

## 30233 - Information Retrieval

### Información del Plan Docente

<b>Academic Year</b>	2017/18
<b>Faculty / School</b>	110 - Escuela de Ingeniería y Arquitectura
<b>Degree</b>	439 - Bachelor's Degree in Informatics Engineering
<b>ECTS</b>	6.0
<b>Year</b>	4
<b>Semester</b>	Indeterminate
<b>Subject Type</b>	Compulsory
<b>Module</b>	---

### **1.General information**

#### **1.1.Introduction**

##### **Brief presentation of the subject**

This course aims to introduce students to the techniques and algorithms that facilitate information retrieval. Information Retrieval is a discipline within Computer Science focused on the development of computer based search tools. This discipline provides the models and algorithms to address aspects as diverse as the representation, storage, organization and access to information elements.

#### **1.2.Recommendations to take this course**

Students who take this course must have training in methods and techniques of artificial intelligence at the level of the subject of Artificial Intelligence.

#### **1.3.Context and importance of this course in the degree**

The course of Information Retrieval is a compulsory subject within the set of subjects on specific technology about "Machine Learning and Information Retrieval" in the Computer Science intensification. Although Information Retrieval is presented as an application example of Artificial Intelligence techniques in the third year compulsory subject of "Artificial Intelligence", this Information Retrieval course gives an insight into the Computer Science discipline that allows developing information retrieval systems for different types of big and unstructured data sources.

#### **1.4.Activities and key dates**

The exam schedule and deadlines of work will be announced in advance.

### **2.Learning goals**

#### **2.1.Learning goals**

The student, for passing this subject, should demonstrate the following results:

## **30233 - Information Retrieval**

- Knowledge and use of information retrieval techniques on data collections stored in different repositories (including hypermedia and multimedia repositories).
- Applicability of information retrieval techniques to face new discovery problems.
- Management of ontology based techniques to represent the information available in a specific domain.
- Applicability of semantic retrieval techniques to develop search applications.

### **2.2.Importance of learning goals**

Currently, the dissemination and sharing of data in many different formats through the Web allows any type of software application to process and exploit a set of information resources that until a few years ago was unthinkable.

However, the syntactic and semantic heterogeneity of the data that can be downloaded requires also the application of a series of technical and formal procedures to enable the appropriate processing and extraction of information in order to take advantage of all these resources. Therefore, training in the concepts, techniques and methods presented in this course will be very important to deal with any information search problem.

### **3.Aims of the course and competences**

#### **3.1.Aims of the course**

After having completed a number of courses of programming, databases and information systems, the student is competent to apply data recovery strategies. Data recovery is mainly oriented to identify those records from a repository (file, database, etc.) containing the terms specified in the user's queries. However, when working with heterogeneous and unstructured data sources (e.g., Web or large repositories of text or multimedia data) such searches are not accurate enough to meet the user information needs. The aim of this course is to learn to apply a set of information retrieval techniques that are more focused on retrieving information about a topic than on the recovery of data that exactly match with a query.

The course has a clear applied nature. The student will learn the main information retrieval techniques, applying them to a set of examples of information retrieval systems integrated in Digital Libraries and other document repositories, to provide information search functionality. These systems will be presented in classroom and in the sessions of laboratory work.

#### **3.2.Competences**

- Know and apply the basic algorithmic procedures of computer technology to design solutions to problems, analysing the suitability and complexity of the proposed algorithms.
- Know and apply the necessary tools for storage, processing and access to Information systems, including web-based.
- Acquire, obtain, formalize and represent human knowledge in a computable form for problem solving using a computer in any area, particularly those related to aspects of computing, perception and action in intelligent environments.
- Develop and evaluate interactive systems and those that allow the presentation of complex information. To know how to use these systems for solving computer-person interaction problems.
- Understand and develop computational learning techniques and design and implement applications and systems that use them, including those dedicated to automatic extraction of information and knowledge from large volumes of data.

### **4.Assessment (1st and 2nd call)**

#### **4.1.Assessment tasks (description of tasks, marking system and assessment criteria)**

**The student must demonstrate that it has achieved the intended learning results through the following evaluation activities**

## 30233 - Information Retrieval

**Evaluation in June** . The overall assessment of the subject is done through two evaluation activities:

**P1.** Written exam in which the student will have to answer short questions and solve small exercises related to the subject. A minimum score of 5.0 points is required to pass the course. If this minimum score is obtained, this score will represent 50% of the final course grade. The date of this exam will be scheduled by the Faculty of Engineering and Architecture Board.

**P2.** Practical teamwork project. A minimum score of 5.0 points is required in this activity to pass the course. If this minimum score is obtained, this score will represent 50% of the final course grade. The project final deliverable will be sent electronically prior to the date established by the Faculty Board for the written exam. Previously, during the semester, students will have to deliver some of the elements that form part of the project, accompanied by presentations in class, which will provide students with the necessary teacher's feedback. If a team of students does not make these partial deliveries and presentations, in addition to submitting all deliverables, they must pass an exam of the teamwork project.

