Natural Language Processing Systems: Extracting Entities from Real-world Data

COSI 217B
Fall 2019
Version 1.0.0 (8/26/2019)

Instructor

Prof. Constantine Lignos (he/him/his)
Email: lignos@brandeis.edu
Office: Volen 139
Office hours: TBD

Teaching Assistant

TBD

Class Information

Meeting time: MW 5-6:20PM
Location: Carl J. Shapiro Science Center LL16
Prerequisites: COSI 114 and COSI 140, or instructor permission
Textbook: None, readings will be linked or distributed through LATTE

Description

This course explores the construction of natural language processing systems to extract and link entities found in real-world, user-created data for the purpose of representing how users interact with those entities. For example, consider building a system that reads restaurant reviews and extracts the menu items (entities) being served at restaurants and identifies which items are the most popular across restaurants (entity linking) and best-loved by customers (sentiment analysis). The challenges of building such a system are different from common information extraction research, which has focused on identifying standard types of entities (people, organizations, locations, etc.) from newswire text. Students will adapt a provided Python framework for entity extraction to develop an end-to-end system. Through assignments
and a final project, students will perform every step required to develop their system, including annotation, error analysis, testing, and evaluation.

Objectives

The objective of this course is to explore the process of building end-to-end natural language processing (NLP) systems and enable learners to build an entity extraction system for user-generated data. Upon completion of the course, learners will be able to construct a full NLP pipeline for a given task and be able to address the specific problems introduced by processing user-generated data. Learners will be able to write a conference-style paper describing an end-to-end system and the experiments performed to develop it.

Course Plan

Homework assignments will be due approximately every two weeks. Each assignment will consist of programming, annotation, and/or data analysis portions. Assignments will be submitted electronically and be graded using a combination of automated testing and manual review. Late assignments will be subject to a "grace days" policy (details to be announced). Assignments will be completed individually subject to a collaboration policy (details to be announced). Students will work in groups on a final project of their choice related to the course topic. There is no final exam.

Lecture attendance is an essential part of learning. Excessive absences will result in a reduced engagement grade. Please contact the instructor in advance if you plan to miss class due to an athletic, religious, or other planned event (e.g., job interview). If your wellness or unforeseen circumstances prevent you from coming to class, please inform the instructor as soon as is feasible.

Success in this four-credit course is based on the expectation that students will spend a minimum of nine hours of study time per week in preparation for class (readings, homework, final project, etc.).

Evaluation

HW assignments: 75% (most likely 5 HWs, 15% each)
Final project: 20%
Engagement: 5% (attendance, participation in forums, contribution to discussion)
Academic Integrity

You are expected to be honest in all of your academic work. Please consult Brandeis University Rights and Responsibilities for all policies and procedures related to academic integrity. Students may be required to submit work to TurnItIn.com software to verify originality. Allegations of alleged academic dishonesty will be forwarded to the director of academic integrity. Sanctions for academic dishonesty can include failing grades and/or suspension from the university. Citation and research assistance can be found on the university library website.

Accomodations

Brandeis seeks to welcome and include all students. If you are a student who needs accommodations as outlined in an accommodations letter, please talk with me and present your letter of accommodation as soon as you can. I want to support you. In order to provide test accommodations, I need the letter more than 48 hours in advance.

I want to provide your accommodations, but cannot do so retroactively. If you have questions about documenting a disability or requesting accommodations, please contact Student Accessibility Support at 781-736-3470 or access@brandeis.edu.

Privacy

This class requires the use of tools (e.g., GitHub) that may disclose your coursework and identity to parties outside the class. To protect your privacy, you may choose to use a pseudonym/alias rather than your name in submitting such work. You must share the pseudonym/alias with me and any teaching assistants as needed. Alternatively, with prior consultation, you may submit such work directly to me.