AAMOD KHATIWADA

Boston, Massachusetts +1 (806)-401-2973 khatiwada.a@northeastern.edu https://aamodkh.github.io
https://www.linkedin.com/in/aamod-khatiwada https://scholar.google.com/citations?user=OnkSeCsAAAAJ

RESEARCH INTERESTS

Data Discovery and Integration, Knowledge Graphs, Data lake Management, Large Language Models

EDUCATION

Northeastern University, Khoury College of Computer Sciences, Boston, MA, USA *Sep 2020 - Present*
PhD Candidate, Computer Science, Current GPA: 3.939
Advised by: Dr. Renée J. Miller
Research summary: *Develop novel tools and algorithms to add semantics to the open data tables and web tables, and use such semantics for data management, table cleansing, table discovery and table integration.*

Northeastern University, Khoury College of Computer Sciences, Boston, MA, USA *Sep 2020 - Apr 2023*
Masters in Computer Science

Tribhuvan University, Institute of Engineering (IOE), Kathmandu, Nepal *Nov 2014 - Nov 2018*
Bachelor's Degree in Electronics and Communication Engineering
Distinction, Department Topper

PUBLICATIONS

- **A. Khatiwada**, R. Shraga and R. J. Miller, *"DIALITE: Discover, Align and Integrate Open Data Tables"*, in SIGMOD-Companion, 2023. https://doi.org/10.1145/3555041.3589732
- **A. Khatiwada**, G. Fan, R. Shraga, Z. Chen, W. Gatterbauer, R. J. Miller and M. Riedewald, *"SANTOS: Relationship-based Semantic Table Union Search"*, in Proc. ACM Manag. Data (PACMMOD) 1, 1, Article 9 (May 2023), 2023. https://doi.org/10.1145/3588689
- **A. Khatiwada**, R. Shraga, W. Gatterbauer and R.J. Miller, *"Integrating Data Lake Tables"*, in PVLDB, 16(4):932-945, 2022. https://doi.org/10.14778/3574245.3574274
- **A.Khatiwada**, S. Shirai, K. Srinivas and O. Hassanzadeh, *"Knowledge Graph Embeddings for Causal Relation Prediction"*, in Deep Learning for Knowledge Graphs Workshop (DL4KG@ISWC), 2022. https://ceur-ws.org/Vol-3342/paper-8.pdf
- S. Shirai, **A.Khatiwada**, D. Bhattacharjya and O. Hassanzadeh, *"Rule-Based Link Prediction over Event-Related Causal Knowledge in Wikidata"*, in 3rd Wikidata Workshop (Wikidata@ISWC), 2022. https://ceur-ws.org/Vol-3262/paper14.pdf
- O. Hassanzadeh, P. Awasthy, K. Barker, O. Bhardwaj, D. Bhattacharjya, M. Feblowitz, **A. Khatiwada**, L. Martie, S. F. Mbouadeu, J. Ni, A. Saha, S. Shirai, K. Srinivas and L. Yip, *"Knowledge-Based News Event Analysis Toolkit"*, in ISWC, 2022. https://ceur-ws.org/Vol-3254/paper399.pdf
- **A. Khatiwada**, P. Kadariya, S. Agrahari and R. Dhakal, *"Big Data and Deep Learning Based Sentiment Analysis System for Sales Prediction"*, in IEEE International Conference on Innovating Technology for Humanity, Pune, 2019. pp. 1-6. https://doi.org/10.1109/PuneCon46936.2019.9105719

PATENTS

- O. Hassanzadeh, **A.Khatiwada** and S. Shirai, *"Link Prediction Using an Ensemble of Representations and Rules* (under revision)

TALKS

- **A.Khatiwada** and G. Fan, *"Table Discovery and Integration in Data Lakes: Challenges and Solutions "*, in Northeast Database Day, 2023. https://northeastern-datalab.github.io/nedbday/2023/

EXPERIENCE

Data lab, Northeastern University, Boston, MA, USA *Sept 2020 - Present*
Graduate Research Assistant

- Developed a principled way of using LLMs to generate benchmarks for tabular tasks.
- Led a group project entitled *SANTOS* on finding the unionable open data tables by detecting their semantic types.
- Developing novel techniques and algorithms to integrate open data tables and web tables in a principled way.
- Created open data benchmarks for data discovery tasks.

IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY, USA *May 2023 - Aug 2023*
AI Research Scientist Intern
Mentors: Dr. Udayan Khurrana, Dr. Kavitha Srinivas and Dr. Oktie Hassanzadeh

- Led a project on building large language models for data management tasks.
- Developed systematic ways of creating benchmarks to fine tune large language models for data discovery tasks.
- Developed an algorithm for searching joinable tables from the data lake.

Northeastern University, Boston, MA, USA *Sept 2022 - Dec 2022*
Teaching Assistant (Khoury College of Computer Sciences)

- Contributed to the development of course materials and course projects for a graduate-level course *CS7290: Special Topics in Data Science* (`https://northeastern-datalab.github.io/cs7290.f22/`)
- Taught three lectures covering state-of-the-art systems for data discovery and management.

IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY, USA *May 2022 - Aug 2022*
Exploratory Science Research Intern
Mentor: Dr. Oktie Hassanzadeh

- Led a project on detecting the causal relations in Knowledge graphs using embeddings and GNN-based models.

Tribhuvan University, Kathmandu, Nepal *Nov 2018 - Jan 2020*
Teaching Assistant (Department of Electronics and Computer Engineering)

- Carried out lab classes on C programming, Digital Logic and Big Data Analytics.
- Assisted final year students to design and debug their major projects.

AWARDS AND EXTRA CURRICULAR

- SIGMOD Student Travel Award from ACM SIGMOD. (Apr 2023)
- PhD Startup Fund from Khoury College of Computer Sciences, Northeastern University, MA, USA. (Sept 2020)
- Leadership And Mentorship Program (LAMP) Scholarship from the Dean of the Whitacre College of Engineering, Texas Tech University, TX, USA. (Jan 2020)
- Awarded as **the best undergraduate student** at Electronics and Computer Department, Tribhuvan University, Nepal for three consecutive years (Sophomore, Junior and Senior). (2016, 2017, 2018)

PROFESSIONAL SERVICE

- PC Member at SIGMOD, Availability and Reproducibility, 2023
- PC Member at SemTab Challenge@ISWC, 2023
- Reviewer at SIGIR Conference, 2023
- Member at Computer Science PhD Admission Committee, Northeastern University, 2021

SKILLS

- **Programming:** SPARQL, Python, C, C++, PHP, SQL, JavaScript, Java, Solidity, Assembly Programming
- **Tools and frameworks:** Knowledge Graphs, Large language models, Hadoop MapReduce, Spark, Laravel, jQuery, D3, Tensorflow, PyTorch LaTeX

GRADUATE COURSEWORK

Information Visualization (project: *https://aamodkh.github.io/theta-join-visualization*), Distributed Systems (project: *https://github.com/aamodkh/distributed-datalake-tapestry*), Principle of Scalable Database Management, Advanced Algorithm, Large Scale and Parallel Data Processing, Special Topics in Database Management