

Course Title	Natural Language Processing				
Course Code	AI650				
Course Type	Elective				
Level	Master (2 nd cycle)				
Year / Semester	1 st Year/2 nd Semester or 2 nd Year/1 st Semester				
Teacher's Name	Grigorios Tsoumakas				
ECTS	7	Lectures / week	Up to 6 Teleconferences	Laboratories / week	None
Course Purpose and Objectives	<p>This course synthesizes ideas from linguistics and computer science to provide students with a deep understanding of natural language processing. The course will focus on the most widely-used theoretical and computational models of language, including nonstatistical, statistical and neural network ones. The course will cover methods for learning linguistic representations at all levels of analysis: words, syntax and semantics. It will discuss a range of basic NLP tasks such as language modeling, part-of-speech tagging, probabilistic parsing, dependency parsing, text categorization, as well as more advanced tasks, such as named entity recognition, relation extraction, question answering and natural language conversation.</p>				
Learning Outcomes	<p>Upon successful completion of this course students should be able to:</p> <ul style="list-style-type: none"> • Identify, construct, and analyze examples of different kinds of ambiguity in natural language (e.g., in part-of-speech, word sense, syntactic attachment), explain how ambiguity presents a problem for computational analysis, and some of the ways it can be addressed. • Describe and apply standard sequence and classification models; describe parsing and search algorithms for different levels of analysis (e.g. morphology, syntax, and semantics). • For a range of NLP tasks, outline a processing pipeline for that task, including standard data sets, models, algorithms, and evaluation methods. • Identify and discuss the main linguistic, machine learning, and ethical challenges involved in the development and use of natural language processing systems • Provide a well-justified selection of informative features given a particular NLP task and a description of the available training resources • Discuss the computational and engineering challenges that arise in the use of different models for natural language processing, and discuss the pros and cons of different models for a given task 				

	<ul style="list-style-type: none"> • Design NLP systems for information extraction, question answering, and conversational systems. 		
Prerequisites	None	Co-requisites	AI600
Course Content	<p>1) <u>Text Processing</u>: text normalization (sentence segmentation, word segmentation, word normalization), regular expressions, minimum edit distance</p> <p>2) <u>N-gram Language Models</u>: language modeling, n-grams, evaluating language models, handling unknown words, smoothing.</p> <p>3) <u>Part-of-Speech Tagging</u>: part-of-speech (POS) word classes, hidden Markov models, maximum entropy Markov models.</p> <p>4, 5) <u>Grammars</u>: (probabilistic) context-free grammars, (probabilistic) Cocke-Kasami-Younger (CKY) parsing, dependency grammars, dependency parsing</p> <p>6, 7) <u>Vector Semantics</u>: lexical semantics, word sense disambiguation, vector semantics, term-document matrix, term-term matrix, tf-idf, pointwise mutual information, (probabilistic) latent semantic indexing, latent dirichlet allocation, Word2Vec, bias in vector representations</p> <p>8) <u>Text categorization</u>: multinomial naive Bayes, sentiment analysis</p> <p>9, 10) <u>Neural Models</u>: Neural language models, neural text categorization, recurrent neural networks (RNNs), long short-term memory networks, gated recurrent units, stacked and bidirectional RNNs, convolutional neural networks, character embeddings, sequence to sequence models.</p> <p>11) <u>Information Extraction</u>: named entity recognition, relation extraction</p> <p>12) <u>Question Answering</u>: information retrieval based approaches, knowledge-based approaches, hybrid approaches</p> <p>13) <u>Conversational Agents</u>: task-based dialog systems, chatbots</p>		
Teaching Methodology	E-Learning		
Bibliography	<p>Daniel Jurafsky and James H. Martin (2019). Speech and Language Processing, Latest Edition</p> <p>Christopher D. Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing. MIT Press, Latest Edition.</p> <p>Manning, C. D., Raghavan, P., & Schutze, H. Introduction to Information Retrieval, Latest Edition.</p>		

	Yoav Goldberg, Neural Network Methods in Natural Language Processing, Morgan & Claypool Publishers, Latest Edition	
Assessment	Final Examination	50%
	Assignments/On-going evaluation	50%
		100%
Language	English	