AIT 590/IT 499: Natural Language Processing
Spring 2018
Dr. Ozlem Uzuner
Email: ouzuner@gmu.edu
Phone: 703-993-8633
Office location: Nguyen 5359
Schedule: WF from 16:30-17:45
Room: IN 129

I. Course Objectives

This is an introductory course on natural language processing. It will focus on studies of textual data using rule-based and statistical methods. Our goal will be to create computer programs that analyze, interpret, and even generate human language. Topics include:

- Lexical, syntactic, and semantic elements of language.
- Statistical properties of language.
- Rule-based and data-driven approaches to building language models.
- Machine learning for understanding language.
- Application of NLP to real-world problems such as spam filtering, authorship identification, and spell checking.

Prerequisites: Python programming. Statistics or probability.
Desirable: Machine learning.

IT499 Additional information:
- Prerequisites: IT 106 Python or CS 112 or approved Python course and STAT 250.
- This course will be a WDM or DTP concentration core option (or INFS, HIT, ITE, NTEL concentration elective).

To register, students must submit an override request at ist.gmu.edu/go/override.

II. Textbook and Required Materials

For this course, we will be using two required textbooks.
Speech and Language Processing, by Jurafsky and Martin (2nd edition). Other editions are not a good substitute.

Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, by Bird, Klein and Loper (1st edition). This book is freely available online.

III. Course Learning Activities and Grade Distribution

This course will include 4 programming assignments, several lecture reports, classroom presentations and discussions, and one final project. All activities will be done in teams of two. Class participation will be graded individually.

The course is graded out of 100 points. Your grade will be based on the following breakdown:

Item	Total Points (Percent of Grade)
Programming Assignments	30
Lecture Reports	20
Class Participation	20
Final Project, Paper & Presentation	30
TOTAL	**100**

IV. Grade Distribution
Letter grades will be assigned as follows:

Letter Grade	Numeric Range
A+	97-100
A	93-96
A-	90-92
B+	87-89
B	83-86
B-	77-82
C	70-76
F	0-69

V. Tentative Course Outline
We will cover the following topics in this course (subject to change). There will be readings assigned from the two books and from external sources. Readings will be announced in advance and will need to be completed before the corresponding class meeting.

What is NLP? Applications and Methods
Regular Expressions, FSA and Formal Languages
Ngrams and Smoothing
Information Theory
Part of Speech Tagging
Word Sense Disambiguation
Lexical Semantics
Spelling Correction
Information Retrieval
Named Entity Recognition, Relation Extraction, and Temporal Relations
Summarization
Question Answering
Sentiment Analysis

VI. Communication, Writing and Submissions

Communication: Course announcements will be made through BlackBoard.

Writing: All discussions, and assignments for this course must be in standard English. Do not use slang or texting abbreviations (i.e., lol). Capitalize and use complete sentences in your discussion responses and in your paper. You can use bulleted lists if they make sense as a way to convey the information. Emoticons are acceptable as long as they are not over used and help with communication.

Before submitting work, be sure to proof read your writing and make sure that any references that you include are correct.

Submission of Work: All work for this class must be submitted as the assignment states.

ASA Style Guide: ASA Style Guides are easy to locate using an internet search. The following link is one that should work well for this class, you can access it by clicking here :
http://personal.monm.edu/jkessler/ASA-Style.htm.

VII. Academic Honesty
An important component in learning is taking on tasks, assignments and exams in an honest effort to do your best possible work. On this note, you are expected to turn in and do original work.

VIII. Etiquette and Disabilities
Please observe proper "etiquette" and "netiquette" – courteous and appropriate forms of communication and interaction – within this course. This means no personal attacks, obscene language, or intolerant expression. All viewpoints should be respected.

Giving Feedback: This course is designed along the principles of synergy and collaborative learning. Therefore, it is important that all students understand how to provide quality feedback to their peers. Here are a few tips for providing, positive, constructive, and useful feedback to peers.

- Be empathetic and remember that this environment is a safe place for making mistakes
- Use nonjudgmental language and phrases that do not attack an individual. One way of doing this is to ask the individual to discuss his/her process for making the final decision.
- Use specific questions, examples, and references to research as a way of making your point.
- Make your feedback useful by providing suggestions that the individual can understand and use to improve her/his work.

Disabilities: Please message me if you have a disability so we can discuss ways to help you succeed in the course. If you need accommodations that would affect the terms of this syllabus, you will need to provide documentation of your disability.