Venkata Vamsikrishna Meduri

CONTACT INFORMATION	650 Harry Road San Jose, CA 95120	*Webpage:* `https://vamsikrishna1902.github.io` *E-mail:* Vamsi.Meduri@ibm.com *Ph:* (480) 434-5255

RESEARCH INTERESTS

Human-Database Interaction, Data Integration, Database Query Processing and Optimization, Information Retrieval and Data Mining

CURRENT EMPLOYMENT

IBM Almaden, San Jose, California

Research Scientist   **May 2022 - Present**
Databases and Storage Group
Manager: Dr. Berthold Reinwald

EDUCATION

Arizona State University, Tempe, Arizona, USA

Ph.D., Computer Science (Jan 2015-May 2022)

- Dissertation Title: "Human-in-the-Loop Machine Learning Systems for Data Integration and Predictive Analytics"
- GPA: 3.9 / 4.0
- Advisor: Prof. Mohamed Sarwat

National University of Singapore, Singapore

M.S., Computer Science (Databases), (Jan 2008-Dec 2010)

- Thesis Title: "Exhaustive Reuse of Subquery Plans to stretch Iterative Dynamic Programming for Complex Query Optimization"
- GPA: 4.06 / 5.0
- Advisor: Prof. Kian-Lee Tan

Vellore Institute of Technology, Vellore, India

B.Tech., Computer Science, (2003-2007)

- GPA: 9.26 / 10.0

RESEARCH PROJECTS

Active Learning for Ontology Mapping

We reconcile diverse ontologies into a unified ontology by performing schema mapping. In order to alleviate the need for large amounts of training data, we selectively query a human-in-the-loop via active learning to label informative concept pairs from the ontologies. We evaluate our proposed active learning strategies against state-of-the-art baselines.

Guided Data Analysis for Conversational Business Intelligence

We predict and recommend next query patterns to Business Intelligence (BI) analysts who issue Natural Language (NL) queries on the top of a conversational system. Here we use a high-level abstraction for both the queries and the underlying database schema. The schema is abstracted as a BI ontology graph that is used to create ontology-aware embeddings for NL queries and the corresponding next query search space representation.

AI for Human-Database Interaction

We develop large-scale systems applying concepts from the AI literature such as reinforcement learning (RL) to Human-Database Interaction. We compare state-of-the-art ML techniques for SQL

query prediction during an interaction session against our adaptation of synthesis-based RNNs and Q-Learning.

A Unified Active Learning Framework for Entity Resolution

We build a unified, generalized active learning framework for entity matching that can mix-and-match different learners with applicable example selection algorithms. We compare various example selectors and learners on a variety of quality and latency metrics.

Rule Discovery in Knowledge Bases: RuDiK

We discover rules for the positive and negative relationships between entity pairs in KBs while being robust to their errors and incompleteness. The rules are mined from the training examples by employing scalable, disk-based generation and validation techniques.

Program Synthesis for Entity Matching

We use a powerful technique called *program synthesis* and a solver named SKETCH to generate concise and interpretable boolean expressions (rules) satisfying matching and non-matching assertions on the training data to perform entity matching.

Data Integration of the Electric System Schemata

We integrate real world schemata with a lot of inconsistencies and apply approximate entity matching and schema alignment techniques to reconcile electric system transmission, distribution and location data with diverse format. This project was a collaborative effort between the CASCADE team at ASU and Salt River Project (SRP) which is one of the primary electricity distributors in Arizona.

PRIOR EXPERIENCE

IBM Almaden, San Jose, California

Summer Research Intern   **May-August 2021**
"Active Learning for Ontology Mapping" in the Database group
Mentors: Dr. Abdul Quamar, Dr. Chuan Lei, Dr. Xiao Qin
Manager: Dr. Berthold Reinwald

IBM Almaden, San Jose, California

Summer Research Intern   **June-September 2020**
"Conversational BI Recommendation" in the Database group
Mentors: Dr. Abdul Quamar, Dr. Chuan Lei, Dr. Vasilis Efthymiou
Manager: Dr. Fatma Özcan

IBM Almaden, San Jose, California

Summer Research Intern   **May-August 2017 & 2018**
"Unified Active Learning" in the Scalable NLP group
Mentors: Dr. Lucian Popa, Dr. Min Li, Dr. Prithviraj Sen, *Manager*: Dr. Yunyao Li

SRM Research Institute, Bangalore, India

Member of the Research Group   **April, 2013 - December, 2014**
Participated in the design and development of an E-learning portal, *Atiha*, and an energy management framework for a smart grid using applied machine learning.

