



PhD Course “Text mining and analytics”

Period and duration

November 2017 - 15 hours

Detailed schedule

DATE	TIME	LOCATION
Tuesday November 14, 2017	From 8.30 am to 12.30 am	SALA C
Thursday November 16, 2017	From 8.30 am to 12.30 am	SALA C
Tuesday November 21, 2017	From 4 pm to 7 pm	LABINF
Thursday November 23, 2017	From 8.30 am to 12.30 am	SALA C

Description

The diffusion of digital libraries and social platforms has produced a huge amount of textual data written in different languages, with different styles, and stored in various formats, structured and not. The analysis of textual data coming from heterogeneous application domains has as common objective the automatic extraction of knowledge useful for analysts and domain experts. Examples of extracted knowledge are (i) summaries of news published by different online newspapers and abstracts of scientific books or regulations, (ii) subsets of keywords or groups of “semantically related” terms occurring in textual content published on social platforms, (iii) opinions (sentiment) of analysts and domain experts. The goal of the course is to overview the main techniques aimed at analyzing textual data as well as to introduce the main open-source instruments nowadays available for text preparation and analysis.

Covered topics

- Introduction to text mining
- Text transformation techniques and representation models (e.g. Principal Component Analysis, Latent Semantic Analysis)
- Text preparation and cleaning
- Entity recognition and disambiguation
- Association analysis of textual data
- Topic detection
- Opinion mining
- Text summarization and validation of the generated summaries
- Overview of the main open-source libraries and software for textual data analyses (e.g. RapidMiner, Lucene, Yago, WordNet)

For more information contact **Luca Cagliero** (name.surname@polito.it)