UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Natural Language Processing

2425-2-FDS01Q011

Aims

The aim of the course is to provide an introduction to the fundamental concepts related to Natural Language Processing (NLP) as well as an
overview of the main tools used in the field. Moreover, some NLP applications will be presented, e.g. information retrieval, machine translation and hate speech detection.

Contents

The course content includes fundamental principles of Natural Language Processing (NLP) and offers an overview of the key tools utilized in this field. The course will cover a range of topics, ranging from statistical techniques to recent advancements in neural approaches. Moreover, the course incorporates practical demonstrations of different NLP applications, including information retrieval, machine translation, and hate speech detection.

Detailed program

Course introduction

- Rationalist and Empiricist Approaches to Language
- The Ambiguity of Language: Why NLP Is Difficult
  Linguistic Essentials
- Lexical resources
- Zipf's laws
- Collocations
- Concordances

- Syntax
  Frequentist Representation of Text (TF, TF-IDF, etc..) and Word Embeddings
    - Word2Vec
    - FastText
    - Glove
      Visualization of embeddings:
    - Principal Components Analysis
    - T-distributed stochastic neighbor embedding
    - Uniform Manifold Approximation and Projection
      Sequence-to-Sequence (RNN, LSTM)
      Transformers and Large Language Models
    - Attention Mechanisms: Self and Multi Head Attention
      Contextualized Language Models:
    - ELMO
    - BERT
    - GPT
    - LLAMA
      Prompting and Instruct Tuning
      Transformers and Large Language Models
- Interpretability and Explainability of Language Models

Prerequisites

Basic knowledge of statistics and programming languages.

Teaching form

The course will be taught in English, and it will consist of both lectures introducing the main topics and tutorial sessions where open-source tools will be explained.
Seminars held by experts at national and international levels may be part of the course.

Textbook and teaching resource

Daniel Jurafsky and James Martin, "Speech and Language Processing, 2nd Edition", Prentice Hall, 2008.

Emily M. Bender, "Linguistic Fundamentals for Natural Language Processing", Synthesis lectures on human language technologies, Morgan&Claypool Publishers, 2013.

Yoav Goldberg, "Neural Network Methods for Natural Language Processing", Synthesis lectures on human language technologies, Morgan&Claypool Publishers, 2017.

Mohammad Taher Pilehvar and Jose Camacho-collados, "Embeddings in Natural Language Processing", Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2021.

Semester

First Semester

Assessment method

Project
• The project consists in the development of a natural language processing tool based on methods and models presented during the course.
• Each group must identify a domain of interest and dataset for which it intends to address specific NLP tasks.
• The project must be presented orally
• The project is evaluated in the range [0-24]

Oral Exam
• The oral exam can have an outcome between [-8; +8]
• It consists of 4 questions about topics addressed during the course
-2 will be given for an incorrect answer or no answer, +2 for a correct answer.

Office hours

To be agreed with the teacher

Sustainable Development Goals