COURSE: <u>Text Mining and Document Analysis</u>

INSTRUCTORS: Fabio Massimo Zanzotto (Ph.D. 2002), Danilo Croce (Ph.D. 2012)
EMAIL: fabio.massimo.zanzotto@uniroma2.it
croce@info.uniroma2.it
WEB PAGE:
http://directory.uniroma2.it/index.php/chart/dettagliDocente/5312
http://directory.uniroma2.it/index.php/chart/dettagliDocente/8152

COURSE DESCRIPTION

The course provides an introduction to natural language processing and to its applications to text mining and document analysis. Empirical applications will be illustrated using updated software tools.

LEARNING OUTCOMES

- ✓ Understand the different levels of a stack for natural language processing
- ✓ Understand the theories behind the natural language processing models
- ✓ Gain practical experience in using natural language processing tools.

METHODOLOGY

Theoretical lessons and practice using CoreNLP (in Java) and NLTK (in Python).

ASSESSMENT

Written exam; weighting: 80%
Project; weighting: 20%

OUTLINE

- ➢ 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
    - ➢ 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

TEXTBOOKS

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

ADDITIONAL SUGGESTED READING

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