Information Sciences and Technology Department

AIT 526:
Introduction to Natural Language Processing
Course Syllabus

Lindi Liao, PhD
dliao2@gmu.edu

For each section, a customized syllabus with information specific to that section will be made available to registered students via the Blackboard Learning System.

I. Course Description

Objectives

This course emphasizes on both theory and practice to build a foundation of Natural Language Processing (NLP). It covers a broad set of fundamental NLP topics that are building blocks of many applications and teaches skills for creating 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 analyzing and understanding language.
- Introducing neural networks/deep learning for NLP.
- Applications to real-world problems.

Furthermore, this course emphasizes both **NLP theory** and **industry-ready hands-on technical skills**. There are a lot of NLP code examples and tutorials as well as hands-on labs and programming assignments that we specially designed to help students learn effectively. Particularly, the hands-on labs and programming assignments focus on solving the following important NLP applications:

- o Chatbots, Sentence Generation, Text Summarization
- o Information Extraction, De-Identification
- o Information Retrieval, Question Answering
- o Sentiment Analysis, Sentiment Classification
- o Building Neural Networks with PyTorch, etc.

Prerequisites

- No NLP background needed.
- Familiarity with **Python** programming, or have Java/C/C++ programming experience.
- Knowledge of statistics or probability.
- Desirable: machine learning and/or algorithm design.

Special Topics Description

This course content includes the following topics but extends their wings of knowledge to some advanced techniques:

- **Chatbots**, Regular Expression, Finite State Automata (FSA) and Formal Languages
- N-Gram **Language Models** & **Smoothing** Techniques

- **Sentence Generation**, **Text Summarization**
- **Hidden Markov Models**, **Part-Of-Speech (POS) Tagging**
- Lexical Semantics, **Word Sense Disambiguation** (WSD), Decision Lists
- **Vector Semantics**, TF-IDF, **Document Similarity**
- **Information Extraction**, **Named Entity Recognition** (NER) & **De-Identification**
- **Information Retrieval**, **Question Answering**, **Context-Free Grammar** (CFG)
- **Sentiment Classification**, Naïve Bayes
- **Neural Networks/Deep Learning for NLP**
- Text Data **Preprocessing** and Cleansing, **Web Scraping**
- Text Data **Visualization** (World Clouds, Dependencies, NER, etc.)
- **Python**, **PyTorch**, **NLTK**, **SpaCy**, and more NLP Programming Tools
- NLP **Algorithm Design** & NLP Project Management

II. Textbook and Required Materials

For this course, we will be using the following required textbooks.
- **Textbook JM2**
  - *Speech and Language Processing*, by Jurafsky and Martin (**2nd** edition). Other editions are *not* a good substitute. ISBN: 9780131873216.
- **Textbook JM3**
  - *Speech and Language Processing*, by Jurafsky and Martin (3rd edition). The draft is available [online](.).
- **Textbook NLP with Python**
  - *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. Faculty

AIT 526 Instructor: Lindi Liao, PhD
AIT 526 Course Coordinator: Lindi Liao, PhD & Ozlem Uzuner, PhD

IV. Course Learning Activities and Grade Distribution

This course includes several components as shown in the following table. The labs and tests need to be done individually. The programming assignments and a final project need to be done in groups. Your grade will be based on the following breakdown*:

Component	Quantity	Percent of Grade	Extra Credit	On-Time Submission (Extra Credit)
Online Activities *(Participation, team evaluations, peer grading...)*	∞	10%	No	No
Individual Labs	3	15%	No	No
**Programming Assignments	3	30%	Yes	Total 3 points: 1 point each
Tests *(Open-Book/Notes/Internet)*	2	20%	Yes	Up to 6 points
**Final Project	1	25%	No	No
TOTAL	**100% +** Extra Credit			

* Subject to revision before and throughout the course.
** Teamwork

VI. Course Duration, Class Participation, and Schedule

Course Duration – 8 weeks or **16 weeks**

Dates: GMU Academic Calendar: https://registrar.gmu.edu/calendars/

Class Participation

This course is designed for in-person, or distance learning ("**asynchronous**" or "**synchronous**") delivery (https://masononline.gmu.edu/course-delivery-methods):

Schedule

A detailed schedule will be published on Blackboard. As many factors may affect the development and progress of a class, the instructor reserves the right to alter the schedule as may be required to assure attainment of course objectives. The schedule is subject to revision before and throughout the course.

Registered students can access the latest class schedule on Blackboard.

VII. Grading Guidelines and Grade Scale

Grading Guidelines

Some grade components are evaluated subjectively
A: consistently above and beyond the course/assignment requirements
B: meets and occasionally exceeds the course/assignment requirements
C: minimally meets the course/assignment requirements
F: fails to meet the course/assignment requirements

Grades will be awarded in accordance with the Mason Grading System for graduate students. See the university catalog for policies: http://catalog.gmu.edu for more information.

Grading Scale

- The grading scale for this course, is:

97 – above	A+	Passing
93 – 96%	A	Passing
90 – 92%	A-	Passing
87 – 89%	B+	Passing
83 – 86%	B	Passing
77 – 82%	B-	Passing
70 – 76%	C	Passing
0 – 69%	F	Failing

NOTE: Academic success takes constant effort!

Instructor will <u>double check</u> all students' coursework graded by GTA throughout the course.

Raw scores may be <u>adjusted</u> by the Instructor to calculate final grades.

Students are responsible for checking the currency of their grade books. Grade discrepancies must be brought to instructor's attention within one week of assignment submission and 48 hours of exam submission.

No make-up for any activity, unless arranged in advance. Only in special cases, such as medical problems and family emergency, make-ups and late assignments may be allowed with verifiable proof.

Final grades will be posted to PatriotWeb, which is the only vehicle for students to obtain those grades. A student with a "hold" on his/her PatriotWeb account will be unable to access final grades until the hold has been removed by the Registrar.

VIII. 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 overused 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.

VI. Academic Honesty
An important component in learning is taking on tasks, assignments and exams in an honest effort to do your best possible work. You are expected to turn in and do **original** work.

VII. 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 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.