Natural Language Processing
IST 664

Description:

Linguistic and computational aspects of natural language processing technologies. Lectures, readings, and projects in the computational techniques required to perform all levels of linguistic processing of text.

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 of the computational techniques required to perform these levels of natural language processing of text.
- Studies of real-world applications that incorporate substantive NLP modules.

The course primarily covers the techniques of NLP in the levels of linguistic analysis, going through tokenization, word-level semantics, part-of-speech tagging, syntax, semantics, and on up to the discourse level. It also includes the use of the NLP techniques, such as information retrieval, question answering, sentiment analysis, summarization, and dialogue systems, in applications.

Learning Objectives

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

- Demonstrate 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.
- Describe how NLP is used in many types of real-world applications.

Course Organization:

Approximately half of each weekly session will be lectures on the levels of NLP and half will be demonstrations of computing techniques, primarily using the software package NLTK (Natural Language Toolkit). References to the text and other materials will be provided for further reading but are not required. Each week there will be questions in

which students apply NLP techniques to text examples, either by hand or through the NLTK. The live sessions will include discussions of the lessons learned from these questions.

The demo sessions will use 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 e-mail, blogs, and social media; and customer and product reviews.

Assignments:

In addition to the weekly questions and participation in live-session discussions, five other assignments occur through the semester.

- There will be three homework assignments at intervals of two weeks that will set a particular analysis task and text examples. Some homework problems will have the option to 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, conversational 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 Will Be Determined As Follows:

Attendance in the live sessions is mandatory. Please e-mail the professor in the case of illness or other possible excused absences.

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

Textbook:

The required textbook is available online:
Bird, S., Klein, E., & Loper, E. *Natural language processing with Python*. Available from http://www.nltk.org/book/

Please refer to this online version for reading (instead of the older version) as it is updated with Python 3 and NLTK 3.

The following textbook will be referred to but is not required:
Jurafsky, D., & Martin, J. H. *Speech and language processing* (3rd ed. draft). Available from https://web.stanford.edu/~jurafsky/slp3/

Additional supplementary readings will be assigned during the semester and will be available online.

Course Outline:

Week 1: Introduction to NLP and the processing of text words
Week 2: Corpus statistics and language modeling
Week 3: Regular expressions, morphology, and processing text files
Week 4: Part of speech (POS) tagging and introduction to machine learning (ML)
Week 5: Context-free grammars and parsing
Week 6: Semantics
Week 7: Discourse and dialog
Week 8: Sentiment analysis
Week 9: NLP applications: information extraction (IE), summarization, and machine translation (MT)
Week 10: NLP applications: information retrieval (IR), question answering (QA), and conversational agents
Week 11: Complete final projects and optional lecture on deep learning for NLP

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 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.**