Xiaozhen (Jamie) Liu

2069 Donald Bren Hall, Irvine, CA 92697-3435

xiaozl3@uci.edu xiao-zhen-liu.github.io xiaozhen-liu-jamie xiao-zhen-liu

Education

University of California, Irvine Sep. 2021 - Current

Ph.D. student in Computer Science, member of UCI Information Systems Group (ISG) *a.k.a.* Database Group *Irvine, CA*

Currently working on the Texera project, advised by *Prof. Chen Li* GPA: 4.0/4.0

Southeast University Sep. 2017 – June 2021

Bachelor of Engineering in Computer Science & Technology *Nanjing, China*

Thesis: "Natural Language to SQL Conversion Based on Deep Learning" (*Outstanding Thesis Award*) GPA: 3.8/4.0

Technical Skills & Interests

Languages & Tools: Java, Scala, C/C++, Python, TypeScript, HTML/CSS, SQL, x86 Assembly, LaTeX, and Git.
Libraries & Frameworks: Angular, Yjs, Kubernetes, Apache Spark, Apache Arrow, Apache Kafka, PyTorch, and Tensorflow.
Interests: Big Data Processing, Collaborative Data Analytics, Distributed Systems, Cloud Computing, and Databases in general.

Experience

Graduate Student Researcher & Graduate Teaching Assistant Sep. 2021 – Current

UCI ISG | Bren School of Information & Computer Sciences (ICS) | University of California, Irvine *Irvine, CA*

- Worked as a Graduate Teaching Assistant (TA) or Reader for multiple ICS courses.
- Currently working on **Texera**, an open-source GUI-and-workflow-based collaborative big data analytics service.
- Programmed in **Java**, **Scala**, **Python**, and **TypeScript** to improved Texera's user experience by introducing multiple new features; regularly maintained Texera's production servers.
- Developed a **real-time collaborative** workflow editor for Texera's frontend, enabling a shared-editing experience similar to Google Doc and Overleaf for data analytics workflow construction.
- Enhanced Texera's collaborative capabilities by enabling both shared editing and shared workflow execution; work published at **VLDB 2022** as a demo paper.
- **Mentored** 4 undergraduate students who worked on multiple features in Texera.
- Currently working on Texera's backend, specifically on containerization, multi-tenancy and cluster resource management for big data systems.

Summer Research Intern June 2020 – Sep. 2020

UCI ISG | Bren School of ICS | University of California, Irvine *Remote*

- Developed the initial version of Texera's **Python User-defined Function (UDF) Operator** which allows users to input scripts as part of a workflow; the operator integrated seamlessly into the system with **minimal serialization cost**.
- Explored and used **Apache Arrow** for data serialization and **Arrow Flight RPC** for IPC between JVM and Python.
- Migrated the UDF Operator and **visualization operators** from Texera's old engine to a new **distributed engine**, ***Amber***.

Undergraduate Research Assistant Mar. 2019 – June 2021

Advisor: Prof. Guilin Qi | Knowledge Science & Engineering Lab | Southeast University *Nanjing, China*

- Programmed in Python to do data **cleaning**, **processing**, **analysis**, and **visualization**.
- Explored various ways to improve a state-of-the-art Natural Language to SQL conversion (**NL2SQL**) model, HydraNet, and successfully surpassed the baseline model in performance; implementation done in **PyTorch**.
- Helped the creation of the ***MLPQ*** dataset containing 300K parallel questions in 3 natural languages; work facilitates research on **multi-lingual question answering** methods.
- Explored ML-based QA methods like **IRN** and multilingual knowledge graph embedding models like **MTransE**.
- Proposed and used **Tensorflow** to implement ***MIRN***, a baseline for MLPQ by integrating IRN and MTransE.

Publications

- **Xiaozhen Liu**, Zuozhi Wang, Shengquan Ni, Sadeem Alsudais, Yicong Huang, Avinash Kumar, and Chen Li. "Demonstration of Collaborative and Interactive Workflow-Based Data Analytics in Texera." PVLDB, 15(12), 2022. [PDF] [Video]

Talks

- **Collaborative and Interactive Workflow-Based Data Analytics in Texera**, *ISG Talk* Oct. 2022

Projects

- **Collaborative Workflow Editor**: Led, designed and implemented in **TypeScript** a real-time collaborative workflow editor for Texera's frontend, the first and only open-source implementation of a shared workflow editor. Used the **CRDT**-based shared editing library **Yjs**. Solved core engineering challenges involving **architectural redesign** of the frontend. Implementation details published in **Texera's blog**.
- **PeterDB**: Implemented a database management system (DBMS) in C++ involving a **Record-based File Manager**, a **Relation Manager**, a **B+ Tree** -based **Index Manager** and a pull-based **Query Engine**. The DBMS supports creation and deletion of relations and indexes, insertion, deletion and updating of records, operations like scanning, selection, projection, and **joining** (block-nested loop join, index-nested loop join, and grace-hash join), and querying using SQL.
- **Big Active Inventory**: Created a prototype of a middleware for a city-wise real-time product tracking system based on **Apache Kafka** and Spring, built the frontend based on Angular and Leaflet and ran simulations to prove the high throughput and low latency of the system.
- **MiniCC-ts**: Implemented the intermediate/object code generation and optimizations for a C Compiler written in **TypeScript**.
- **SEULex & SEUYacc**: Implemented in **C++** the scanner generator Lex and the parser generator Yacc; focused on the conversion of regular expressions to NFAs and the generation and visualization of ASTs.

Scholarships & Awards

- **VLDB 2022 NSF Student Travel Award**, *National Science Foundation (NSF)* Sep. 2022
- **CS Travel Grant Award**, *Department of Computer Science, University of California, Irvine* Sep. 2022
- **CS Department Dean's Fellowship**, *Department of Computer Science, University of California, Irvine* 2021-2022
- **Outstanding Graduate**, *Southeast University* June 2021
- **Outstanding Undergraduate Thesis**, *Southeast University* June 2021
- **Goldcard Smart Group's Scholarship**, *Southeast University Education Foundation* 2020-2021
- **Guosheng Scholarship**, *Southeast University Education Foundation* 2019-2020

TAships & Readerships

CS220P: *Databases and Data Management*, Fall 2021 · CS295P: *Keystone Project for Computer Science*, Winter 2022 · ICS51: *Introduction to Computer Organization*, Spring 2022 · CS122A: *Introduction to Data Management*, Fall 2022

Graduate Courses Taken

CS222: *Principles in Data Management* (**A+**) · CS223: *Transaction Processing and Distributed Data Management* · CS230: *Distributed Computer Systems* (**A+**) · CS232: *Computer and Communication Networks* (**A+**) · CS237: *Middleware for Networked and Distributed Systems* · CS256: *Systems and Machine Learning* · CS272: *Statistical Natural Language Processing* · CS273A: *Machine Learning*

Leadership & Civil Engagement

- **Member**, *QT STEM @ UCI (UCI LGBTQ+ Club for STEM students)* 2022 - Current
- **Director**, *New Media Studio (Official social media maintainer of School of CSE, Southeast University)* 2018 - 2019