Dwija Parikh

+1 (713) 715-9407 | Seattle, WA
dwija@uw.edu | linkedin.com/in/dwija-parikh | dwijap.github.io

EDUCATION

University of Washington Sep 2022 - March 2024
M.S. Computational Linguistics Seattle, WA
Coursework: Shallow & Deep Processing for NLP, Statistical NLP, Syntax Engineering, Phonetics

University of Houston Aug 2017 - May 2022
B.S. Computer Science, B.S. Mathematics (Data Science Option) Houston, TX
Coursework: Advanced NLP, Data Structures & Algorithms, Stochastic Processes

TECHNICAL SKILLS

Languages: Python | R | SQL | Java | C++ | C
Libraries & Technologies: PyTorch | Tensorflow | Huggingface | Scikit-learn | NLTK | SpaCy | Git/Github

EXPERIENCE

Hewlett Packard Data Science Institute Jun 2021 - Aug 2021
Data Science Intern Houston, TX

- Performed data analysis and modeling to deliver insights for optimization of ovarian and prostate cancer treatment processes on a dataset of 65,000 patients' electronic health insurance claims and structured electronic medical records
- Engineered a framework using network analysis tools such as DAGs (Directed Acyclic Graphs), predictive analysis, and community detection to model patient pathways across healthcare providers to visualize treatment narratives across diverse patients
- Designed and optimized data pipelines and workflows for data collection, preprocessing, and analysis for over 1 million data points

RiTUAL Lab at the University of Houston Aug 2018 - Aug 2021
Research Assistant, supervised by Prof. Thamar Solorio Houston, TX

- Implemented part-of-speech tagging and named entity recognition by fine-tuning multilingual BERT on noisy Spanish-English data sourced from Twitter leading to over 91% accuracy for POS tagging and 65% for named entity recognition
- Conducted comprehensive research to investigate language models (BERT, ELMo, and GloVe) in handling code-switched text for language identification and named entity recognition through supervised analysis of learned representations

PROJECTS

Probing Multilingual LLMs for Typological Signals
Master's Thesis, supervised by Prof. Shane Steinert-Threlkeld

- Implementing an in-depth analysis of large pre-trained multilingual language models to probe for phylogenetic and geographic signals, aiming to uncover language-specific features and cross-lingual patterns across diverse languages
- Developing a comprehensive probing methodology, including linguistic probe tasks such as case marking, possession, and morphological tag count, to investigate the capabilities of multilingual language models in capturing typological properties

Ontology-Based Recommender System for E-Commerce
Independent Study, supervised by Prof. Christoph Eick

- Designed and implemented an ontology-based recommender system for an e-commerce website, leveraging Graph Neural Networks (GNNs) to enhance the accuracy of product recommendations achieving an accuracy rate of 72%
- Constructed a comprehensive ontology that captured domain-specific knowledge and relationships between products, enabling an understanding of user preferences and item characteristics thereby improving customer engagement

PUBLICATIONS

Normalization and Back-transliteration for Code-Switched Text, *CALCS (NAACL 2021)*
Dwija Parikh and Thamar Solorio

- Developed a preprocessing module specifically designed for code-switched data, utilizing a hybrid approach that combined rule-based phonemic transcription methods with machine learning techniques, including a seq2seq model employing LSTM networks, resulting in an accuracy rate of 78.6%
- Engineered a novel grapheme-to-phoneme (G2P) conversion technique specifically tailored for Romanized Hindi data, enhancing the processing and analysis of code-switched text in social media contexts
- Contributed to the field by releasing a valuable dataset of script-corrected Hindi-English code-switched sentences, meticulously labeled for named entity recognition and part-of-speech tagging tasks, fostering further advancements in code-switching research