Knight Foundation School of Computing and Information Sciences

Course Title: Graduate Introduction to Natural Language Processing **Date:**

Course Number: CAP 5640

Number of Credits: 3

Taught by:
Mark A. Finlayson
CASE Room 362
markaf@fiu.edu / (305) 348-7988

Summary:
This class is a graduate introduction to fundamental principles and techniques of natural language processing (NLP). In this class students will be familiarized with NLP principles, techniques, and algorithms. In homework they will tackle key questions and problems in the NLP field, as well as implement a number of fundamental NLP algorithms. Students will also pursue a term-long research project, the goal being to understand in more depth an area of NLP of interest to the team. The ultimate goal of this course is to the stage for students to engage in research that advances the state of the art of NLP.

Textbook:
Daniel Jurafsky & James H. Martin, *Speech and Language Processing*, 2nd Edition, Prentice Hall, ISBN-13: 978-0131873216

Times & Locations:
Lecture will be held twice a week for 1 hour, 15 minutes;
Office hours will be held once a week.

Grading

Homework (10)	24%
Midterm Exam	24%
Final Exam	24%
Project	24%
Class Participation	4% (attendance will be noted)

Reading
Reading assignments will be distributed via the website. Readings will be associated with each lecture, and these should be completed before lecture begins, as we will rely on content in the reading during lecture.

Version: August 17, 2021, 11:49 AM

Knight Foundation School of Computing and Information Sciences
CAP 5640
Graduate Introduction to Natural Language Processing

Homework

There will be 10 homework assignments. HomeWorks will be available on the website by the beginning of the first class of the week. Homeworks must be submitted electronically as a single pdf file via the website. Late homeworks will be penalized 50%, and not accepted after the start of the first class of the week. Homeworks will be graded out of 10 points each, with point distributions noted in the assignment.

Exams

There will be two exams: a midterm and a final.

Project (100 points)

The class project will proceed in five stages (numbers in parentheses indicate possible points):

1. Proposal (10 pts; due at the end of the 4th week):
   2 pages outlining an NLP task of interest to the student, with proposed seed references and final implementation deliverables. This will set the stage for the literature review.
2. Literature Review (20 pts; due at the end of the 8th week)
   10 pages (minimum) reviewing the literature relevant to the proposed project. This review will give the student the necessary background to proceed to project implementation.
3. Summary (5 pts; due at the end of the 10th week)
   1 paragraph summarizing project deliverables. After the review has been returned with feedback, students may wish to assemble into teams, and the purpose of the summary is to allow merging of individual student's project work so far. Each team (or individual student, if working alone) should prepare a 1 paragraph summary clearly outlining what exactly the project will entail, and laying out clear and unambiguous criteria for success.
4. Presentation (15 pts; last week of class)
   A presentation to the class on the project, its goals, and results. Presentation length will be determined by the total number of students, divided equally across the 2 final classes.
5. Implementation & Report (50 pts; due last day of the normal classes)
   20 pages (minimum) describing the project implementation, what was done, what was achieved, and how it was measured. Portions of the student's literature review should be incorporated as appropriate.

Schedule of Topics

Lecture #	Topic
1	Logistics, Introduction
2	Natural Language Processing Application: Question Answering
3	Finite-State Automata
4	Tokenization
5	Morphology and Lemmatization
6	N-Grams
7	Part of Speech Tagging
8	Hidden Markov Models
9	Maximum Entropy
10,11	Grammar

12,13	Parsing
14,15	Statistical Parsing
16	Semantics
17	Lexical Semantics
18	Word Sense Disambiguation
19	Semantic Role Labeling
20	Discourse
21	Named Entities and Relations
22	Time and Events
23,24	Question Answering
25,25	Final Project Presentations