Dwija Parikh

dwija@uw.edu | 713-715-9407 | linkedin.com/in/dwija-parikh

EDUCATION

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

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

TECHNICAL SKILLS

Languages: Python | R | SQL | Java | C++ | C
Libraries & Technologies: PyTorch | scikit-learn | tensorflow | pandas | NumPy | HuggingFace | NLTK | spaCy | matplotlib | regex | Git

EXPERIENCE

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

- Conducted data analysis and engineered a framework using network analysis tools (DAGs) to model cancer patient pathways delivering insights to optimize cancer treatment processes on a dataset of 65,000 patients' health insurance claims
- Contributed to the design and optimization of data pipelines and workflows for data collection, preprocessing, and analysis
- Collaborated with a multidisciplinary team to integrate the framework into healthcare systems, improving the understanding and care of cancer patients

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

- Conducted research in multilingual Natural Language Processing (NLP), specializing in language models such as M-BERT and ELMo
- Implemented named entity recognition and POS tagging of noisy Spanish-English and Hindi-English data sourced from Twitter
- Conducted comprehensive research to investigate the efficacy and robustness of transformer-based language models in handling noisy code-switched text

RESEARCH & PROJECT

Probing Large Pre-trained Multilingual Language Models for Typological Signals
Master's Thesis supervised by Prof. Shane Steinert-Threlkeld

- Conducted an in-depth analysis of large pre-trained multilingual language models to probe typological signals, aiming to uncover language-specific features and patterns across diverse languages
- Developed and implemented a comprehensive probing methodology, including novel linguistic tasks and evaluation metrics, 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
- Constructed a comprehensive ontology that captured domain-specific knowledge and relationships between products, enabling an understanding of user preferences and item characteristics
- Employed GNNs to model and learn complex interactions within the product ontology, achieving a accuracy rate of 72% in generating personalized recommendations for users, 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 within NLP