# 270709 - AHLT - Advanced Human Language Technologies

Coordinating unit: 270 - FIB - Barcelona School of Informatics  
Teaching unit: 723 - CS - Department of Computer Science  
Academic year: 2019  
Degree:  
- MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)  
- MASTER'S DEGREE IN ADVANCED TELECOMMUNICATION TECHNOLOGIES (Syllabus 2019).  
  (Teaching unit Optional)  
- MASTER'S DEGREE IN TELECOMMUNICATIONS ENGINEERING (Syllabus 2013). (Teaching unit Optional)  
- MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2017). (Teaching unit Optional)  
ECTS credits: 5

## Degree competences to which the subject contributes

### Basic:
- CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.  
- CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.  
- CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

### Specific:
- CEA3. Capability to understand the basic operation principles of Machine Learning main techniques, and to know how to use on the environment of an intelligent system or service.  
- CEA5. Capability to understand the basic operation principles of Natural Language Processing main techniques, and to know how to use in the environment of an intelligent system or service.

### Generical:
- CG3. Capacity for modeling, calculation, simulation, development and implementation in technology and company engineering centers, particularly in research, development and innovation in all areas related to Artificial Intelligence.

### 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.  
- CT6. REASONING: Capability to evaluate and analyze on a reasoned and critical way about situations, projects, proposals, reports and scientific-technical surveys. Capability to argue the reasons that explain or justify such situations, proposals, etc.  
- CT7. ANALYSIS Y SYNTHESIS: Capability to analyze and solve complex technical problems.
270709 - AHLT - Advanced Human Language Technologies

**Teaching methodology**

The course will be structured around five main blocks of lectures. In each theory lecture, we will present fundamental algorithmic and statistical techniques for NLP. This will be followed by problem lectures, where we will look in detail to derivations of algorithms and mathematical proofs that are necessary in order to understand statistical methods in NLP.

Furthermore, there will be four problem sets that students need to solve at home. Each problem set will consist of three or four problems that will require the student to understand the elements behind statistical NLP methods. In some cases these problems will involve writing small programs to analyze data and perform some computation.

Finally, students will develop a practical project in teams of two or three students. The goal of the project is to put into practice the methods learned in class, and learn how the experimental methodology that is used in the NLP field. Students have to identify existing components (i.e. data and tools) that can be used to build a system, and perform experiments in order to perform empirical analysis of some statistical NLP method.

**Learning objectives of the subject**

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**

<table>
<thead>
<tr>
<th><strong>Total learning time:</strong> 60h</th>
<th>Hours large group:</th>
<th>Hours medium group:</th>
<th>Hours small group:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30h</td>
<td>15h</td>
<td>15h</td>
</tr>
<tr>
<td></td>
<td>50.00%</td>
<td>25.00%</td>
<td>25.00%</td>
</tr>
</tbody>
</table>
# Content

## Statistical Models for NLP

**Degree competences to which the content contributes:**

**Description:**

## Distances and Similarities

**Degree competences to which the content contributes:**

**Description:**

## Sequence Prediction

**Degree competences to which the content contributes:**

**Description:**
Prediction in word sequences: PoS tagging, NERC. Local classifiers, HMM, global predictors, Log-linear models.

## Syntactic Parsing

**Degree competences to which the content contributes:**

**Description:**
Parsing constituent trees: PCFG, CKY vs Inside/outside
Parsing dependency trees: CRFs for parsing. Earley algorithm

## Document-level modelling

**Degree competences to which the content contributes:**

**Description:**
Document representation: from BoW to NLU.
Document similarities.
Document classification.

## Deep Learning approaches - Introduction

**Degree competences to which the content contributes:**
# Deep Learning approaches - Word Sequences

**Degree competences to which the content contributes:**

**Description:**
- PoS tagging, NERC

# Deep Learning Approaches - Sentences

**Degree competences to which the content contributes:**

**Description:**
- Sentence similarity, sentence classification. LSTM, BERT. Sentence embeddings

# Deep Learning approaches - Document Level

**Degree competences to which the content contributes:**

**Description:**
- Document similarity, document classification, document embeddings - doc2vec

# Deep Learning Approaches - Machine Translation

**Degree competences to which the content contributes:**

**Description:**
- Neural Machine Translation
### Planning of activities

**Course Introduction**

- **Hours:** 3h  
  - Theory classes: 2h  
  - Practical classes: 1h  
  - Laboratory classes: 0h  
  - Guided activities: 0h  
  - Self study: 0h

**Description:**

**Specific objectives:** 2, 4

---

**Distances and Similarities**

- **Hours:** 8h  
  - Theory classes: 5h  
  - Practical classes: 3h  
  - Laboratory classes: 0h  
  - Guided activities: 0h  
  - Self study: 0h

**Description:**

**Specific objectives:** 2, 4

---

**Sequence Models in NLP**

- **Hours:** 10h  
  - Theory classes: 6h  
  - Practical classes: 4h  
  - Laboratory classes: 0h  
  - Guided activities: 0h  
  - Self study: 0h

**Description:**
These lectures will present sequence models, an important set of tools that is used for sequential tasks. We will present this in the framework of structured prediction (later in the course we will see that the same framework is used for parsing and translation). We will focus on machine learning aspects, as well as algorithmic aspects. We will give special emphasis to Conditional Random Fields. Also Deep Learning models will be presented

**Specific objectives:** 2, 4
### Syntax and Parsing

**Hours:** 9h  
Theory classes: 6h  
Practical classes: 3h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 0h

**Description:**  
We will present statistical models for syntactic structure, and in general tree structures. The focus will be on probabilistic context-free grammars and dependency grammars, two standard formalisms. We will see relevant algorithms, as well as methods to learn grammars from data based on the structured prediction framework. Sentence similarity, sentence classification. LSTM. BERT. Sentence embeddings

**Specific objectives:**  
2, 4

### Document-level modelling

**Hours:** 6h  
Theory classes: 4h  
Practical classes: 2h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 0h

**Description:**  
Document representation: from BoW to NLU.  
Document similarities.  
Document classification  
document embeddings - doc2vec

**Specific objectives:**  
2, 4

### Neural Machine Translation

**Hours:** 6h  
Theory classes: 4h  
Practical classes: 2h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 0h

**Description:**  
Neural Machine Translation

**Specific objectives:**  
2, 4
## Final Exam

<table>
<thead>
<tr>
<th>Specific objectives:</th>
<th>Hours: 13h 30m</th>
</tr>
</thead>
<tbody>
<tr>
<td>2, 3, 4</td>
<td>Guided activities: 3h</td>
</tr>
<tr>
<td></td>
<td>Self study: 10h 30m</td>
</tr>
</tbody>
</table>

## Project

<table>
<thead>
<tr>
<th>Specific objectives:</th>
<th>Hours: 45h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 4</td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td>Self study: 45h</td>
</tr>
</tbody>
</table>

## 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:


Manning, C.; See, A. Natural language processing with deep learning. Stanford University.


Others resources:

Hyperlink

http://www.lsi.upc.edu/~ageno/anlp