Michael Prappas
EMIS 8331

Abstract
"Natural Language Processing: Making Sense of Text with NLTK"

Natural language processing (NLP) is increasingly becoming a part of our daily lives. Apple's Siri kicked off the digital assistant arms race in 2011, and while digital assistants are great microcosms of natural language processing, they are hardly the only application of natural language processing. NLP encompasses everything from hearing human speech and reading human writing to answering questions about that text, perhaps in another language. Decades of research has been spent improving the methods and algorithms of every step along the way; while they are far from perfect, it is a large area of study to cover in one tutorial. This tutorial will introduce participants to the spectrum of problems in NLP and the resources available to them. Specifically, this tutorial introduces the Natural Language Toolkit (NLTK), an open-source collection of linguistic corpora and Python packages that is easy for pick-up-and-play NLP. The tutorial will walk through how to think about language and text data and how it differs from traditional "data." After working with some simple operations, the tutorial will discuss how to categorize words, specifically by their parts of speech. The tutorial will walk through different ways to iteratively improve the solutions to this problem. After discussing some more abstract concepts about language, the tutorial will discuss what "grammar" means in the NLP context. The tutorial will conclude with methods for extracting the grammatical structure of text.