Minguk Choi

mgchoi@dankook.ac.kr | https://min-guk.github.io/

Research Interests

- **Systems for ML**: Resource-efficient Training/Serving on Cloud/Edge, AI Agent
- **ML for Systems**: Learned Index, Learned Cache, Learned Scheduler

Education

Dankook University — Yongin, Korea
– Master of Science in AI-based Convergence — Mar 2023 – Aug 2024

Dankook University — Yongin, Korea
– Bachelor of Science in Software Science — Mar 2017 – Feb 2023
– GPA: 4.1 / 4.5 – **3rd Place** in Department

Publications

International Conference

[SIGMOD 2024] Can Learned Indexes be Built Efficiently? A Deep Dive into Sampling Trade-Offs — **Top Conference**
Minguk Choi, Seehwan Yoo, and Jongmoo Choi — [Paper] [Slides] [Poster] [Code]

International Journal

[Electronics 2023] An Empirical Study of Segmented Linear Regression Search in LevelDB
Ramadhan Agung Rahmat, Minguk Choi, Yoojin Chung, and Jongmoo Choi.

Selected Domestic Conference **(+7 more)**

[KCC 2024] Analysis of RMI Using CPU-Optimized Search Algorithms — **Best Paper Award**
Yeojin Oh, Minguk Choi, Boseung Kim, Yongjie Zhu, Seehwan Yoo, and Jongmoo Choi

[KCC 2024] Breakdown Internal Operations in Updatable Learned Index — **Best Presentation Award**
Suhwan Shin, Minguk Choi, Nakyeong Kim, Seehwan Yoo, and Jongmoo Choi

On-going Research

Towards Optimal Federated Learning Plan under Privacy Constraints: Compile the optimal federated runtime plan for end-to-end ML pipelines (e.g., data preparation, debugging, and training) using a cost model based on the different privacy constraints of data, workers, and tasks in Apache SystemDS

Exploring the Design Space for SIMD Acceleration in Learned Indexes: Introduce novel approaches that accelerate learned indexes by leveraging SIMD and data parallelism in internal operations (e.g., error-bound estimation, model-biased insert). Additionally, it extends the SIMD design space of index structures from horizontal to vertical vectorization.

Experiences

Remote Research Internship — Technische Universität Berlin, Germany
DAMS Lab (Data Management for Data Science Laboratory) — Aug 2024 – Present
– Adviser: Prof. Matthias Boehm

Conducted remote research from Korea based on Apache SystemDS, focusing on compiler and runtime backend for local, distributed, and federated environments. Currently working on a federated learning project under privacy constraints.

Student Researcher — Dankook University, Korea
System Software Laboratory — Jun 2021 – Aug 2024
– Advisor: Prof. Seehwan Yoo and Prof. Jongmoo Choi

Researched Linux Kernel (CPU Scheduler, Memory Allocator, Block I/O Stack), Key-Value Store (Skiplist, KV Cache, Compaction, SST File, Bloom Filter), and Learned & Traditional Index Structures (Sample Learning, Semi-Ordered Structure, SIMD)

Compulsory Military Service — Korean Air Force
Air Force Operations Command Service & Support Group — Jan 2019 – Oct 2020
• Honorably discharged as a sergeant after completing mandatory service

Honors and Awards

Best Research Award, Dankook University — Aug 2024
Top 3 Research Achievements Among PhD and Master's Students in Science at Dankook University

Best Paper Award, Korea Computer Congress 2024 — Jun 2024

Best Presentation Award, Korea Computer Congress 2024 — Jun 2024

Certificate of Appreciation, Korea Computer Congress 2024 — Jun 2024

Academic Excellence Award, Dankook University — Feb 2023
Graduated with high honors (3rd in Software Science department)

Battalion Lead Soldier, Korean Air Force — Mar 2020
Represented and supervised soldiers, and supported recruit adaptation

Services

Availability & Reproducibility Committee, ACM SIGMOD 2024
Evaluated two SIGMOD 2024 artifacts: one for Availability and another for both Availability & Reproducibility

Open-source Contribution

Apache SystemDS, An end-to-end ML system with hybrid local and distributed execution compilation — Contributer
Conducting a Federated Learning project under privacy constraints; implemented the roll function in local, distributed (Spark), and federated versions; fixed incorrect varID reference when requesting federated data; corrected a bug that miscounted element count when writing mtx headers

BASIL, Benchmark for Sampling Applied Learned Indexes — Owner
Created a benchmark that fairly evaluates the build time and lookup performance of 8 state-of-the-art sampling applied learned and traditional indexes on 12 sorted datasets (SIGMOD 2024 Artifact)

LevelDB WIKI, Analysis document about LevelDB (Key-Value Store) — Owner
Co-authored a document with 13 undergraduate students analyzing LevelDB, detailing theory, code, and including results from benchmark experiments and the YCSB tuning contest.

YCSB-CPP, Yahoo! Cloud Serving Benchmark (YCSB) written in C++ — Contributer
Fixed underflow and compilation bugs, and added support for LevelDB property options

Uftrace, Function graph tracer for C/C++/Rust/Python — Wiki Contributer
Introduction to How to Analyze Key-Value Stores (LevelDB, RocksDB) Using uftrace

Teaching Assistant

Operating Systems, Dankook University — Spring 2024
Developed and graded assignments on 8 different CPU scheduler simulations, concurrent data structures (Queue, BST), and Ext2 file system forensics. Coding assignments were auto-graded using Google Test and MOSS. Answered student questions and proctored exams.

Index Structure, Dankook University — Winter 2023
Delivered a 4-hour introductory lecture on traditional and learned index structures. Mentored 5 papers for domestic conferences, resulting in one best paper award and one best presentation award.

System Programming, Dankook University — Fall 2023
Created and graded assignments on file management and shell development (myLs, myCreate, myCat, myCopy, myShell), answered student questions, and proctored exams.

Operating System Practice, Dankook University — Summer 2023
Delivered a 40-hour lecture on operating system. Developed and supervised exercises on basic tools (vim, ssh, gcc, gdb), file handling (myCat, myCreate), concurrency problems (Dining Philosophers, Readers-Writers), and shell implementation. Answered student questions, created, proctored, and graded exams.

Key-value Store, Dankook University — Summer 2022
Delivered a 6-hour introductory lecture on Key-value Stores (LevelDB), mentored 4 papers for domestic conferences, organized a YCSB tuning contest, and co-authored open-source documentation with students.

Language and Technical Skills

Languages: Korean - Native, Englsih - TOEFL 104/120 (RC: 29, LC: 29, SP: 22, WR: 24)
Programming Languages: C/C++, Java, Python

Personal interests

Fitness: Running, Weight lifting
Cooking: Korean Cuisine, Japanese Cuisine, Grilling

References

Prof. Jongmoo Choi choijm@dankook.ac.kr — Software Science, Dankook University, Korea

Prof. Seehwan Yoo seehwan.yoo@dankook.ac.kr — Mobile Systems Engineering, Dankook University, Korea

Prof. Seong-je Cho sjcho@dankook.ac.kr — Software Science, Dankook University, Korea

Prof. Matthias Boehm matthias.boehm@tu-berlin.de — Big Data Engineering, Technische Universität Berlin, Germany