

Natural Language Processing
Fall 2014
IST 400/ 664
CIS 468/668

Nancy McCracken
Research Associate Professor

Contact Information: Office: 322 Hinds Hall
njmccrac@syr.edu
443-3955 (office)

Office Hours: Wednesdays 2:15 – 3:15pm

Class Sessions:

Lecture/Lab Mon/Wed 12:45 – 2:05pm 013 Hinds Hall

Course Description:

This course is designed to develop an understanding of how Natural Language Processing (NLP) can process written text and produce a linguistic analysis that can be used in other applications. This goal will be achieved by:

- Readings, lectures, and class discussions of the multiple levels of linguistic analysis required for a computer to accept natural language input, interpret it, and carry out a particular application;
- Lab exercises and assignments in analyzing or implementing some computational techniques required to perform these levels of natural language processing of text, and,
- Reports on real world applications which incorporate substantive NLP modules.

The topics of the course will cover the techniques of NLP in the levels of linguistic analysis, going through tokenization, Part-Of-Speech tagging, syntax, semantics and on up to the discourse level. It will also focus on the use of the NLP techniques in applications. These will include Information Retrieval, Question Answering, Sentiment Analysis, Summarization and Dialogue Systems.

Learning Objectives

At the end of the course the student will be able to

- understand the levels of linguistic analysis, the computational techniques used to understand text at each level and what the challenges are for those techniques
- process text through the language levels using the resources of the Natural Language Toolkit (NLTK) and some rudimentary use of the programming language Python
- understand how NLP is used in many types of real world applications.

Course Organization:

The format of the course will divide the time approximately with half for classroom lecture and discussions, and half for lab investigations and exercises.

The lab investigations will analyze text using computational processing techniques in the open-source Natural Language Toolkit <http://nltk.sourceforge.net/>. While no programming experience is assumed, students will be provided with small scripts in the Python programming language in using this resource and will run them as tools in their analysis of text. Text examples will include news articles, current and historical literature, informal text from email and blogs, and customer and product reviews.

Assignments:

Due to the typically heterogeneous mix of student backgrounds in terms of linguistic knowledge and computational skills, the coursework will be accomplished in a variety of modes:

- Lab exercises will be done in small groups in-class and will accommodate the variety of student backgrounds
- Homework assignments (tentatively 3) will set a particular analysis task and text examples, but will have options that can focus on either the analysis of the task or the computational technique. While no programming is required for assignments, students who choose the focus on computational technique will have the opportunity to learn more of the programming language Python. Small homework groups will be allowed, but not required.
- Student investigations will allow students to choose NLP applications such as speech understanding, information retrieval, question-answering, information extraction, text-mining, natural language generation, dialogue agents, machine translation, or summarization for further investigation.
- The final project will allow the same types of options as the homework assignments.

Graduate students will be required to do both the final project and the final class investigation, but undergraduate students will only be required to do the final project and a short report on the NLP applications.

Grading (Graduates) - Grades will be determined (tentatively) as follows:

Participation in labs and in-class exercises, and contributions to class discussion	20 %
Homework Assignments (3)	45 %
Final Project	20 %
NLP Application Investigations	15 %

Textbook:

The following textbook is recommended but not required:

Speech and Language Processing. Daniel Jurafsky & James H. Martin, 2nd ed. 2008. Prentice-Hall.

The first edition of this book is also o.k. and is available used for a much cheaper price.

Additional supplementary readings will be assigned during the semester and will be available on-line or on Blackboard.

Tentative Course Outline of Topics:

This is a list of topics by week from a previous version of the course:

Introduction to NLP, Corpus Linguistics

N-gram Analysis, Morphology

Regular Expressions

Part-of-Speech Tagging

Context Free Grammars

Parsing, including statistical parsing and dependency parsing

Semantic Representations, Lexical Semantics and WordNet

Case Grammar, Semantic Role Labeling

Classification and Machine Learning, Information Extraction

Information Retrieval and Question Answering

Sentiment and Opinion Analysis, Summarization, Machine Translation

Discourse Linguistics, Anaphora and Coherence

Pragmatics, Dialogue Theory

Additional Applications and Student Investigations

Educational Use of Student Work

I intend to use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will either get your written permission or render the work anonymous by removing all your personal identification.

Academic Integrity

Syracuse University sets high standards for academic integrity. Those standards are supported and enforced by students, including those who serve as academic integrity hearing panel members and hearing officers. The presumptive sanction for a first offense is course failure, accompanied by the transcript notation "Violation of the Academic Integrity Policy." The standard sanction for a first offense by graduate students is suspension or expulsion. Students should review the Office of Academic Integrity online

resource “[Twenty Questions and Answers About the Syracuse University Academic Integrity Policy](#)” and confer with instructors about course-specific citation methods, permitted collaboration (if any), and rules for examinations. [The Policy](#) also governs the veracity of signatures on attendance sheets and other verification of participation in class activities. Additional guidance for students can be found in the Office of Academic Integrity resource: [What does academic integrity mean?](#)

Disabilities

If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), <http://disabilityservices.syr.edu>, located at 804 University Avenue, room 309, or call 315-443-4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue students with documented disabilities “Accommodation Authorization Letters,” as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible.