270654 - ANLP - Advanced Natural Language Processing

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science
Academic year: 2018
Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: Catalan

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:
CEC1. Ability to apply scientific methodologies in the study and analysis of phenomena and systems in any field of Information Technology as well as in the conception, design and implementation of innovative and original computing solutions.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.

General:
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

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.
270654 - ANLP - Advanced Natural Language Processing

**Learning objectives of the subject**

1. Understand fundamental methods of Natural Language Processing from a computational perspective
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. Learn to apply statistical methods for NLP in a practical application

**Study load**

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Theory classes: 24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 6h</td>
<td>4.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
### Course Introduction

**Degree competences to which the content contributes:**  
**Description:**  
Fundamental tasks in NLP. Main challenges in NLP. Review of statistical paradigms. Review of language modeling techniques.

### Classification in NLP

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

### Sequence Models

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

### Syntax and Parsing

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

### Machine Translation

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

### Unsupervised and Semisupervised methods in NLP

**Degree competences to which the content contributes:**
270654 - ANLP - Advanced Natural Language Processing

**Description:**

**Qualification system**
Final grade = 0.6 final exam + 0.4 project

where
final exam is the grade of the final exam
project is the grade of the project

**Bibliography**

**Basic:**
Collins, Michael. Lecture Notes for Coursera Course "Natural Language Processing". 2013.