

Zhongjun Jin

RESEARCH INTERESTS Query engines and large-scale databases. Interactive data preparation, visualization, integration, and analytics systems.

EDUCATION **University of Michigan**, Ann Arbor, MI, USA **Aug. 2014 - Aug. 2020**
Ph.D., Computer Science and Engineering

- Advisor: Prof. Michael Cafarella and Prof. H. V. Jagadish
- Dissertation: Democratizing Self-Service Data Preparation through Example Guided Program Synthesis

Purdue University, West Lafayette, IN, USA **Aug. 2011 - May 2014**
B.S. in Computer Science, Mathematics

Tianjin University, Tianjin, China **Aug. 2009 - Jul. 2011**
Electronic Information Science

PROFESSIONAL EXPERIENCE **TikTok**, Mountain View, CA **Jul 2020 - Now**
Senior Software Engineer
Query Engine and Large-scale Database Technologies

Microsoft Research, Redmond, WA **Feb 2019 - May 2019**
Research Intern (Mentored by Yeye He)
Designed and implemented a system (like the seminal **FlashFill** system) recommending string data transformation programs which standardize or normalize regular-expression-like data patterns for a given set of string data (like phone numbers, or dates) with heterogeneous data patterns. Unlike most existing systems, which synthesize programs in real time, the recommended programs in our system are learned offline from a large corpus of web data and of higher quality. The technology has been patented in Dec. 2021 (Patent # US 20210374134A1 12/2021).

Trifacta, San Francisco, CA **May 2017 - Sep. 2017**
Software Engineering Intern (Mentored by Sean Kandel, Michael Minar, and Joe Hellerstein)
Designed and implemented **CLX**, an interactive data cleaning system. **CLX** 1) automatically identifies regular-expression-like data patterns for a given set of string data with heterogeneous data patterns for non-expert users to understand, and 2) suggests pattern-based transformation programs to unify various data patterns. The work was integrated to Trifacta Cloud Wrangler as a main feature in Aug 2018 and available at <https://cloud.trifacta.com/>.

CONFERENCE AND WORKSHOP PAPERS

1. Identifying Insufficient Data Coverage for Ordinal Continuous-Valued Attributes
Abolfazl Asudeh, Nima Shahbazi, **Zhongjun Jin**, H. V. Jagadish
Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data (SIGMOD 2021), Xi'an, China, 2021
2. Auto-Transform: Learning-to-Transform by Patterns
Zhongjun Jin, Yeye He, Surajit Chauduri
Proceedings of the VLDB Endowment 13 (11) (VLDB 2020), Tokyo, Japan, 2020
3. Duoquest: A Dual-Specification System for Expressive SQL Queries
Christopher Baik, **Zhongjun Jin**, Michael Cafarella, and H. V. Jagadish

- Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data* (SIGMOD 2020), Portland, OR, 2020
4. MithraCoverage: A System for Investigating Population Bias for Intersectional Fairness
Zhongjun Jin, Mengjing Xu, Chenkai Sun, Abolfazl Asudeh, H. V. Jagadish
Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD 2020), Portland, OR, 2020
 5. Constructing Expressive Relational Queries with Dual-Specification Synthesis.
Christopher Baik, **Zhongjun Jin**, Michael Cafarella, H. V. Jagadish.
10th Conference on Innovative Data Systems Research (CIDR 2020), Amsterdam, Netherlands, 2020
 6. Disambiguating Queries in Conversational Interface.
Christopher Baik, **Zhongjun Jin**, Michael Cafarella.
Proceedings of the Conversational Access to Data Workshop (CAST @ VLDB 2019), Los Angeles, 2019
 7. Assessing and Remediating Coverage for a Given Dataset.
Abolfazl Asudeh, **Zhongjun Jin**, H. V. Jagadish.
35th IEEE International Conference on Data Engineering (ICDE 2019), Macau, China, 2019
 8. CLX: Towards verifiable PBE data transformation.
Zhongjun Jin, Michael Cafarella, H. V. Jagadish, Sean Kandel, Michael Minar, Joseph M. Hellerstein.
22nd International Conference on Extending Database Technology (EDBT 2019), Lisbon, Portugal, 2019
 9. Demonstration of a Schema Mapping System Using Multiresolution Constraints.
Zhongjun Jin, Christopher Baik, Michael Cafarella, H. V. Jagadish, Yuze Lou.
9th Biennial Conference on Innovative Data Systems Research (CIDR 2019), Asilomar, CA, 2019
 10. Beaver: Towards a Declarative Schema Mapping.
Zhongjun Jin, Christopher Baik, Michael Cafarella, H. V. Jagadish.
Proceedings of the 3rd Workshop on Human-In-the-Loop Data Analytics (HILDA @ SIGMOD 2018), Houston, TX, 2018
 11. Foofah: Data Transformation By Example.
Zhongjun Jin, Michael R Anderson, Michael Cafarella, H. V. Jagadish.
Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data (SIGMOD 2017), Chicago, IL, 2017
 12. Foofah: A Programming-By-Example System for Synthesizing Data Transformation Programs. (demo, selected as Best of Demos)
Zhongjun Jin, Michael R Anderson, Michael Cafarella, H. V. Jagadish.
Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data (SIGMOD 2017), Chicago, IL, 2017
 13. Privacy Preserving Access Control in Service-Oriented Architecture.
Rohit Ranchal, Bharat K. Bhargava, Ruchith Fernando, Hui Lei, **Zhongjun Jin**.
IEEE International Conference on Web Services (ICWS 2016), San Francisco, CA, 2016.
 14. A Self-Cloning Agents Based Model for High-Performance Mobile-Cloud Computing.
Pelin Angin, Bharat Bhargava, **Zhongjun Jin**.

Cloud Computing, 2015 IEEE 8th International Conference (CLOUD 2015), New York, 2015.

PATENTS

1. Automatic transformation of data by patterns.
Yeye He, Surajit Chaudhuri, **Zhongjun Jin**.
Patent # WO 2021/242435 12/2021

HONORS AND AWARDS

- 1st Place in “Systems, Software Engineering and Computer Science” session in *Michigan Engineering Graduate Symposium 2017 (EGS 2017)*, 2017.
- Selected as “Best of Demos” at SIGMOD 2017.
- Sigmod Travel Award, 2017.
- University of Michigan Departmental PhD Fellowship, 2014.
- Outstanding Undergraduate Research Endeavor Award, Purdue Computer Science Dept, 2014
- Purdue Computer Science Neel Memorial Scholarship, 2013
- Purdue Computer Science Departmental Scholarship, 2012

INVITED TALKS

- “Intelligent Self-service Data Preparation: Problems and Solutions”, 11/15/2018, Llamasoft Inc., USA.

SERVICE

- Program Committee: ICDE’23
- Reviewer: SoCC’19

REFERENCES

Michael Cafarella <i>Associate Professor</i> University of Michigan EECS Department michjc@umich.edu	H. V. Jagadish <i>Professor</i> University of Michigan EECS Department jag@umich.edu
---	---