Heikal Badrulhisham

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

Skills

Technologies: Python, TensorFlow, Scikit-learn, NLTK, SpaCy, Java, Julia, HTML, CSS, JavaScript

General skills: natural language processing, machine learning, web scraping, data cleaning, Git

Coursework: programming, data structures, artificial intelligence, optimization, natural language processing

Experience

Corpus linguistics Master's thesis project
Simon Fraser University, 2018/09—2019/11
- Planned 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 Turkish National Corpus 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 an 82-page thesis to communicate the story within the data while accommodating readers with a linguistics background.
- Prepared detailed, user-friendly documentation of project to facilitate 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.
- 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.
- Classified English speech errors for a database by making decisions on dozens of annotation variables.

Newengland-mad
Individual 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.
- Wrote a program to apply syntactic tests on each data point to exclude false cases of adverbial 'mad'.

Drop-in tutor
University Housing, University of Wisconsin-Madison, 2013/09—2015/05
- Tutored students on a drop-in basis in subjects including precalculus, calculus and linear algebra.
- Encouraged students to discover the solution by drawing connections with elements they already knew.
- Walked students through a solution step-by-step. Highlighted general methods and reasoning skills applicable to other problems.

Education

Simon Fraser University, 2017/09—2019/12
M.A. in Linguistics
4.11/4.33 CGPA

University of Wisconsin-Madison, 2012/09—2016/05
B.A. in Economics, Linguistics
3.96/4.00 CGPA