Course Description

Text as Data focuses on applications of natural language processing, guided by psychological theories, for identifying various social and cognitive properties evident in textual data. In this course, we will survey state-of-the-art techniques, and applications of such techniques, for investigating various aspects of human cognition. The intended audience for this course is advanced psychology and cognitive science undergraduate students, and more broadly students in social sciences, who are interested in learning text analysis.

Learning Objectives

This course is designed to survey current state of research in text analysis. In order to achieve this objective, each week several papers/books will be read and presented by students. Also, there will be a final project and written report.

- **Prerequisite(s):** Psych 274L, CSCI 103L (or a similar programming course)
- **Recommended Preparation:** Psych 421 or a similar course

Course Notes

Students are not allowed to use laptops or smartphones during class, unless used for lab sessions or class presentations. Homework assignments will be posted on Blackboard. Students are also highly encouraged to use the course forum on Blackboard.

Required Readings and Supplementary Materials


• Wickham, H. and Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. O’Reilly Media, Inc. (Available free online: https://r4ds.had.co.nz/)


**Description and Assessment of Assignments**

Successful students:

1. Homework assignments. Each week students will complete programming problems from one of the required books. The assignments will be graded based on both output and style of the code. The homework material will be reviewed during lab clinics.

2. Paper presentation. Each student will present a set of papers related to one of the topics discussed in class.

3. Reaction paragraphs. Students are asked to write a short note, one or two paragraphs in length, about their reaction to the reading assignments of the week. These can be a quick summary of the material, comments about the subject area, or a critique of a particular theory or experiment. I will read these paragraphs before each class, and will use them to guide the discussion in class.

4. Class Projects. Students will complete two individual projects, and one group project. The individual projects will be assigned from the reading material. The final project will involve three phases: 1. proposal 2. update 3. final presentation and write up. The students need to work actively with the instructor and/or TA to design a project, and update the instructor and the TA on frequently about the status of their project.

**Grading Policy**

- 5% Project 1
- 5% Project 2
- 5% Final Project Status Update
- 5% Final Project Presentation
- 20% Final Project Write up
- 20% Homework
- 10% Participation
- 10% Paper Presentations
Assignment Submission Policy

Reaction paragraphs are due on Tuesdays at 10am, and the programming assignments and projects due on Thursdays at 10am, each before the start of class submitted on Blackboard. All homework turned in any later than 10:10am will be considered late. Students will be allowed a total of seven late days that can be used on the assignments. In exceptional circumstances, arrangements must be made in advance of the due date to obtain an extension. Once you have used up your seven late days, one additional day late will result in a 25% reduction in the total score, two additional days late will yield a 50% reduction, and no credit will be given for three or more additional days late. Late days are in units of days, not hours, so using up part of a day uses up the whole day. The final project report, plus the R code used, will be due on the day of the final exam. All assignments, including the projects, need to be written using `sweave` or `knitR`. Copied and pasted code/results will not be accepted.

Schedule and weekly learning goals

The schedule is tentative and subject to change.

**Week 01, 01/07 - 01/11:** Introduction to R
- Install R/Rstudio
- Install `swirl`
- Run `R_Programming_E` swirl course
- Homework & Reading:
  3. Finish `R_Programming_E` swirl course
  4. Salganik (2017): Questions 2.3 & 2.4

**Week 02, 01/14 - 01/18:** Introduction to Computational Social Sciences & introduction to `dplyr`
- Run `Getting_and_Cleaning_Data` swirl course
- Discuss `dplyr` packages
- Homework & Reading:
1. Wickham and Grolemund (2016): Chapters 5


4. (optional) Salganik (2017): Chapter 3

5. Finish *Getting_and_Cleaning_Data* swirl course

6. Wickham and Grolemund (2016): Exercises 5.2.4.2, 5.3.1.2, 5.4.1.3, 5.5.2.4, 5.6.7.5, 5.7.1.7

Week 03, 01/21 - 01/25: Introduction to tidyR & RNotebook

- Discuss tidyR & RNotebooks
- Discuss Iliev et al. (2015) & Chen and Wojcik (2016)
- Homework & Reading:
  1. Wickham and Grolemund (2016): Chapters 12 & 27
  3. Wickham and Grolemund (2016): Exercises 12.2.1.1, 12.2.1.2, 12.2.1.3, 12.4.3.1, 12.4.3.2, 12.6.1.1, 12.6.1.2, 12.6.1.4. These exercises need to be done in RNotebook.

