LINGUINE: A WEB APPLICATION FOR LINGUISTIC ANALYSIS

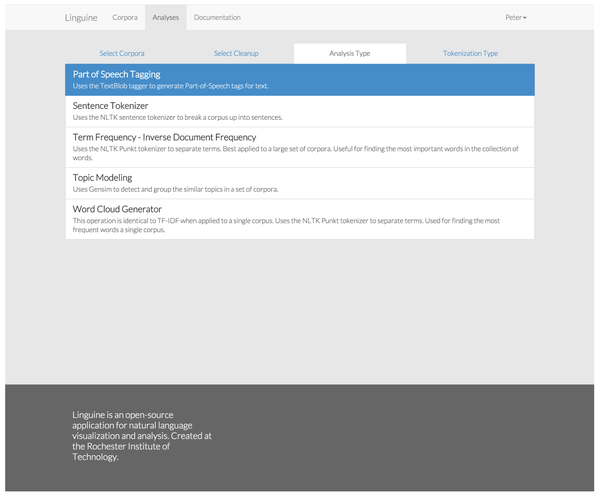
Team Rigatoni: Dan Lavoie, Mike Wideman, Kristen Mills, Peter Mikitsh
Coaches: Mei Nagappan, Jim Vallino
Sponsor: Cecilia Ovesdotter Alm
Senior Project 2015

Software Engineering
Rochester Institute of Technology

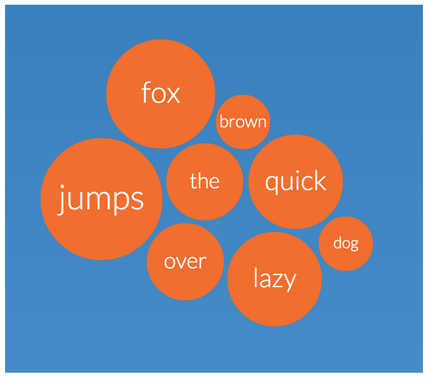
R·I·T
College of Liberal Arts

About Linguine

Linguine is an easy-to-use web application that helps students and other users perform natural language processing (NLP) tasks without knowledge of programming.



Visualize natural language analysis



Standard Protocol

```
{
  "transaction_id": "transactionId", // An ID associated with the current request.
  "operation": "tfidf", // The analytic operation to be performed.
  "library": "nltk", // The library to use when executing the analysis.
  "corpora_ids": ["id1", "id2", "etc"] // The corpora ID's to run the analysis on.
  "user_id": "user1", // The user who requested the analysis.
  "cleanup": ["removeCapsGreedy","removePunct", "etc"] // The cleanup operations to perform on the text.
  "tokenizer": "whitespaceTokenizer" // The tokenizer to use to break text into word tokens if needed.
}
```

Architecture

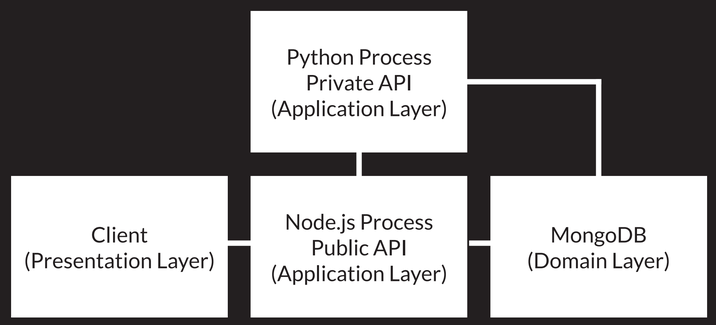
Our application layer exposes a public API using Node.js and a private Python API for corpora analysis, which uses NLTK, Gensim for topic modeling, and TextBlob. The presentation layer utilizes Angular.js and HTML, CSS, and JS. The data layer is MongoDB, which is accessed via the Mongoose ORM.

Technologies

Background

Automated natural language analysis often requires some level of programming experience and lacks visualizations. Through our simple interface, users can perform advanced analytics operations with visual output without that knowledge.

Architecture



Process Methodology

Our team applied an process model of two week iterations, alternating between requirement elicitation and a deliverable. We applied a commitment-based planning approach to each iteration.