

Name of Module: Information Extraction, Retrieval and Integration	ECTS: 4.5	Module-ID:
<b>Person Responsible for Module (Name, Mail address):</b> (C) David Pérez del Rey, <a href="mailto:dperezdelrey@fi.upm.es">dperezdelrey@fi.upm.es</a> Víctor Maojo García, <a href="mailto:vmaojo@fi.upm.es">vmaojo@fi.upm.es</a> Miguel García Remesal, <a href="mailto:mgresal@fi.upm.es">mgresal@fi.upm.es</a> Raúl Alonso Calvo, <a href="mailto:ralonso@fi.upm.es">ralonso@fi.upm.es</a> María del Carmen Suárez de Figueroa Baonza, <a href="mailto:mcsuarez@fi.upm.es">mcsuarez@fi.upm.es</a>		
University: UPM	Departments: DIA	

### 1. Prerequisites for Participation

Students should have finished their Degree Project and completed the first semester of the ICT-Labs Master

### 2. Intended Learning Outcomes

Students after finishing the course will have:

- The ability to integrate technologies, applications, services and computer systems in general and in multidisciplinary contexts
- The ability to analyze information needs to build an information system
- To understand database and data representation language foundations
- To understand and interact with information retrieval systems
- To understand and interact with data extraction systems
- To understand and interact with integration systems
- To apply information retrieval, data extraction and integration to the biomedical field

### 3. Content

Description of main issues to be covered:

1. Basic Concepts
  - 1.1 Introduction
  - 1.2 Data, information and knowledge
  - 1.3 Data types
2. Extraction and information retrieval
  - 2.1 Information extraction
  - 2.2 Free text data
  - 2.3 Information retrieval models
  - 2.4 Search engines
  - 2.5 Non-textual data
3. Data integration
  - 3.1 Integration architectures
  - 3.2 Query rewriting
  - 3.3 Semantic interoperability
  - 3.4 Data provenance
4. Applications in biomedicine
  - 4.1 Clinical interoperability standards
  - 4.2 Biomedical vocabularies
  - 4.3 Biomedical Information Systems

#### 4. Teaching and Learning Methods

Teaching and learning approach:

- Theoretical lectures will be based on verbal explanations supported by presentations. With a balance between discussion and individual learning.
- Types of academic activities include:
  - Formal lectures
  - Preparation of individual and group practical assignments
  - Evaluation of knowledge through exams, homework or classroom participation
- Individual work assignments will be performed outside the lecture schedule. In these works, students will apply content explained during theoretical lectures and consult external bibliographical references
- Group work assignments share the same methodology of individual work, including cooperative work, conflict resolution, etc.

ECTS distribution (4.5 ECTS)

- Individual/group work: 1.5
- Applied lectures: 3

#### 5. Calendar

<b>Week</b>	<b>Classroom activities</b>	<b>Individual work</b>	<b>Group work</b>
1 (6h)	- Introduction to the course (1h) - Lecture on topic 1.1 (1h) - Supervised individual / group work (1h)	Self-study (3h)	
2 (6h)	- Lecture on topic 1.2 (1h) - Lecture on topic 1.3 (1h) - Supervised individual / group work (1h)	Self-study (3h)	
3 (6h)	- Lecture on topic 2.1 (2h) - Supervised individual / group work (1h)	Self-study (3h)	
4 (6h)	- Lecture on topic 2.2 (2h) - Supervised individual / group work (1h)	Individual assignment (2h) Self-study (1h)	
5 (7h)	- Lecture on topic 2.3 (2h) - Supervised individual / group work (1h)	Individual assignment (3h) Self-study (1h)	
6 (7h)	- Lecture on topic 2.4 (2h) - Supervised individual / group work (1h)	Self-study (4h)	
7 (7h)	- Lecture on topic 2.5 (2h) - Supervised individual / group work (1h)	Self-study (1h)	Group assignment (3h)
8 (8h)	- Lecture on topic 2.5 (2h) - Assignment presentation (1h)	Self-study (2h)	Group assignment (3h)
9 (8h)	- Lecture on topic 3.1 (2h) - Supervised individual / group work (1h)	Self-study (5h)	
10 (7h)	- Lecture on topic 3.2 (2h) - Supervised individual / group work (1h)	Self-study (4h)	
11 (7h)	- Lecture on topic 3.3 (2h) - Supervised individual / group work (1h)	Self-study (4h)	
12 (7h)	- Lecture on topic 3.3 (1h) - Lecture on topic 3.4 (1h) - Supervised individual / group work (1h)	Self-study (1h)	Group assignment (3h)
13 (8h)	- Lecture on topic 3.4 (2h) - Assignment presentation (1h)	Self-study (2h)	Group assignment (3h)
14 (8h)	- Lecture on topic 4.1 (2h) - Supervised individual / group work (1h)	Self-study (5h)	
15 (8h)	- Lecture on topic 4.2 (2h) - Supervised individual / group work (1h)	Self-study (1h)	Group assignment (4h)
16 (7h)	- Lecture on topic 4.3 (2h) - Assignment presentation (2h)	Self-study (1h)	Group assignment (2h)
Total (112h)	48h	46h	18h

## 6. Grading

The evaluation of the students will be based on:

- 4 assignments = 100% (25% each)

## 7. Assignments

Individual and group assignments will be provided during the course

## 8. Workload calculation (contact hours, homework, exam preparation,..)

48 hours lecturing

46 hours for individual work

18 hours for group work

Personal tuition will be offered to students or teams (up to 6 hour/week)

## 9. Frequency and dates

This course will be organised during the second semester of the 1<sup>st</sup> year.

3 hours activity per week

## 10. Max. Number of Participants

Students will be distributed in teams of 2-3 (depending on the number of students in the course) for developing the planned group-based activities.

## 11. Enrolment Procedure

Enrolment is not independent of the general enrolment process of the ICT-Labs master.

## 12. Recommended Reading, Course Material

Students will use the slides and open software used in the lectures and the following selected bibliography:

1. Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. Modern information retrieval. New York: ACM press, 1999.
2. Kimball, Ralph, and Margy Ross. The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons, 2011.
3. Doan, AnHai, Alon Halevy, and Zachary Ives. Principles of data integration. Elsevier, 2012.
4. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge University Press. 2008
5. Witten IH, Moffat A, Bell TC. Managing Gigabytes: Compressing and Indexing Documents and Images, 2nd Edition. Morgan Kaufmann. 1999.
6. Korfhage, R. Information Storage and Retrieval. Wiley. 1997.
7. Bird S, Klein E, Loper E. Natural Language Processing with Python. O'Reilly 2009. In successive academic years the individual work prepared by E1 students will be also available for other students' cohorts.

#### **11. Other Information (e.g. home page of module)**

--