Week 04, 01/28 - 02/01: Word count I & Tidy text format

- Discuss Pennebaker (2011, chap. 1-5)
- Silge and Robinson (2017): Chapter 1
- Word Count in R
- Homework & Reading:
  1. Pennebaker (2011): Chapters 5-10
  2. Silge and Robinson (2017): Chapter 1
  3. Project 1: Salganik (2017), Question 2.6. Due on 2/15

Week 05, 02/04 - 02/08: Word count II, tf-idf & sentiment analysis

- Discuss Pennebaker (2011, chap. 6-10)
- Silge and Robinson (2017): Chapter 2-3
- Homework & Reading:


**Week 06, 02/11 - 02/15:** Word count III & introduction to topic modeling

- Project 1 due
- Discuss Back et al. (2010); Pury (2011); Back et al. (2011); Dehghani et al. (2013); Iliev et al. (2016)
- Silge and Robinson (2017): Chapter 4, 6
- Homework & Reading:
  1. Project 2: Salganik (2017), Question 2.7. Due on 3/01
  2. Silge and Robinson (2017): Chapter 4, 5, 6

**Week 07, 02/18 - 02/22:** Topic modeling based methods & introduction to latent semantic analysis

- Discuss Schwartz et al. (2013); Park et al. (2015); Eichstaedt et al. (2015)
- LSA handout
• *lsa* package in R

• **Homework & Reading:**


**Week 08, 02/25 - 03/01:**  Latent semantic analysis & introduction to neural networks

• Project 2 due

• Discuss Dam and Kaufmann (2008); Sagi and Dehghani (2014); Dehghani et al. (2016)

• Introduction to neural networks

• **Homework & Reading:**


**Week 09, 03/04 - 03/08:**  Neural networks for text analysis & *keras*

• Final project proposal presentations

• Discuss Garten et al. (2018); Lin et al. (2018); Mooijman et al. (2018)

• *keras* in R

• **Homework & Reading:**

1. Play with *keras*


**Week 10, 03/11 - 03/15:** Spring Break

**Week 11, 03/18 - 03/22:** Other methods

- Discuss Caliskan et al. (2017); Garg et al. (2018); Voigt et al. (2017); Boghrati et al. (2018); Garten et al. (2019)

- More on *keras*

- Homework & Reading:


**Week 12, 03/25 - 03/29:** Clinical & cognitive applications

- Discuss Resnik et al. (2015); Walsh et al. (2017); Mitchell et al. (2008); Dehghani et al. (2017)

- Project updates
Homework & Reading:

1. Salganik (2018): Chapters 6, 7
4. more on Cambridge Analytica

**Week 13, 04/01 - 04/05:** Ethics

- Discuss Salganik (2017, chap. 6-7) & Wienberg and Gordon (2015); Persily (2017)
- Watch in class: Friends You Haven’t Met Yet

**Week 14, 04/08 - 04/12:** Final project presentations
Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism — presenting someone else’s ideas as your own, either verbatim or recast in your own words — is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Section 11, Behavior Violating University Standards https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct/.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity http://equity.usc.edu/ or to the Department of Public Safety http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety whole USC community. Another member of the university community — such as a friend, classmate, advisor, or faculty member — can help initiate the report, or can initiate the report on behalf of another person. The Center for Women and Men http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems

A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the American Language Institute http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. The Office of Disability Services and Programs http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, USC Emergency Information http://emergency.usc.edu/will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.
References


