherland, July 2019.
- Invited talk at ICML Workshop on Learning with Limited Labeled Data (LLD), New Orleans, LA, May 2019.
- Keynote at SigKDD Workshop on Mining and Learning with Graphs (MLG), London, UK, August 2018.

Self-Driving Product Understanding for Thousands of Categories.

- Keynote at Personal Health Knowledge Graph (PHKG) Workshop at Knowledge Graph Conference (KGC), Virtual, 2021
- Keynote at KDD Workshop on Data-Efficient Machine Learning (DeMaL), Virtual, 2021.
- Invited talk at NorthEast Univ. DATA lab speaker series, Virtual, August 2020
- Keynote at KDD Workshop on KG & E-Commerce (KGEcom), Virtual, August 2020
- Invited talk at NeurIPS Workshop on Knowledge Representation & Reasoning Meets Machine Learning (KR2ML), Vancouver, Canada, December 2019.

Knowledge Graph and Machine Learning: A Natural Synergy.

- Lecture at Stanford CS520 Knowledge Graphs--How should AI explicitly represent knowledge, Virtual, April 2020

Building A Broad Knowledge Graph for Products.

- Keynote at IEEE Int. Conf. On Data Engineering (ICDE), Macau, April 2019.
- Keynote at SigIR Workshop on eCommerce (eCOM), Paris, France, July 2019.
- Invited talk at AI NEXTCon Seattle, WA, 2019.
- Invited talk at EMNLP Workshop on Fact Extraction and VERification (FEVER), Brussels, Belgium, November 2018
- Invited talk at Duke CS Colloquium, Durham, NC, October 2018
- Invited talk at Berkeley RISE lab Seminar, Berkeley, CA, 2018.
- Keynote at VLDB PhD workshop, Rio de Janeiro, Brazil, August 2018
- Invited talk at SigKDD ADS Invited Speakers Series, London, UK, August 2018.

“Challenges and Innovations in Building a Product Knowledge Graph.”

- Invited talk at Applied Data Science Invited Talks at SigKDD, London, UK, August 2018.

- Invited talk at Graph Data Management Experiences & Systems (GRADES) Workshop, Houston, TX, June 2018.
- Invited talk at BigData Innovators Gathering (BIG) at WWW, Lyon, France, April 2018.
- Invited talk at MoDas Workshop, Eilat, Isreal, February 2018.
- Invited talk at Workshop on Knowledge Base Construction, Reasoning, and Mining (KBCOM), Los Angeles, CA, February 2018.
- Invited talk at Northwest DB Day, Seattle, WA, Jan 2018.
- Invited talk at Workshop on Automated Knowledge Based Construction (AKBC), Long Beach, CA, December 2017.

“Leaving No Valuable Data Behind: the Crazy Ideas and the Business.”

- Keynote at Alberto Mendelzon International Workshop on Foundations of Data Management (AMW), Montevideo, Uruguay, June 2017.
- Invited talk at Machine Learning Conference in Seattle, USA, May 2017.
- Invited talk in Distinguished speaker series: Oxford Women in Computer Science, Oxford, UK, March 2017.
- Keynote at Methods to Manag Heterogenous Big Data and Polystore Databases Workshop, Washington DC, USA, December 2016.
- Talk for VLDB Early Career Research Contribution Award, New Delhi, India, September 2016

“How Far Are We from Collecting the Knowledge in the World?”

- Keynote at International Workshop on the Web and Databases (WebDB), San Francisco, CA, June 2016.
- Keynote at International Conference on Web Engineering (ICWE), Lugano, Switzerland, June 2016.
- Invited talk at Microsoft Research, March 2016.
- Invited talk at Allen Institute for Artificial Intelligence, March 2016.

“Knowledge Fusion and Knowledge-Based Trust.”

- Invited talk at Quora, May 2015.
- Invited talk at Stanford Computer Systems Colloquium (EE380), May 2015.
- Invited talk at NorCal DB Day, May 2015

“From Data Fusion to Knowledge Fusion.”

