

KEXIN RONG

KACB 3322
266 Ferst Dr NW
Atlanta, GA 30332

<http://kexinrong.github.io/>
kronr@gatech.edu
Revised 08/2023

EMPLOYMENT

Georgia Institute of Technology
Assistant Professor, School of Computer Science 2022-present

VMware Research Group
Affiliated Researcher 2023-present
Postdoctoral Researcher 2021-2022

EDUCATION

Stanford University 2015-2021
Ph.D. in Computer Science.
Advisors: Peter Bailis, Philip Levis

California Institute of Technology 2011-2015
B.S. in Computer Science.
GPA: 4.0

RESEARCH INTERESTS

data management, databases, data-intensive computing, human-centered data science

AWARDS and HONORS

PVLDB Vol16 Distinguished Reviewer, 2023

The Catherine M. and James E. Allchin Early Career Professorship, 2022-2024

ACM SIGMOD Jim Gray Doctoral Dissertation Award Honorable Mention, 2022

Rising Star in EECS, 2020

“Best of SIGMOD 2017” invitation to ACM TODS

PUBLICATIONS

Journal Articles

Clara E. Yoon, Karianne J. Bergen, Kexin Rong, Hashem Elezabi, William L. Ellsworth, Gregory C. Beroza, Peter Bailis, Philip Levis. Unsupervised LargeScale Search for Similar Earthquake Signals. *Bulletin of the Seismological Society of America*, 2019.

Firas Abuzaid, Peter Bailis, Jialin Ding, Edward Gan, Samuel Madden, Deepak Narayanan, Kexin Rong, Sahaana Suri. MacroBase: Prioritizing attention in fast data. *ACM TODS*, 43(4), 12, 2018. “Best of SIGMOD 2017” Special Issue.

Refereed Conference Proceedings

Kexin Rong, Mihai Budiu, Athinagoras Skiadopoulos, Lalith Suresh, Amy Tai. Scaling a Declarative Cluster Manager Architecture with Query Optimization Techniques. *VLDB*, 2023.

Peng Li, Zhiyi Chen, Xu Chu, Kexin Rong. DiffPrep: Differentiable Data Preprocessing Pipeline Search for Learning over Tabular Data. *SIGMOD*, 2023.

Kexin Rong, Yao Lu, Peter Bailis, Srikanth Kandula, Philip Levis. Approximate Partition Selection for Big-Data Workloads using Summary Statistics. *VLDB*, 2020.

Paris Siminelakis*, Kexin Rong*, Peter Bailis, Moses Charikar, Philip Levis. Rehashing Kernel Evaluation in High Dimensions. *ICML*, 2019. (Long talk)

Kexin Rong, Clara Yoon, Karianne Bergen, Hashem Elezabi, Peter Bailis, Philip Levis, Gregory Beroza. Locality-Sensitive Hashing for Earthquake Detection: A Case Study of Scaling Data-Driven Science. *VLDB*, 2018.

Kexin Rong and Peter Bailis. ASAP: Prioritizing Attention via Time Series Smoothing. *VLDB*, 2017.

Peter Bailis, Edward Gan, Samuel Madden, Deepak Narayanan, Kexin Rong, and Sahaana Suri. MacroBase: Prioritizing Attention in Fast Data. *SIGMOD*, 2017.

Peter Bailis, Edward Gan, Kexin Rong, and Sahaana Suri. Prioritizing Attention in Fast Data: Challenges and Opportunities. *CIDR*, 2017.

Conference Demonstrations

Gaurav Tarlok Kakkar, Aryan Rajoria, Myna Prasanna Kalluraya, Ashmita Raju, Jiashen Cao, Kexin Rong, Joy Arulraj. Interactive Demonstration of EVA. *VLDB Demo*, 2023.

Peter Bailis, Edward Gan, Kexin Rong, and Sahaana Suri. Demonstration: MacroBase, A Fast Data Analysis Engine. *SIGMOD Demo*, 2017

Refereed Workshop Proceedings

Justin Chen, Edward Gan, Kexin Rong, Sahaana Suri, Peter Bailis. CrossTrainer: Practical Domain Adaptation with Loss Reweighting. *SIGMOD DEEM Workshop*, 2019.

