



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros de  
Telecomunicacion

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

95000346 - Nlp Y Recuperación De Información

### PLAN DE ESTUDIOS

09BM - Grado En Ingenieria Biomedica

### CURSO ACADÉMICO Y SEMESTRE

2021/22 - Segundo semestre

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Conocimientos previos recomendados.....	2
4. Competencias y resultados de aprendizaje.....	2
5. Descripción de la asignatura y temario.....	3
6. Cronograma.....	5
7. Actividades y criterios de evaluación.....	7
8. Recursos didácticos.....	8
9. Otra información.....	9

## 1. Datos descriptivos

---

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	95000346 - Nlp y Recuperación de Información
<b>No de créditos</b>	4 ECTS
<b>Carácter</b>	Optativa
<b>Curso</b>	Cuarto curso
<b>Semestre</b>	Octavo semestre
<b>Período de impartición</b>	Febrero-Junio
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	09BM - Grado en Ingeniería Biomedica
<b>Centro responsable de la titulación</b>	09 - Escuela Tecnica Superior De Ingenieros De Telecomunicacion
<b>Curso académico</b>	2021-22

## 2. Profesorado

---

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías *</b>
Carlos Angel Iglesias Fernandez (Coordinador/a)		carlosangel.iglesias@upm.es	- -
Luis Fernando D'haro Enriquez		luisfernando.dharo@upm.es	Sin horario.
Alejandro Rodriguez Gonzalez		alejandro.rg@upm.es	Sin horario.

Ernestina Menasalvas Ruiz		ernestina.menasalvas@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. Conocimientos previos recomendados

---

#### 3.1. Asignaturas previas que se recomienda haber cursado

- Algoritmos Y Estructuras De Datos
- Bases De Datos
- Fundamentos De Programación

#### 3.2. Otros conocimientos previos recomendados para cursar la asignatura

El plan de estudios Grado en Ingeniería Biomedica no tiene definidos otros conocimientos previos para esta asignatura.

### 4. Competencias y resultados de aprendizaje

---

#### 4.1. Competencias

CE23 - Capacidad para conocer, utilizar y diseñar sistemas de información y comunicaciones en sanidad y biomedicina

CG02 - Aplicar de forma profesional a su trabajo los conocimientos adquiridos.

## 4.2. Resultados del aprendizaje

RA196 - Conocimiento de modelos de recuperación de información.

RA198 - Conocimiento de técnicas de motores de búsqueda y crawlers

RA194 - Conocimiento de las técnicas de procesado de lenguaje natural, minería de textos y su aplicación en ingeniería biomédica

RA197 - Capacidad de desarrollo práctico de un sistema de recuperación de información

RA195 - Conocimiento de fundamentos de recuperación de información

## 5. Descripción de la asignatura y temario

---

### 5.1. Descripción de la asignatura

1. NLP y recuperación de información.

Introducción a la necesidad de técnicas de análisis y búsqueda de información masiva. Introducción al procesamiento de lenguaje natural (PLN), sus técnicas básicas y aplicaciones. Búsqueda y análisis vs bases de datos. Proceso de recuperación de información.

2. Fundamentos de procesamiento de lenguaje Natural.

Análisis léxico, sintáctico y semántico. Reconocimiento de entidades.

3. Modelos de recuperación de información

Definición formal y conceptos básicos. Modelos abstractos de documentos y lenguajes de consulta. Modelo booleano. Modelo vectorial. Indexado semántico latente.

4. Implementación de indexado y búsqueda.

Ficheros de firma e inverso. Compresión de índices. Implementación eficiente de la regla de medida del coseno con tf-idf. Ejemplo con motor de indexación y búsqueda (ej. Lucene).

5. Evaluación en recuperación de información:

Precisión y cobertura (recall). Otras medidas de prestaciones. Colecciones de referencia. Realimentación de relevancia y expansión de la consulta.

6. Motores de búsqueda Web y arañas (crawlers)

Ranking relevancia en la web. El algoritmo page rank. Crawling. Arquitectura de un buscador web.

7. Aplicaciones. Bibliotecas médicas y aplicaciones en ingeniería biomédica.

Herramientas para NLP en biomedicina (ej. SemRep, MetaMap, Saber).

Extracción de información biomédica en Wikipedia (ej. caso de uso de MetaMap)

Procesamiento de texto en historias clínicas en español. (ej. CliKES)

## 5.2. Temario de la asignatura

1. NLP y recuperación de información.
2. Fundamentos de procesamiento de lenguaje Natural
3. Modelos de recuperación de información
4. Implementación de indexado y búsqueda
5. Evaluación en recuperación de información
6. Motores de búsqueda Web y arañas (crawlers)
7. Aplicaciones. Bibliotecas médicas y aplicaciones en ingeniería biomédica

