

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, SQL, 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 (Java), data structures (Java), artificial intelligence (Java), optimization (Julia), research techniques and experimental design, natural language processing (Python)

Experience

Researcher, Simon Fraser University, 2018/09 – 2019/11

- Wrote Python and Java programs to process and analyze Turkish corpus data.
- Automated querying and downloading 1.49 million data points from 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 Python to uncover patterns within data and test hypotheses. Supported arguments with visualizations.
- Prepared detailed, user-friendly project documentation for replication by other researchers.

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

- Planned a pipeline of Python programs to process data from a corpus of Cantonese for a research project.
- 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.

New England 'Mad' project

- Used Python to build a dataset of instances of 'mad' as an adverb in New England dialect in Twitter data.
- Automated Twitter data collection by interacting with a web browser and scraping data from HTML files.
- Programmatically applied syntactic tests that filters out false positives in data.

Van Gogh on Twitter project

- Wrote a Python program that generates texts mimicking Vincent Van Gogh using a Markov chain.
- Web-scraped Van Gogh's letters to build a dataset that underlies the Markov chain.

Birds and Topics project

- Used Python to test using topic modelling to categorize the taxonomic family of birds in Wikipedia articles.
- Web-scraped text data from Wikipedia pages and crawled through related pages.
- Preprocess text data to build an LDA topic model.

Education

Simon Fraser University	MA in Linguistics	4.11/4.33 CGPA	2017/09 – 2019/12
University of Wisconsin	BA in Economics, Linguistics	3.96/4.00 CGPA	2012/09 – 2016/05