Last update: 06-05-2019

230717 - AHLT - Advanced Human Language Technologies

Coordinating unit: 230 - ETSETB - Barcelona School of Telecommunications Engineering

Teaching unit: 739 - TSC - Department of Signal Theory and Communications

Academic year: 2019

Degree: MASTER'S DEGREE IN TELECOMMUNICATIONS ENGINEERING (Syllabus 2013). (Teaching unit Optional)
MASTER'S DEGREE IN ADVANCED TELECOMMUNICATION TECHNOLOGIES (Syllabus 2019). (Teaching unit Optional)

ECTS credits: 5 Teaching languages: English

Teaching staff

Coordinator: Marta Ruiz Costa-jussà

Opening hours

Timetable: During the course the hour after the class

Prior skills

Introductory concepts and methods of Natural Language Processing.
Introductory concepts and methods of Machine Learning.
Programming.

Requirements

Student will benefit from experience in programming in Python

Degree competences to which the subject contributes

Transversal:

CT3. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.

CT4. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.

CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

Teaching methodology

-Lectures
-Programming/Laboratory
-Problems

Learning objectives of the subject

Can a machine learn to correct the grammaticality of text? Can a machine learn to answer questions we make in plain English? Can a machine learn to translate languages, using Wikipedia as a training set?

Last update: 06-05-2019

230717 - AHLT - Advanced Human Language Technologies

This course offers an in depth coverage of methods for Natural Language Processing. We will present fundamental models and tools to approach a variety of Natural Language Processing tasks, ranging from syntactic processing, to semantic processing, to final applications such as information extraction, human-machine dialogue systems, and machine translation. The flow of the course is along two main axis: (1) computational formalisms to describe natural language processes, and (2) statistical and machine learning methods to acquire linguistic models from large data collections.

1. Learn to apply statistical methods for NLP in a practical application

2. Understand statistical and machine learning techniques applied to NLP

3. Develop the ability to solve technical problems related to statistical and algorithmic problems in NLP

4. Understand fundamental methods of Natural Language Processing from a computational perspective

Study load

Total learning time: 125h	Hours large group:	26h	20.80%
	Hours small group:	13h	10.40%
	Self study:	86h	68.80%

Last update: 06-05-2019

230717 - AHLT - Advanced Human Language Technologies

Content

Advanced Human Language Technology

Learning time: 2h 05m

Theory classes: 0h 26m
Laboratory classes: 0h 13m
Self study : 1h 26m

Description:

1. Course Introduction (1h Theory)
2. Syntactic Parsing (5h Theory)

Three lectures of the course will be devoted to syntactic parsing:

1.- Statistical parsing. The core are SCFG. Learning (supervised from treebanks or unsupervised using the inside/outside algorithm), parsing (Viterbi). Pros & Cons of SCFG. Other probabilistic approaches.

2.- Dependency Parsing. Projective and non projective dependency trees. Eisner & Chu, Liu, Edmonds algorithms. Transition-based parsing.

3.- Robust parsing. Chunking. HMM-based chunkers. Cascaded FSM chunkers, grammars for chunking.

3. Distances and Similarities (3h Theory)

Distances (and similarities) between linguistic units. Textual, Semantic, and Distributional distances. Semantic spaces (WN, Wikipedia, Freebase, Dbpedia).

4. Semantic Role Labelling (2h Theory)
5. Semantic Parsing (6h Theory)

Semantic Representation. Semantic parsing. Building semantic grammars. Learning semantic parsers.

6. Distributional models (2h Theory)

Distributional models of semantics. Vector Space Model (VSM). Dimensionality reduction. Latent Semantic Indexing (LSI). Using Topic models: Latent Dirichlet Allocation (LDA).

7. Linguistic Inference (2h Theory)

Detecting inference between linguistic units. Recognizing Textual Entailment. The case of paraphrasing.

8. Deep Learning for NLP (6h Theory)

Three lectures will be devoted to Deep Learning for NLP

1.-Linear models. Feed Forward NN. Simple Perceptron. Multilayer Perceptron (MLP).

2.- Neural language modeling and Word embeddings. Use of words embeddings

3.- More advanced NN. Convolutional NN, Embeddings more complex units. Recurrent NN (RNN): GRU, LSTM. NLP applications. Libraries and languages for NN: PyTorch

9. Laboratory assignment (13h)

230717 - AHLT - Advanced Human Language Technologies

Qualification system

Final grade = 0.5*FE + 0.5*LP

where

FE is the grade of the final exam

LP is the grade of the lab project

Bibliography

Basic:

Goldberg, Y.; Hirst, G. Neural network methods in natural language processing. Morgan & Clypool Publishers, 2017. ISBN 9781627052986.

Jurafsky, D.; Martin, J.H. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. 2nd ed. Upper Saddle River: Prentice Hall, 2008. ISBN 9789332518414.

Manning, C.D.; Schütze, H. Foundations of statistical natural language processing. Cambridge, Mass.: MIT Press, 1999. ISBN 0262133601.