CS 585: Natural Language Processing
Fall 2019
Stuart Hall, Room 111
Thursdays, 6:25–9:05

Instructor: Derrick Higgins
dhiggins1@iit.edu
Office: Stuart 228A

TA: Ruo Yang
ryang23@hawk.iit.edu

Overview

This course is about how to build systems that analyze unstructured natural language texts and extract useful information from them. Students should expect to gain familiarity with the most common types of natural language processing (NLP) tasks, including text classification, sequence labeling, and structure prediction—and to learn appropriate frameworks for performing these tasks. The course will cover the technical methodology in sufficient detail to allow students to apply these frameworks in an informed way, and to make current research accessible.

Prerequisite knowledge

To succeed in the course, you will need to have knowledge of programming, probability theory, algorithm design and linear algebra. Previous knowledge of natural language processing and machine learning will be helpful. Practical programming exercises will be done in Python 3. Some Unix shell commands will be introduced, but shell-scripting experience is not a prerequisite.

Course expectations

You are expected to do the reading for each lecture before class. There are occasional quizzes on the material we cover. Five programming assignments will be given, to be done in Python. Your code must be readable as well as run correctly—it will be scored for both functionality and style. There will also be a midterm and a final exam, which will be open book/notes and multiple choice.

Grading

Online quizzes (through Blackboard)	10%
Practical programming homework assignments (Python; through Autolab)	60%
Midterm exam	10%
Final exam	20%

Course reading

- **M&S**: *Foundations of Statistical Natural Language Processing*, by Christopher Manning and Hinrich Schutze, MIT Press, 1999.
- **E-NLP**: *Natural Language Processing*, by Jacob Eisenstein. `https://github.com/jacobeisenstein/gt-nlp-class/tree/master/notes`
- **UNIXWiki**: `https://en.wikibooks.org/wiki/Guide_to_Unix/Commands/Text_Processing`

Communication

- **Slack:** *Do* post (and answer) questions about the course; *don't* expect immediate or comprehensive answers
- **Instructor office hours:** Virtual via Slack, Tuesdays 5:15–6:15, or by special arrangement
- **TA office hours:** Tuesdays 9-11 AM, SB019

Course resources

- **Course Website:** `http://www.cs.iit.edu/~cs585/`
- **Slack Channel:** `https://iitcs585fall2019.slack.com/`
- **Autolab:** To be shared
- **Blackboard**

Academic integrity

- Please read IIT's Academic Honesty Policy: `https://web.iit.edu/sites/web/files/departments/student-affairs/pdfs/Handbook_18-19.pdf`
- All work you turn in must be done by you alone
- You may not look at the solution of any other student prior to the due date.

Lecture schedule

Date		Topic	Reading	Homework
8/22/2019	1 2	Welcome, linguistic concepts Mathematics review 1: probability and linear algebra	M&S 2.1, 3	HW 1 out
8/29/2019	3 4	Mathematics review 2: information theory and frequency distributions Practical text processing	M&S 2.2; UNIXWiki	HW 2 out
		Words		
9/5/2019	5 6	Words and pattern matching Lexical representations for NLP	M&S 1.4, 4.2, E-NLP 4.3	HW 1 due
9/12/2019	7 8	Neural nets 1: neural word embeddings Word sense disambiguation	M&S 7.1–7.3, E-NLP 14	
		Texts		
9/19/2019	9 10	Text categorization and naïve Bayes Generalized linear models	E-NLP 2.1–2.2, 2.5–2.6, 4.4	HW 2 due HW 3 out
9/26/2019	11 12	Neural nets 2: feedforward networks Sentiment analysis	E-NLP 3.1–3.3.3, 4.1	
		Sequences		
10/3/2019	13 14	Language models Parts of speech and sequence tagging	E-NLP 6.1, 6.2.1, 6.2.2, 6.4, 6.5, 7.1, 7.2	HW 3 due
10/10/2019		**Midterm**		
10/17/2019	15 16	Hidden Markov models and the Viterbi algorithm Unsupervised sequence labeling (EM)	E-NLP 7.3, 7.4, 7.7	HW 4 out
10/24/2019	17 18	Structured prediction Neural nets 3: neural models for sequence labeling	E-NLP 7.5, 3.4, 6.3, 7.6	
		Trees		
10/31/2019	19 20	Context-free grammars and syntax CKY parsing	E-NLP 9.2, 10.1–10.2	HW 5 out
11/7/2019	21 22	Probabilistic CFG parsing Dependency grammar	E-NLP 10.3–10.4	
		Tasks		
11/14/2019	23 24	Semantic role labeling Machine Translation	E-NLP 13.1–13.2, 18	HW 5 due
11/21/2019	25 26	Other tasks Final review	E-NLP 16, 19.2	HW 4 due
11/28/2019		**Thanksgiving Break—No Class**		
12/5/2019		**Final Exam**		