Project Synopsis

Team Rigatoni

Linguine is an open-source application for natural language processing that aims to make language analysis easier to complete without knowledge of programming or computational language analysis techniques. It allows users with any level of domain expertise to import data, analyze it, and visualize it. Linguine began development as a senior project in 2013, but lacks many features that are necessary for language analysis. Its development was also challenged by difficulties with the technologies used to implement it.

Our team plans to rework the existing code to overcome these difficulties, extend Linguine to be more feature-rich, perform more thorough user testing, and add the ability to import corpora such as social media data. In addition to user testing, we plan to improve the coverage of unit tests. Specific analyses that we plan to implement include text preprocessing, sentence parsing, and topic modeling. We also plan to add and rework visualizations for Linguine, including parse trees and word clouds. Corpus importing will be implemented for Twitter and Project Gutenberg. Additional analyses and visualizations will continue being developed as time allows to provide as much access as possible to the NLP libraries that we will utilize (including NLTK and Stanford CoreNLP).