Course guides
270709 - AHLT - Advanced Human Language Technologies

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.

Degree:
MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2017). (Optional subject).
MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).
MASTER'S DEGREE IN ADVANCED TELECOMMUNICATION TECHNOLOGIES (Syllabus 2019). (Optional subject).
MASTER'S DEGREE IN TELECOMMUNICATIONS ENGINEERING (Syllabus 2013). (Optional subject).

Academic year: 2019  ECTS Credits: 5.0  Languages:

LECTURER

Coordinating lecturer:

Others:

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

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.

General:
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.

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.
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>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours medium group</td>
<td>15,0</td>
<td>25.00</td>
</tr>
<tr>
<td>Hours large group</td>
<td>30,0</td>
<td>50.00</td>
</tr>
<tr>
<td>Hours small group</td>
<td>15,0</td>
<td>25.00</td>
</tr>
</tbody>
</table>

Total learning time: 60 h

CONTENTS

Statistical Models for NLP

Description:

Distances and Similarities

Description:

Sequence Prediction

Description:
Prediction in word sequences: PoS tagging, NERC. Local classifiers, HMM, global predictors, Log-linear models.
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syntactic Parsing</strong></td>
<td>Parsing constituent trees: PCFG, CKY vs Inside/outside</td>
</tr>
<tr>
<td></td>
<td>Parsing dependency trees: CRFs for parsing, Earley algorithm</td>
</tr>
<tr>
<td><strong>Document-level modelling</strong></td>
<td>Document representation: from BoW to NLU.</td>
</tr>
<tr>
<td></td>
<td>Document similarities.</td>
</tr>
<tr>
<td></td>
<td>Document classification.</td>
</tr>
<tr>
<td><strong>Deep Learning approaches - Introduction</strong></td>
<td>Introduction to ANN for NLP</td>
</tr>
<tr>
<td></td>
<td>Lexical semantics. Word Embeddings</td>
</tr>
<tr>
<td><strong>Deep Learning approaches - Word Sequences</strong></td>
<td>PoS tagging, NERC</td>
</tr>
<tr>
<td><strong>Deep Learning Approaches - Sentences</strong></td>
<td>Sentence similarity, sentence classification. LSTM. BERT. Sentence embeddings</td>
</tr>
<tr>
<td><strong>Deep Learning approaches - Document Level</strong></td>
<td>Document similarity, document classification, document embeddings - doc2vec</td>
</tr>
<tr>
<td><strong>Deep Learning Approaches - Machine Translation</strong></td>
<td>Neural Machine Translation</td>
</tr>
</tbody>
</table>
ACTIVITIES

Course Introduction

Description:

Specific objectives:
2, 4

Related competencies:
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.
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.
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.
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.
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..
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.

Full-or-part-time: 3 h
Theory classes: 2h
Practical classes: 1h

Distances and Similarities

Description:

Specific objectives:
2, 4

Related competencies:
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.
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.
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.
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.
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..
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.

Full-or-part-time: 8 h
Theory classes: 5h
Practical classes: 3h
Sequence Models in NLP

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

Related competencies:
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.
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.
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.
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.
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..
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.

Full-or-part-time: 10 h
Theory classes: 6h
Practical classes: 4h

Syntax and Parsing

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

Related competencies:
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.
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.
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.
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.
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..
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.

Full-or-part-time: 9 h
Theory classes: 6h
Practical classes: 3h
### Document-level modelling

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

**Specific objectives:**
2, 4

**Related competencies:**
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.  
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.  
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.  
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.  
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..  
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.

**Full-or-part-time:** 6 h  
Theory classes: 4h  
Practical classes: 2h

### Neural Machine Translation

**Description:**
Neural Machine Translation  

**Specific objectives:**  
2, 4

**Related competencies:**
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.  
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.  
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.  
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.  
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..  
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.

**Full-or-part-time:** 6 h  
Theory classes: 4h  
Practical classes: 2h
Final Exam

Specific objectives:
2, 3, 4

Related competencies:
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.
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.
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.
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.
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..
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.

Full-or-part-time: 13 h
Guided activities: 3h
Self study: 10h 30m

Project

Specific objectives:
1, 2, 4

Related competencies:
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.
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.
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.
CT7. ANALISIS Y SINTESIS: Capability to analyze and solve complex technical problems.
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..
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.
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.

Full-or-part-time: 45 h
Self study: 45h
GRADING 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:

RESOURCES

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