

# Human Language Technologies

Federico Matteoni

A.A. 2021/22

# Index

0.1	Introduction . . . . .	2
-----	------------------------	---

## 0.1 Introduction

Prof. Giuseppe Attardi

Prerequisites are: proficiency in Python, basic probability and statistics, calculus and linear algebra and notions of machine learning.

**What will we learn** Understanding of and ability to use effective modern methods for **Natural Language Processing**. From traditional methods to current advanced ones like RNN, Attentions...

Understanding the difficulties in dealing with NL and the capabilities of current technologies, with experience with **modern tools** and aiming towards the ability to build systems for some major NLP tasks: word similarities, parsing, machine translation, entity recognition, question answering, sentiment analysis, dialogue system...

**Books** Speech and Language Processing (Jurafsky, Martin), Deep Learning (Goodfellow, Bengio, Courville), Natural Language Processing in Python (Bird, Klein, Loper)

**Exam** Project (alone or team of 2-3 people) with the aim to experiment with techniques in a realistic setting using data from competitions (Kaggle, CoNLL, SemEval, Evalita...). The topic will be proposed by the team or chosen from a list of suggestions.

### Experimental Approach

1. Formulate hypothesis
2. Implement technique
3. Train and test
4. Apply evaluation metric
5. If not improved:
  - Perform error analysis
  - Revise hypothesis
6. Repeat!

**Motivations** Language is the most distinctive feature of human intelligence, **it shapes thought**. Emulating language capabilities is a scientific challenge, a **keystone for intelligent systems** (see: Turing test)

**Structured vs unstructured data** The largest amount of information shared with each other is unstructured, primarily text. Information is mostly communicated by e-mails, reports, articles, conversations, media... and attempts to turn text to structured (HTML) or microformat only scratched the surface.

Problems: requires universal agreed **ontologies** and additional effort. Entity linking attempts to provide a bridge.