Introduction to the Text Analytics course

Andrea Esuli

Hello!

Teacher of this course: Andrea Esuli

Senior Researcher @ Consiglio Nazionale delle Ricerche

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

Office:

Area della Ricerca di Pisa,

Edificio B, Ingresso 16, Room I-31

Contact:

andrea.esuli@isti.cnr.it   050 621 2895

Text Analytics course

Course code: 635AA - 6 CFU - First semester

Master programme in Data Science and Business Informatics (WDS-LM).

Mutuated by the Master programme in Digital humanities (WFU)

[Course home page](#)

Lessons 2021/22:

9 - 11 Monday   - Aula C   -   [Microsoft Teams](#)

11 - 13 Friday   - Aula M1 -   [Microsoft Teams](#)

Office hours:

by appointment (send email) - [Microsoft Teams](#)/[Skype (aesuli)](#)

Contents

Disciplinary background: Natural Language Processing, Information Retrieval and Machine Learning

- Mathematical background: Probability, Statistics and Algebra
- Linguistic essentials: words, lemmas, morphology, PoS, syntax
- Basic text processing: regular expression, tokenization
- Data gathering: data collection, twitter API, scraping, dataset annotation
- Basic modelling: collocations, language models
- Statistical Machine Learning for text analytics
- Deep/Neural Machine Learning for text analytics
- Libraries and tools: NLTK, Spacy, scikit-learn, Tensorflow/Keras, Pytorch
- Applications: Classification, Clustering, Regression, Language Modeling, Sentiment Analysis, Opinion Mining, Information Extraction…

Tentative calendar

First two weeks: Introduction to course, probability and python.

October: From strings to NLP and text analytics, statistical ML.

November: From Statistical ML to Deep/Neural ML,
Neural Language Models.

December: Advanced applications.

Textbooks

- D. Jurafsky, J.H. Martin, Speech and Language Processing. 3nd edition, Prentice-Hall, 2018.
- S. Bird, E. Klein, E. Loper. Natural Language Processing with Python.

Further readings:

- J. Eisenstein. Introduction to NLP. MIT Press, 2019.
- I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. MIT Press, 2016.
- B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.

Other material

A substantial part of slides are derived from the previous editions of the course held by Professor Giuseppe Attardi.

Green text (and sometimes also images) in slides are **hyperlinks to additional info** that enriches the discussion of the topic (papers, news, websites, tools).

Python notebooks will provide practical examples of the presented topics.

Exam: project

PROJECT: Take a Text Analytics task and implement, test, and discuss, a solution for it.

- Topic: a challenge ([SEMEVAL](), [EVALITA](), [Kaggle]()), a research paper, propose your own.

- **Students MUST contact the teacher at least one month before the date set for the exam session, to agree on the topic of the project.**

- **The date set for the exam session is the deadline for submission.**

- Submission is a paper reporting on the activity (4-10 pages) + code.

- Oral exam is a discussion of your project.

- Groups: **max 3 persons**. I suggest working in pairs, mixing skills.