VRAJ SHAH

Phone: (+1) 858-291-2697
Email: vps002@eng.ucsd.edu
Web: linkedin.com/vraj

3230 EBU3B CSE
9500 Gilman Drive
La Jolla, CA 92093

EDUCATION

University of California, San Diego, CA — Sept 2016 - Present
PhD, Computer Science.
Thesis Advisor: Arun Kumar

Indian Institute of Technology, Indore, India — Aug 2012 - June 2016
Bachelor of Technology, Computer Science & Engineering.

EXPERIENCE

Graduate Student Researcher — Sept 2016 - Present
University of California, San Diego

- **SortingHat:** Developed a tool in Python that semi-automates the inference of ML feature types. This allow users to quickly dispose of easy features, and repriortize their effort towards features that need more human attention.
  Project Website: `https://adalabucsd.github.io/sortinghat`

- **SpeakQL:** Developed a system for making spoken SQL querying effective and efficient. The speech driven interface allows the users to query in any domain with infinite vocabulary using interactive query correction.
  Project Website: `https://adalabucsd.github.io/speakql`

- **Hamlet:** Analyzed the accuracy effects of joins on machine learning algorithms, when learning over normalized data to improve performance and usability for analysts.
  Project Website: `https://adalabucsd.github.io/hamlet`

Research Intern — June 2018 - Sept 2018
Microsoft

- Implemented computational graph-level optimizations inside Microsoft's deep learning system for the inference task.

- Implemented analytical cost model for Machine Learning operators in the inference serving system and used learning to predict the operator execution time.

Research Intern — June 2017 - Sept 2017
Infor Corporation

- Worked with the analytics team to integrate machine learning algorithms inside the existing LogicBlox forecasting engine.

Research Assistant — May 2015 - July 2015
University of Alberta, Canada

- Developed a Big Data Adaptor as Eclipse plugin which efficiently handles GitHub's data dump. The tool makes the data more appropriate for a variety of downstream purposes like analytics.

- Developed a tool in python that identifies influential communities in social network platforms like Twitter, GitHub and Stack Overflow to design effective advertisement campaigns.

PUBLICATIONS

Demonstration of SpeakQL: Speech-driven Multimodal Querying of Structured Data.
Vraj Shah, Side Li, Kevin Yang, Arun Kumar, Lawrence Saul.
SIGMOD 2019.

The ML Data Prep Zoo: Towards Semi-Automatic Data Preparation for ML.
Vraj Shah, Arun Kumar.
DEEM Workshop, SIGMOD 2019.

SpeakQL: Towards Speech-driven Multi-modal Querying.
Vraj Shah.
Student Research Competition, SIGMOD 2019.

Are Key-Foreign Key Joins Safe to Avoid when Learning High Capacity Classifiers?
Vraj Shah, Arun Kumar, Xiaojin Zhu.
In Proceedings of Very Large Data Bases Conferences, VLDB 2018.

SpeakQL: Towards Speech-driven Multi-modal Querying.
Dharmil Chandarana, Vraj Shah, Arun Kumar, Lawrence Saul.
In Proceedings of Human In Loop Data Analytics, SIGMOD 2017.

GitHub's Big Data Adaptor: An Eclipse Plugin.
Ali Sajedi, Vraj Shah, Eleni Stroulia.
In Proceedings of 25th Annual International Conference on Computer Science and
Software Engineering. IBM CASCON 2015.

SKILLS	**Languages**: C, C++, Java, Python, R , SQL.
	Web Development: HTML, CSS, JavaScript, PHP.
	Tools and Libraries: Scikit-learn, Keras, Tensorflow, Matlab, Eclipse, Git.

AWARDS

NSF Travel Award to attend VLDB 2018.

Microsoft Travel Award to attend SIGMOD 2017.

University of Alberta Research Experience program honor for poster presentation at
University of Alberta International Research Symposium 2015.

MITACS Globalink Research Award 2015.

DAAD WISE fellowship 2015.

SERVICE   *External Reviewer:* VLDB 2019, VLDB Demo 2018

POSTERS

SpeakQL: Towards Speech-driven Multi-modal Querying of Structured Data.
SoCal DB 2018

Are Key-Foreign Key Joins Safe to Avoid when Learning High Capacity Classifiers?
VLDB 2018, UCSD Research Expo 2017

SpeakQL: Towards Speech-driven Multi-modal Querying.
ACM SIGMOD 2017

GitHub's Big Data Adaptor: An Eclipse Plugin.
University of Alberta Research Symposium 2015

TALKS

Towards Semi-Automatic ML Feature Type Inference
UCSD Database Seminar   May 2019

Are Key-Foreign Key Joins Safe to Avoid when Learning High Capacity Classifiers?
VLDB   August 2018

SpeakQL: Towards Speech-driven Multi-modal Querying.
ACM SIGMOD, UCSD Database Seminar   June 2017

REFERENCES   Arun Kumar
Assistant Professor, University of California, San Diego

arunkk@eng.ucsd.edu

Konstantinos Karanasos
Senior Scientist, Microsoft
kokarana@microsoft.com

Lawrence Saul
Professor, University of California, San Diego
saul@cs.ucsd.edu

Ndapa Nakashole
Assistant Professor, University of California, San Diego
nnakashole@eng.ucsd.edu