Natural Language Processing and Large Language Models

Corso di Laurea Magistrale in Ingegneria Informatica

Lesson 0

Course Introduction

Nicola Capuano and Antonio Greco

DIEM – University of Salerno

Objectives

Knowledge:

- Basic concepts of **Natural Language Processing (NLP)**
- **Natural Language Understanding** and **Generation**
- **Statistical Approaches** to NLP
- **Large Language Models (LLM)** based on **Transformers**
- NLP **applications** with LLM
- **Prompt Engineering** and **Fine Tuning** of LLM

Abilities:

- **Design and implementation of a NLP system** based on LLMs, integrating existing technologies and tools

Fundamentals of NLP

- **Basic concepts**, Evolution and Applications of NLP
- **Representing text**: Tokenization, Stemming, Lemmatization, POS tagging
- **Math with Words**: Bag of Words, Vector Space Model, TF-IDF, Search Engines
- **Text Classification**: Topic Labelling, Sentiment Analysis
- **Word Embeddings**: Word2Vec, CBOW, Skip-Gram, GloVe, FastText
- **Neural Networks for NLP**: RNN, LSTM, GRU, CNN, Introduction to Text Generation
- **Information Extraction**: Parsing, Named Entity Recognition
- **Question Answering** and **Dialog Engines** (chatbots)

Transformers

- **Self-Attention**, Multi-Head Attention, Positional Encoding, Masking
- **Encoder** and **Decoder** of a Transformer
- Introduction to **HuggingFace**
- **Encoder-Decoder** or Seq2Seq models (translation and summarization)
- **Encoder-only Models** (sentence classification and named entity recognition)
- **Decoder-only Models** (text generation)
- Definition and training of a **Large Language Model**

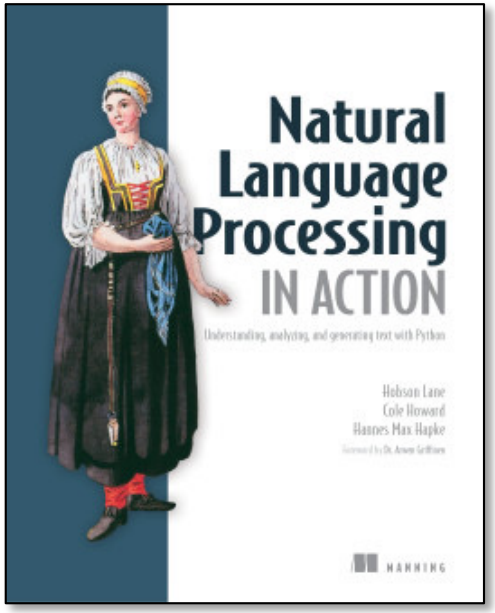
Prompt Engineering

- **Zero-shot** and **Few-shot** Prompting
- **Chain-of-Thought**, **Self-Consistency**, **Prompt Chaining**
- **Role** Prompting, **Structured** Prompts, **System** Prompts
- **Retrieval Augmented Generation**

LLM Fine Tuning

- **Feature-Based** Fine Tuning
- **Parameter Efficient** Fine Tuning and **Low Rank Adaptation**
- **Reinforcement Learning with Human Feedback**

Textbook



H. Lane, C. Howard, H. M. Hapke

Natural Language Processing IN ACTION

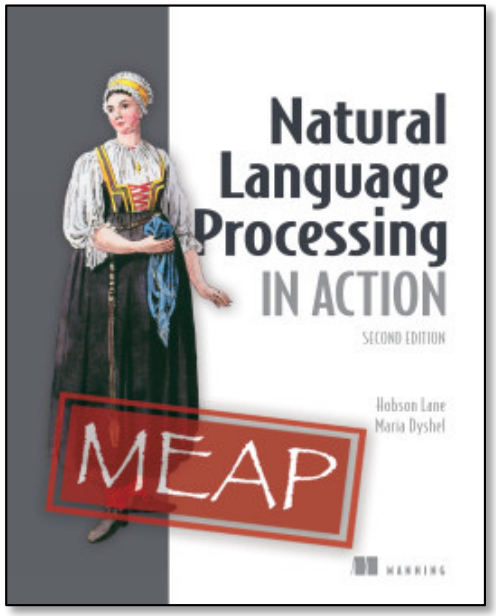
Understanding, analyzing, and generating text with Python

Manning, *2019*

Second Edition in **fall 2024**

Early Access version available online:

https://www.manning.com/books/natural-language-processing-in-action-second-edition



Further Info

Teachers

- **Nicola Capuano**

  DIEM, FSTEC-05P02007

  ncapuano@unisa.it

  089 964292

- **Antonio Greco**

  DIEM, FSTEC-05P01036

  agreco@unisa.it

  089 963003

Online Material

- https://elearning.unisa.it/

Exam

- Realization of a **project work**
- **Oral exam** (including the discussion of the project work)

Natural Language Processing and Large Language Models

Corso di Laurea Magistrale in Ingegneria Informatica

Lesson 0

Course Introduction

Nicola Capuano and Antonio Greco

DIEM – University of Salerno