uc3m | Universidad **Carlos III** de Madrid

Natural Language Processing

Academic Year: (2020 / 2021)   **Review date: 06/07/2020 22:08:28**

Department assigned to the subject: Signal and Communications Theory Department

Coordinating teacher: MARTÍNEZ OLMOS, PABLO

Type: Electives ECTS Credits : 3.0

Year :   Semester :

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

The students are expected to have basic knowledge of
- Calculus
- Programming skills
- Statistics

OBJECTIVES

- Know the basic techniques of text pre-processing.
- Use software tools for pre-processing text.
- Know the techniques of topic modeling.
- Use topic modeling software tools in corpus of documents.
- Use topic models for information retrieval in corpus of documents.
- Learn how to train models of semantic representation in a vector space.
- Learn to train language models using recursive neural networks.
- Know basic translation structures based on recursive neural networks.
- Use optimization tools to build language models with recursive neural networks.

DESCRIPTION OF CONTENTS: PROGRAMME

- Document preprocessing techniques
- Topic Modeling
- Recurrent Neural Networks
- Language Models with RNNs
- Sequence to Sequence for Machine Translation
- Attention Models

LEARNING ACTIVITIES AND METHODOLOGY

All sessions will be theoretical / practical, in which each session introduces a theoretical aspect and is developed using specific software libraries. It is important to highlight that these classes will require initiative and personal and group work on the part of the student (there will be concepts that they will have to study personally based on some indications, particular cases they will have to develop, etc.) These practices, on the one hand, allow the student apply the theoretical knowledge acquired to try practical solutions, so that they can consolidate and critically analyze such knowledge.

ECTS credits include in all cases the corresponding part of personal or team work by the student.

ASSESSMENT SYSTEM

% end-of-term-examination/test: 0

% of continuous assessment (assigments, laboratory, practicals…):   100

La evaluación continua (100%) se basará en proyectos de programación para implementar las herramientas vistas en clase. La evaluación continua también incluirá un test de prácticas que se realizará al final del cuatrimestre.

BASIC BIBLIOGRAPHY

- Cristopher Bishop Pattern Recognition and Machine Learning, Springer, 2006

- Ian Goodfellow and Yoshua Bengio and Aaron Courville Deep Learning, MIT Press, 2017

- Steven Bird, Ewan Klein, Edward Loper Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, O'Reilly, 2009