

Natural Language Processing Bots in the field of Digital Humanities

Name: Rohan Ben Joseph

SFU Faculty/Major:

Faculty of Applied Science, Computing Science and

Faculty of Arts and Social Sciences, Linguistics

Presentation Description:

This presentation considers the programming and usage of bots and daemons that parse through data. It discusses the practical applications of Computational Linguistics and Natural Language Processing techniques (such as word prediction and supervised learning) in improving Data Analysis and Representation in the context of Social Sciences.

Abstract:

My presentation deals with applications of Language Technology for digitization and dissemination of fragile literary material. I focus on my experience working with Special Collections at the SFU Library as part of the SpokenWeb project. While dealing with metadata entry from collections of thousands of poetry readings from tapes stored in the SFU archives, I developed a bot to traverse through the metadata ingestion system in order to make batch edits. My presentation deals with improvements I made to this bot, specifically involving improvements to the Python code and Bash scripts to facilitate large volumes of data, usage of NLP word prediction techniques to fill in gaps in areas with missing data and supervised learning to automate the data entry process. Besides this, I also include considerations of more experimental applications of the field of Computational Linguistics such as lip reading and sentiment analysis. I base this on my experience programming chatbots (with wit.ai) and working with Keras implementations of LipNet as an alternative to text-to-speech in environments with lots of background noise. I believe that these technologies may have a drastic impact on a world that is getting more and more digital every day including accessibility for the deaf, automated language generation and leaps in global connectivity through improvements to translation technology. Through my presentation, I hope to offer a well rounded view into the interdisciplinary field of Computational Linguistics using my own experience as a case study.