

## **COURSE SYLLABUS**

### **Natural Language Processing, IST 664**

**Class Location and Time:** Mondays, 9:30 am – 12:15 pm, Hinds Hall 027

Instructor: Professor Lu Xiao

Phone: 315-443-1707

Office: 213 Hinds Hall

Email: [lxiao04@syr.edu](mailto:lxiao04@syr.edu)

Office Hours: Fridays, 4:30 - 5:30 pm. or by appointment

**Prerequisite / Co-requisite:** None

**Audience:** The course is intended primarily for Masters' students in Information Management, Computer Science, Computational Journalism, or Linguistics, though it can be of interest and value to students in other programs.

#### **Description:**

Linguistic and computational aspect of natural language processing technologies. Lectures, readings, and projects in the computational techniques required to perform all levels of linguistic processing of text. Additional work required of graduate students.

#### **Additional 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 that incorporate substantive NLP modules.

The topics of the course will cover 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 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.

**Credits:** 3 credits

#### **Learning Objectives:**

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

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

## Recommended Course Materials and Textbooks

The following textbooks are recommended but not required:

*Speech and Language Processing*. Daniel Jurafsky & James H. Martin, 2<sup>nd</sup> ed. 2008. Prentice-Hall.

The third edition draft is available online at <https://web.stanford.edu/~jurafsky/slp3/>

Bird, S., Klein, E., & Loper, E. Natural language processing with Python, <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.

Additional supplementary readings will be available on Blackboard.

## Grading

Grades are assigned based on the quality of the work, not upon how well others performed. As shown below, several methods are used to assess a student's learning outcome resulting in 12 letter grades. In general, an A category means you've done an outstanding job in this course work, a B means your work is about what would be expected of a serious graduate student, and a C means that your work is below what is expected for this course. If you get a C or below in any assessment, you should contact the professor to discuss how to improve your learning outcome in this course (please note that this is applicable to participation grade as it is not provided until the end of the course).

Your final letter grade is calculated by first converting all the letter grades into numeric grades as follows:

A: 4  
A-: 3.6  
B+: 3.4  
B: 3.2  
B-: 3.0  
C+: 2.8  
C: 2.6  
C-: 0

Then, the average of all the numeric grades will be converted back to the letter grade using the above formula. The average falls right in between the letter categories or higher then it will be rounded to the higher letter grade (e.g., average grade being 3.5, 3.32 and 3.1 will receive grade A-, B+, and B). Otherwise, it will be rounded to the lower letter grade (e.g., average grade being 3.46, 3.28, and 3.06 will receive grade B+, B, and B-).

No.	Letter graded Assessment	Due Dates (11:59 pm. of the day)
1	Individual Assignment 1	2/18
2	Individual Assignment 2	3/18
3	Individual Assignment 3	4/1
4	Individual Assignment 4	4/22
5	NLP Investigation (Group Project): Report	4/29
6	NLP Investigation (Group Project): Poster	4/29
7	NLP Investigation (Group Project): Poster Presentation (tentative)	4/30
8	Class Participation	

- **Individual Assignments**

There are four individual assignments. The first assignment is designed to help you exercise on getting general corpus statistics. The second assignment is about regular expression. The third assignment is about the syntactic structure analysis of the texts. The fourth assignment requires you to do one experiment in the NLTK on the given dataset in which you compare the baseline performance of using just word features with some modified or additional features that you implement.

- **NLP Application Investigations (Report)**

We will introduce main NLP application areas in this course, but there will be other areas and topics that we will not discuss. You will choose a general area or a specific topic to explore, and write a report about it (10 pages long including figures, tables, and references, single column, Times New Roman, 12 points). You will work in groups of four for this investigation.

Your group report needs to include: 1). title and names of the group members; 2) introduction to the topic, related important definitions, etc.; 3) major NLP techniques reported in the researched resources; 4) representative examples in this topic; 5) suggested directions in this topic; and 5) references.

Below is a list of the areas and topics that you may be interested in exploring. Or, you may suggest another area of interest to you. Please email Prof. Xiao your first and second choices

- Machine Translation
- Information Retrieval/Search Engines
- Mining Content of Social Media sites (some possible topics are listed below)
  - Detection of online influential users
  - Twitter data analysis
  - Facebook content analysis
  - Reddit content analysis
- Human Computer Interfaces / Dialogues, including “chatbots”
- Summarization
- Language Generation

- Question/Answering Systems
- Image to Text

You need to form the group and choose first and second choices of your groups' topic before the spring break ends. Please note that your group needs to have two computer science (CS) students and two non-CS students (for the exception, please talk to me). Please form the group and identify the topics by Mar. 15. Based on your input in the document, I will decide the topic for your group to make sure that there is no duplicate topic among the groups.

- **NLP Application Investigations (Poster)**

On April 30<sup>th</sup>, 2018 (last class), your group will present the NLP application investigation as a poster presentation.

The evaluation of the poster presentation includes three aspects: the *poster content*, the *poster presentation* itself, and the *poster design*.

