Fall Semester 2016
Iowa State University

ENGL 520 - Computational Analysis of English

Assignment 4
Submission Deadline: 29 OCT 2016, end of the day

Instructions: This assignment consists of two questions. First question carries 5 marks and the Second one carries 10 marks. First question does not involve programming. Second does not involve in you writing a program as such, but involves studying an existing software and understanding how to use it for text classification. Create a zip file with two components (Two PDF files). If any of the programs does not run and throws errors, you cannot get a credit for that unless you defend it during the office hours. Some times, a programming question may have different instructions for perl and python programmers.

Question 1

Analyse any two existing spelling and grammar checking tools (e.g., cyWrite, grammarly, languagetool.org etc.) and write a short (1-2 page) summary of your observations about when the tools work and when they do not, with examples. You don't have to write any programs for this. Access the web interfaces of these tools, test with a few sentences, and draw some conclusions. Your report can be upto 2 pages long. Write a few lines on what tools you chose, and what sentences you tried them with. Then, analyse the performance of each tool, and compare their performances.

Question 2

Download LightSide researcher work bench (`http://ankara.lti.cs.cmu.edu/side/download.html`). Figure out the installation issues and make it work (the UI should show up). Use the sentiment_sentences.csv file in data/ folder, and learn to train sentiment classification models using Naive Bayes classifier or Logistic Regression in LightSide. You can use Word N-gram features, or use any other feature configuration lightside provides, as long as you have a conceptual understanding of what they mean and write a short analysis of the performance of your models, some useful "features", why the classification works or doesn't work etc. Your final writeup can be upto 2 pages long.