Description of Course Unit

Master in Informatics and Computing Engineering
Information Description, Storage and Retrieval
Instance: 2014/2015
—

Institutional page

General Information

Course Unit: Information Description, Storage and Retrieval
Code: EIC0108
Programmes: MIEIC, 5º, xx students/PRODEI yy students
Academic Year: 2013/2014
Semester: 1S
Credits: 6/7,5 ECTS
Hours/Weeks: 3 TP
Teachers: Cristina Ribeiro | João Correia Lopes

Teaching language

Suitable for English-speaking students

Objectives

1 - BACKGROUND

The "Information Description, Storage and Retrieval" unit assumes as its context the existence of large collections of heterogeneous information which needs to be organized, described, stored and retrieved.

2 - SPECIFIC OBJECTIVES

1. Make the students aware of the main issues in the organization and storage of large data collections.
2. Make the students familiar with the main concepts in textual information retrieval and their application in retrieval tools.
3. Explore the semantic web methods and tools, and use web resources and their descriptions in applications that make use of data semantics.

Skills and learning outcomes

On completion of this course, the student should be able to:

- Identify data sources in data repositories, online services APIs and user logs;
- Decide on the quality of the data sources and briefly characterize a selected dataset;
- Choose the document granularity and a storage model for the dataset;
- Use data manipulation tools to select appropriate data subsets and to fit the data to their intended applications;
- Describe the models used in information retrieval, specifically in web retrieval;
- Recognize the various tasks considered in information retrieval;
- Apply information retrieval evaluation measures to the comparison of web retrieval tools;
- Relate web documents with the metadata that describes or links them;
- Treat ontologies as providers of description tools;
- Explore the applications which manipulate semantic web information descriptions and create metadata sets for a chosen domain;
- Compare semantic web- based services with simpler approaches to resource description.

Work mode

In attendance

Previous knowledge

Knowledge and practice in programming languages for application development.

Program

- Introduction to datasets; tools for dataset collection, preparation and access; data models and dataset storage.
- Text information retrieval; retrieval models; evaluation; web information retrieval.
- Information description: semantic web languages; RDF, RDF-Schema, OWL; ontologies for data in a domain.

Main bibliography

- Anders Møller, Michael I. Schwartzbach;An Introduction to XML and Web Technologies, Addison Wesley Professional, 2006. ISBN: 0321269667 Biblioteca
- Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze;Introduction to Information Retrieval, Cambridge University Press, 2008. ISBN: 0521865719 Biblioteca

Teaching procedures

Lectures include theoretical presentation of the course subjects and practical sessions where proposed research topics are discussed with the students and practical coursework reported.

Software

- Apache Lucene + Solr
- oXygen XML editor
- Protégé

Keywords

Physical sciences > Computer science > Informatics

Evaluation type

Distributed evaluation without final exam

Registered evaluation and occupation components

Description	Type	Time (hours)	Date of conclusion
Attendance (estimated)	Lectures	39	
1st Project Delivery	Project Work	15	9/10/2014
2nd Project Delivery	Project Work	30	13/11/2014
3rd Project Delivery	Project Work	30	05/01/2015
Test 1	Test/Examination	1	20/11/2014
Test 2	Test/Examination	1	n/a
Study	Test/Examination	46/84	n/a
	Total:	162/200	

Eligibility for exams

The course has a practical component which results from the execution of projects, to be delivered up to the due dates established in the course plan. The students are admitted to the final exam if they achieve 50% in each component of the project work. Success in the course also requires 40% in each intermediate written test.

Working students and students with similar status who are not required to participate in class must present the evolution of their work in the time periods defined with teachers. These students are also required to take the mini-tests, deliver their practical work and participate in the presentation sessions.

Final grade

The final grade is computed using the formula: GRADE= 60% Projects + 40% Tests.

The Projects component is the result of the practical evaluation and can be obtained:

- completing three practical assignments according to the proposed scripts;
- proposing a semester-long project and reporting its results in the same sessions as the assignments.

The project and its workplan must be validated by the course instructors.

Special assignments

None. All students have to complete the projects and present them as scheduled.

Special evaluation (TE, DA, ...)

Distributed evaluation, performed during the semester, is required of all students, regardless of their enrolment status.

Improvement of final/distributed classification

Improving the classification requires a new enrolment in the course, taking the course projects and tests again.

– MCR, JCL