

Yikun Han

yikunhan@umich.edu | [homepage](#) | [linkedin](#) | [github](#)

Education

University of Michigan

Master of Data Science

2023/08 – 2025/05

Ann Arbor, United States

Sichuan University

Bachelor of Information Resources Management

2019/09 – 2023/06

Chengdu, China

- Rank 2/76 - WES GPA 3.87/4.0 - Major GPA 3.95/4.0

Papers Under Review

Yikun Han, Chunjiang Liu, Pengfei Wang. *A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge* [In submission]

Other Papers

A MatSciBERT and LLM-based Approach for Materials Domain Knowledge Graph Construction [In preparation]

Research Experience

Machine olfaction

2023/08 – Now

Supervised by [Ambuj Tewari](#)

[LSA Statistics](#)

- Aiming to find the mapping relationship between molecular structures and odors.
- Pre-training of graph neural network on pyrume dataset and transfer learning between different datasets.

Retrieval-enhanced LLM for patent

2023/06 – Now

Co-supervised by [Chunjiang Liu](#)

[Chinese Academy of Sciences](#)

- Aiming to build a Retrieval-enhanced LLM for patent.
- Incorporated knowledge graphs into the LLM fine-tuning process which enhances the ability to extract domain knowledge.

Aasen's algorithm for solving symmetric matrix functions

2022/05 – 2022/11

Supervised by [Gang Chen](#)

[Tianyuan Mathematical Center](#)

- Optimized the LAPACK function by setting up threads, dichotomizing, adjusting the matrix chunk size, and setting up OpenMP task.
- Parallelized the factorization and back substitution parts of matrices, programmatically split the key steps of the Aasen algorithm and achieved a 40x improvement in computer performance on large-scale matrices.

Knowledge base and publication of retrieval language terms

2021/10 – 2022/01

Supervised by [Wei Fan](#)

- Developed a Chinese knowledge base for information retrieval to integrate into the Linked Open Data Cloud and further improve the theoretical system with Chinese elements and visions.
- Semantically processed lexical entries following the principles coined by Tim Burners-Lee, gave metadata encoding proposals for different cases, and deployed a mirror website of VocBench by using Java and Xshell.

Semantically modelling and publishing lexicon with VocBench

2021/03 – 2021/10

Supervised by [Wei Fan](#)

- Modeled the lexicon using the Simple Knowledge Organization System and published it with a web platform that makes SPARQL semantic search possible by transforming tabular data into RDF metadata.
- Participated in the [DCMI Virtual 2021 Student Forum](#) and shared the research about Linked Data and Semantic Web in progress.

Professional Experience

Data analyst

2022/07 – 2023/03

[Tencent](#)

Shenzhen, China

- Developed more than 5 interactive dashboards and implemented 60+ components with Javascript and SQL, letting users get information without writing queries thus reply about 400% faster.
- Wrote 30+ Python scripts to crawl websites and replace Excel for data pre-processing, increasing the speed by nearly 25 times.

Teaching

STATS 315 / DATASCI 315 Statistics & AI (course development)

Winter, 2024

Community Service & Volunteering

Datawhale

2022/07 – Now

- Led project [video-clip-extraction-by-description](#), deeply involved in projects like [prompt-engineering-for-developers](#), [d2l-ai-solutions-manual](#), [whale-paper](#).
- Wrote installation and implementation tutorials, prepared learning roadmaps, and organized relevant free courses as a teaching assistant for people who didn't have access to AI learning resources, such as [dive into deep learning](#).

STATCOM

2023/09 – 2023/12

- Deeply involved in NLP project [OLHSA](#) (Oakland Livingston Human Service Agency).

Awards

Outstanding Graduate of Sichuan University

2022

Sichuan University Second Prize Scholarship

2022

Outstanding Student of Sichuan University

2021

Outstanding Student of Sichuan University

2020