Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

Master in Applied Linguistics (LM-39)

Course title	Computational Linguistics
Course code	54113
Scientific sector	L-LIN/01
Degree	Master in Applied Linguistics
Semester	1st
Course year	2nd
Credits	6 + 2
Modular	*Yes*

Total lecturing hours	30
Total lab hours	30
Attendance	according to the regulation
Prerequisites	

Specific educational objectives	The objective of the course is to provide the students with a practical understanding of a) some aspects of speech and text data computational processing b) the computational tools and techniques available to achieve so. The computational linguistics module will focus on how to build a pipeline, how to process significant quantities of data, how to train language models and how to test them. The speech technology module will focus in the first part on the implementation of typical speech processing techniques, in particular for feature extraction and speech segmentation. In the second part, the course will address more articulated tasks as speaker identification and keyword spotting as a whole.

Module 1	Computational linguistics
Lecturer	Luca Ducceschi
Scientific sector	L-LIN/01
Teaching language	English
Office hours	from Monday to Friday on request
List of topics covered	
Teaching format	*Laboratorial activities and personal projects*
Total lecturing hours	*30*
Credits	*6*

Module 2	Speech Technologies - LAB
Lecturer	Alessio Brutti
Scientific sector	L-LIN/01
Teaching language	English
Office hours	from Monday to Friday on request
List of topics covered	
Teaching format	*Frontal lectures, laboratorial activities and personal projects*
Total lecturing hours	*30*
Credits	*2*

Learning outcomes	Computational linguistics The course's goal is to teach the students how to understand and address non-trivial tasks relative to natural language

	processing, such as a) accessing and managing gigabytes of textual data (e.g. the whole English Wikipedia), b) processing strings of characters using the basic Python modules and regular expressions, c) train some language models to solve problems like part-of-speech tagging, or text classification, evaluate such models using statistical tools and gold standard datasets. In order to stimulate their independent thinking and problem-solving skills students will be required to experiment with different solutions to a given problem and to evaluate the different options (e.g. rule based approaches vs trained models). Given the nature of the exam (an oral presentation of a computational project), some of the lab sessions will be dedicated to presenting and discussing the project ideas. The goal of the speech technology lab is to introduce the students to digital processing of non-stationary signals and to some common speech processing tasks as: speaker identification, keyword spotting and language recognition. Students will earn hands-on experience on: 1) the basic characteristics of speech signals; 2) how to process speech signals in python, extracting and evaluating the most relevant features (e.g. pitch, spectrograms, MFCC) 3) how to tackle from scratch some simple speech processing tasks (speaker identification and keyword detection) using neural models in Keras. The course will also provide an overview of speech recognition systems, focusing in particular on current state-of-the-art end-to-end frameworks

Assessment	*Oral Exam: students are required to discuss the details of a project that encompasses and develops the subjects covered during the course. The project will be defined and agreed upon before the exam. The oral exam will include 1 or 2 questions on the topics addressed during the course and not covered by the project.* *Each module will have a separate project/exam.*
Assessment language	*English*
Evaluation criteria and criteria for awarding marks	*The outcome of the exam will be determined by: a) the level of understanding of the topics covered during the course, b) the computational skills of the candidate, c) the clarity of the exposition and of the project.* *The final mark will be the weighted average of the marks obtained in each module.*

Required readings	Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit, Steven Bird, Ewan Klein, and Edward Loper, available here: https://www.nltk.org/book/
Supplementary readings	Speech Communications: human and machine 2nd edition, O'Shaughnessy Introduction to Digital Speech Processing by Rabiner and Schafer, 2007