COMSC 341-NL: Natural Language Processing

Syllabus Fall 2018

Meeting Information

- Lecture
  - Monday/Wednesday 9:30-10:45 AM, Kendade 203
- Fourth hour
  - Friday 9:30-10:20 AM, Kendade 203

Instructor

Heather Pon-Barry
Email: ponbarry@mtholyoke.edu
Website: www.ponbarry.com
Office: Clapp 226

Office Hours

- General drop-in office hours: Monday 1:30-2:30, Thursday 1:30-2:30
- NLP-only office hours: Friday 9:30-10:20 when there is no scheduled 4th hour meeting (drop-in or appointments, check piazza)

Overview

Natural Language Processing (NLP) is the scientific and engineering discipline of how to get computers to understand and process human language. Speech recognition, machine translation, and search engines are all NLP systems that have revolutionized how we work with information.

This course introduces the fundamental techniques for automated text and speech analysis and understanding. It covers computational algorithms, hands-on practice, and insights from linguistics.

Provisional topics include: language modeling, part-of-speech tagging, speech recognition, speech synthesis, prosodic analysis, conversational dialogue, context-free grammars, syntactic parsing, coreference, text classification, sentiment analysis, and machine translation.

Learning Objectives

- Understand the challenges of processing natural language; appreciate the basic linguistic issues that underlie NLP problems.
- Recognize and understand NLP terminology and methods in widespread use in modern NLP systems.
- Gain experience implementing several components of NLP systems.
- Be able to read and understand current research papers published in conferences such as ACL, Interspeech, and SIGDial.

Prerequisites

CS 211 or permission from the instructor. You should be comfortable designing programs, writing code, and learning new programming languages. You should be comfortable with mathematical reasoning. If you have any questions about whether you have the right background, please talk to me.

Class Format

This class is a mixture of traditional lectures, small group activities, and hands-on lab activities. In the lab activities, you will gain experience with unix tools, text analysis with python, and software for speech analysis, recognition, and synthesis.

Reading Materials

All readings will be available online or on moodle. Most of these are chapters from Speech and Language Processing (3rd edition draft) by Dan Jurafsky and James Martin.

Piazza Q&A Forum

We will use piazza for announcements, Q&A, discussions, and reading responses: piazza.com/mtholyoke/fall2018/cs341nl/.

Homework

There will be 4 or 5 homework assignments. They will be a mixture of conceptual and programming exercises. The programming exercises are in python.

Final Project

The final project is an integral part of this course. It will give you an opportunity to creatively explore a topic of interest to you. There will be milestones throughout the semester to guide your project development. You will work with a partner of your choice on the project.

Participation

This includes attendance, constructive participation in small group and whole-class discussions, informal presentations throughout the semester, giving structured feedback to your peers, and arriving to class on-time.

Grading

- Homework assignments: 40%
- Literature review: 10%
- Lab activities: 5%
- Participation: 10%
- Final Project: 35%

Late Policy

You may use up to 3 late days on the homework assignments ($<=$ 24 hours counts as 1 day), excluding Homework 0. After your late days are exhausted, late assignments will not be accepted. You do not need to email me to use a late day; moodle records the submission time.

Attendance

Class and lab attendance is extremely important and counts towards the participation component of the course grade. You are responsible for material presented in class and lab; be sure to ask questions if there are concepts you do not understand. If you miss a class or lab, please ask a classmate for notes. You may see me with any subsequent questions on covered material, but do not come to office hours and ask me what you missed in class. If you have an emergency that results in missing more than one class, you should let me know as soon as possible.

Classroom Expectations

Cell phones should not be out during lectures. Laptop computers and tablets are allowed only for taking notes, following along with course materials, and working on lab activities. If you choose to use a laptop or tablet during lecture days, you must sit in the front row(s) of the classroom. Violations of these policies will affect the participation component of your final grade.

Accommodations

If you have a disability for which you require accommodations, please make an appointment to see the instructor within the first two weeks of classes so that we can make appropriate arrangements. You will need to have a letter from the AccessAbility Services Office (phone: 413-538-2634, Accessability-services@mtholyoke.edu).

