CS/Ling 581 Computational Linguistics

San Diego State University
Schedule # 13721 (CS), 11492 (Ling)
Spring 2023
TTh 11:00am–12:15pm
LH-340

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 text processing. Other topics covered: n-gram language models, classifiers (Naive Bayes and Logistic Regression), sentiment analysis, dependency parsing, and distributional semantics.

Prof. Rob Malouf
Website: malouf.sdsu.edu
Email: rmalouf@sdsu.edu
Office hours: TTh 1:00–2:00 or by appointment
Real office: SHW 244
Zoom office: SDSU.zoom.us/j/88663149131

Student Learning Outcomes

Upon successful completion, students will have the knowledge and skills 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 and logistic regression classifiers
5. Identify the major components in a transition-based dependency parser

Course Materials

The textbook for this course is:

- Jurafsky, Daniel and Martin, James H. 2023. *Speech and Language Processing*. Third edition draft (Jan 7, 2023). web.stanford.edu/~jurafsky/slp3/

This book is only available as a downloadable PDF. There are hard copy 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!**

All other course information, additional readings, assignments, slides, etc. will be available on Canvas.

Course Design

The grade for the course will be based on readings and homework assignments (50%), a midterm exam (25%), and a final exam (25%). Programming assignments should be completed in Python. Students with no programming background will find this course extremely challenging!

Schedule

Week	Topic	Reading
1	Introduction	
2—3	Finite state machines	Chapter 2
4—5	Language models	Chapter 3
6—7	Classifiers	Chapter 4, 5
8	**Midterm**	
9	Word meanings	Chapter 6
10–11	Sequence labeling	Chapter 8
12–13	Neural nets	Chapter 7, 9
14	Dependency parsing	Chapter 18
15	Future prospects	

Revised: 14 January 2023