- Keynote at WISA Conference, September 2014.
- Keynote at WACCK Workshop, 2014.
- Keynote at DEOS Workshop, 2014.

“Truth Finding on the Deep Web.”

- Machine learning group, Princeton, NJ, Nov 2012
- Database group, MIT, MA, Oct 2012
- Computer Science Dept, University of Massachusetts, Amherst, MA, Oct 2012
- Database group, Brown University, RH, Oct 2012
- Database group, Columbia University, NY, Oct, 2012
- Qatar Computing Research Institute, Doha, Qatar, June, 2012.
- AT&T Labs Cookie talk, Florham Park, NJ, May 2011.
- Computer Science Dept, Hongkong University of Science and Technology, Hongkong, China, April 2012.

“Develop Your Big Ideas.”

- New-researcher symposium, Sigmod, Athens, Greece, June 2011.

“Solomon: Seeking the Truth Via Copying Detection.”

- Computer Science Dept, Tianjin University of Technology, Tianjin, China, January 2012.
- Computer Science Dept, Univ. of Washington, WA, August 2011.
- Microsoft Research, WA, August 2011.
- BEWEB invited talk, Uppsala, Sweden, March 2011.
- Computer Science Dept, SUNY Binghamton, NY, March 2011.
- QDB keynote talk, Singapore, September 2010.
- AT&T Labs Cookie talk, Florham Park, NJ, August 2010.

“Data Fusion--Resolving Data Conflicts for Integration.”

- NDBC tutorial, Nanchang, China, October 2009.
- Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, September 2009.

“Sailing the Information Ocean with Awareness of Currents: Discovery and Application of Source Dependence.”

- Invited talk at Person Validation and Entity Resolution Conference, US Census Bureau, May 2011.
- Invited talk at ISAT “What’s Data Worth?” Workshop, February 2010.
- Database group, UC Irvine, CA, February 2010.
- NDBC invited talk, Nanchang, China, October 2009.
- SKG tutorial, Zhuhai, China, October 2009.
- Microsoft Research at Asia, Beijing, China, September 2009.
- AT&T Labs Cookie talk, Florham Park NJ, July 2009.

“Data Integration with Uncertainty.”

- SKG panal talk, Zhuhai, China, October 2009.
- Database group, Renmin University, Beijing, China, September 2009.
- Computer Science Dept, New Jersey's Science & Technology University, Newark, NJ, February 2009.
- Database Group, University of Pennsylvania, Philadelphia, PA, November 2008.
- Computer Science Dept, Stevens Institute of Technology, Hoboken, NJ, September 2008.
- AT&T Labs Cookie talk, Florham Park, NJ, May 2008.

“Managing a Space of Heterogeneous Data.”

- Database Group, Renmin University, Beijing, China, June 2007.
- Database Group, University of Wisconsin, Madison WI, February 2007.

“Semex: A platform for personal information management and integration.”

- Adaptive System & Interaction Core Group, Microsoft Research, Seattle, WA, November 2005.
- Microsoft Research at Asia, Beijing, China, June 2005.

Patents

“Similar but Different (SBD): Presenting Item Recommendations in Dynamically Generated Groups with Explanations,” Andrey Kan, Christos Faloutsos, and Xin Dong, United States Patent 10.891.676, issued 1/2021.

“Providing User-Interactive Graphical Timelines,” Xin Dong, Tim Althoff, Kevin Murphy, Safa Alai, Van Dang, and Wei Zhang. United States Patent 20160313876, issued 10/27/2016.

“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, 20130138480, issued 5/30/2013.

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

“Securing Database Content,” Su Chen, Xin Dong, Laks Lakshmana and Divesh Srivastava, United States Patent 20120036136, 2/9/2012.

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

“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 Exploring and Selecting Data Sources,” Xin Dong and Divesh Srivastava, United States Patent 20050165866, issued 7/28/2005.