

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

103000829 - Information Retrieval Extraction And Integration

PLAN DE ESTUDIOS

10AZ - Master Universitario En Innovación Digital

CURSO ACADÉMICO Y SEMESTRE

2023/24 - Segundo semestre

Índice

Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Competencias y resultados de aprendizaje.....	2
4. Descripción de la asignatura y temario.....	3
5. Cronograma.....	5
6. Actividades y criterios de evaluación.....	7
7. Recursos didácticos.....	10
8. Otra información.....	11

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	103000829 - Information Retrieval Extraction And Integration
No de créditos	4.5 ECTS
Carácter	Optativa
Curso	Primer curso
Semestre	Segundo semestre
Período de impartición	Febrero-Junio
Idioma de impartición	Inglés/Castellano
Titulación	10AZ - Master Universitario en Innovación Digital
Centro responsable de la titulación	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Curso académico	2023-24

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Miguel Garcia Remesal (Coordinador/a)	2206	miguel.garcia.remesal@upm.es	M - 11:00 - 14:00 J - 11:00 - 14:00
David Perez Del Rey	2104	david.perez.rey@upm.es	L - 14:00 - 16:00 X - 14:00 - 16:00 V - 11:00 - 13:00

M. Carmen Suarez De Figueroa Baonza	3205	mdelcarmen.suarezdefiguero a@upm.es	M - 14:00 - 16:00 X - 11:00 - 13:00 J - 14:00 - 16:00
Victor Manuel Maojo Garcia	2102	victormanuel.maojo@upm.es	M - 12:30 - 15:30 X - 12:30 - 15:30
Raul Alonso Calvo		raul.alonso@upm.es	Sin horario.

* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

3. Competencias y resultados de aprendizaje

3.1. Competencias

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.

CE-CD06 - Capacidad para extraer, integrar y consultar datos heterogéneos en diferentes escenarios

CG01 - Que los estudiantes sean capaces de predecir y controlar la evolución de situaciones complejas mediante el desarrollo de nuevas e innovadoras metodologías de trabajo adaptadas al ámbito científico/investigador, tecnológico o profesional concreto, en general multidisciplinar, en el que se desarrolle su actividad.

CG03 - La capacidad de usar la lengua inglesa de manera competente, es decir, con capacitación para tareas complejas de trabajo y estudio.

3.2. Resultados del aprendizaje

RA95 - Understand and apply information retrieval systems

RA49 - apply IA techniques in real data scenarios

RA94 - Understand and design information extraction systems

RA34 - Conocer cómo se aplican las técnicas de computación científica en algún campo específico de ciencia o ingeniería.

4. Descripción de la asignatura y temario

4.1. Descripción de la asignatura

La cantidad de datos disponibles en la mayoría de las áreas científicas ha crecido exponencialmente durante los últimos años. Sin embargo, este incremento no ha tenido un impacto equiparable en el conocimiento disponible para la toma de decisiones. Hay actualmente una necesidad de modelos automatizados para gestionar dichos datos, teniendo en cuenta que los seres humanos nunca utilizarán directamente la mayor parte de ellos. El curso Recuperación, extracción e integración de información está enfocado en proporcionar los métodos y herramientas necesarios para extraer información y modelos para recuperar datos de manera eficiente para una mayor integración en procesos de decisión. Estas son tareas críticas para brindar información relevante para la toma de decisiones, cuya complejidad aumenta con la cantidad de datos disponibles. Como áreas de aplicación, nos centraremos principalmente en la biomedicina debido a la complejidad y los requisitos específicos.

4.2. Temario de la asignatura

1. Basic concepts
 - 1.1. Introduction
 - 1.2. Data, information and knowledge
 - 1.3. Data types
2. Handling textual data / Information retrieval
 - 2.1. Modern information retrieval
 - 2.2. IR systems evaluation
3. Handling non-textual data
 - 3.1. Introduction and basic descriptors
 - 3.2. Content-based information retrieval
4. Data Integration
 - 4.1. Introduction
 - 4.2. Bias challenges
 - 4.3. Fairness challenges
5. Search engines
 - 5.1. Web search engines
 - 5.2. Machine learning-based ranking
6. Applications in biomedicine
 - 6.1. Biomedical information systems
 - 6.2. Biomedical vocabularies
 - 6.3. Standards for clinical interoperability
 - 6.4. Systems for retrieving scientific literature

5. Cronograma

5.1. Cronograma de la asignatura *

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	Presentación del curso Duración: 01:00 LM: Actividad del tipo Lección Magistral Desarrollo del Tema 1 Duración: 01:00 LM: Actividad del tipo Lección Magistral Desarrollo del Tema 2 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	Desarrollo del Tema 2 Duración: 04:00 LM: Actividad del tipo Lección Magistral			Entrega de trabajo en grupo (tema 2) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo) TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 03:00
3	Desarrollo del Tema 3 Duración: 04:00 LM: Actividad del tipo Lección Magistral			
4	Desarrollo del Tema 3 Duración: 02:00 LM: Actividad del tipo Lección Magistral Desarrollo del Tema 4 Duración: 02:00 LM: Actividad del tipo Lección Magistral			Entrega de trabajo en grupo (tema 3) y presentación en clase TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 03:00
5	Desarrollo del Tema 4 Duración: 04:00 LM: Actividad del tipo Lección Magistral			Entrega de trabajo en grupo (tema 4) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo) TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 03:00
6	Desarrollo del Tema 5 Duración: 04:00 LM: Actividad del tipo Lección Magistral			
7	Desarrollo del Tema 5 Duración: 02:00 LM: Actividad del tipo Lección Magistral Desarrollo del Tema 6 Duración: 02:00 LM: Actividad del tipo Lección Magistral			Entrega de trabajo en grupo (tema 5) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo) TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 03:00

8	Desarrollo del Tema 6 Duración: 04:00 LM: Actividad del tipo Lección Magistral			
9				
10				
11				
12				
13				
14				
15				
16				
17				

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.