- a). The poster content needs to include essential information of all the five parts of the report
- b). The poster presentation itself will evaluate 1) the group members' familiarity with the project; and 2) the group members' professionalism in the presentation (e.g., seriousness, confident, eye contact, presentation skill, and communication skill)
- c). The poster design focuses on the information design of the poster, e.g., whether it is easy to identify the presented information, the organization of the presented information, etc.

**Note:** Your group needs to give one or two demos of representative techniques for your area/topic at the poster session.

- **Class participation**

Class attendance is mandatory. Active participation in the class is expected. The professor judges a student's class participation based on the following guidelines:

- In this course, being prepared, attentive, communicative, and respectful of your peers and the professor are all required. Participation may take many forms, verbal or written. It is expected that the students bring a positive presence to the class discussion. This involves an original idea or concept, elaborating but not simply repeating the ideas of other class members. In verbal form, the students should allow time for others to speak as well. Monopolizing the conversation are not encouraged.
- Class participation means not only offering opinions and information, but also asking questions, working constructively with other students, and listening.
- Python exercises are mandatory. During the lab session of the lecture, we will exercise Python to work on some language problems. You will also be given exercises to complete outside of the class and to upload your results to the course web site
- Attendance at all classes is mandatory. Students are required to participate in all class activities. A student who contributes positively to all or almost all classes throughout the course should expect appropriate credit. A student who persistently monopolizes the discussions should expect to earn less credit. A student who does not contribute to class discussions at all during class should not expect credit for the class participation grade.

## **University Attendance Policy**

Attendance in classes is expected in all courses at Syracuse University. Students are expected to arrive on campus in time to attend the first meeting of all classes for which they are registered. Students who do not attend classes starting with the first scheduled meeting may be academically withdrawn as not making progress toward degree by failure to attend. Instructors set course-specific policies for absences from scheduled class meetings in their syllabi.

It is a federal requirement that students who do not attend or cease to attend a class to be reported at the time of determination by the faculty. Students should also review the university's religious observance policy and make the required arrangements at the beginning of each semester.

## **Specific Course Policies**

- Written submissions will use Times New Roman, 12 pt., single column unless otherwise stated.
- Late submission will NOT be graded unless a written explanation is provided before the due date (a doctor's note is needed if applicable, e.g., late submission due to medical reasons). With a written explanation, a late submission will be accepted and graded if it is delivered to the professor by the new deadline arranged with the professor. However, a penalty will be taken by having the letter grade one level down (e.g., A -> A-, B+ -> B).
- A student who misses two or more classes will lose participation grade by one letter grade (e.g., A -> A-, B+ -> B, etc.). For each absent class, the student needs to write an essay (~1000 words) about the content covered in the class. The essay is due on the next class.
- If there is any question/concern about a grade, please contact the professor within 1 week. No re-grading requests will be accepted after the one-week period. In rare situations where a student fails to grasp the concept or skill from the individual assignment, the professor will contact the student to re-do the assignment. A grade penalty will apply in the re-do situation by one letter grade (e.g., A-> A-, B+ -> B, etc.).

## **University Policies:**

Students should review the University's policies regarding Disability-Related Accommodation; Diversity and Disability; the Religious Observances Notification and Policy; and Orange SUccess, which can be accessed via the Office of the Provost's website at:  
<http://provost.syr.edu/>.

## **Academic Integrity Policy**

Syracuse University's academic integrity policy reflects the high value that we, as a university community, place on honesty in academic work. The policy defines our expectations for academic honesty and holds students accountable for the integrity of all work they submit. Students should understand that it is their responsibility to learn about course-specific expectations, as well as about university-wide academic integrity expectations. 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 professors. The presumptive penalty for a first instance of academic dishonesty 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 presumptive penalty for a first instance of academic dishonesty by a graduate student is suspension or expulsion. SU students are required to read an online summary of the university's academic integrity expectations and provide an electronic signature agreeing to abide by them twice a year during pre-term check-in on MySlice. For more information and the complete policy, see <http://academicintegrity.syr.edu/>.

In this course, while students are encouraged to discuss the course content and form study groups on their own, individual assignments are expected to be completed by the students individually. In the group assignments, all group members are expected to participate actively and make sure that no plagiarism or other violations of the Academic Integrity occurs in the group reports.

**Note:** If I suspect any type of plagiarism or cheating, I will consult with the academic integrity panel.

### **Disability-Related Accommodations**

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), [disabilityservices.syr.edu](mailto:disabilityservices.syr.edu), located at 804 University Avenue, third floor, 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 "Accommodation Letters" to students as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible. Our goal at the iSchool is to create learning environments that are useable, equitable, inclusive and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, please meet with me to discuss additional strategies beyond official accommodations that may be helpful to your success.

### **Diversity**

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. My goal is to create learning environments that are useable, equitable, inclusive and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, I invite any student to meet with me to discuss additional strategies beyond accommodations that may be helpful to your success.

