Linguine Project Plan

Team Rigatoni

[Project Synopsis](#h.u6uiuhpy664k)

[Scope](#h.xxc7mmkgr6nb)

[Stakeholders](#h.rpggvgarqvko)

[Functional Requirements](#h.3h2txgg1cw18)

# Project Synopsis

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).

# Scope

## 2.1 In-Scope Functionality

* Text Preprocessing
* Sentence parsing
* Topic modelling
* Part-of-speech tagging
* Named entity recognition
* Coreference
* Corpus importing (incl. Twitter and Project Gutenberg)
* Parse tree visualization
* NLTK and Stanford NLP integration
* All functionality implemented in Linguine version 1
* Additional language analysis functionality identified throughout the course of the project

## 2.2 Out-of-Scope Functionality

* All other functionality not specifically mentioned within the in-scope functionality

# Stakeholders

Primary stakeholders include the Senior Project team, Dr. Cecilia Ovesdotter Alm, students taking introductory linguistics courses, undergraduate and graduate students specializing in linguistics, and linguistic researchers.

# Functional Requirements

See requirements document.

# Nonfunctional Requirements

See requirements document.

# Deliverables and Artifacts

## 6.1 Deliverables

1. Project Website (rigatoni.se.rit.edu)

2. Project Plan (this document)

3. Time and Effort Tracking Report

4. Metric Report

5. Interim and Final Project presentations

6. Project Poster

7. Technical Report

8. Self-assessment (Interim and Final)

9. Curriculum Reflection Report

10. Copy of all project artifacts

11. End user manual

1. Usability evaluation report
2. User training materials

## 6.2 Artifacts

1. Github repository

2. Risks and Recommendations document

3. Project Plan

4. Architecture document

5. Test Plan

6. Usability document

# Development Process

We plan to utilize an incremental delivery process, for reasons outlined in the Risks and Recommendations document. The process will consist of 2-week iterations. At the end of each iteration, the team will deliver a functional product, including unit tests and documentation. Further details can be found below in the estimated schedule.

# Estimated Schedule

Ramp-up and Planning: 8/28/14 - 10/9/14

Iteration 1: 10/9/14 - 10/23/14

Iteration 2: 10/23/14 -11/6/14

Iteration 3: 11/6/14 -11/20/14

Al Dente Iteration: 11/20/14 - 12/2/14

Presentation Prep: 12/2/14 - Presentation Date

Interim Finalization: Presentation Date - 12/18/14

Intersession Iteration (potential): 1/6/15 - 1/20/15

The details of the spring semester schedule will be determined once the project team and sponsor know the details of their spring class schedules.

# Actual Schedule

The actual schedule will be updated as it progresses.

# Metrics

The three primary project metrics will be Hours Spent, Github Issue Closure Rate, and Test Coverage Percentage.

# Revisions

# Appendix

## Risks and Recommendations

Please see the included risks and recommendations document.