Introduction to Natural Language Processing

Spring 2023

Instructor: Yeabin Moon ([yeabinmoon@brandeis.edu](mailto:yeabinmoon@brandeis.edu))

**Course Page:**

**Office Hours:** After class, or by appointment

**Course Description:**

**Assignments:**

**Exams:**

**Main References:** This is a partial list of various interesting and useful books that will be touched during the course.

* Dipanjan Sarkar, *Text Analytics with Python* (DS)
* Steven Bird, Ewan Klein, Edward Loper, *Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit* (BKL)
* Daniel Jurafsky and James H. Martin, *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition* (JM)
* Chris Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing* (MS)
* Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana, *Practical Natural Language Processing* (VMGS)

Other useful reference is *Introduction to Natural Language Processing* by Jacob Eisenstein for avid students of mathematical exposition.

**Prerequisites:**

**Course Outline**

1. NLP Primer
2. Introduction to NLP

* Chapter 1, JM
* Preface, BKL
* Turing, Alan M. “*Computing machinery and intelligence.*” Parsing the turing test. Springer, Dordrecht, 2009. 23-65.

1. NLP Pipeline

* Chapter 2, VMGS

1. Text Representation
2. Vector space models

* Chapter 3, VMGS

1. Use of linguistic data, text corpora

* Chapter 2, BKL
* Chomsky hierarchy: <https://people.cs.umass.edu/~mccallum/courses/inlp2007/lect2-regex.ppt.pdf>
* Regular Expressions: Chapter 2, JM

1. Basic Vectorization Approaches
   * The bag of words: Chapter 2.1, Eisenstein

* Tokenization, N-grams: Chapter 3, BKL
* Sproat, R., W. Gale, C. Shih, and N. Chang (1996). “*A stochastic finite-state word segmentation algorithm for Chinese.”* Computational linguistics 22(3), 377–404.
* TF-IDF Classification: Chapter 6, BKL

1. Probability and Optimization theory
2. Search and Learning

* Combinatorial optimization
* Numerical optimization

1. Conditional probability and Bayes’ rule
   * Modeling and estimation
2. Text Classification
3. Language modeling and Naive Bayes

* Chapter 4.1-4.9, JM
* Maximum Likelihood estimation
* Smoothing (discounting): Laplace smoothing

1. XOR problem
   * Discriminative learning: Chapter 7.2
   * Minsky, M. and Papert, S. (1969). *Perceptrons*. MIT Press.

(C) Learning in Logistic Regression

* Loss functions and large-margin classification
* Regularization and Gradients
* Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. New York: Cambridge University Press.

(D) Optimization in practice (*time permitted*)

* Batch Optimization: quasi-Newton algorithm
* Online learning algorithm
  + Dynet: The dynamic neural network toolkit