

Course guide

270707 - IHLT - Introduction to Human Language Technology

Last modified: 21/07/2022

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.
Degree: MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2017). (Compulsory subject).
Academic year: 2022 **ECTS Credits:** 5.0 **Languages:** English

LECTURER

Coordinating lecturer: JORGE TURMO BORRÁS

Others: Primer quadrimestre:
GERARD ESCUDERO BAKX - 11, 12
SALVADOR MEDINA HERRERA - 12
JORGE TURMO BORRÁS - 11, 12

PRIOR SKILLS

Those acquired in the course of Artificial Intelligence (AI) (degree in Computer Engineering)

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

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.
CEP4. Capability to design, write and report about computer science projects in the specific area of Artificial Intelligence.
CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.
CEP7. Capability to respect the legal rules and deontology in professional practice.

Generical:

CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of Artificial Intelligence.
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.

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.

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..

TEACHING METHODOLOGY

There are two types of sessions: theory/exercise and laboratory.

In each theory/exercise session we will introduce new concepts together with the challenges they present and the approaches to face them. In addition, we will solve some exercises to fix those concepts, techniques and algorithms introduced in the session.

In the laboratory sessions small practices will be developed using the appropriate NLP tools to practice and reinforce the knowledge learned in the theory classes.

LEARNING OBJECTIVES OF THE SUBJECT

1. Understand the fundamental concepts of Natural Language Processing, most well-known techniques and theories as well as most relevant existing resources.
2. Understand most relevant applications of NLP and the theories, techniques and resources they use.
3. Design and development of programs to solve specific problems in the NLP context, involving the selection of most appropriate techniques and resources as well as the use of existing resources. There would be one larger programs to be developed in groups of two students.
5. Reason (occasionally, in group) about several problems in the NLP context that imply considering different techniques and resources.

STUDY LOAD

Type	Hours	Percentage
Guided activities	5,0	11.11
Hours medium group	10,0	22.22
Hours large group	25,0	55.56
Hours small group	5,0	11.11

Total learning time: 45 h

CONTENTS

Document Structure and Language

Description:

Text selection, Tokenization, Sentence splitting, Language Identifiers

Words

Description:

Morphology, Finite States Automata, Finite States Transducers.
PoS tagging, Hidden Markov Models.
Lexical semantics, Semantic resources.
Word Sense Disambiguation.



Word sequences

Description:

Recognition and classification of word sequences with meaning.
BIO discriminative models. Conditional Random Fields (CRF).
Named Entity Recognition and Classification (NERC).
Noun-phrase Chunking.

Sentences

Description:

Syntactic grammars, typology. Context free grammars. Probabilistic context free grammars. Chomsky normal form grammars.
Syntactic parsers, properties and strategies. CKY and probabilistic CKY parsers.

Sentence sequences

Description:

Coreference resolution. Mention detection. Types of techniques for the generation of coreferents chains. Mention-pair model.
Entity-mention model. Rankers model.

ACTIVITIES

Introduction

Specific objectives:

1, 2

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.
CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.
CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.
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..
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.

Full-or-part-time: 6h

Theory classes: 1h 30m
Laboratory classes: 1h 30m
Self study: 3h



Document structure and language

Specific objectives:

1, 3

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.

CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.

CEP4. Capability to design, write and report about computer science projects in the specific area of Artificial Intelligence.

CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capability to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 11h

Theory classes: 1h 30m

Laboratory classes: 1h 30m

Self study: 8h

Morphological analysis

Description:

Finite States Automata. Finite States Transducers.

Specific objectives:

1, 2

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.

CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.

CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

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..

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.

Full-or-part-time: 22h

Theory classes: 3h

Laboratory classes: 3h

Self study: 16h



Lexical semantics, Semantic resources.

Specific objectives:

1, 2, 5

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.

CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.

CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capability to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 19h

Theory classes: 1h 30m

Laboratory classes: 1h 30m

Self study: 16h

Word Sense Diambiguation.

Specific objectives:

1, 2, 5

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.

CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.

CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capability to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 19h

Theory classes: 1h 30m

Laboratory classes: 1h 30m

Self study: 16h



Recognition and classification of word sequences with meaning.

Description:

BIO discriminative models. Conditional Random Fields (CRF).
Named Entity Recognition and Classification (NERC).
Noun-phrase Chunking.

Specific objectives:

1, 3, 5

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.
CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.
CEP4. Capability to design, write and report about computer science projects in the specific area of Artificial Intelligence.
CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.
CEP7. Capability to respect the legal rules and deontology in professional practice.
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..
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.

Full-or-part-time: 22h

Theory classes: 3h

Laboratory classes: 3h

Self study: 16h



Syntactic parsing: Syntactic grammars

Description:

Typology. Context free grammars. Probabilistic context free grammars. Chomsky normal form grammars.

Specific objectives:

1, 2, 5

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.

CG1. Capacity to plan, design and implement products, processes, services and facilities in all areas of Artificial Intelligence.

CEA5. Capacity 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.

CEP6. Capacity to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capacity to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 22h

Theory classes: 3h

Laboratory classes: 3h

Self study: 16h

Final Exam

Specific objectives:

1, 2, 5

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.

CG1. Capacity to plan, design and implement products, processes, services and facilities in all areas of Artificial Intelligence.

CEA5. Capacity 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.

CEP6. Capacity to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capacity to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 2h

Guided activities: 2h

Syntactic parsing: parsers

Description:

Syntactic parsers, properties and strategies. CKY and probabilistic CKY parsers.

Specific objectives:

1, 2, 5

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.

CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.

CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capability to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 22h

Theory classes: 3h

Laboratory classes: 3h

Self study: 16h

Coreference resolution

Specific objectives:

1, 2

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.

CG1. Capability to plan, design and implement products, processes, services and facilities in all areas of 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.

CEP6. Capability to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

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..

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.

Full-or-part-time: 11h

Theory classes: 1h 30m

Laboratory classes: 1h 30m

Self study: 8h



PoS tagging

Description:

Hidden Markov Models

Specific objectives:

1, 2, 5

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.

CG1. Capacity to plan, design and implement products, processes, services and facilities in all areas of Artificial Intelligence.

CEA5. Capacity 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.

CEP6. Capacity to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capacity to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 11h

Theory classes: 1h 30m

Laboratory classes: 1h 30m

Self study: 8h

Project presentation

Specific objectives:

3, 5

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.

CG1. Capacity to plan, design and implement products, processes, services and facilities in all areas of Artificial Intelligence.

CEA5. Capacity 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.

CEP4. Capacity to design, write and report about computer science projects in the specific area of Artificial Intelligence.

CEP6. Capacity to assimilate and integrate the changing economic, social and technological environment to the objectives and procedures of informatic work in intelligent systems.

CEP7. Capacity to respect the legal rules and deontology in professional practice.

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..

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.

Full-or-part-time: 2h 30m

Guided activities: 0h 30m

Self study: 2h



GRADING SYSTEM

There will be a unique exam at the end of the course, one project and a deliverable for each lab session. The exam will include all the course contents.

The mark of the project and deliverables will be computed by considering the documents presented by the students.

The final mark of the course will be calculated as follows:

Course mark = final exam mark * 0.45 + project mark * 0.45 + deliverables of lab sessions * 0.1

BIBLIOGRAPHY

Basic:

- Jurafsky, D.; Martin, J.H. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. 2nd ed. Prentice-Hall, Inc., 2008. ISBN 9332518416.
- Dale, R.; Moisl, H.; Somers, H. Handbook of natural language processing. Marcel Dekker, 2000. ISBN 0824790006.
- Manning, C.D.; Schütze, H. Foundations of statistical natural language processing. MIT Press, 1999. ISBN 0262133601.
- Mitkov, R. (ed.). The Oxford handbook of computational linguistics. Oxford University Press, 2003. ISBN 0198238827.

RESOURCES

Hyperlink:

- <http://www.cs.upc.edu/~turmo/IHLT.html>