Thesis

Kexin Rong. *Improving Computational and Human Efficiency in Large-Scale Data Analytics*, Stanford University, 2021. PhD Dissertation.

In Preparation and Under Review

Renzhi Wu, Jingfan Meng, Jie Jeff Xu, Huayi Wang, Kexin Rong. Rethinking Similarity Search: Embracing Smarter Mechanisms over Smarter Data. *Preprint*, 2023.

Amey Agrawal, Sameer Reddy, Satwik Bhattamishra, Venkata Prabhakara Sarath Nookala, Vidushi Vashishth, Kexin Rong, Alexey Tumanov. DynaQuant: Compressing Deep Learning Training Checkpoints via Dynamic Quantization. *Preprint*, 2023.

INVITED TALKS

Towards a Human-Centric Approach to Machine Learning Lifecycle Management at *UCSD Database Lab Research Seminar*, May 2023.

Learned Indexing and Sampling for Improving Query Performance in Big-Data Analytics at *Stanford MLSys Seminar*, April 2022.

Prioritizing Computation and Analyst Resources in Large-scale Data Analytics

University of Waterloo, Jan 2021

Microsoft Research New York City, Feb 2021

Hong Kong University of Science and Technology, Feb 2021

The University of Texas at Austin, Feb 2021

Cornell University, Feb 2021

VMware Research Group, Feb 2021

Georgia Institute of Technology, Feb 2021

Rice University, March 2021

University of Toronto, March 2021

Simon Fraser University, March 2021

University of Michigan, March 2021

The Chinese University of Hong Kong, March 2021

Yale University, March 2021

University of Pennsylvania, March 2021

Microsoft Research Redmond, March 2021

Brown University, March 2021

University of Wisconsin-Madison, April 2021

Automating Dashboard Displays with ASAP at *Monitorama*, May 2017, Portland, OR.

MacroBase: An Analytics Engine for Prioritizing Attention in Fast Data at *the 43rd Asilomar Micro-computer Workshop*, April 2017, Asilomar, CA.

TEACHING

CS 8803-MDS: Human-in-the-loop Data Analytics 2023 Fall
Instructor, Georgia Tech

CS 8803-MDS: Human-in-the-loop Data Analytics 2022 Fall
Instructor, Georgia Tech

ADVISING

Rajveer Bachkaniwala, PhD CS, Georgia Tech, (co-advised with Ada Gavrilovska; 2022-present)

Jie Jeff Xu, PhD CS, Georgia Tech, (2023-present)

SERVICE

Organization

Organizing Committee, EECS Rising Stars Workshop 2023

Program Committee

ACM SIGMOD: 2024, 2023

VLDB: 2023

ACM SoCC: 2023

SIGMOD Student Research Competition: 2023, 2022

SIGMOD Demonstration: 2022

Institute Contributions

SCS Faculty Recruiting Committee: 2023

SCS Ph.D. Admissions Committee: 2022

FUNDING

The Catherine M. and James E. Allchin **Early Career Professorship**, from the College of Computing at the Georgia Institute of Technology, 2022-2024.

Bosch Research **Research Gift**, 2022.

INDUSTRY EXPERIENCE

Research Intern

Microsoft Research

Jun 2019 - Jan 2020

Redmond, WA

- Mentors: Dr. Srikanth Kandula and Dr. Yao Lu
- Worked on a research project that makes novel use of summary statistics to inform block-level sampling for approximate query processing.

Software Engineering Intern

Pinterest

Apr 2015 - July 2015

San Francisco, CA

- Collected training data, built and evaluated models that classify soft 404 pages with high precision.

Server Platform Intern

OpenX

Jan 2015 - Mar 2015

Pasadena, CA

- Evaluated HBase as an alternative to their ad quality database based on MariaDB, and suggested ways to significantly cut down the database volume.

Software Engineering Intern

Facebook Inc.

Jun 2014 - Sep 2014

Menlo Park, CA

- Reduced the update delay for custom audience membership in the Facebook Ads Audience Insights tool from several days down to one hour.

Software Engineering Intern

Lookout Mobile Security

Jun 2013 - Aug 2013

San Francisco, CA

- Built an application crawler for Google Play. Created an internal web service to store, retrieve, edit, delete configurations and schedule tasks for the crawler.