Political conflict often occurs in text—either spoken or written. Candidates debate during elections. Representatives write laws. Nations negotiate peace treaties. Clerics issue Fatwas. These examples, and many others, suggest that to understand what politics is about, we need to know what political actors are saying.

This class will provide a set of tools for exploring the content of texts. The focus is applied. Students will learn about tools for analyzing texts quantitatively and intuition for why the tools are useful. Proofs, however, will be avoided.

Students will be evaluated using homework and a final paper. Home work will be assigned weekly and should be completed with a data set of text that students plan to use for their final paper. In the final paper, students should apply techniques learned in this course to analyze texts quantitatively. Homework and the final paper will be given equal weight in grades.

Books

There are no required books for the class. But there are many books on Text Analysis and Machine Learning you may find useful.

Natural Language Processing


Machine Learning


Computer Languages
Class Outline

Introduction and Acquiring Text Data


Representing Texts as Data


Dictionary Methods: Inference about Classes with Knowledge About Words


Methods for Identifying Distinctive Words and Phrases: Inference About Words with Knowledge About Classes

- Monroe, Burt, Michael Colaresi, and Kevin Quinn. 2008. “Fightin’ Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict”. 16(4) [coursework]
- MRS, Section 13.5
- Yano, Tae, Philip Resnik, and Noah Smith. 2010. “Shedding (a Thousand Points of) Light on Biased Language” Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk [coursework]
Vector Spaces, Term Weighting, Distance Measures, and Projection

- MRS 6.2 and 6.3

Unsupervised Learning: Clustering Models

- 14.3. Hastie, Tibshirani, and Friedman. *The Elements of Statistical Learning* Springer. [Coursework]

Unsupervised Learning: Mixed Membership Models

- Quinn, Kevin; Burt Monroe, Mike Colaresi, Mike Crespin, and Drago Radev. 2010 “How to Analyze Political Attention with Minimal Assumptions and Costs”. AJPS, 54, 1 209-228. [coursework]

Supervised Learning: An Introduction

- Hopkins, Dan and Gary King. 2010. “A Method of Automated Nonparametric Content Analysis for Social Science” AJPS, 54, 1

Supervised Learning: Ensemble Methods

  [Coursework]


Applications/Wildcard