



POLITÉCNICA

INTERNATIONAL
CAMPUS OF
EXCELLENCE

COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001



E.T.S. de Ingenieros
Informáticos

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000666 - Information retrieval, extraction and integration

DEGREE PROGRAMME

10AP - Eit Digital Master's Programme In Data Science

ACADEMIC YEAR & SEMESTER

2018/19 - Semester 2

Index

Learning guide

1. Description.....	1
2. Faculty.....	1
3. Skills and learning outcomes	2
4. Brief description of the subject and syllabus.....	3
5. Schedule.....	5
6. Activities and assessment criteria.....	9
7. Teaching resources.....	12

1. Description

1.1. Subject details

Name of the subject	103000666 - Information retrieval, extraction and integration
No of credits	4.5 ECTS
Type	Compulsory
Academic year of the programme	First year
Semester of tuition	Semester 2
Tuition period	February-June
Tuition languages	English
Degree programme	10AP - Eit digital master's programme in data science
Centre	10 - Escuela Tecnica Superior de Ingenieros Informaticos
Academic year	2018-19

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
David Perez Del Rey (Subject coordinator)	2104	david.perez.rey@upm.es	M - 14:00 - 16:00 W - 14:00 - 16:00 F - 11:00 - 13:00
Victor Manuel Maojo Garcia	2102	victormanuel.maojo@upm.es	Tu - 12:30 - 15:30 W - 12:30 - 15:30
Miguel Garcia Remesal	2206	miguel.garcia.remesal@upm.es	Tu - 11:00 - 14:00 Th - 11:00 - 14:00

M. Carmen Suarez De Figueroa Baonza	3205	mdelcarmen.suarezdefiguero a@upm.es	Tu - 14:00 - 16:00 W - 11:00 - 13:00 Th - 14:00 - 16:00
--	------	--	---

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

2.3. External faculty

Name and surname	Email	Institution
Raul Alonso Calvo	ralonso@fi.upm.es	ETSIIInf - DLSIIS

3. Skills and learning outcomes *

3.1. Skills to be learned

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CE01 - Capacidad para la integración de tecnologías, aplicaciones, servicios y sistemas propios de la Ingeniería Informática, con carácter generalista, y en contextos más amplios y multidisciplinares.

CE08 - Capacidad para analizar las necesidades de información que se plantean en un entorno y llevar a cabo en todas sus etapas el proceso de construcción de un sistema de información.

CG04 - Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos

CG05 - Aplicación de los métodos de resolución de problemas más recientes o innovadores y que puedan implicar el uso de otras disciplinas

CG07 - Integración del conocimiento a partir de disciplinas diferentes, así como el manejo de la complejidad

CG09 - Apreciación de los límites del conocimiento actual y de la aplicación práctica de la tecnología más reciente

3.2. Learning outcomes

RA79 - Conocer y diseñar sistemas de extracción de información

RA81 - Comprender y aplicar los fundamentos de los sistemas de recuperación de información

RA82 - Conocer los requisitos específicos de sistemas de información en biomedicina

RA80 - Ser capaz de diseñar la arquitectura de un sistema de información

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

4. Brief description of the subject and syllabus

4.1. Brief description of the subject

The amount of available data in any area has grown dramatically during the last years. However, this increment did not have a proportional impact in the knowledge available for decision making. There is a need of automatic models to manage the data, taking into account that the majority of the data will never be used by a human being. The course Information Retrieval, Extraction and Integration is focused on the necessary tasks to extract information, models to efficiently retrieve data for further integration. These are critical tasks to provide relevant information for decision making, which complexity increases with the amount of data available. As application areas, we focus on biomedicine, due to the complexity and to the specific requirements.

4.2. Syllabus

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. Information Retrieval Models
 - 2.3. Natural Language Processing
 - 2.4. Web Search Engines
 - 2.5. Non-textual Data
3. Data Integration
 - 3.1. Integration Architectures
 - 3.2. Semantic Interoperability
 - 3.3. Data Provenance
4. Applications in Biomedicine
 - 4.1. Sistemas de información biomédica
 - 4.2. Estándares de interoperabilidad clínica
 - 4.3. Vocabularios biomédicos
 - 4.4. Sistemas de recuperación de literatura científica

5. Schedule

5.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Other face-to-face activities	Assessment activities
1	<p>Presentación de la asignatura Duration: 01:00 Lecture</p> <p>Desarrollo del tema 1.1 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
2	<p>Desarrollo del Tema 1.2 Duration: 01:00 Lecture</p> <p>Desarrollo del Tema 1.3 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
3	<p>Desarrollo del Tema 1.3 Duration: 01:00 Lecture</p> <p>Desarrollo del Tema 2.1 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
4	<p>Desarrollo del Tema 2.1 Duration: 01:00 Lecture</p> <p>Desarrollo del Tema 2.2 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 05:00</p>

5	<p>Desarrollo del Tema 2.2 Duration: 01:00 Lecture</p> <p>Desarrollo del Tema 2.3 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Entrega trabajo individual Individual work Continuous assessment and final examination Duration: 06:00</p>
6	<p>Desarrollo del Tema 2.3 Duration: 02:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 05:00</p>
7	<p>Desarrollo del Tema 2.3 Duration: 01:00 Lecture</p> <p>Desarrollo del Tema 2.4 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
8	<p>Desarrollo del Tema 2.5 Duration: 01:00 Lecture</p> <p>Trabajo en grupo supervisado Duration: 01:00 Cooperative activities</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
9	<p>Desarrollo del Tema 2.5 Duration: 01:00 Lecture</p> <p>Trabajo en grupo supervisado Duration: 01:00 Cooperative activities</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Entrega trabajo en grupo Group work Continuous assessment and final examination Duration: 06:00</p>
10	<p>Desarrollo del Tema 3.1 Duration: 01:00 Lecture</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 03:00</p>