It is mandatory to do both evaluation activities to pass the course. If the mark of one of the activities, or both, is lower than 5.0, the final course grade will be the weighted average of the two grades (50% P1 and 50% P2), with a maximum of 4.0.

**Evaluation in September** . The overall assessment of the subject is done through two evaluation activities, which will be similar to those of June, with the same weights and minimum requirements. P2 activity in September will consist of the delivery of work and, if the students had not made partial deliveries and evaluations during the teaching of the subject, an additional exam should be done. The scores obtained in the June for P1 and P2 are held in September, unless the student chooses to be evaluated again. In case of being evaluated for a second time, the new rating will prevail.

## 5. Methodology, learning tasks, syllabus and resources

### 5.1. Methodological overview

The learning process that is designed for this subject is based on the following:

- Continuous study and work done since the first day of class.
- Learning of concepts and techniques through lectures, in which student participation is encouraged.
- Application of previous knowledge for solving problems. In the classes related to problems, students will play an active role in the discussion of cases and solving of the problems.
- Practical classes in laboratory where students learn how to implement the algorithms and strategies presented in lectures.
- Teamwork projects for the development and evaluation of two information retrieval systems that facilitate searches on a downloadable web document collection. The first system will be a recovery system where traditional information retrieval techniques are applied. The second system will be a semantic recovery system that will transform the collection documents in semantic descriptions of resources (RDF) stored in a triplestore to facilitate its search with a semantic query language.

### 5.2. Learning tasks

The learning activities that are offered to students to help them achieve the expected results are the following ones:

- In the classroom, the syllabus of the course will be developed through lectures, case analysis and problem solving that apply the concepts and techniques presented in the course syllabus.
- The practical sessions will take place in a computer lab. Throughout the different sessions each student must do, individually or in teams, work directly related to the topics studied in the course.
- In addition, teamwork projects under the tutorship of professors will be realized. In these projects each team must develop and evaluate different types of information retrieval systems over a document collection accessible through

the Web.

### 5.3.Syllabus

#### Subject program

##### Block I - Traditional Information Retrieval

- Introduction to information retrieval: the Boolean model
- The indexing process
- The vector space model
- Evaluation of search engines
- The probabilistic information retrieval model
- Relevance feedback and query expansion

##### Block II - Hypermedia and multimedia systems

- Search the Web
- User interface and visualization

##### Block III - Semantic Retrieval

- Introduction to ontologies and the Semantic Web
- The RDF representation language
- The SPARQL query language
- The OWL representation language

### 5.4.Course planning and calendar

Schedule of sessions and presentation of works:

The educational organization of the subject is as follows.

- Classroom classes (lectures and case problems) (3 hours in an ordinary week, which will correspond approximately, in the overall calculation of the course, to 2 hours of lectures and 1 hour of problem solving per week according to the academic calendar established by the Faculty Board).
- Laboratory practical sessions (one 2-hour session every two weeks, depending on the academic calendar established by the Faculty Board and the availability for booking laboratories). They are working sessions in the use of technologies, supervised by a teacher.
- Teamwork under the tutorship of professors where students develop and evaluate different types of information retrieval systems over document collections accessible on the Web.

Presentation of works under evaluation:

- The deadline for submitting the deliverables of the teamwork project will be the date established by the Faculty Board for the written exam (P1 evaluation activity). The deadlines of partial deliveries and presentations of the teamwork project depend on the academic calendar and will be announced in class on the first day devoted to the presentation of the subject, and on the Moodle platform within the practical work description.

### 5.5.Bibliography and recommended resources

[BB: Bibliografía básica / BC: Bibliografía complementaria]

- [BB] A semantic Web primer / Grigoris Antoniou...[et al.] . 3rd ed. Cambridge [etc.] : MIT Press, 2012
- [BB] Baeza-Yates, Ricardo. Modern information retrieval : the concepts and technology behind search / Ricardo Baeza-Yates, Berthier Ribeiro-Neto . 2nd ed. Harlow [etc.] : Addison-Wesley, 2011

## 30233 - Information Retrieval

- [BB] Gómez-Pérez, Asunción. Ontological engineering : with examples from the areas of knowledge management, e-Commerce and the Semantic Web / Asunción Gómez-Pérez, Mariano Fernández-López and Óscar Corcho . London ; New York : Springer, cop. 2010
- [BB] Hearst, Marti A. Search user interfaces / Marti A. Hearst . 1st pub. Cambridge [etc.] : Cambridge University Press, 2009
- [BB] Manning, Christopher D.. Introduction to information retrieval / Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze . 1st ed., repr. New York : Cambridge University Press, 2009
- [BB] Witten, Ian H. Managing gigabytes : compressing and indexing documents and images / Ian H. Witten, Alistair Moffat, Timothy C. Bell . 2nd ed. San Francisco, Calif. : Morgan Kaufmann Publishers, 1999