

LING 410: Language as Data

Semester: Spring '18

Instructor: Sowmya Vajjala

Iowa State University, USA

26 Apr 2018

To Do

- ▶ Upload your presentations on Canvas by today. I cannot grade you until you submit. One submission per group is fine.
- ▶ Guidelines for final submission:
 - ▶ Deadline: May 3, end of the day (15% of the final grade)
 - ▶ What it should contain: All code you wrote, any visuals you generated, and a report (not exceeding 5 pages) describing your research question, what you did with R, what is the data, what conclusions you can draw from your results etc.
- ▶ Submit your course evals - do them today in the remaining time!!

Quick Summary of what we learnt

In terms of working with textual data, we looked at:

- ▶ How to work with different formats/sources of data (txt, html, tweets, guardian/times etc)
- ▶ How to calculate some basic statistics (frequencies, lexical richness etc)
- ▶ How to look for words in context (KWIC, ngrams)
- ▶ How to create some basic plots and visualizations (plot function, stylo, wordclouds etc)
- ▶ How to work with text classification, clustering, and topic modeling

Far from complete ..

Remember: We did only a basics course.

- ▶ There is lot more to explore, and there are more things you can do for your discipline's research problems or in jobs, with this knowledge.
- ▶ I hope you will continue being interested on this and will enhance your knowledge in future.
- ▶ It is not easy, but it is certainly doable, with some patience, consistent effort, and with a mentor at your workplace in future.
- ▶ Check: tidytextmining.com

Good luck!

- ▶ Have fun with data and good luck for your future!
- ▶ My email: sowmya@iastate.edu, vbsowmya@gmail.com
- ▶ If you need a reference in future, you can contact me (there is no guarantee I will write a letter for each person who asks me, though!)

It is a challenging course, and you are all reaching the finish line. So, Congratulations!!