11	<p>Desarrollo del Tema 3.1 Duration: 02:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
12	<p>Desarrollo del Tema 3.1 Duration: 01:00 Lecture</p> <p>Desarrollo del Tema 3.2 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
13	<p>Desarrollo del Tema 3.2 Duration: 01:00 Lecture</p> <p>Desarrollo del Tema 3.3 Duration: 01:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Entrega de trabajo en grupo Group work Continuous assessment and final examination Duration: 06:00</p>
14	<p>Desarrollo del Tema 3.3 Duration: 02:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 02:00</p>
15	<p>Desarrollo del Tema 3.4 Duration: 02:00 Lecture</p> <p>Estudio supervisado Duration: 01:00 Additional activities</p>			<p>Estudio autónomo Other assessment Continuous assessment Duration: 04:00</p>
16	<p>Estudio supervisado Duration: 02:00 Additional activities</p> <p>Desarrollo del Tema 3.4 Duration: 01:00 Lecture</p>			<p>Entrega de Trabajo en grupo Group work Continuous assessment and final examination Duration: 04:00</p>
17				<p>Entrega de trabajo en grupo Group work Continuous assessment and final examination Duration: 06:00</p>

The independent study hours are training activities during which students should spend time on individual study or individual assignments.

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27

hours of student face-to-face contact and independent study time.

* The subject schedule is based on a previous theoretical planning of the subject plan and might go through experience some unexpected changes along throughout the academic year.

6. Activities and assessment criteria

6.1. Assessment activities

6.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
1	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
2	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
3	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
4	Estudio autónomo	Other assessment	No Presential	05:00	%	3 / 10	
5	Entrega trabajo individual	Individual work	No Presential	06:00	25%	3 / 10	CG05 CE01 CG09 CB10 CE08
6	Estudio autónomo	Other assessment	No Presential	05:00	%	3 / 10	
7	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
8	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
9	Entrega trabajo en grupo	Group work	No Presential	06:00	20%	3 / 10	CE01 CG07 CB07
10	Estudio autónomo	Other assessment	No Presential	03:00	%	3 / 10	
11	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
12	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
13	Entrega de trabajo en grupo	Group work	Face-to-face	06:00	20%	3 / 10	CG04 CG09 CE08

14	Estudio autónomo	Other assessment	No Presential	02:00	%	3 / 10	
15	Estudio autónomo	Other assessment	No Presential	04:00	%	3 / 10	
16	Entrega de Trabajo en grupo	Group work	Face-to-face	04:00	10%	3 / 10	
17	Entrega de trabajo en grupo	Group work	Face-to-face	06:00	25%	3 / 10	CG07 CB07

6.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
5	Entrega trabajo individual	Individual work	No Presential	06:00	25%	3 / 10	CG05 CE01 CG09 CB10 CE08
9	Entrega trabajo en grupo	Group work	No Presential	06:00	20%	3 / 10	CE01 CG07 CB07
13	Entrega de trabajo en grupo	Group work	Face-to-face	06:00	20%	3 / 10	CG04 CG09 CE08
16	Entrega de Trabajo en grupo	Group work	Face-to-face	04:00	10%	3 / 10	
17	Entrega de trabajo en grupo	Group work	Face-to-face	06:00	25%	3 / 10	CG07 CB07

6.1.3. Referred (re-sit) examination

No se ha definido la evaluación extraordinaria.

6.2. Assessment criteria

Modalidad de evaluación continua (ordinaria)

La calificación final de la asignatura se calculará a partir de las calificaciones de un trabajo individual y 3 trabajos en grupo.

Primer trabajo individual. Tras la presentación del primer tema de minería de textos, los alumnos deberán realizar un trabajo. La calificación de este trabajo será un 25% sobre el contenido.

Trabajo en grupo. Trabajo en grupo sobre la aplicación de sistemas de recuperación de información en el campo de la biomedicina. Se deberá realizar una implementación, un breve informe y una exposición en clase. La calificación de este trabajo será un 20% sobre el informe y un 5% sobre la presentación.

Segundo en grupo. Trabajo sobre interoperabilidad semántica. La calificación de este trabajo será un 20% sobre el informe y un 5% sobre la presentación.

Tercer trabajo en grupo. Dividido en dos partes, sobre optimización en buscadores y ontologías biomédicas. La calificación de este trabajo será un 20% sobre el informe y un 5% sobre la presentación.

Modalidad de evaluación extraordinaria

En la convocatoria ordinaria, la elección entre el sistema de evaluación continua o el sistema de evaluación mediante prueba final corresponde al estudiante. Quien desee seguir el sistema de evaluación mediante sólo prueba final, deberá OBLIGATORIAMENTE comunicarlo DURANTE LOS 15 PRIMEROS DÍAS a contar desde el inicio de la actividad docente de la asignatura, mediante escrito dirigido al Prof. Coordinador de la asignatura que entregará dentro del plazo establecido. La calificación en caso de evaluación final será a través de examen.

7. Teaching resources

7.1. Teaching resources for the subject

Name	Type	Notes
Modern Information Retrieval	Bibliography	Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. Modern information retrieval. New York: ACM press, 1999.
The data warehouse toolkit	Bibliography	Kimball, Ralph, and Margy Ross. The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons, 2011.
Introduction to Information Retrieval	Bibliography	Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge University Press. 2008
Managing Gigabytes	Bibliography	Witten IH, Moffat A, Bell TC. Managing Gigabytes: Compressing and Indexing Documents and Images, 2nd Edition. Morgan Kaufmann. 1999.
Natural Language Processing with Python	Bibliography	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.