ENGL 516: Methods of Formal Linguistic Analysis

Semester: Spring '18

Instructor: Sowmya Vajjala

Iowa State University, USA

12 Apr 2018

Last class' exercise

I want a program, that takes as input a .txt file path from the user, and prints the following:

- ▶ Number of unique words/total number of words
- ▶ Percentage of words in this document that appeared only once.
- ▶ Average word frequency for top 25 words.

Solutions on the forum - discussion

Project presentations schedule

- 24th:
  1. Brody, Emily
  2. Kristin, Afnan
  3. Tim, Taichi, Fatemeh
- 26th:
  1. Sondoss, Hardi, William
  2. Yasin, Haeyun, Nazli

Next class

I compiled some readings which (I think) will give a perspective before starting 520. This time, I am going to pick three people to present these together with insights you may have obtained with 516 (time: 15 min max, + discussion):

- ▶ "The promise of NLP and speech processing technologies in language assessment" (LTJ, Chapelle and Chung, 2010)
  `http://journals.sagepub.com/doi/abs/10.1177/0265532210364405`
  (My chosen person: Fatemeh)

- ▶ "Natural Language and Language Learning" (EAL, Meurers, 2013)
  `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781405198431.wbeal0858`
  (My chosen person: Yasin)

- ▶ Intelligent Computer Assisted Language Learning (EAL, Heift, 2013)
  `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781405198431.wbeal0548`
  (My chosen person: William)

Rest of today

- You know know how to collect some basic corpus statistics.
- Incorporate this into a web UI.
- i.e., user uploads a text, and your program calculates these statistics (ttr, avg. word freq. for top 25 words, precentage of rare words) and shows back the output.
- You have all the pieces in what we discussed earlier - you have to think through the approach, and how to put what you learnt so far together.