**COURSE : TEXT MINING AND DOCUMENT ANALYSIS**

INSTRUCTOR : Fabio Massimo Zanzotto, Danilo Croce

CONTACTS:

email : Fabio.massimo.zanzotto@uniroma2.it

**COURSE BACKGROUND**

**LEARNING OBJECTIVES**

The course introduces to the common practices and the common models of natural language processing. As a result of the learning, students will be able to implement a model for natural language processing.

**METHODOLOGY**

**EXAM**

Oral exam:

**CONTENTS**

The language: linguistic models and theories

Linguistic models and systems.

- Morpholgy: Finite state automaton and transducers

- Syntactic analysis with context-free grammars

- - Parsing with context-free grammars

- - Feature Structures and Unification

- - Tree Adjoining Grammars

- - Modular and Lexicalized Parsing

- - Probabilistic context-free grammar

- Semantics

- - Symbolic Semantic Representation: WordNet and FrameNet

- - Lambda Calcolus for natural languagte semantics

- - Distributional semantics

- Textual Entailment Recognition

**TEACHING MATERIAL**

**SUGGESTED READING**

- Daniel Jurafsky and James H. Martin, SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Second Edition)

- I.Dagan, D.Roth, M.Sammons, F.M.Zanzotto, Recognizing Textual Entailment: Models and Applications, Synthesis Lectures on Human Language Technologies #23, Morgan&Claypool Publishers, 2013

**ADDITIONAL SUGGESTED TEXTBOOKS**