Avinash Kumar

avinask1@uci.edu | https://www.linkedin.com/in/avinash0161 | Sunnyvale, CA
https://github.com/avinash0161

Interests: Scalable Systems and Infrastructure, Cloud Computing, Distributed Systems, Intuitive Applications Design, Big Data Management, Databases

EDUCATION

University of California, Irvine *Sep 17 - Dec 22*

PhD, Computer Science *(Systems and Databases)* **(GPA 4.0/4.0)**

Worked on the Texera project, advised by Prof. Chen Li

Towards Interactive, Adaptive and Result-aware Big Data Analytics, Phd Thesis [PDF]

Indian Institute of Technology (IIT), Roorkee *May 10 - May 15*

Integrated Dual Degree (Bachelors + Masters) in Computer Science and Engineering **(CGPA 8.7/10)**

Dissertation: “System for Mention Detection in Multilingual News Headlines”

COURSES

- **Graduate Courses:** Principles of Data Management, Transaction Processing and Distributed Data Management, Information Retrieval, Software Architecture, Statistical NLP, Artificial Intelligence
- **Undergraduate Courses:** All core undergraduate Computer Science courses

PROFESSIONAL EXPERIENCE

Software Development Engineer, Google Since *Jan 23*

Team: F1 Query Optimizer (Core Data Infrastructure)

Software Development Engineer, Microsoft *July 15 - Aug 17*

Team: Word/Excel/Powerpoint for Android, Kaizala

- Worked on the infrastructure for building and testing of Android WXP apps
- Developed the monitoring and reporting infrastructure for the Kaizala product

INTERNSHIPS

PhD Intern, Google *June 21 - Sep 21*

Team: F1 Query Optimizer (Core Data Infrastructure)

- Enabled foreign key constraints in the logical tree of the query plan
- Used the foreign key constrains to optimize FK-PK joins using techniques such as aggregation push-down

PhD Intern, Facebook *June 20 - Sep 20*

Team: Core Speed (Systems & Infrastructure)

- Worked on the Signal Boosting project for early detection and debugging of performance regressions
- Enabled virtual alpha pipeline and integrated it with Incidents Tracker system

Software Development Engineer, Salesforce *June 19 - Sep 19*

Team: Schema Services (Core Platform)

- Investigated and proposed a solution for the complex and slow running SQL of Salesforce formulas
- Devised and prototyped a weighing strategy for the formula functions

Software Development Engineer, Microsoft *May 14 - Jul 14*

Product: Modern Office Content for Mobile

- Developed Sprightly app which can be used to make presentations, brochures etc. on mobile
- Investigated the possible integration of Sprightly with existing Microsoft cloud-based products

PROJECTS

- **Texera and Amber (since 2017):** Texera is a workflow-based data analytics service that allows people from different technical backgrounds to collaborate and analyze data. We are developing Amber, a debuggable dataflow engine based on the actor model for the purpose. Our experiments show that Amber has performance comparable to Spark and allows for quick interaction and real-time debugging.
- **DBMS (2017):** Built a database management system in C++ consisting of Record-Based File Manager, Relation Manager, Index (B+ tree) Manager and Query Engine. The database supported insertion, deletion and updating of records along with various operations like scanning, selection and projection. The query engine implemented a pull-based approach to execute a chain of operators.
- **Reverse dictionary using Neural Networks (2018):** Implemented the approach suggested by Hill et al 2015 in their paper "Learning to understand phrases by embedding the dictionary". The basic suggested approach is modified in three ways - 1. Expansion of Vocabulary, 2. Using GRU with Attention and 3. Stemming. The project was implemented in TensorFlow and run on Google cloud. The project report can be found at https://github.com/avinash0161/ReverseDictionary/blob/master/reverse-dictionary-neural.pdf.
- **Eventing System for Tippers (2018):** Tippers, an IOT project at UCI, collects data from various sensors across UCI. We developed an eventing system using which developers/users could subscribe to specific events (predicates on the data being collected). Spark Streaming was used to process the incoming data streams and Apache Active MQ pub-sub framework was used to publish the generated events to the subscribers. The project is at https://github.com/avinash0161/cs237_GeneralEventingSystem.

PUBLICATIONS

- A. Kumar, “Towards Interactive, Adaptive and Result-aware Big Data Analytics”. Phd Thesis [PDF]
- Z. Wang, S. Ni, A. Kumar, C. Li, “Fries: Fast and Consistent Runtime Reconfiguration in Dataflow Systems with Transactional Guarantees”. PVLDB 2023 [PDF]
- X. Liu, Z. Wang, S. Ni, S. Alsudais, Y. Huang, A. Kumar, C. Li, “Demonstration of Collaborative and Interactive Workflow-Based Data Analytics in Texera”. PVLDB 2022 [PDF]
- Z. Wang, A. Kumar, S. Ni, C. Li, “Demonstration of Interactive Runtime Debugging of Distributed Dataflows in Texera”. PVLDB 2020 [PDF]
- A. Kumar, Z. Wang, S. Ni, C. Li, “Amber: A Debuggable Dataflow System Based on the Actor Model”. PVLDB 2020 [PDF]
- A. Kumar, D. Patel, N. Jain, “Lightweight System for NE-tagged News Headlines corpus creation”. BigNLP, IEEE Big Data, 2016 [PDF]

OTHER ACHIEVEMENTS

- One of 6 finalists for Facebook PhD fellowship in Database Systems 2021 [Official Announcement Link]

REFERENCES

Dr. Chen Li
Professor
School of ICS
UC Irvine
chenli[at]ics.uci.edu

Dr. Sharad Mehrotra
Professor
School of ICS
UC Irvine
sharad[at]ics.uci.edu