## Course 9

### Text Mining

The course’ goal is to provide a high-level overview and practical exposure to state-of-the-art techniques in Statistical Natural Language Processing, and will focus on applied rather than theoretic R&D, not the least due to the limited timeframe.

### Program

**Information Retrieval 101**

- a) The Inverted Index and Word Vectors
- b) TF-IDF Weighting
- c) Locality Sensitive Hashing

**Unsupervised Text Processing**

- d) Sentence Boundary Detection
- e) Latent Semantic Indexing
- f) Text Summarization
- g) Latent Dirichlet Allocation [if time permits]

**Representation Learning**

- h) Word Representations
- i) Neural Embeddings
- j) Paragraph Vectors

**Information Extraction**

- k) Part-of-Speech (PoS) Tagging
- l) Collocations & Keyword Extraction
- m) Named Entity Recognition (NER) & Linking

**Natural Language Understanding**

- n) Deep (Dependency) Parsing
- o) Relationship/Event Extraction

### Bibliography

| Prerequisites | Firm knowledge of undergraduate **algebra and statistics** will be assumed, and prior exposure to machine/deep learning might be helpful (e.g., via other courses of the summer school).  

The practical examples will be provided as **Jupyter Notebooks** ([http://jupyter.org/](http://jupyter.org/)). In particular, we will be using **Python** to go through the examples, so an understanding of Python and Jupyter will be required *if* you want to work along or run the examples on your own laptop.  

| Preparations | If you are not familiar with Python:  
Or with Jupyter Notebooks: [https://jupyter.readthedocs.io/](https://jupyter.readthedocs.io/)  

**Installing Python 3 (not 2.7!) and Jupyter:**  
In a nutshell, install Python (3) using the pre-built Anaconda stack:  
[https://www.continuum.io/downloads](https://www.continuum.io/downloads) and then execute `conda install jupyter` in your terminal/shell.  

During the course we will mostly focus on the Python packages **spaCy** [https://spacy.io/](https://spacy.io/) and **gensim** [https://radimrehurek.com/gensim/index.html](https://radimrehurek.com/gensim/index.html).  
Have them installed **before arriving at the Summer School** with `conda install spacy gensim` if you wish to work along (but is not required).