Centrum Wiskunde en Informatica, Amsterdam, The Netherlands

Ph.D. Student   **Jan, 2012 - Jan, 2013**
Developed parallel query processing algorithms for pre-clustered cache-conscious column store database on multi-core architecture

National University of Singapore, Singapore

Research Assistant   **Dec, 2010 - Dec, 2011**

Design and development of database query processing algorithms for Phase Change Memory

Oracle India Private Limited, Bangalore, India

Associate Applications Engineer   **July, 2007 - December, 2007**

Developed Web applications in Java for Oracle 10g

CODING SKILLS   Java, Python, C++, and some usage of Shell scripts

COURSEWORK   Statistical Machine Learning (Spring 2015), Semantic Web Mining (Spring 2015), Game Theory (Fall 2015), Database Management Systems Implementation (Spring 2016), Fundamentals of Statistical Learning (Fall 2016), Advanced Topics on Social Media Analysis (Fall 2017)

TEACHING   CSE 511 - Data Processing at Scale (Spring 2021, Instructor)
ASSISTANT   CSE 412 - Database Management Systems (Spring 2017 & 2020)
                    CSE 310 - Data Structures and Algorithms (Fall 2019)
                    CSE 494 - Information Retrieval, Mining and Integration on the Internet (Spring 2015 & 2016)
                    CSE 110 - Principles of Programming in Java (Fall 2015)

AWARDS AND   • Recipient of best reviewer award for the VLDB 2023 PhD Workshop.
SCHOLARSHIPS   • Recipient of Engineering Grad Fellowship from the Ira A. Fulton School of Engineering and the Polytechnic School at Arizona State University for the academic years 2018-19 and 2021-22.
                    • Recipient of a travel award for SIGMOD 2017 from ASU Graduate & Professional Student Association.
                    • Recipient of SIGMOD 2017 Student Travel Award from the conference committee.
                    • Recipient of Research Scholarship from National University of Singapore.
                    • Merit Scholarship holder at Vellore Institute of Technology in the final year of B.Tech coursework for best academic performance.
                    • Secured All India Rank of 570 (97 percentile) in GATE exam conducted for admission to post-graduate courses.
                    • Received a Merit Certificate from Central Board of Secondary Education, New Delhi for being among the top 0.1% successful students of All India Secondary School Exam 2001 in Sanskrit.
                    • Received "Vidya Vikas Award" for being a meritorious student in 7th class final examination in high school.
                    • Received various awards for the recitation of Bhagavad Gita and won prizes from Ramakrishna Mission for the recitation and elocution competitions conducted over consecutive years in English, Telugu and Sanskrit.

PUBLICATIONS   Venkata Vamsikrishna Meduri, Abdul Quamar, Chuan Lei, Xiao Qin, Berthold Reinwald: ALFA: Active Learning for Graph Neural Network-based Semantic Schema Alignment. The VLDB Journal: Special issue on Machine Learning and Databases, Volume 32, Issue 6, Article No.: 4, 2023.

                    Kanchan Chowdhury, Venkata Vamsikrishna Meduri, Mohamed Sarwat: A Machine Learning-Aware Data Re-partitioning Framework for Spatial Datasets. ICDE 2022 (research track)

                    Setu Shah, Venkata Vamsikrishna Meduri, Mohamed Sarwat: GEM: An Efficient Entity Matching Framework for Geospatial Data. ACM SIGSPATIAL 2021 (poster paper)

                    Venkata Vamsikrishna Meduri, Kanchan Chowdhury, Mohamed Sarwat: Evaluation of Machine Learning Algorithms in Predicting the Next SQL Query From the Future. ACM TODS 2021, Volume 46, Issue 1, Article No.: 4

Venkata Vamsikrishna Meduri, Lucian Popa, Prithviraj Sen, Mohamed Sarwat: A Comprehensive Benchmark Framework for Active Learning Methods in Entity Matching. ACM SIGMOD 2020 (research track)

Venkata Vamsikrishna Meduri, Kanchan Chowdhury, Mohamed Sarwat: Recurrent Neural Networks for Dynamic User Intent Prediction in Human-Database Interaction. EDBT 2019 (short paper)

