

USC Viterbi School of Engineering

CSCI 544: Applied Natural Language Processing

Units: 4

Term—Day—Time:

Spring 2025 – Tue/Thur – 5:00-6:50 PM

Location: THH 201

Piazza: <https://piazza.com/usc/spring2025/csci544>

Instructor: Xuezhe Ma

Office Hours: After each class or by appointment

Contact Info: xuezhema@usc.edu

Teaching Assistants:

Name: Nan Xu

Office Hours: book a 15 min slot upon request

Contact Info: nanx@usc.edu

Name: Linghao Jin

Office Hours: Fri 10:00-11:00 AM at

<https://usc.zoom.us/j/92557004633>

Contact Info: linghaoj@usc.edu

Name: Joshua Robinson

Office Hours: TBA

Contact Info: joshuar5@usc.edu

Name: Muzi Tao

Office Hours: Wed 2:15-3:15 PM at

<https://usc.zoom.us/j/96876003850>

Contact Info: muzitao@usc.edu

Name: Yubo Zhang

Office Hours: book a 15 min slot upon request

Contact Info: yuboz@usc.edu

Name: Nuan Wen

Office Hours: Mon 1:30-2:30 PM at
<https://usc.zoom.us/j/92049232838?pwd=Ync1DS05nX9DmRaJp1VkoG3s4amYHa.1>

Contact Info: nuanwen@usc.edu

Graders: TBD

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 here.

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 8: Project proposals (2 pages)
- Week 11: Project status update due (one page status report)
- Week 14: 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.

Midterm 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.

Grading Schema:

Quizzes	5%
Homework	40%
Class Project	35%
Midterm	10%
Final	10%
<hr/>	
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
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	01/14/2025	Introduction	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 2: Regular Expressions, Text Normalization, and Edit Distance.
2	01/16/2025	Naive Bayes, Linear Classifier, Logistic Regression, etc.	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 4: Naive Bayes Classification and Sentiment Analysis 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
3	01/21/2025	Word Embedding	Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013). Mikolov, Tomas, et al. "Distributed Representations of Words and Phrases and their Compositionality". Mikolov, Yih and Zweig (2013): Linguistic Regularities in Continuous Space Word Representations Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 7 Notes on Noise Contrastive Estimation and Negative Sampling. Chris Dyer. Quiz1
4	01/23/2025	Deep Learning for NLP, MLP & CNNs	Deep Learning Book Part II: Chapter 6-9
5	01/28/2025	Sequence Labeling: HMMs, MEMMs & CRFs	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 Notes from Michael Collins Quiz2
6	01/30/2025	PyTorch & Basic Concepts in DL	

			HW1 Deadline HW2 Release
7	02/04/2025	Sequence Labeling: RNN & LSTM	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 7 LSTM-CNNs-CRF Quiz3 Group Formation Deadline
8	02/06/2025	Syntactic Parsing	Jurafsky and Martin, 14.1-14.4 Deep BiAffine Parser NeuroMST Parser Stack-LSTM Parser Stack-Pointer Parser
9	02/11/2025	Sequence-to-sequence Learning & Machine Translation	Jurafsky and Martin, Speech and Language Processing (3rd edition draft), Chapter 10: Machine Translation and Encoder-Decoder Models. Quiz4
10	02/13/2025	Advanced Topics in Machine Translation	BPE Context-Aware NMT HW2 Deadline HW3 Release
11	02/18/2025	Transformers-I	Attention is All You Need Quiz5
12	02/20/2025	Transformers-II	Residual Connection, Normalization Layers, Positional Embeddings
13	02/25/2025	Neural Language Models & Contextualized Embeddings	BERT, GPT2, BART Quiz6
14	02/27/2025	Deep Generative Models	VAE, Generative Flows, Diffusion Models HW3 Deadline HW4 Release
15	03/04/2025	Advanced Transformers-I	Longformer, Mega Quiz7
16	03/06/2025	Reinforcement Learning from Human Feedback (Guess lecture from Dr. Jon May)	Project Proposal Deadline
17	03/11/2025	Midterm Exam	
18	03/13/2025	Spring Recess (No Class)	HW4 Deadline

19	03/25/2025	Guest Lecture (Jieyu Zhao)	
20	03/27/2025	Advanced Transformers-II	Quiz8
21	04/01/2025	Advanced Techniques in Large-Scale Pre-training	
22	04/03/2025	Guest Lecture	Project Status Report Deadline
23	04/08/2025	Project Presentation	
24	04/10/2025	Project Presentation	
25	04/15/2025	Project Presentation	
26	04/17/2025	Project Presentation	
27	04/22/2025	Project Presentation	
28	04/24/2025	Project Presentation	Project Final Report Deadline
29	04/29/2025	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/departement-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.