



UNIVERSITÀ DI PISA

TEXT ANALYTICS

ANDREA ESULI

Academic year	2020/21
Course	DATA SCIENCE AND BUSINESS INFORMATICS
Code	635AA
Credits	6

Modules	Area	Type	Hours	Teacher(s)
TEXT ANALYTICS	INF/01	LEZIONI	48	ANDREA ESULI

Obiettivi di apprendimento

Conoscenze

Apprendimento di tecniche, algoritmi e modelli fondamentali utilizzati nell'elaborazione del linguaggio naturale. Comprensione delle architetture delle tipiche applicazioni di analisi del testo e delle librerie per la loro realizzazione. Competenza nella progettazione, implementazione e valutazione di applicazioni che sfruttano l'analisi, l'interpretazione e la trasformazione dei testi.

Modalità di verifica delle conoscenze

Gli studenti applicheranno i concetti e i metodi presentati durante il corso nello sviluppo di un progetto.

Capacità

Capacità di progettare, implementare e valutare applicazioni che sfruttano l'analisi, l'interpretazione e la trasformazione dei testi.

Modalità di verifica delle capacità

Gli studenti applicheranno i concetti e i metodi presentati durante il corso nello sviluppo di un progetto.

Comportamenti

Gli studenti saranno in grado di analizzare un problema di elaborazione del testo, selezionare i metodi corretti per risolverlo e implementare una soluzione funzionante.

Modalità di verifica dei comportamenti

Gli studenti applicheranno i concetti e i metodi presentati durante il corso nello sviluppo di un progetto.

Prerequisiti (conoscenze iniziali)

Prerequisiti utili:

- Coding (python)
- Teoria della probabilità
- Teoria dell'informazione

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 collection: twitter API, scraping
6. Basic modelling: collocations, language models
7. Introduction to Machine Learning: theory and practical tips
8. Libraries and tools: NLTK, Spacy, Keras, pytorch
9. Classification/Clustering



UNIVERSITÀ DI PISA

10. Sentiment Analysis/Opinion Mining
11. Information Extraction/Relation Extraction/Entity Linking
12. Transfer learning
13. Quantification

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. Bird, E. Klein, E. Loper. [Natural Language Processing with Python](#).

Indicazioni per non frequentanti

Gli studenti non frequentanti devono contattare il docente per concordare il progetto d'esame.

Modalità d'esame

L'esame consisterà in un progetto da concordare con l'insegnante e una prova orale. Il risultato del progetto sarà del codice e una relazione sull'attività (4-10 pagine è l'intervallo di lunghezza tipico).

L'esame orale consisterà nella presentazione e discussione del progetto.

Pagina web del corso

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

Ultimo aggiornamento 05/08/2020 17:50