

Course 9	<p>Text Mining</p> <p>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.</p>
Program	<p>Information Retrieval 101</p> <ul style="list-style-type: none"> a) The Inverted Index and Word Vectors b) TF-IDF Weighting c) Locality Sensitive Hashing <p>Unsupervised Text Processing</p> <ul style="list-style-type: none"> d) Sentence Boundary Detection e) Latent Semantic Indexing f) Text Summarization g) Latent Dirichlet Allocation [if time permits] <p>Representation Learning</p> <ul style="list-style-type: none"> h) Word Representations i) Neural Embeddings j) Paragraph Vectors <p>Information Extraction</p> <ul style="list-style-type: none"> k) Part-of-Speech (PoS) Tagging l) Collocations & Keyword Extraction m) Named Entity Recognition (NER) & Linking <p>Natural Language Understanding</p> <ul style="list-style-type: none"> n) Deep (Dependency) Parsing o) Relationship/Event Extraction
Bibliography	<ul style="list-style-type: none"> • Christopher D Manning, and Hinrich Schütze. <i>Foundations of Statistical Natural Language Processing</i>. MIT Press, 1999. • Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. <i>Introduction to Information Retrieval</i>. Cambridge University Press, 2008. http://nlp.stanford.edu/IR-book/ • Steven Bird, Ewan Klein, and Edward Loper. <i>Natural Language Processing with Python</i>. O'Reilly Safari Books, 2009. http://www.nltk.org/book/ • Yoav Goldberg. <i>A Primer on Neural Network Models for Natural Language Processing</i>, 2015. https://arxiv.org/abs/1510.00726

Prerequisites	<p>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).</p> <p>The practical examples will be provided as Jupyter Notebooks (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 <i>if</i> you want to work along or run the examples on your own laptop.</p>
Preparations	<p>If you are not familiar with Python: http://docs.python-guide.org/en/latest/intro/learning/ Or with Jupyter Notebooks: https://jupyter.readthedocs.io/</p> <p>Installing Python 3 (<i>not 2.7!</i>) and Jupyter: http://jupyter.readthedocs.org/en/latest/install.html In a nutshell, install Python (3) using the pre-built Anaconda stack: https://www.continuum.io/downloads and then execute <code>conda install jupyter</code> in your terminal/shell.</p> <p>During the course we will mostly focus on the Python packages spaCy https://spacy.io/ and gensim https://radimrehurek.com/gensim/index.html. Have them installed before arriving at the Summer School with <code>conda install spacy gensim</code> if you wish to work along (but is not required).</p>