

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 <i>PhD</i> , Computer Science. Thesis Advisor: Prof. Arun Kumar	Sept 2016 - Present
	Indian Institute of Technology , Indore, India <i>Bachelor of Technology</i> , Computer Science & Engineering.	Aug 2012 - June 2016
RESEARCH EXPERIENCE	Graduate Student Researcher <i>University of California, San Diego</i>	Sept 2016 - Present
	<ul style="list-style-type: none">• SortingHat: Automate data preparation tasks for AutoML by formalizing them and creating benchmark labeled datasets. This would help objectively validate and improve AutoML platforms. 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 Assistant <i>University of Alberta, Canada</i>	May 2015 - July 2015
INDUSTRY EXPERIENCE	<ul style="list-style-type: none">• 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.	
	Research Intern <i>Microsoft</i>	June 2018 - Sept 2018
	<ul style="list-style-type: none">• 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.	
PUBLICATIONS	Research Intern <i>Infor Corporation</i>	June 2017 - Sept 2017
	<ul style="list-style-type: none">• Worked with the analytics team to integrate machine learning algorithms inside the existing LogicBlox forecasting engine.	
	<i>Towards Benchmarking Feature Type Inference for AutoML Platforms.</i> Vraj Shah , Jonathan Lacanlale, Premanand Kumar, Kevin Yang, Arun Kumar. SIGMOD 2021.	
	<i>SpeakQL: Towards Speech-driven Multimodal Querying of Structured Data.</i> Vraj Shah , Side Li, Arun Kumar, Lawrence Saul. SIGMOD 2020.	

Demonstration of SpeakQL: Speech-driven Multimodal Querying of Structured Data.
Vraj Shah, Side Li, Kevin Yang, Arun Kumar, Lawrence Saul.
SIGMOD (*Demo track*) 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.
VLDB 2018.

SpeakQL: Towards Speech-driven Multi-modal Querying.
Dharmil Chandarana, **Vraj Shah**, Arun Kumar, Lawrence Saul.
HILDA Workshop, SIGMOD 2017.

GitHub's Big Data Adaptor: An Eclipse Plugin.
Ali Sajedi, **Vraj Shah**, Eleni Stroulia.
IBM CASCON 2015.

AWARDS

Second Runner-up, SIGMOD Student Research Competition 2019.
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

Program Committe: VLDB 2022
External Reviewer: VLDB 2019, VLDB Demo 2018

SKILLS

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