Honor Code

Mount Holyoke College Computer Science Department Honor Code Statement
The Computer Science Department follows the Mount Holyoke College Honor Code. Work submitted for grading must be entirely your own, unless you were instructed to work in groups. The purpose of course assignments is to practice skills, gain a deeper understanding of the course material, and apply that knowledge to new situations. Assignments are designed to challenge you, stimulate critical thinking, and help you understand the concepts related to the course. Your grade is a reflection of your understanding of the material. We recognize that collaboration can help you master course material. In fact, there are certain ways in which we will encourage you to collaborate. These include: discussing course content at a high level, getting hints or debugging help, talking about problem-solving strategies, and discussing ideas together. However, you must do **all coding and write-ups on your own**. Writing code and solutions on your own will test and demonstrate your mastery of course material. **Looking at solutions from other students or any other source (including the web), or collaborating to write solutions to individual work, is considered a violation of the honor code**. All suspected violations will be referred to the academic honor board. If you are uncertain whether something is allowed, it is your responsibility to ask.

If you have engaged in any of the above acceptable collaboration activities for an assignment, you MUST acknowledge the classmates or TAs with whom you spoke – this should be done in a comment at the top of your main submission file.

Note that the Association for Computing Machinery has a strong Code of Ethics and Professional Conduct. At this site you can read the new 2018 version.

Internet sources:
The internet is a useful resource when learning to solve computer science problems, and in some cases you will be expected to use reference material found online (e.g., documentation for a programming language or library). In general, it's OK to look at resources for a broad topic such as a programming language, but it is not OK to look at solutions for specific programming or written problems. If you are unsure whether something is allowed, ask. **You must cite all online sources used while working on an assignment**. Instructors will clarify more specific expectations or deviations from this policy, but it is always the student's responsibility to ask if they are unsure.

Dos and Don'ts:
These lists are intended to clarify what types of behaviors are and are not generally permissible. Follow these guidelines unless specifically directed otherwise. (clarify if uncertain)

Do:

- Organize study groups.
- Clarify ambiguities or vague points in class handouts, textbooks, assignments, and labs.
- Discuss assignments at a high level to understand what is being asked for, and to discuss related concepts and the high-level approach.
- Refine high-level ideas/concepts for projects (i.e., brainstorming).
- Outline solutions to assignments with others using diagrams or pseudocode, but not actual code.
- Walk away from the computer or write-up to discuss conceptual issues if you get stuck.
- Get or give help on how to operate the computer, terminal, or course software.
- Get or give limited debugging help. Debugging includes identifying a syntax or logical error but not helping to write or rewrite code.
- Submit the result of collaborative coding work if and only if group work is explicitly permitted (or required).

Don't:

- Look at another student's solutions.
- Use solutions to same or similar problems found online or elsewhere.
- Search for homework solutions online.
- Turn in any part of someone else's work as your own (with or without their knowledge).
- Share your code or written solutions with another student.
- Share your code or snippets of your own code online.
- Allow someone else to turn in your work as their own. (Be sure to disconnect your network drive when you logout and remove any printouts promptly from printers.)
- Collaborate while writing programs or solutions to written problems. (But see above about specific ways to give or get debugging help.) Write homework assignments together unless it is specified as a group assignment.
- Collaborate with anyone outside your group for a group assignment.
- Use resources during a quiz or exam beyond those explicitly allowed in the quiz/exam instructions. (If it is not listed, don't use it. Ask if you are unsure.)
- Submit the same or similar work in more than one course. (Always ask the instructor if it is OK to reuse any part of a different project in their course.)

Specific guidelines for CS341-NL

- You MUST acknowledge classmates or TAs that you worked with for each assignment – this should be done at the top of your main homework writeup.
- All the homework assignments that you submit must be yours: **it is against the honor code to have somebody else do the assignment for you or to copy it from somewhere else (including books or the internet).**