CS/Ling 581 Computational Linguistics

San Diego State University
Schedule # 21215 (CS)
Schedule # 22207 (Ling)
Spring 2022
TTh 11:00am–12:15pm
SH-316[*]

> [*]Following campus guidelines, this class will meet synchronously on Zoom until Feb 4, 2022. After that, class will be held in person in SH-316. Updated information will be provided via Canvas as it becomes available.

This course will serve as an introduction to the field of computational linguistics, which includes aspects of speech recognition, natural language processing, information retrieval, and information extraction.The course begins with an introduction to finite-state automata and some basic natural language applications; this is extended to finite-state transducers with applications in morphology (word structure). Other topics covered: ngram language models, classifiers (Naive Bayes and Logistic Regression), sentiment analysis, part of speech tagging, context-free grammars and context-free parsing (with statistical extensions), and distributuional semantics.

Prof. Rob Malouf	
Website:	malouf.sdsu.edu
Email:	rmalouf@sdsu.edu
Office hours:	TTh 2:00–3:00 or by appointment
Real office:	SHW 244
Zoom office:	SDSU.zoom.us/j/88612436627

Student Learning Outcomes

After completion of this course, students will be able to:

1. Identify and use three different incarnations of Finite-State Natural Language processing methods, Regular Expressions, Finite-State Automata (as language recognizers and generators), and Finite-State Transducers (as morphological analyzers).

2. Apply basic laws of probability to derive three different kinds of conditional probability language models: n-gram models, HMM taggers, and Naive Bayes Classifiers

3. Explain the geometric justification for a variety of semantic similarity models for words.

4. Derive and apply algorithms for Naive Bayes, logistic regression, and perceptron classifiers

5. Identify the major components in a CKY context-free parser and a transition-based dependency parser

Course Materials

The textbook for this course is:

- Jurafsky, Daniel and Martin, James H. 2021. *Speech and Language Processing.* Third edition draft. web.stanford.edu/~jurafsky/slp3/

This book is only available as a downloadble PDF. There are hardcopy versions of the 1st and 2nd editions for sale on, e.g., Amazon, but those are the wrong books! **Do not buy any books for this class!**

Course Design

The grade for the course will be based on readings and problem sets (40%), a take-home midterm exam (30%), and a take-home final exam (30%). Programming assignments should be completed in Python. Students with no programming background will find this course very challenging!

Schedule

Week	Topic	Reading
1	Introduction	Chapter 1
2–3	Finite state models	Chapter 2
4	Language models	Chapter 3
5–7	Sequence labeling models	Chapter 8
8	**Midterm**	
9	Word meanings	Chapter 6
10–11	Classifiers	Chapter 4, 5
12–13	Context free grammars	Chapter 12, 13
14	Dependency parsing	Chapter 14
15	Wrap up and review	

Land Acknowledgement

For millennia, the Kumeyaay people have been a part of this land. This land has nourished, healed, protected and embraced them for many generations in a relationship of balance and harmony. As members of the San Diego State University community, we acknowledge this legacy. We promote this balance and harmony. We find inspiration from this land, the land of the Kumeyaay.

Essential Student Information

- Compliance with CSU / SDSU vaccination and facial covering policies (newscenter.sdsu.edu/student_affairs/srr/covid-policies.aspx) is required.

- Your SDSU email address (gsuite.sdsu.edu) will be used for all course-related communications.

- The Student Conduct Code (newscenter.sdsu.edu/student_affairs/srr/conduct.aspx) prohibits conduct disruptive to instruction, including academic dishonesty and the unauthorized recording, dissemination, or publication (including on websites or social media) of lectures or other course materials.

- SDSU provides disability-related accommodations via the Student Ability Success Center (sascinfo@sdsu.edu | sdsu.edu/sasc). Please allow 10–14 business days for this process.

- The Family Educational Rights and Privacy Act (FERPA) (bfa.sdsu.edu/hr/oerc/students/ferpa.aspx) mandates the protection of student information, including contact information, grades, and graded assignments. I will not post grades or leave graded assignments in public places. Students will be notified at the time of an assignment if copies of student work will be retained beyond the end of the semester or used as examples for future students or the wider public.

- As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I am required to share information regarding sexual violence on SDSU's campus with the Title IX (titleix.sdsu.edu) coordinator, Gail Mendez (619-594-6464), who will contact you to let you know about support services at SDSU and possibilities for holding accountable the person who harmed you. If you do not want the Title IX Officer notified, you can speak confidentially SDSU's Sexual Violence Victim Advocate (619-594-0210) or Counseling and Psychological Services (619-594-5220, psycserv@sdsu.edu).

- Class rosters are provided to the instructor with the student's legal name. Please let me know if you would prefer an alternate name and/or gender pronoun.

- Need help finding an advisor, tutor, counselor, emergency economic assistance, or other support? Contact the SDSU Student Success Help Desk (studentsuccess.sdsu.edu) Monday through Friday, 9:00 AM to 4:30 PM.

- For technical or computing assistance, contact the Library Computing Hub (virtual-academic-help.sdsu.edu/technology).

Absences

- According to the University Policy File, students should notify the instructors of affected courses of planned absences for religious observances by the end of the second week of classes.

- If you are absent more than five days due to illness or injury, you may contact Student Health Services (shs.sdsu.edu/index.asp) for help in communicating your absence.

- If you miss class because you have been diagnosed with or are required to quarantine due to exposure to COVID-19, contact vpsafrontdesk@sdsu.edu to notify the university

Academic Honesty

The University adheres to a strict policy prohibiting cheating and plagiarism, including

- Copying, in part or in whole, from another's test or other examination.

- Obtaining copies of a test, an examination, or other course material without the permission of the instructor.

- Collaborating with another or others in coursework without the permission of the instructor.

- Falsifying records, laboratory work, or other course data.

- Submitting work previously presented in another course, if contrary to the policies of the course.

- Altering or interfering with grading procedures.

- Assisting another student in any of the above.

- Using sources verbatim or paraphrasing without giving proper attribution (this can include phrases, sentences, paragraphs and/or pages of work).

- Copying and pasting work from an online or offline source directly and calling it one's own.

- Using information found from an online or offline source without giving the author credit.

- Replacing words or phrases from another source and inserting one's own words or phrases.

Under CSU policy, instructors must report instances of academic misconduct to the Center for Student Rights and Responsibilities for disciplinary review by the University, which may lead to probation, suspension, or expulsion. Instructors may also, at their discretion, penalize student grades on any assignment or assessment discovered to have been produced in an academically dishonest manner.