### **Religious Observances Notification and Policy**

SU's religious observances policy, found at [supolicies.syr.edu/emp\\_ben/religious\\_observance.htm](http://supolicies.syr.edu/emp_ben/religious_observance.htm), recognizes the diversity of faiths represented in the campus community and protects the rights of students, faculty, and staff to observe religious holy days according to their tradition. Under the policy, students should have 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 no later than the end of the second week of classes for regular session classes and by the submission deadline for flexible-formatted classes. Student deadlines are posted in MySlice under Student Services/Enrollment/My

Religious Observances/Add a Notification. Professors may access a list of their students who have submitted a notification in My Slice Faculty Center.

### **Educational Use of Student Work**

Student work prepared for University courses in any media may be used for educational purposes, if the course syllabus makes clear that such use may occur. You grant permission to have your work used in this manner by registering for, and by continuing to be enrolled in, courses where such use of student work is announced in the course syllabus.

### **Course Evaluation**

There will be an end of course evaluation for you to complete this semester, described below. This evaluation will be conducted online and is entirely anonymous. You will receive a notification from the Syracuse University Office of Institutional Research & Assessment (OIRA) department in your email account with the evaluation website link and your passcode. We faculty work hard to do the best possible job when preparing and delivering courses for our students. Please understand that not only does the school use the course evaluations to make decisions about the curriculum in order to improve where necessary, but they also use them to make decisions about faculty members. Please take the time and fill out this evaluation as your feedback and support of this assessment effort is very much appreciated.

End of semester evaluation will be available for completion approximately week 14. This evaluation is slightly longer and it is used to gauge the instructor performance and make adjustments to the course to ensure it meets our student needs.

### **Blackboard Support**

The iSchool uses Syracuse University's Blackboard system to facilitate distance learning and main campus courses. The environment is composed of a number of elements that will help you be successful in both your current coursework and your lifelong learning opportunities. To access Blackboard, go to the following URL: <http://blackboard.syr.edu>. Use your Syracuse University NetID & Password to log into Blackboard.

Questions regarding technical aspects of Blackboard, please submit a help ticket to the iSchool dashboard at My.iSchool Dashboard at <https://my.ischool.syr.edu>. Log in with your NetID, select "Submit a Helpdesk Ticket," and select Blackboard as the request type. The iSchool Blackboard support team will assist you.

Week	Date	Topic	Optional Reading and Assignment
1	1/22	Course Overview and Introduction to NLP and the NLTK	NLTK book: <a href="#">Preface</a> , <a href="#">Chapter 1: Language Processing and Python (section 1&amp;2)</a>
2	1/29	Levels of Language and Corpus Linguistics	NLTK book: <a href="#">Chapter 2: Accessing Text Corpora and Chapter (section 3)</a> Chapter 1 J&M

3	2/5	N-gram Models and Analysis	NLTK book: <a href="#">Frequency Distributions and skim through section Back to Python</a> Ch. 2, Ch. 4.1 – 4.7 J&M
4	2/12	Regular Expressions, Morphology	NLTK book: <a href="#">Regular Expressions for Tokenizing Text</a> Ch. 3.1 3.9, 3.12, J&M
5	2/19	Part-of-speech Tagging	NLTK book: <a href="#">Categorizing and Tagging Words</a> (no transformation-based tagging) Ch. 5 J&M (no section 5.6)  Advanced reading: Prins, R. (2004, July). <a href="#">Beyond N in N-gram Tagging</a> . In Proceedings of the ACL 2004 workshop on Student research (p. 61). Association for Computational Linguistics.
6	2/26	Context Free Grammars	NLTK book: <a href="#">Context-Free Grammar &amp; Parsing with Context-Free Grammar</a> Ch. 12 J&M
7	3/5	Parsing, Statistical Parsing, Dependency Parsing	NLTK book: <a href="#">Analyzing Sentence Structure</a> Ch. 13, J&M Joakim Nivre, Deterministic Dependency Parsing of English Text.
<b>3/12 Spring break, no class</b>			
<b><i>Note: Please form the group by Mar. 15 including the first and second choice of topics.</i></b>			
8	3/19	Semantic Representations, Lexical Semantics, WordNet	NLTK book: <a href="#">WordNet, Analyzing the Meaning of Sentences</a> Sections 19.1-19.3, 20-1 – 20-8, and Ch. 17 J&M
9	3/26	Semantic Role Labeling	NLTK book: <a href="#">Learning to Classify Text</a> Readings/Videos: Machine Learning for NLP, Videos from Andrew Ng Sections 19.4 and 20.9, J&M, 3 <sup>rd</sup> edit (chapter 22)
10	4/2	Sentiment Analysis	<a href="#">SentiWordNet in NLTK</a> Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). <a href="#">SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining</a> . In LREC (Vol. 10, pp. 2200-2204). Ch. 22 J&M



11	4/9	Discourse Linguistics, Anaphora & Coherence	NLTK: <a href="#">Analyzing the Meaning of Sentences</a> Ch. 21 J&M
12	4/16	Pragmatics, Dialogue Theory, Topic Modeling	<a href="#">MALLET</a> Ch. 24.1-24.3 J&M
13	4/23	NLP Applications: Information Extraction, Summarization, Machine Translation, Information Retrieval, Question Answering	NLTK book: <a href="#">Extracting Information from Text</a> Ch. 22, J&M
14	4/30	Group Poster Session	