Ruiying Ma

| | |
ruiying021120@gmail.com

Education

Tsinghua University, Yao Class *Sep 2021 - Jun 2025*
B.Eng in Computer Science and Technology *Beijing, China*

- **GPA**: 3.89/4.00
- **Awards and Honors**:
  - 2025 Honor Class of Microsoft Research Asia

Research Interests

Databases, Computer Systems.

Research Experiences

Caching Design for Evolving Workloads in Computer Systems *Aug 2024 - May 2025*
Advised by Dr. Chieh-Jan Mike Liang, Dr. Francis Y. Yan, and Dr. Yanjie Gao *Microsoft Research Asia*

- Identified that persistent system problems arises from evolving workloads. Highlighted the potential of designing computer systems capable of generating new algorithms to adapt to new workloads.
- Explored innovative data structures and algorithms for the caching problem, a long-lasting and typical computer system problem. Simulated such designs on various real-world workloads, enhancing hit ratio by up to 36.74%.

Document Structures Extraction and Application for Unstructured Databases *Feb 2024 - Aug 2024*
Advised by Prof. Aditya Parameswaran *UC Berkeley*

- Identified the limitations of existing databases in handling documents, a common category of unstructured data. Highlighted the significance of documents' hierarchical structures, akin to tables of contents, in document understanding.
- Extracted hierarchical structures from various PDF documents efficiently, and incorporated them into RAG for robust information retrieval, improving document-based query answering accuracy by up to 20.52%.
- Optimized structure extraction and built a new document dataset with analytical queries for ZenDB.

Publications

[1] Yiming Lin, Madelon Hulsebos, Ruiying Ma, Shreya Shanker, Sepanta Ziegham, Aditya G. Parameswaran, Eugene Wu. (2025). **Querying Templatized Document Collections with Large Language Models**. IEEE 41st International Conference on Data Engineering (ICDE), Hong Kong, 2025, pp. 2422-2435, doi: 10.1109/ICDE65448.2025.00183.

Selected Course Projects

Distributed Systems *Nov. 2023 – Jan. 2024*
Implemented a Nakamoto blockchain prototype.

Networks *Aug. 2023 – Sep. 2023*
Implemented a 3D torus topology with the *clue* routing algorithm.

Databases *Mar. 2023 – Jun. 2023*
Implemented a B+ tree, an executor, a cost-based optimizer, and a lock manager.

Computer Architecture *Sep. 2022 – Dec. 2022*
Implemented forwarding and branch prediction in a 5-stage pipeline; optimized RISCV assembly.