USC Viterbi School of Engineering

CSCI 544: Applied Natural Language Processing

Units: 4
Term—Day—Time:
Fall 2023 – Tue/Thur – 5:30-7:20 PM

Location: SAL 101

Instructor: Mohammad Rostami
Office Hours: Tuesdays 2-3 PM and virtually at:
https://usc.zoom.us/my/mrostami
Contact Info: mrostami@isi.edu

Teaching Assistants:

Name: Zizhao Hu

Office Hours: TBD virtually at: https://usc.zoom.us/j/4127297114

Contact Info: zizhaoh@usc.edu

Name: Amin Banayeeanzade

Office Hours: TBD at:

Contact Info: banayeea@usc.edu

Name: Mengxi Wu

Office Hours: TBD at:

Contact Info: mengxiwu@usc.edu

Name: Yuliang Cai

Office Hours: TBD, personal meeting id: 923 042 3461

Contact Info: caiyulia@usc.edu

Name: Ehsan Qasemi

Office Hours: TBD,

Contact Info: qasemi@usc.edu

Graders:
Shoumik Gandre (shoumika@usc.edu)

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 syllabus is available at the following link:
https://docs.google.com/document/d/1llcoPTw818iyCL9yEIihOaZHE_zTbfcb
Please check the link continuously because there may be minor changes in the future.

PLEASE CHECK THE LINK CONTINUALLY FOR THE LATEST VERSION

- 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 with 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 class..

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 0-100 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 (2 pages)
- Week 10: Project status update due (one page status report)
- Week 13: Project final report (4 pages)

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 Presentation: 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 five quiz grades** will be considered. Missed quizzes will receive a zero grade, and there will be no make-up quizzes..
Homework: There will be four coding homework based on the topics of the class.
Final Exam: There is a multiple choice midterm exam, covering about the first half of the material covered in 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 the last day of the classes.
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.
Paper Presentation: The research article presentation is an activity where project teams read an article of their choice and present it to the class. It does not have to be related to the team's project. You will read the article, identify the central points of the research, and present that research to the class. You will also comment on research article presentations of fellow students.

Grading Schema:

Quizzes	5%
Homework	40%
Paper Presentation:	10%
Class Project	30%
Midterm	5%
Final	10%
Total	100%

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

	78 – 80.9 = C+	Below 60 is an F
93 – 100 = A	74 – 77.9 = C	
90 – 92.9 = A-	71 – 73.9 = C-	
87 – 89.9 = B+	68 – 70.9 = D+	
84 – 86.9 = B	65 – 67.9 = D	
81 – 83.9 =B-	62 – 64.9 = D-	

Assignment Submission Policy
Homework assignments are due at 2:59pm on the due date and should be submitted on Blackboard. Late submissions without prior approval, e.g., due to medical conditions, will not be graded.

Course Schedule: A Weekly Breakdown

#	Date	Lecture	Reading
1	08/22/2023	Introduction	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 2: Regular Expressions, Text Normalization, and Edit Distance.

2	08/24/2023	Naive Bayes, Linear Classifier	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 4: Naive Bayes Classification and Sentiment Analysis
3	08/29/2023	Perceptron, SVM, Logistic Regression, Model evaluation	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 5: Logistic Regression. Hal Daumé III, A Course in Machine Learning (v. 0.99 draft), Chapter 4: The Perceptron. Tutorial on Support Vector Machine (SVM) by Vikramaditya Jakkula HW1 Release Quiz1
4	08/31/2023	Sequence Labeling & HMMs	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 8: Part-of-Speech Tagging Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Appendix A: Hidden Markov Models Michael Collins: Tagging Problems, and Hidden Markov Models
5	09/05/2023	Discriminative and Unsupervised Sequence Labeling	Lafferty, McCallum and Pereira (2001): Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data Michael Collins: Log-Linear Models, MEMMs, and CRFs Quiz2
6	09/07/2023	Word Embedding	Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013). Mikolov, Yih and Zweig (2013): Linguistic Regularities in Continuous Space Word Representations Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 6
7	09/12/2023	Word Embedding (continued), Neural Network	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 7 HW1 Deadline HW2 Release Quiz3

8	09/14/2023	PyTorch & Basic Concepts in DL	
9	09/19/2023	Paper Presentation	Paper Selection Deadline
10	09/21/2023	Paper Presentation	
11	09/26/2023	Paper Presentation	HW1 Deadline HW2 Release
12	09/28/2023	Paper Presentation	
13	10/03/2023	Paper Presentation	
14	10/05/2023	RNN, Sentence Representation	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 9: Deep Learning Architectures for Sequence Processing Kiros et al, Skip-Thought Vectors Project Proposal Deadline Quiz4 HW2 Deadline HW3 Release
15	10/10/2023	Language Models, smoothing	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 3: N-gram Language Models.
16	10/17/2023	Midterm Exam	
17	10/19/2023	Spelling Correction & Statistical Machine Translation	Spelling Correction and the Noisy Channel, Jurafsky and Martin Notes by Michael Collins Word-Based Models by Philipp Koehn
18	10/24/2023	Expectation Maximization for MT	Michael Collins, The Naive Bayes Model, Maximum-Likelihood Estimation, and the EM Algorithm Notes by Michael Collins Word-Based Models by Philipp Koehn HW3 Deadline HW4 Release

			Quiz5
19	10/26/2023	Phrase-Based Translation, MT Decoding	Statistical Phrase-Based Translation, Koehn et al Project Status Report Deadline
20	10/31/2023	Transformers	Attention is All You Need Quiz6
21	11/02/2023	Neural Language Models & Contextualized Embeddings-I	BERT, GPT2
22	11/07/2023	Neural Language Models & Contextualized Embeddings-II	BERT, GPT2 Quiz7
23	11/09/2023	Project Presentation	HW4 Deadline
24	11/14/2023	Project Presentation	Quiz8
25	11/16/2023	Project Presentation	
26	11/21/2023	Project Presentation	
27	11/28/2023	Project Presentation	Project Final Report Deadline
28	11/30/2023	Final Exam	

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.