

Natural Language Processing
Spring 2010
IST 400/ 664
CIS 400/600

Nancy McCracken
Research Associate Professor

Contact Information: Office: 209 Hinds Hall
njm@ecs.syr.edu
443-3955 (office)

Office Hours: Tuesday 12:30 – 1:30pm
Wednesday 11am – 12noon

Class Sessions:

Lecture/Lab Tues/Thur 11am-12:20pm 110Link/025Hinds
(Tuesday's classes will be in 110Link and Thursday's in 025Hinds)

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,
- Team studies and reports on real world applications which incorporate substantive NLP modules.

While the topics of the course will cover the techniques of NLP in the levels of linguistic analysis, it will also develop two themes around the use of the NLP techniques in applications:

Artificial intelligence: discussing the current state-of-the-art in representing true natural language understanding that can be used in human/computer interactive applications

Business intelligence: discussing the use of text processing techniques in deriving business information from articles written about business and the forms of text that arise in the business context, such as customer reviews, and in preparing text for data mining

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 2) 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 class presentations 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 and presentations.

Graduate students will be required to do both the final project and the final class presentation, but undergraduate students will only be required to do one of them.

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

Participation in labs and in-class exercises, and contributions to class discussion	20 %
Homework Assignments (2)	40 %
Final Project	25 %
NLP Application Presentations	15 %

Text :

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

Additional supplementary readings will be assigned during the semester. Most readings will be available on the web.

Academic Integrity

The academic community of Syracuse University and of the School of Information Studies requires the highest standards of professional ethics and personal integrity from all members of the community. Violations of these standards are violations of a mutual obligation characterized by trust, honesty, and personal honor. As a community, we commit ourselves to standards of academic conduct, impose sanctions against those who violate these standards, and keep appropriate records of violations. The academic integrity statement can be found at: http://supolicies.syr.edu/ethics/acad_integrity.htm.

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 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.

Ownership of Student Work

This course may use course participation and documents created by students for educational purposes. In compliance with the Federal Family Educational Rights and Privacy Act, works in all media produced by students as part of their course participation at Syracuse University may be used for educational purposes, provided that the course syllabus makes clear that such use may occur. It is understood that registration for and continued enrollment in a course where such use of student works is announced constitutes permission by the student. After such a course has been completed, any further use of student works will meet one of the following conditions: (1) the work will be rendered anonymous through the removal of all personal identification of the work's creator/originator(s); or (2) the creator/originator(s)' written permission will be secured. As generally accepted practice, honors theses, graduate theses, graduate research projects, dissertations, or other exit projects submitted in partial fulfillment of degree requirements are placed in the library, University Archives, or academic departments for public reference.