Pranay Mundra

Email : pmundra@ur.rochester.edu

https://mundrapranay.github.io

Education

University of Rochester — Rochester, NY
Ph.D. in Computer Science — *July 2021 - Present*

University of Washington — Seattle, WA
Bachelor of Science in Mathematics — *September 2017 - June 2021*

Research Interests

Database Systems & Data-Centric Machine Learning(ML); Data Discovery and Mining problems to minimize data bias & improve data quality for ML; Differential Privacy Systems; Problems in number theory, group theory, probability, combinatorics, and graph theory with applications in Computer Science.

Programming Skills

Languages: Python, SQL, Go, Java, C++, JavaScript, Rust

Technologies: AWS, Pytorch, SQL-Server, Boost-Python, React, Faiss

Publications

KOIOS : Top-k Semantic Overlap Set Search: **Pranay Mundra**, Jianhao Zhang, Fatemeh Nargesian, and Nikolaus Augsten. Paper Under Review for ICDE 2023.

Mining approximate acyclic schemes from relations: Batya Kenig, **Pranay Mundra**, Guna Prasaad, Babak Salimi, and Dan Suciu. In Proceedings of the 2020 International Conference on Management of Data, **SIGMOD** Conference 2020, June 14-19, 2020, pages 297–312. ACM, 2020.[**Paper**]

Compositional Generalization with Tree Stack Memory Units: Forough Arabshahi, Zhichu Lu, **Pranay Mundra**, Sameer Singh, Animashree Anandkumar. arXiv Preprint.[**Paper**]

Experience (Research & Industry)

University of Rochester — Rochester, NY
Graduate Research Assistant — *July 2021 - Present*

- Developed **KOIOS**, an **exact, efficient, and generic** filter verification system for **top-k** set similarity search using semantic overlap, where the semantic overlap is the **maximum matching score** of a bipartite graph, where an edge weight between two set elements is defined by an user-defined symmetric similarity function. (Paper under review for ICDE 2023)
- Working on the problem of **selecting a coreset**, i.e, a small sample of data from large datasets, based on two data properties: **coverage & fairness**, to **accelerate** ML model training and **reduce compute** with **minimal drop in performance**.
- Developing a system that would help with **aggregate queries over massive knowledge graphs & open world data** with missing values.
- Advisor : Prof. Fatemeh Nargesian

Caltech Tensorlab — Pasadena, CA
Undergraduate Machine Learning Researcher — *June 2020 - April 2021*

- Worked on the problem of **Neural Program Synthesis**.
- Came up with a **language agnostic** parse tree representation of source code in Java & Python.
- Leveraged **TreeLSTMs** models and the **parse tree representation** to solve **masked node prediction** and used language-specific grammar to get a syntactically correct code snippet.
- Advisors : Animashree Anandkumar and Forough Arabshahi.

University of Washington Database Group — Seattle, WA
Undergraduate Research Assistant — *January - December 2020*

- Worked on **LightDB**, a DBMS that allows querying over all forms of video data, including virtual reality, augmented reality videos.

• Mapped a high-level Python API to the low-level constructs (i) without sacrificing performance, (ii) minimizing device transfer, and (iii) enabling query expression to a wider audience.
• Advisor: Brandon Haynes.

University of Washington Database Group Seattle, WA
Undergraduate Research Assistant *April - September 2019*

• Created **Maimon**, a system for discovering approximate acyclic schemas using **MultiValued Dependencies** from relations.
• Optimized the MVDMiner to **reduce number of file scans**; taking advantage of **Information Theory** to prune out MVDs and calculate the entropies using the already discovered MVDs; compared the performance tradeoff between in-memory database system (H2) and MySQL.
• Paper accepted for **SIGMOD 2020**.
• Advisors : Batya Kenig and Dan Suciu.

Qurb Limited London, UK (worked remotely)
Web Developer Intern *July - September 2018*

• Built **chat bots and web applications** from scratch using MERN stack and Microsoft Bot Framework.
• Shifted NameDrop(written using node & express) to the REACT Framework.
• **Improved the whitelisting feature** for UntrackME; also added a new feature in the application using the Have I Been Pwned API to inform users about the strength of their passwords and whether their credentials were part of a data leak.

Teaching Experience

Department of Computer Science, University of Rochester Rochester, NY
Graduate Teaching Assistant

• Assistant to the Professor for CSC 244/444 - **Knowledge Representation in AI**, (Fall 2022)
• Assistant to the Professor for CSC 263/463 - **Data Management Systems** (Spring 2022)

Paul G. Allen School of Computer Science & Technology Seattle, WA
Undergraduate Teaching Assistant

• Assistant to the Professor for CSE 444 - **Database Systems Internals**, (Winter 2021)
• Assistant to the Professor for CSE 414/344 - **Introduction to Database Systems**, (Fall 2020, Winter 2020, & Spring 2019)

Projects

SimpleDB: Implemented a multi-user transactional database server written in Java.

Husky Map Server: Created a google map for the University of Washington campus, which shows the shortest path between two locations.

Flight Booking Application: Implemented a flight booking service with user management, transaction support, itinerary search & reservations.

Spotify Song Explorer: Web Application that allows visualization of different audio features for Top 50 songs, fetched using the Spotify API.

Relevant Courses

Computer Science (Graduate): Advanced Algorithms; Analytical Methods in Computer Science; Machine Learning, Parallel & Distributed Systems; Computer Networks; Data Mining

Mathematics (Undergraduate): Honors Calculus I, II, III; Real Analysis I, II; Linear Analysis; Probability I, II; Differential Equations, Linear Algebra, Numerical Analysis I, II(Winter 2021); Modern Algebra I, II(Winter 2021); Combinatorial Theory I, II(Winter 2021).

Computer Science (Undergraduate): Computer Programming I, II; Introduction to Database Systems; Database Systems Internals; Data Structures & Algorithms; Linux Fundamentals; Introduction to Artificial Intelligence;