Natural Language Processing

Course Administration & Presentation

Master Degree in Computer Engineering
University of Padua
Lecturer : Giorgio Satta

https://stem.elearning.unipd.it/course/view.php?id=15243

Lecturer

Giorgio Satta
satta@dei.unipd.it
http://www.dei.unipd.it/~satta/

Class hours
Thurs 12:30–14:30, **under email appointment**
Building: DEI/G, fourth floor
Zoom meeting: https://unipd.zoom.us/j/82934199797

Schedule

	Thurs	**Fri**
10:30–11:30	room Ce	room Ce
11:30–12:30	room Ce	room Ce
12:30–13:30		class hours
13:30–14:30		class hours

Content outline

LECTURE	WEEK	TOPIC	BOOK & CHAPTERS
		PRELIMINARIES	
1	1	Natural language processing: Introduction	Slides from the lecture
3	1,2	Text normalization	Speech and Language Processing, Jurafsky and Martin, Chapter 2
		LARGE LANGUAGE MODELS	
4	2, 3	Word embeddings	Speech and Language Processing, Jurafsky and Martin, Chapter 5
5	3, 4	Language models	Speech and Language Processing, Jurafsky and Martin, Chapters 3, 7
6	4, 5, 6	Large language models	Speech and Language Processing, Jurafsky and Martin, Chapters 9, 10
7	6, 7	ChatBots	11
		STRUCURED PREDICTION	
8	8	Sequence labeling	Speech and Language Processing, Jurafsky and Martin, Chapter 17 Introduction to Natural Language Processing, Eisenstein, Chapter 7
10	9, 10	Dependency Parsing	Speech and Language Processing, Jurafsky and Martin, Chapter 19
		END-TO-END APPLICATIONS	
12	10, 11	Machine Translation	Speech and Language Processing, Jurafsky and Martin, Chapter 12
13	11, 12	Question Answering	Speech and Language Processing, Jurafsky and Martin, Chapter 14 Introduction to Natural Language Processing, Eisenstein, Chapter 17
		CONCLUSIONS	
15	12	Wrap-up	Slides from the lecture

Lab sessions

SESSION	WEEK	TASK	MODEL
1	3	Static word embedding	Skip-gram
2	5	Contextual word embedding	BERT
3	7	Retrieval augmented generation	ChatBot
4	9	Named-entity recognition	Sequence labeling

Course requirements

Students should have **basic knowledge** of the following subjects

- calculus + linear algebra
- machine learning + deep learning
- probability theory + information theory
- computer algorithms + dynamic programming
- automata theory + rewriting grammars + formal languages
- Python + NumPy + PyTorch

The class also uses basic knowledge from **linguistics**: all of the working notions in linguistics will be properly introduced.

Textbook

Speech and Language Processing (3rd ed., draft)
Dan Jurafsky and James H. Martin
January 12th, 2025
`https://web.stanford.edu/~jurafsky/slp3/`

Several chapters about machine learning, these will be given for granted.

Additional resources

Additional material and resources available through **moodle page** of the class

- slides, videos and notebooks
- three forums for class, lab, and project discussion

Project

Paper presentation:

- choose a scientific article from a given list of 2025 NLP conferences
- read, understand, and record a short video presentation

NLP system:

- for the announced task/dataset, provide
  - exploratory data analysis
  - baseline
  - neural approach
  - result comparison with state-of-the-art (SotA) and discussion
- report your project in a notebook

Coursework

Written test	70%
Project	
1. NLP system	25%
2. article presentation	05%
	100%

Hard rules

Project groups of **maximum** two people.

Project must be presented **before** earliest written test of the students in the group.

To pass the final exam you need to score 18/30 or above **for both** the written test and the project.