## 6. Cronograma

### 6.1. Cronograma de la asignatura \*

Sem	Actividad presencial en aula	Actividad presencial en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<b>Introducción asignatura. Introducción PLN y RI</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral			
2	<b>Introducción PLN y RI</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral			
3	<b>Fundamentos PLN</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral			
4		<b>Fundamentos PLN</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		
5	<b>Modelos de RI</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral			<b>Entrega memoria prácticas temas NLP</b> TI: Técnica del tipo Trabajo Individual Evaluación continua No presencial Duración: 00:00
6		<b>Introducción RI</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		
7	<b>Implementación de RI</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral			
8		<b>Implementación de RI</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		<b>Examen parcial tipo test temas 1-3</b> ET: Técnica del tipo Prueba Telemática Evaluación continua Presencial Duración: 00:30  <b>Entrega memoria prácticas temas IR</b> TI: Técnica del tipo Trabajo Individual Evaluación continua No presencial Duración: 00:00
9	<b>Evaluación de RI</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral			
10		<b>Evaluación de RI</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		

11	<b>Buscadores web y crawlers</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral			
12		<b>Buscadores web y crawlers</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		
13		<b>Aplicación bibliotecas médicas</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		
14		<b>Aplicación bibliotecas médicas</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		
15		<b>Aplicación bibliotecas médicas</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		
16		<b>Aplicación bibliotecas médicas</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		
17				<p><b>Examen segundo parcial temas 4-7</b> ET: Técnica del tipo Prueba Telemática Evaluación continua Presencial Duración: 00:30</p> <p><b>Examen final temas 1-7</b> ET: Técnica del tipo Prueba Telemática Evaluación sólo prueba final Presencial Duración: 01:00</p> <p><b>Entrega memoria prácticas Crawlers y Aplicaciones Biomédicas</b> TI: Técnica del tipo Trabajo Individual Evaluación continua No presencial Duración: 00:00</p> <p><b>Entrega memoria prácticas curso</b> TI: Técnica del tipo Trabajo Individual Evaluación sólo prueba final No presencial Duración: 00:00</p>

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.



## 7. Actividades y criterios de evaluación

### 7.1. Actividades de evaluación de la asignatura

#### 7.1.1. Evaluación continua

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
5	Entrega memoria prácticas temas NLP	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	10%	4 / 10	
8	Examen parcial tipo test temas 1-3	ET: Técnica del tipo Prueba Telemática	Presencial	00:30	20%	4 / 10	CE23
8	Entrega memoria prácticas temas IR	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	20%	4 / 10	CE23
17	Examen segundo parcial temas 4-7	ET: Técnica del tipo Prueba Telemática	Presencial	00:30	20%	4 / 10	CG02 CE23
17	Entrega memoria prácticas Crawlers y Aplicaciones Biomédicas	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	30%	4 / 10	CG02

#### 7.1.2. Evaluación sólo prueba final

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Examen final temas 1-7	ET: Técnica del tipo Prueba Telemática	Presencial	01:00	40%	4 / 10	CG02 CE23
17	Entrega memoria prácticas curso	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	60%	4 / 10	

#### 7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Examen tipo test temas 1-7	ET: Técnica del tipo Prueba Telemática	Presencial	01:00	40%	4 / 10	CG02 CE23
Entrega memoria prácticas temas 1-7	TI: Técnica del tipo Trabajo Individual	Presencial	00:05	60%	4 / 10	CG02 CE23

## 7.2. Criterios de evaluación

La evaluación continua constará de

- dos exámenes tipo test de los temas 1-3 (20% nota) y 4-7 (20% nota).
- dos memorias de las prácticas de los temas NLP (10% nota), IR (20% nota) y Crawlers y Aplicaciones biomédicas (30% nota).

Todas las notas deberán ser mayores o iguales a 4.

La evaluación final y extraordinaria consta de un examen tipo test de la asignatura (40%), y una memoria de las prácticas de la asignatura (60%). Todas las notas deberán ser mayores o iguales a 4.

## 8. Recursos didácticos

### 8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Referencia principal	Bibliografía	Christopher, D.M., Prabhakar, R. and Hinrich, S., 2008. Introduction to information retrieval. An Introduction To Information Retrieval, Cambridge University Press; p.506.

Referencia de apoyo	Bibliografía	Baeza-Yates, R. and Ribeiro-Neto, B., 2011. Modern Information Retrieval: the concepts and technology behind search 2nd Ed. Harlow, England: Pearson Education Ltd.
Referencia Apoyo 2	Bibliografía	Cohen, K.B. and Demner-Fushman, D., 2014. Biomedical natural language processing (Vol. 11). John Benjamins Publishing Company.
Referencia apoyo 3	Bibliografía	Croft, W.B., Metzler, D. and Strohman, T., 2010. Search engines: Information retrieval in practice. Ed. Addison-Wesley.
Referencia apoyo 4	Bibliografía	Hersh WR, Information Retrieval: A Health and Biomedical Perspective, Third Edition, 2009

## 9. Otra información

---

### 9.1. Otra información sobre la asignatura

Esta es una asignatura de introducción a las TIC que, en líneas generales, puede contribuir a los Objetivos de Desarrollo Sostenible (ODS) 4 y 9 de Naciones Unidas, en relación con el aumento del número de personas con competencias técnicas y profesionales (meta 4.4) y el desarrollo de infraestructuras fiables, sostenibles, resilientes y de calidad (meta 9.1).