Computational Methods in Linguistics

Boğaziçi University
Ling 360 – Spring 2024

Time	TT78 - WW34	**Course Website**	Moodle
Place	İB312 - NH202		
Instructor	Ümit Atlamaz	**TA**	Onur Keleş
Email	umit.atlamaz@bogazici.edu.tr	**Email**	onur.keles1@bogazici.edu.tr
Office Hours	By Appointment	**PS Hours**	W4

Course Description

This course is an introduction to the field of computational linguistics. Throughout the semester, we will analyze human language as a computational device, discuss some of the major theoretical aspects of computational linguistics, and implement practical applications of these theories in Python. The intended audience for this course is students with some background in linguistics but not much experience in computational methods or programming. This course is not intended for students with a strong background in computer science. Feel free to contact the instructor if you still want to take the course or audit some of the classes.

Course Learning Goals

- Analyzing natural language from a computational perspective
- Understanding some of the major symbolic and non-symbolic approaches in computational linguistics
- Applying some computational technology (i.e. nlp techniques) to natural language
- Learning some Python to do basic text processing

Programming

In this course, we will use Python 3 as the programming language. You will need a Python 3 interpreter for this course. I will use Google Colab, which requires no installation and works on any machine (mac, pc, linux, even ios and android.). All you need is a Google account. If you'd like to use some other interpreter, feel free to use Jupyter Notebooks or whichever IDE you like. Please note that my support for anything other than Google Colab will be minimal.

Important Note

We do not expect any previous familiarity with programming or Python. Python taught in this class will be sufficient for you to do the assignments and the project. However, in the past few years, some students mentioned that they struggled due to a lack of prior exposure to programming. If you are not willing and ready to devote a consistent amount of time to programming exercises, **you should consider dropping the course** and retaking it after you have familiarized yourself with a programming language.

Reading

Throughout the semester, we will use chapters from the following books and some others. Relevant chapters will be posted on Moodle every week:

- Bird, Steven, Ewan Klein, & Edward Loper. *Natural Language Processing with Python*. URL: https://www.nltk.org/book
- Jurafsky, Daniel & James H. Martin. 2020. *Speech and Language Procesing 3rd Edition*. URL: https://web.stanford.edu/ jurafsky/slp3

Course Requirements & Evaluation

Final grades will be determined by your performance on assignments, in class presentations, a final project, and your participation.

Evaluation Criteria		**Grade Catalog**	
Assignments	60%	AA	$\geq 90\%$
Attendance & Participation	5%	BA	$\geq 85\%$
Project Presentation	10%	BB	$\geq 80\%$
Final Project	25%	CB	$\geq 75\%$
		CC	$\geq 70\%$
		DC	$\geq 65\%$
		DD	$\geq 55\%$
		F	$\leq 54\%$

Final Project

- A python project on a topic covered in the course submitted as a Colab or Jupyter Notebook.
- Final projects are expected to be team projects. Each team should consist of 2-4 people. If you are not willing to work with anyone, **you should consider dropping the course**.

Policies

Attendance
Attendance is mandatory. You miss 6 or more classes without an excuse, you get 0 on your attendance and participation grade.

Late Assignments & Projects
Late assignments and projects will not be accepted, except:

- If you are ill and missed the deadline. You will be asked to provide proof of illness (doctor's report).
- If you know you will miss a deadline due to a family, cultural, or religious observance, it is your responsibility to contact your instructor early in the semester. You should arrange to take an alternative assessment, or to submit the assignment on another date.

Academic Integrity
Students are expected to know and follow Boğaziçi University policies on Academic Integrity:

- Student Rights and Responsibilities

Students may not collaborate on answering questions in homework assignments or exams; homework assignments and exams must be done independently. Collaboration is a violation of the Academic Integrity policy. All instances of plagiarism will be reported.

Disability Services
Students with disabilities requesting accommodations should reach out to the instructor as well as:

- Students with Disabilities Unit
- Engelli Öğrenciler Danışma Ve Koordinasyon Birimi

Please, make sure your email registered in ÖBİKAS is in use. It is your responsibility to keep track of the email announcements.

Tentative Schedule

Disclaimer: The schedule is tentative and subject to change.

Date	Topic
Week 1	Intro, Syllabus, Marr's Three Levels, Python Setup Formal Languages, Grammars
Week 2	Regular Expressions
Week 3	Text Processing
Week 4	String similarity, Edit Distance
Week 5	Finite State Automata
Week 6	Finite State Transducers
Week 7	N-gram Language Modeling
Week 8	Parsing (off the shelf NLP Tools)
Week 9	SPRING BREAK
Week 10	BOW, Document Classification, Naive Bayes
Week 11	Naive Bayes & Document Classification
Week 12	Vector Semantics
Week 13	Generalization Learning, Tolerance Principle
Week 14	Project Presentations