

Rising Star Award co-Winner

I am very honored to win this prestigious award. I would like to take this opportunity to thank TCDE committee members, award committees, and especially my nominators, for recognizing my research contributions to human-in-the-loop data analytics. I also want to thank my research collaborators and my graduate students. Without them, I would not be able to achieve what I achieved today. In particular, I want to thank five people, who have made a tremendous impact on my research career in the past ten years.

I did my Ph.D. in the Database Group at Tsinghua University from 2008 to 2013. Profs. Jianhua Feng and Guoliang Li were my Ph.D. supervisors. I want to say thanks to them from the bottom of my heart. They helped me grow from a fresh graduate, who knew little about SIGMOD, VLDB, or ICDE, to an independent researcher, who was able to independently publish a high-quality research paper in these top conferences. My Ph.D. thesis is entitled “CrowdER: Crowdsourcing Entity Resolution”. It discussed how to build a hybrid human-machine Entity Resolution system, and validated that such a hybrid system can achieve both good efficiency and high accuracy compared to machine-only or human-only alternatives. Due to these contributions, the thesis won the China Computer Federation (CCF) Distinguished Dissertation Award (the most prestigious award for CS Ph.D. students in China).

I visited AMPLab at UC Berkeley for six months from 2011 to 2012, and then did a Postdoc for two years and a half from 2013 to 2015. I was so fortunate to have Prof. Michael Franklin as my supervisor and Prof. Tim Kraska as my mentor. They helped me grow from a fresh Ph.D. graduate, who had little experience to lead a large project, to a mature postdoc, who had known how to think big and do influential work. We started the SampleClean project with the vision of scaling up data cleaning and crowdsourcing to big data. I led a team of two Ph.D. students and two undergraduates to develop the SampleClean system. The system was incorporated into the BADS (Berkeley Data Analytics Stack). It helps users to extract value from dirty data, at significantly reduced time and cost.

I joined the School of Computing Science at Simon Fraser University (SFU) as an assistant professor in 2016. Prof. Jian Pei was my mentor. I feel deeply grateful to Jian. He helped me grow from a postdoc, who had no experience in setting up a research lab, to a junior faculty member, who has known how to launch a research lab and manage a group of students. My lab’s mission is to *speed up data science*. We develop innovative technologies and open-source tools for data scientists such that they can turn raw data into actionable insights in a more efficient manner. We have already made a lot of progress on this mission.

- **Data Exploration.** We have built AQP++, an interactive analytics system that enables data scientists to query a large database interactively when they only have limited hardware.
- **Data Enrichment.** We have built Deeper, a data enrichment system that can help data scientists to reduce their time spent on data enrichment with Deep Web from hours to minutes.
- **Data Labeling.** We have built TARS, a label cleaning advisor that can provide valuable advice for data scientists when they need to train or/and test a model using noisy labels.

I demonstrated these systems in my Big Data Science course (<http://tiny.cc/sfu-ds>). Many students found them quite useful and tried to use them in their final projects.

In the past ten years, I feel lucky since I got so much support from so many people in the database community. A new journey has begun. In the next ten years, I hope that I can be the one who helps students grow to what they want to be and leads them to conduct database research that can make a big impact.

Jinnan Wang
Simon Fraser University
Vancouver, Canada