Natural Language Processing
Fall 2015
IST 400 (M003)/ 664
CIS 468/668

Nancy McCracken
Research Associate Professor

Contact Information: njmccrac@syr.edu (preferred)
                    Office: 322 Hinds Hall
                    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 using some computational techniques required to perform these levels of natural language processing of text, and,
- Studies of 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. References to the text and other materials will be provided for further reading, but are not required.

The course will also include lab sessions to 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:

In addition to the weekly exercises and participation, there are five other assignments that occur through the semester.

- Weekly lab exercises will be done in small groups in-class and will accommodate the variety of student backgrounds
- There will be three homework assignments at intervals of approximately 2-3 weeks, that will set a particular analysis task and text examples. Some homework problems will have options that can focus on either the analysis of the task or the computational technique. While no original programming is required for assignments, students who choose to focus on computational techniques will have the opportunity to learn more of the programming language Python.
- Near the end of the semester, students will do an NLP Application Investigation, where they choose an NLP application such as speech understanding, information retrieval, question-answering, information extraction, text-mining, natural language generation, dialogue agents, machine translation, or summarization for further investigation. These investigations will typically report on some examples of such systems found on the web.
- The final assignment will be a final project where students will focus on a text classification task, which may be something like sentiment analysis, and conduct and report on a series of experiments.

The homework assignments and the final project will all take the form of carrying out the required tasks and then writing a report on the process and on the results.

Grades (Graduates) will be determined as follows:

Participation in weekly lecture and lab exercises, and contributions to class discussions	25 %
Homework Assignments (3)	45 %
Final Project	20 %
NLP Application Investigations	10 %

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, Language Models
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, Coreference Resolution
Pragmatics, Dialogue Theory
Reports on NLP Applications

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's Academic Integrity Policy holds students accountable for the integrity of the work they submit. Students should be familiar with the policy and know that it is their responsibility to learn about course-specific expectations, as well as about university policy. The university policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The policy also prohibits students from submitting the same written work in more than one class without receiving written authorization in advance from both instructors. The presumptive penalty for a first offense by an undergraduate student is course failure, accompanied by a transcript notation indicating that the failure resulted from a violation of Academic Integrity Policy. The standard sanction for a first offense by a graduate student is suspension or expulsion. For more information and the complete policy, see http://academicintegrity.syr.edu/academic-integrity-policy/

Disabilities

If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), http://disabilityservices.syr.edu, located in Room 309 of 804 University Avenue, or call (315) 443-4498, TDD: (315) 443-1371 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.

Religious Observances Policy

SU religious observances policy, found at http://supolicies.syr.edu/emp_ben/religious_observance.htm, recognizes the diversity of faiths represented among the campus community and protects the rights of students, faculty, and staff to observe religious holidays according to their tradition. Under the policy, students are provided an opportunity to make up any examination, study, or work requirements that may be missed due to a religious observance provided they notify their instructors before the end of the second week of classes for regular session classes and by the submission deadline for flexibly formatted classes.

For fall and spring semesters, an online notification process is available through **MySlice/StudentServices/Enrollment/MyReligiousObservances.**