

LING 410X: Language as Data

Semester: Spring '18

Instructor: Sowmya Vajjala

Iowa State University, USA

5 Apr 2018

Class Outline

- ▶ Update on Assignment 5/mallet/topicmodels
- ▶ Introduction to Stylo R package
- ▶ Exercise with Stylo
- ▶ Reminder: Assignment 5 - submit this weekend.
- ▶ Reminder: Project initial report - submit this weekend.

topic models

- ▶ IT support did not manage to make mallet work on lab computers!
- ▶ So, work with topic models, reduce those numbers in my tutorial (iter=50, thin=25 - this will perhaps work)
- ▶ If you complete this, move on to the next slides.

For those who already finished Assignment 5: Work with Stylo package in R by reading the documentation/based on the papers you read on Tuesday

<https://sites.google.com/site/computationalstylistics/stylo>

Working with stylo

- ▶ Installation instructions:
`https://sites.google.com/site/computationalstylistics/stylo`
- ▶ Corpus to work with: There is a zip file on Canvas, `stylocorpus.zip`.
- ▶ Task: Learning to cluster texts together
- ▶ Requirement: In the working directory where you call `stylo()` function in the `stylo` library, it expects to see a folder called `corpus` (lowercase) containing the text files you want to cluster
- ▶ Useful links:
`https://cran.r-project.org/web/packages/stylo/stylo.pdf`
`stylo_howto.pdf` on Canvas.

Things to think about

- ▶ How do the cluster arrangements change when you change: words to ngrams of different sizes? when you change other settings (MFW)? What are those other settings?
- ▶ What is the best way of doing cluster visualizations in Stylo? (If you read those two articles from Tuesday, you will know)
- ▶ Post your observations in the forum for today's date.

Next week

- ▶ Discussion of your observations from working with stylo
- ▶ Discussion on what calculations stylo performs to get those visuals
- ▶ Clustering - quick overview (using stylo)