SYLLABUS DEL CORSO

Natural Language Processing

2324-1-F551MI004

Obiettivi

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 automatic misogyny identification.

Contenuti sintetici

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 automated misogyny detection.

Programma esteso

Introduction to data pre-processing and to some NLP tasks, such as part of speech tagging, and named entity recognition.
Text representation (eg. tf-idf)
Statistical LM (eg. n-gram model)
Dense vector representation (eg. Word2Vec, FastText, etc.)
Dense contextualized word vectors (eg. Neural Language Model)
Sequence2sequence models for NLP (eg. Encoder-Decoder)
Applications of NLP:
Information Retrieval

Automatic Misogyny Identification
Machine Translation

Prerequisiti

Basic knowledge of statistics and programming languages.

Modalità didattica

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.

Materiale didattico

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.

Periodo di erogazione dell'insegnamento

First Semester

Modalità di verifica del profitto e valutazione

Written and optional oral individual examination.

The written examination is aimed at assessing the level of understanding of the basic aspects taught during the course; it is constituted by a set of open questions.

Orario di ricevimento

To be agreed with the teacher

Sustainable Development Goals