6. Actividades y criterios de evaluación

6.1. Actividades de evaluación de la asignatura

6.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
2	Entrega de trabajo en grupo (tema 2) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo)	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CG01 CG03
4	Entrega de trabajo en grupo (tema 3) y presentación en clase	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CB10 CG03
5	Entrega de trabajo en grupo (tema 4) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo)	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CB07 CG03
7	Entrega de trabajo en grupo (tema 5) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo)	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CG03 CE-CD06

6.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
2	Entrega de trabajo en grupo (tema 2) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo)	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CG01 CG03
4	Entrega de trabajo en grupo (tema 3) y presentación en clase	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CB10 CG03
5	Entrega de trabajo en grupo (tema 4) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo)	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CB07 CG03

7	Entrega de trabajo en grupo (tema 5) y posible presentación del trabajo en clase (bajo demanda del profesor responsable del trabajo)	TG: Técnica del tipo Trabajo en Grupo	Presencial	03:00	25%	5 / 10	CG03 CE-CD06
---	--	---------------------------------------	------------	-------	-----	--------	-----------------

6.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Examen final	EX: Técnica del tipo Examen Escrito	Presencial	01:00	100%	5 / 10	CB07 CB10 CG01 CG03 CE-CD06

6.2. Criterios de evaluación

Evaluación Progresiva (durante el período docente)

La calificación final se calculará a partir de las notas de cuatro trabajos prácticos grupales obligatorios sobre los temas 2, 3, 4 y 5. Cada trabajo práctico tiene el mismo peso (25%) sobre la calificación final, por lo que esta se calculará como la media aritmética de las calificaciones de los trabajos. Por otro lado, los profesores de la asignatura pueden solicitar a los diferentes grupos de prácticas que lleven a cabo una presentación GRUPAL en el aula sobre los trabajos realizados. Esta presentación es obligatoria para uno de los trabajos (tema 3) y a discrección del profesor encargado del trabajo para el resto de los trabajos. La nota de la presentación estará integrada en la calificación del trabajo práctico y por ello esta puede diferir entre los miembros de un mismo grupo.

Para superar la asignatura, la calificación individual de TODOS los trabajos prácticos propuestos debe ser igual o superior a 5 puntos. No obstante, será posible también superar la asignatura si como máximo en uno de los trabajos se ha obtenido una calificación inferior a 5 puntos e igual o superior a 3 puntos. En este último caso, la nota final se calculará restando 1 punto de la media aritmética calculada con un suelo de 5 puntos. Es decir, si la calificación resultante tras la deducción fuese inferior a 5 puntos, entonces la nota final del alumno sería de 5 puntos (aprobado). En caso de haber obtenido una calificación inferior a 5 puntos en dos o más trabajos, entonces la calificación se calculará como el mínimo entre 4 puntos y la media aritmética de las prácticas (suspense en ambos casos)

Cualquier trabajo práctico puede entregarse en la fecha programada incluso si no se ha entregado alguno de los trabajos previos. Los trabajos suspensos o no entregados pueden recuperarse en la convocatoria ordinaria, aunque nótese que ya no será posible obtener la máxima nota (ver reglamento de la convocatoria ordinaria)

Convocatoria ordinaria

Si el alumno no ha superado el curso por evaluación progresiva, entonces este podrá realizar una nueva entrega de TODOS los trabajos que no haya entregado durante la evaluación progresiva, así como de aquellos en los que haya obtenido una calificación inferior a 5 puntos. No será posible realizar una nueva entrega de trabajos superados durante la evaluación progresiva (calificación igual o superior a 5 puntos). La fecha límite de entrega será la fijada en el calendario oficial de la titulación para la prueba global de la convocatoria ordinaria. El alumno tendrá que realizar una presentación INDIVIDUAL en el aula de cada uno de los trabajos presentados en esta convocatoria si el profesor responsable del trabajo así lo considera. Estas presentaciones, si finalmente son requeridas, se realizarán de forma presencial el día fijado en el calendario oficial de la titulación para la prueba global de la convocatoria ordinaria. Al igual que en la evaluación progresiva, la calificación de la presentación, si se requiere, estará integrada en la nota final del trabajo. Será posible únicamente superar el curso si la calificación de TODOS los trabajos entregados en esta convocatoria, junto con aquellos superados durante la evaluación progresiva, es igual o superior a 5 puntos. La calificación final se calculará como la media aritmética de las notas de todos los trabajos deduciendo 0.5 puntos de esta media aritmética por cada trabajo entregado en esta convocatoria con un suelo de 5 puntos. En caso de haber obtenido una calificación inferior a 5 puntos en uno o más trabajos, entonces la calificación se calculará como el mínimo entre 4 puntos y la media aritmética de las prácticas (suspense en ambos casos)

Convocatoria extraordinaria

En la convocatoria extraordinaria se realizará un examen final escrito que cubrirá todos los contenidos de la asignatura. Para superar el curso será necesario obtener una calificación igual o superior a 5 puntos. La calificación final es la nota del examen, calificado sobre 10 puntos.

7. Recursos didácticos

7.1. Recursos didácticos de la asignatura

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

8. Otra información

8.1. Otra información sobre la asignatura

El curso se impartirá de forma intensiva durante medio semestre (8 semanas)