Naser Ahmadi, Viet-Phi Huynh, Venkata Vamsikrishna Meduri, Stefano Ortona, Paolo Papotti: Mining Expressive Rules in Knowledge Graphs. ACM JDIQ Special Issue on Quality Assessment of Knowledge Graphs 2019

Mohamed Sarwat, Venkata Vamsikrishna Meduri: Human-Database Interaction - A Holistic Approach. ICDE 2018 (lightning talk abstract)

Paolo Papotti, Stefano Ortona, Venkata Vamsikrishna Meduri: RuDiK: Rule Discovery in Knowledge Bases. PVLDB 2018 (demo track)

Stefano Ortona, Venkata Vamsikrishna Meduri, Paolo Papotti: Robust Discovery of Positive and Negative Rules in Knowledge-Bases. ICDE 2018 (research track)

Rohit Singh, Venkata Vamsikrishna Meduri, Ahmed Elmagarmid, Samuel Madden, Paolo Papotti, Jorge-Arnulfo Quiane-Ruiz, Armando Solar-Lezama, Nan Tang: Synthesizing Entity Matching Rules by Examples. PVLDB 2017 (research track)

Rohit Singh, Venkata Vamsikrishna Meduri, Ahmed Elmagarmid, Samuel Madden, Paolo Papotti, Jorge-Arnulfo Quiane-Ruiz, Armando Solar-Lezama, Nan Tang: Generating Concise Entity Matching Rules. ACM SIGMOD 2017 (demo track)

Venkata Vamsikrishna Meduri, Paolo Papotti: Towards User-Aware Rule Discovery. ISIP 2016

Sushovan De, Yuheng Hu, Venkata Vamsikrishna Meduri, Yi Chen, Subbarao Kambhampati: BayesWipe: A Scalable Probabilistic Framework for Cleaning BigData. ACM JDIQ Special Issue on Web Data Quality 2016

Lydia Manikonda, Venkata Vamsikrishna Meduri & Subbarao Kambhampati: Tweeting the Mind and Instagramming the Heart: Exploring Differentiated Content Sharing on Social Media. ICWSM 2016 (short paper)

Tathagata Chakraborti, Venkata Vamsikrishna Meduri, Vivek Dondeti & Subbarao Kambhampati: A Game Theoretic Approach to Ad-hoc Coalitions in Human-Robot Societies. AAAI MIPC 2016

Venkata Vamsikrishna Meduri, Zhan Su, Kian-Lee Tan: A Write Efficient PCM-Aware Sort. DEXA (1) 2012: 86-100.

Venkata Vamsikrishna Meduri, Kian-Lee Tan: Subquery Plan Reuse based Query Optimization. COMAD 2011: 35-46.

PATENTS

Abdul Quamar, Fatma Özcan, Chuan Lei, Vasilis Efthymiou, Venkata Vamsikrishna Meduri: Guided Exploration for Conversational Business Intelligence. US Patent App. 17/188,405.
`https://patents.google.com/patent/US20220277031A1/en`

ACADEMIC SERVICE

PC Member

- EDBT Industrial Track 2024
- ICDE Industrial Track 2024
- ACM Symposium on Cloud Computing (SoCC) 2023
- VLDB Industrial Track 2023
- VLDB PhD Workshop 2023
- Tabular Data Analysis (TaDA) Workshop @ VLDB 2023

Journal Reviewer

- VLDB Journal
- Neural Networks (NEUNET)
- Journal of Web Semantics (JoWS)
- ICT Express
- Information Sciences

External Reviewer

- The Web Conference (formerly known as WWW) Industry Track 2024 (served as a subreviewer)
- VLDB Research Track 2018-2020, 2023, 2024
- SIGMOD 2018
- SIGSPATIAL 2018
- HILDA @ SIGMOD 2019
- ICDE 2020

REFEREES
- Prof. Mohamed Sarwat (msarwat@asu.edu)
- Dr. Berthold Reinwald (reinwald@us.ibm.com)
- Dr. Fatma Özcan (fozcan@google.com)
- Dr. Lucian Popa (lpopa@us.ibm.com)
- Dr. Abdul Quamar (abdulquamar@google.com)
- Prof. Paolo Papotti (Paolo.Papotti@eurecom.fr)
- Dr. Dragan Boscovic (dboscovi@asu.edu)
- Mr. Nunn H. Stewart (Stew.Nunn@srpnet.com)