CSCI 544: Applied Natural Language Processing
Units: 4
Term—Day—Time:
Fall 2021 – Tuesday/Thursday – 2:00-3:50pm

Location:

Instructor: Xuezhe Ma
Office Hours: After each class virtually, or by appointment
Contact Info: xuezhema@isi.edu
Instructor: Mohammad Rostami
Office Hours: After each class virtually, or by appointment
Contact Info: mrostami@isi.edu

Teaching Assistant: TBD
Office Hours: TBD
Contact Info:

Grader:
Contact Info: (please CC the TA)

Catalogue Course Description
This course covers both fundamental and cutting-edge topics in Natural Language Processing (NLP) and provides students with hands-on experience in NLP applications.

Learning Objectives
The learning objectives for this course are:
- Read technical literature in Natural Language Processing (including original research articles) and answer questions about such readings.
- Implement language processing algorithms and test them on natural language data.
- Solve language processing problems and explain the reasoning behind their solution

Required Preparation:
Experience programming in Python

Course Notes
The course will be run as a lecture class with student participation strongly encouraged. There are weekly readings and students are encouraged to do the readings prior to the discussion in class. All of the course materials, including the readings, lecture slides, and homeworks will be posted online.   The class project is a significant aspect of this course and at the end of the semester students will present their projects in the form of short videos.

Required Readings and Supplementary Materials
Textbook:
Foundations of Statistical Natural Language Processing by Manning and Schutze
Speech and Language Processing by Jurafsky and Martin (3rd edition draft),
We use a set of technical papers and book chapters that are all available online. All of the required readings are listed in the course schedule.

Description and Assessment of Assignments

Homework Assignments
There will be four coding homework assignments. The assignments must be done individually. Each assignment is graded on a scale of 0-10 and the specific rubric for each assignment is given in the assignment.
Grading inquiries and questions about the grading of the homeworks and the quizzes can be asked (to the TA) within two weeks from the grading date.

Course Project
An integral part of this course is the course project, which builds on the topics and techniques covered in the class. Students can work in teams of five people on their project.
Project Timeline:
  ▪   Week 6:   Project proposals (team members, topic)
  ▪   Week 10: Project status update due (1 page status report)
  ▪   Week 13: Project final report (4 pages) and short videos (2 minutes)

Project description: Each project team will select a topic of their choice. The project types can include NLP prototype design, presenting the design of a novel, original NLP application.

Grading breakdown of the course project:
  ▪   Proposal: 10%

  ▪   Status Reports: 10%

  ▪   Project video: 10%
  ▪   Final Write-up: 70%

Grading Breakdown
Quizzes: There will be weekly quizzes at the start of class based on the material from the week before. The **highest ten quiz grades** will be considered. Missed quizzes will receive a zero grade, and there will be no make-up quizzes for any reason.
Midterm: There is a mid-term exam.
Homework: There will be four coding homework based on the topics of the class.
Final Exam: There is a multiple choice final exam at the end of the semester covering all of the material covered in the class. The final exam will be held on **December 9th 2021**, which is the date designated by USC
Class Project: Each student will do a group class project based on the topics covered in the class. Students will propose their own project, do the research and build a proof-of-concept, create a video demonstration of the proof-of-concept, and present the project in their report.

Grading Schema:

Quizzes	10%
Homework	40%
Midterm:	20%
Class Project	25%
Final	5%
Total	100%

Grades will range from A through F. The following is the breakdown for grading:

94 - 100 = A+   74 – 76.9 = C+   Below 60 is an F
90 – 93.9 = A   70 – 73.9 = C

87 – 89.9 = A-   67 – 69.9 = C-
84 – 86.9 = B+   64 – 66.9 = D+
80 – 83.9 = B   62 – 63.9 = D
77 – 79.9 =B-   60 – 61.9 = D-

Assignment Submission Policy
Homework assignments are due at 11:59pm on the due date and should be submitted on Blackboard. You can submit homework up to one week late, but you will lose 40% of the possible points for the assignment. After one week, the assignment cannot be submitted.

Course Schedule: A Weekly Breakdown

#	Date	Lecture	Reading	Instructor
1	08/24/2021	Introduction	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 2: Regular Expressions, Text Normalization, and Edit Distance.	MR
2	08/26/2021	Naive Bayes, Linear Classifier & Feature Design	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 4: Naive Bayes Classification and Sentiment HW1 Release	MR
3	08/31/2021	Word Embedding	Mikolov, Yih and Zweig (2013): Linguistic Regularities in Continuous Space Word Representations	MR
4	09/02/2021	Word Embedding	Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).	MR
	09/07/2021	Labor Day		
5	09/09/2021	Sentence Representation	Kiros et al, Skip-Thought Vectors HW1 Deadline	MR
6	09/14/2021	PyTorch & Basic Concepts in DL	HW2 Release	TA
7	09/16/2021	Sequence Labeling & HHMs	Jurafsky and Martin, 8.1-8.4 Notes from Michael Collins	XM
8	09/21/2021	MEMMs & CRFs	Notes from Michael Collins	XM
9	09/23/2021	Constituent Parsing, PCFG & CKY algorithm	Jurafsky and Martin, 12.1-12.4, 13.1-13.2 Notes from Michael Collins	XM

10	09/28/2021	Dependency Parsing, Transition-based & Graph-based Parsing	Jurafsky and Martin, 14.1-14.4 Notes from Michael Collins HW2 Deadline	XM
11	09/30/2021	Dependency Parsing, Transition-based & Graph-based Parsing	Jurafsky and Martin, 14.1-14.4 Notes from Michael Collins	XM
12	10/05/2021	Statistical Machine Translation	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 11: Machine Translation and Encoder-Decoder Models. HW3 Release	MR
13	10/07/2021	Expectation Maximization for MT	Michael Collins, The Naive Bayes Model, Maximum-Likelihood Estimation, and the EM Algorithm Project Proposal Deadline	MR
14	10/12/2021	Sequence-to-sequence models	Sutskever et al, Sequence to Sequence Learning with Neural Networks	MR
	10/14/2021	Fall Recess		
15	10/19/2021	Transformers	Attention is All You Need HW3 Deadline	XM
16	10/21/2021	Transformers	TBA	XM
17	10/26/2021	Midterm	HW4 Release	
18	10/28/2021	Advanced topics in MT	TBA	XM
19	11/02/2021	N-gram Language Models, Smoothing	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 3: N-gram Language Models.	MR
20	11/04/2021	Neural Language Models & Contextualized Embeddings	BERT, GPT2 Project Status Report Deadline	XM
21	11/09/2021	Pre-training & Natural language inference	BERT, GPT2 HW4 Deadline	XM

22	11/11/2021	NLP for low-resource learning regimes	1. Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach 2. An embarrassingly simple approach to zero-shot learning	MR
23	11/16/2021	Information Extraction: Semantic Role Labeling, Relation Extraction, Events		MC
24	11/18/2021	Language Grounding		JT
	11/23/2021	Thanksgiving Day		
25	11/25/2021	Question Answering		FI
26	11/30/2021	Event NLU		MC
27	12/02/2021	Review	Project Final Report & Video Deadline	TA

Statement on Academic Conduct and Support Systems

Academic Conduct
Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, Behavior Violating University Standards https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu or to the *Department of Public Safety* http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems
A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not

English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* *http://emergency.usc.edu* will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.