

## SYLLABUS DEL CORSO

### Natural Language Processing

2324-2-FDS01Q011

---

#### Obiettivi

Il corso si propone di introdurre gli elementi fondazionali e i più recenti modelli computazionali avanzati relativi al processamento del linguaggio naturale. Lo studente, al termine dell'attività formativa, avrà acquisito conoscenze e competenze relative ad algoritmi, strumenti e modelli per l'elaborazione e l'analisi del linguaggio naturale, al fine di sfruttare i più recenti sistemi di processamento presenti allo stato dell'arte.

#### Contenuti sintetici

Elementi fondazionali di rappresentazione del linguaggio naturale  
Semantica delle parole  
Large Language Models  
Applicazioni di NLP

#### Programma esteso

##### 1. Fundamentals

- Rationalist and Empiricist Approaches to Language
- The Ambiguity of Language: Why NLP Is Difficult
- Linguistic Essentials
  - Words, Tokens, Lemmas, Stems
  - Parts of Speech and Morphology
  - Phrase Structure
- Dirty Hands-on Text

- 1.4.1 Lexical resources
- 1.4.2 Word counts
- 1.4.3 Zipf's laws
- 1.4.4 Collocations
- 1.4.5 Concordances

## 2. Vector Semantics

- Frequentist Representation of Text (TF, TF-IDF, etc..)
- Word Embeddings
  - Word2Vec
  - FastText
  - Glove
- Visualization of Embeddings:
  - Principal Components Analysis
  - T-distributed stochastic neighbor embedding
  - Uniform Manifold Approximation and Projection

## 3. Transformers and Large Language Models

- Attention Mechanisms: Self and Multi Head Attention
- Positional Embeddings
- Transformers as Language Models
- Pretraining Large Language Models
- Prompting and Instruct Tuning
- Interpretability and Explainability of Language Models

## 4. NLP Applications

- Text and Token Classification
- Chatbots and Dialog Systems
- Word Sense Disambiguation
- Topic Modeling
- Machine Translation

## **Prerequisiti**

Utile, ma non obbligatorio: apprendimento automatico, programmazione in python

## **Modalità didattica**

Lezioni ed esercitazioni in aula.

Il corso verrà erogato in lingua inglese.

## **Materiale didattico**

Cristopher MANNING and Hinrich SCHÜTZE. Foundations of Statistical Natural Language Processing. MIT Press.  
Dan JURAFSKY and James H. MARTIN. Speech and Language Processing. Prentice Hall.

## **Periodo di erogazione dell'insegnamento**

Secondo semestre

## **Modalità di verifica del profitto e valutazione**

Progetto ed Esame Orale. Sono assenti prove in itinere intermedie.

Il progetto consisterà nello sviluppo di uno strumento di natural language processing basato su metodi e modelli presentati a lezione. Il progetto prevede una valutazione espressa in un range 0-24.

L'orale prevede 4 domande di teoria tra gli argomenti del corso elencati nel programma dettagliato. Per ciascuna domanda verrà data una valutazione compresa pari a -2, per una risposta errata o mancata risposta, e +2 punti per una risposta corretta.

## **Orario di ricevimento**

Su appuntamento da concordare via email con il docente.

## **Sustainable Development Goals**

---