

CS445 - NATURAL LANGUAGE PROCESSING

Fall 2024

Instructor:	Dilara Keküllüoğlu	Email:	dilara.kekulluoglu@sabanciuniv.edu
Time:	M 9:40 - 11:30, F 14:40 - 15:30	Place:	FASS G062

TAs: Ayşegül Rana Erdemli, Semih Gülüm, and Kerem Aydın

Office Hours:

Dilara Keküllüoğlu	Tuesday 15:00-17:00	FENS 1089
Ayşegül Rana Erdemli	Wednesday 13:40-14:30	https://sabanciuniv.zoom.us/j/95271697406
Semih Gülüm	Tuesday 9:40-10:30	https://sabanciuniv.zoom.us/j/93845280002
Kerem Aydın	Thursday 12:40-13:30	https://sabanciuniv.zoom.us/j/95444315184

You can also come visit the instructor in their office by appointment through email. Please add CS445 in the subject in your emails so we do not miss them.

Main Reference: Daniel Jurafsky and James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd edition online). <https://web.stanford.edu/~jurafsky/slp3/>

Optional Reference: Steven Bird, Ewan Klein, and Edward Loper, *Natural Language Processing with Python*. <https://www.nltk.org/book/>

Contents: This course studies the theory, design and implementation of natural language processing systems. Topics include text processing, regular expressions, statistical properties of text, edit distance, language modeling, text classification, sequence modeling, topic modeling, computational morphology, neural networks for NLP, chatbots, transfer learning for NLP.

Learning Objectives: This course aims to give foundational knowledge of the Natural Language Processing field. A student who successfully fulfills the course requirements will be able to:

1. **describe** the statistical properties of text in natural language.
2. **implement** programs that can process textual data and extract valuable information from it.
3. **apply** well-known language processing techniques to text.
4. **explain** the significance and principles of language modeling.
5. **develop** machine learning models to classify documents, sub-documents or terms.
6. **assess** the quality of natural language processing models applied to text.

Prerequisites: CS 204 and (CS 210 or DSA 210)

Course Outline:

- Week 1: Introduction to Natural Language Processing
- Week 2: Regular Expressions and Text Normalization
- Week 3: Text Analysis and Edit Distance
- Week 4: Language Models
- Week 5: Text Classifications
- Week 6: Semantics, Embeddings, and Sentiment
- Week 7: Morphology, POS Tagging and Named Entities
- Week 8: Hidden Markov Models
- Week 9: Context-Free Grammars
- Week 10: Parsing
- Week 11: Content Review and Midterm
- Week 12: Deep Learning for NLP
- Week 13: Conversational Agents
- Week 14: NLP as a tool in Current Research

Grading Policy: Coding Assignments (20%), Midterm (40%), Project (40%).

Important Dates:

Midterm	Week 11
Project Deadline	Week 14

Class Policy:

- Regular attendance is essential and expected. Attendance will be taken but will not count towards your final grade.
- You have three points you can use for late submissions. Every day you are late for a coding assignment will remove a point from your balance. You can use these three points as you wish for the coding assignments. If you do not have any points left, you will receive zero marks for missed deadlines.
- Late submissions are not accepted for the project deadlines.

Grading:

- Midterm exam, coding assignments, and the projects will be counted for the last letter grade.
- Letter grades will be assigned based on curve, considering the distribution of the grades in the class.
- Any one of the followings lead to failing the course: (1) overall total grade below 40 (over 100), (2) final exam grade below 30 (over 100).

Academic Honesty: All students must follow the university guidelines of academic integrity. <https://www.sabanciuniv.edu/en/academic-integrity-statement>

Flexibility: The contents of the syllabus are subject to change. The students will be notified when there is an update.