## **COURSE DETAIL**

## NATURAL LANGUAGE PROCESSING

**Country** United Kingdom - England

Host Institution Imperial College London

**Program(s)** Imperial College London

**UCEAP Course Level** Upper Division

UCEAP Subject Area(s) Computer Science

**UCEAP Course Number** 159

**UCEAP Course Suffix** 

UCEAP Official Title NATURAL LANGUAGE PROCESSING

**UCEAP Transcript Title** NATURAL LANG PROC

UCEAP Quarter Units 5.00

UCEAP Semester Units

3.30

## **Course Description**

This class will provide students the techniques and tools to devise and develop Natural Language Processing (NLP) components and applications. The course will cover the foundations, building blocks and applications of NLP, with an emphasis on the necessary linguistic intuitions as well as a broad coverage of statistical and deep learning models that can be used for language tasks. NLP is an important topic in Artificial Intelligence with a wide range of applications, from sentiment analysis to machine translation. Modern NLP is primarily based on statistical methods and machine learning algorithms, where linguistic information is provided by instances of uses of language. For most NLP tasks, state of the art approaches are based on neural models, which will be at the core of this module. However, significant attention will be given to the linguistic principles that underpin the field. More specifically, students will: Gain familiarity with important linguistic concepts involved in language understanding and generation, from morphological analysis to pragmatics Gain familiarity with, devise, implement and apply relevant pre-processing steps for natural language processing components and applications Critically compare statistical and deep learning approaches for natural language processing Map various well established techniques in machine learning to specific problems in natural language processing Build, evaluate, critically analyze and improve models using ex-isting machine learning algorithms and frameworks (such as TensorFlow) for a range of natural language processing tasks, including: classification, structured prediction, sequence to se-quence labeling and generation Devise, implement and evaluate classifiers for a range of natu-ral language processing tasks. Learning outcomes After the course, students should be able to: (ILO1) Identify and automatically pre-process texts that can be useful for language processing tasks (ILO2) Devise and evaluate solutions for a range of natural lan-guage components using existing algorithms, techniques and frameworks, including part-of-speech tagging, language mod-eling, parsing and semantic role labeling (ILO3) Devise, implement and evaluate algorithms for single and multi-class classification problems (ILO4) Apply existing statistical and deep learning techniques to language applications such as machine translation. ILO1 and IL3 will be assessed mainly through the coursework, while ILO2 and ILO4 will be assessed mainly via exam. Module syllabus 1 Introduction to NLP (language

challenges, applications, clas-sical vs statistical vs deep learning-based) 2 Basic concepts in Linguistics (including morphology, syntax, semantics, pragmatics) 3 Pre-processing techniques, word meaning (TF-IDF, distributional models, word2vec, glove, etc) 4 Lab/tutorial session on pre-processing and word meaning 5-6 Classification tasks with simple classification models (Naïve Bayes, perceptron): SPAM detection, part-of-speech tagging, word sense disambiguation 7 Classification tasks with CNN models 8 Lab/tutorial session on classification 9 Coursework specification and discussion 10 Ngram language models 11 Neural language models (RNNs, LSTMs, GRUs) 12 Lab/tutorial session on language models 13 Structured prediction - POS tagging with HMM 14 Structured prediction - POS tagging with neural models (RNN) 15 Syntax and parsing 16 Lab/tutorial session on POS tagging 17 Rules-based and probabilistic parsing 18 Neural models for parsing 19 Semantic role labeling 20 Lab/tutorial session on parsing 21-23 Sequence to sequence modelling - machine translation (SMT, NMT, attention) 24 Lab/tutorial session on sequence to sequence modelling 25 Guest lecture on advanced NLP topics 26 Guest lecture on advanced NLP topics 27 Revision lecture 28 Revision lecture

Language(s) of Instruction English

Host Institution Course Number CO490

Host Institution Course Title NATURAL LANGUAGE PROCESSING

Host Institution Campus Imperial College

**Host Institution Faculty** 

Host Institution Degree

Host Institution Department Computing

Print