



UNIVERSITÀ DI PISA TEXT ANALYTICS

GIUSEPPE ATTARDI

Anno accademico
CdS

2017/18
INFORMATICA PER L'ECONOMIA E
PER L'AZIENDA (BUSINESS
INFORMATICS)
635AA
6

Codice
CFU

Moduli	Settore/i	Tipo	Ore	Docente/i
TEXT ANALYTICS	INF/01	LEZIONI	48	GIUSEPPE ATTARDI ANDREA ESULI

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

Progetto o seminario.

Capacità

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

Modalità di verifica delle capacità

Progetto o seminario.

Modalità di verifica dei comportamenti

Progetto o seminario.

Prerequisiti (conoscenze iniziali)

Programmazione.

Calcolo delle probabilità e statistica.

Programma (contenuti dell'insegnamento)

1. Disciplinary background: Natural Language Processing, Information Retrieval and Machine Learning
2. Mathematical background: Probability, Statistics and Algebra
3. Linguistic essentials: words, lemmas, morphology, PoS, syntax
4. Basic text processing: regular expression, tokenisation
5. Data gathering: twitter API, scraping
6. Basic modelling: collocations, language models
7. Introduction to Machine Learning: theory and practical tips
8. Libraries and tools: NLTK, Keras
9. Applications
 - Classification/Clustering
 - Sentiment Analysis/Opinion Mining
 - Information Extraction/Relation Extraction
 - Entity Linking



UNIVERSITÀ DI PISA

- Spam Detection: mail spam & phishing, blog spam, review spam

Bibliografia e materiale didattico

1. C. Manning, H. Schütze. [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](#).
6. I. Goodfellow, Y. Bengio, A. Courville. [Deep Learning](#). MIT Press, 2016.
7. M. Nielsen. [Neural Networks and Deep Learning](#).

Modalità d'esame

Progetto.

Pagina web del corso

<http://didawiki.di.unipi.it/doku.php/mds/txa/start>

Ultimo aggiornamento 05/09/2017 10:54