COURSE PRESENTATION FORM – ACADEMIC YEAR 2010/2011

COURSE NAME	**Text Processing**
COURSE CODE	72100 (MSc 270) / 70178 (BSc, MSc 509)
LECTURER	Bernardo Magnini
TEACHING ASSISTANTS	Yashar Mehdad
TEACHING LANGUAGE	English
CREDIT POINTS	4
LECTURE HOURS	24
EXERCISE HOURS	12
TIME SPAN	21.02.2011 - 11.06.2011
TIME TABLE	See Timetable Page
OFFICE HOURS LECTURER	During the lecture time, Fridays by previous appointment via e-mail, Faculty of CS, POS Building, piazza Domenicani 3, office 2.10
OFFICE HOURS TEACHING ASSISTANT	During the lecture time, Thursdays by previous appointment via e-mail, Faculty of CS, POS Building, piazza Domenicani 3, office 2.10
PREREQUISITES	Basics in Computational Linguistics, Formal Languages, Statistics and Machine Learning.
OBJECTIVES	Understanding the content expressed by written texts is one the more challenging topic in Artificial Intelligence as well as a crucial area of technological development in Information Access. The course will provide basic notions in Text Processing according to both data-driven and knowledge-based methodologies and technologies. Students will be introduced to text processing technologies, from morpho-syntactic analysis to content extraction. Implemented tools and application scenarios will serve as exemplification of concrete use of fundamental techniques.
SYLLABUS	The course will review basic methods and technological achievements in text processing and content extraction from texts. State of art approaches in Part of Speech Tagging, Shallow Parsing, Terminology Recognition, Named Entities Recognition, Word Sense Disambiguation and Textual Entailment will be addressed in depth. • Introduction to Text Processing • Corpora and markup languages • Part of Speech Tagging

- Information Retrieval and Question Answering
- Terminology Recognition
- Named Entities Recognition
- Word Sense Disambiguation
- Textual Entailment
- Text Categorization and Clustering

TEACHING FORMAT

Frontal lectures, labs.

ASSESSMENT

The exam consists of:

- a project [50 % of mark]
- a final written exam [50 % of mark]

Both parts have to be passed to pass the exam, but they can be taken independently of each other. In case of a positive mark, the part that has been passed will count for all 3 regular exam sessions of the Academic Year (i.e., if the student fails or does not take, e.g., the oral exam, he keeps the project and only needs to retake the oral exam).

READING LIST

Textbook:

- D. Jurafsky and J. Martin: Speech and Language Processing, Prentice Hall, 2000.

Additional readings:

- C. Manning and H. Schutze: Foundations of Statistical Natural language Processing, MIT Press, 1999.
- R. Mitkov (ed.): The Oxford Handbook of Computational Linguistics, Oxford University Press, 2003.
- Manning, Raghavan, Schutze: An Introduction to Information Retrieval, CUP (2008), http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html
- Fellbaum: WordNet, MIT Press, 1998.
- E. Agirre and P. Edmonds: Word sense disambiguation, Springer

SOFTWARE USED

- MultiWordNet
- WEKA
- EDITS

LEARNING OUTCOME

In depth knowledge about the main steps in text processing and content extraction.
Ability to manage simple tools and execute text processing tasks.

COURSE PAGE

http://web.me.com/bmagnini/Bolzano/Text_Processing_2011.html