



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros  
Informaticos

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

**105001014 - Procesamiento De Lenguaje Natural**

### PLAN DE ESTUDIOS

10CD - Grado En Ciencia De Datos E Inteligencia Artificial

### CURSO ACADÉMICO Y SEMESTRE

2022/23 - Primer 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.....	8
7. Recursos didácticos.....	10
8. Otra información.....	11

## 1. Datos descriptivos

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	105001014 - Procesamiento de Lenguaje Natural
<b>No de créditos</b>	3 ECTS
<b>Carácter</b>	Obligatoria
<b>Curso</b>	Segundo curso
<b>Semestre</b>	Tercer semestre
<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10CD - Grado en Ciencia de Datos e Inteligencia Artificial
<b>Centro responsable de la titulación</b>	10 - Escuela Tecnica Superior De Ingenieros Informaticos
<b>Curso académico</b>	2022-23

## 2. Profesorado

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías *</b>
Mariano Rico Almodovar (Coordinador/a)	3205	mariano.rico@upm.es	M - 14:00 - 17:00 J - 12:00 - 15:00 Escribir email para concertar cita.
Emilio Serrano Fernandez	2201	emilio.serrano@upm.es	M - 15:00 - 18:00 X - 09:00 - 12:00 Escribir email para concertar cita.

\* 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

CB01 - Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio

CB02 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio

CB03 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética

CB04 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado

CB05 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CE14 - Capacidad para describir las técnicas de adquisición y representación del conocimiento, y modelos de razonamiento en entornos centralizados y distribuidos, y utilizarlas para desarrollar sistemas basados en el conocimiento orientados a la resolución de problemas y toma de decisiones que requieran conducta inteligente.

CE17 - Capacidad para describir y aplicar los mecanismos de interacción en sociedades es artificiales e híbridas, incluyendo aspectos relacionados con el procesamiento de lenguaje natural, la decisión colectiva, la negociación y la coordinación.

CG01 - Capacidad de trabajo en equipo, en entornos interdisciplinarios y complejos, negociando y resolviendo conflictos, diseñando soluciones eficientes, fiables, robustas y responsables.

CG02 - Capacidad para organizar y planificar tareas y proyectos, identificando objetivos, prioridades, plazos, recursos y riesgos, y controlando los procesos establecidos.

CG04 - Capacidad para innovar y encontrar soluciones creativas en situaciones complejas o de incertidumbre en

el ámbito de la ingeniería.

CG06 - Identificar y utilizar las tecnologías de la información y las comunicaciones más adecuadas en el ámbito de la ingeniería.

CG07 - Capacidad para integrar aspectos sociales, ambientales, económicos y éticos inherentes a la ingeniería, analizando sus impactos, y comprometiéndose con la búsqueda de soluciones a retos del desarrollo sostenible.

## 3.2. Resultados del aprendizaje

RA82 - RA-IA-15 Conocer las técnicas de ?minería de textos?, extracción y recuperación de información y sistemas pregunta-respuesta

RA81 - RA-IA-14 Conocer los fundamentos de las técnicas de procesamiento de lenguaje natural y los sistemas que requieren tratamiento de textos, buscadores semánticos y el tratamiento de la multilingüalidad

RA83 - RA-IA-16 Conocer las áreas de mercado que requieren sistemas de procesamiento de lenguaje natural y herramientas específicas, o la creación automática o semi-automática de recursos lingüísticos

## 4. Descripción de la asignatura y temario

---

### 4.1. Descripción de la asignatura

En este curso de introducción al Procesamiento del Lenguaje Natural se aprenderán los fundamentos de las técnicas básicas para realizar análisis morfológicos y sintácticos de forma automática, así como tratamiento de textos en distintos idiomas (como español e inglés) y aspectos básicos de la semántica computacional. También se estudiarán diversos corpus, cómo extraer y recuperar información, y sus aplicaciones a sistemas pregunta-respuesta, herramientas para creación de recursos lingüísticos de forma automática o semi-automática.

Como entorno de programación se usará RStudio (que permite el uso de los lenguajes R y Python). Como lenguaje de programación preferente se usará R.

## 4.2. Temario de la asignatura

1. Unidad 1. Introducción
  - 1.1. Tema 1.1. Introducción al procesamiento del lenguaje natural
2. Unidad 2. Recursos lingüísticos
  - 2.1. Tema 2.1. Corpus digitales
  - 2.2. Tema 2.2. Anotación de corpus
  - 2.3. Tema 2.3. Visualización
3. Unidad 3. Tratamiento de textos
  - 3.1. Tema 3.1. Expresiones regulares
  - 3.2. Tema 3.2. