Heikal Badrulhisham

Port Moody, Greater Vancouver, BC • heikal93@gmail.com • +1604-652-8894 • github.com/heikalb

Skills

Python, TensorFlow, scikit-learn, NLTK, spaCy, pandas, Git, Java, Julia, HTML, CSS, JavaScript, Django

General skills: natural language processing, machine learning, web scraping, data cleaning, visualization, statistics, documentation, technical communication, team/pair programming

Coursework: programming, data structures, artificial intelligence, optimization, research techniques and experimental design, natural language processing

Experience

Corpus linguistics thesis project
Simon Fraser University, 2018/09 – 2019/11

- Designed a pipeline to process corpus data, including data retrieval, data cleaning, morphological parsing and deriving statistical association data.
- Wrote programs in Python and Java to carry out corpus data processing, including a program to query and download 1.49 million data points from a corpus of Turkish by interacting with a web browser over 7 hours without crashing.
- Applied transformations on language data to optimize the performance of a morphological parser.
- Used a Python script to slice association data to uncover suffix distribution patterns. Supported linguistic arguments with program-generated visualizations.
- Prepared detailed, user-friendly project documentation for replication by other researchers.

Research assistantship
Language Production Lab, Simon Fraser University, 2017/09 – 2019/08

- Planned a pipeline of Python programs to process data from a corpus of Cantonese to derive word frequency data for a research project (to be published).
- Diagnosed, documented and corrected quirks of corpus data that could impact latter processing.
- Trained a general part-of-speech tagger to work on Cantonese input.
- Coded English speech errors from audio recordings by making decisions on dozens of annotation variables for the Simon Fraser University Speech Error Database.

Newengland-mad
Personal project, 2018/06 – 2018/07

- Wrote Python programs to build a corpus of the usage of 'mad' as an intensifier adverb in New England English in Twitter data.
- Devised a method of retrieving Twitter data by automatically interacting with a web browser and scraping data from HTML files to get around limited free Twitter API.
- Wrote a program to apply syntactic tests to exclude false cases of adverbial 'mad'.

Education

Simon Fraser University, M.A. in Linguistics, 4.11/4.33 CGPA, 2017/09 – 2019/12
University of Wisconsin-Madison, B.A. in Economics, Linguistics; 3.96/4.00 CGPA, 2012/09 – 2016/05