UNIVERSITÀ DI PISA

HUMAN LANGUAGE TECHNOLOGIES

GIUSEPPE ATTARDI

Anno accademico	2021/22
CdS	INFORMATICA
Codice	649AA
CFU	9

Moduli	Settore/i	Tipo	Ore	Docente/i
HUMAN LANGUAGE TECHNOLOGIES	INF/01	LEZIONI	72	GIUSEPPE ATTARDI

Obiettivi di apprendimento

Conoscenze
Learning fundamental techniques, algorithms and models used in natural language processing. Understanding of the architectures of typical text analytics applications and of libraries for building them. Expertise in design, implementation and evaluation of applications that exploit analysis, interpretation and transformation of texts.

Modalità di verifica delle conoscenze
Homeworks and final project.

Capacità
Ability to design, implement and evaluate applications that exploit analysis, interpretation and transformation of texts.

Prerequisiti (conoscenze iniziali)

- programming skills, proficiency in the programming language Python
- elementary Calculus and Linear Algebra (e.g. course “Computational Mathematics for learning and data analysis” (646AA))
- elements of probability and statistics (e.g. course “Calcolo delle Probabilità e Statistica” (269AA))
- machine learning (e.g. course “Machine Learning” (654AA))

Programma (contenuti dell'insegnamento)

The course presents principles, models and the state of the art techniques for the analysis of natural language, focusing mainly on statistical machine learning approaches and Deep Learning in particular. Students will learn how to apply these techniques in a wide range of applications using modern programming libraries.Formal and statistical approaches to NLP.

- Statistical methods: Language Model, Hidden Markov Model, Viterbi Algorithm, Generative vs Discriminative Models
- Linguistic essentials: words, lemmas, morphology, PoS, phrases.
- Parsing: constituency and dependency parsing.
- Processing Pipelines: UIMA, Tanl
- Lexical semantics: collocations, corpora, thesauri, gazetteers.
- Distributional Semantics: Word embeddings, Character embeddings.
- Deep Learning for natural language.
- Applications: Entity recognition, Entity linking, Classification, Summarization.
- Opinion mining, Sentiment Analysis.
- Question answering, Language inference, Dialogic interfaces (chatbots)
- Statistical Machine Translation.
- NLP libraries: NLTK, Theano, Tensorflow, Keras

Bibliografia e materiale didattico

UNIVERSITÀ DI PISA

1. C. Manning, H. Schutze. Foundations of Statistical Natural Language Processing. MIT Press, 2000.
2. D. Jurafsky, J.H. Martin, Speech and Language Processing. 2nd edition, Prentice-Hall, 2008.
3. S. Kubler, R. McDonald, J. Nivre. Dependency Parsing. 2010.
4. P. Koehn. Statistical Machine Translation. Cambridge University Press, 2010.
5. S. Bird, E. Klein, E. Loper. Natural Language Processing with Python.

Modalità d'esame

Project.

Ultimo aggiornamento 02/01/2022 10:24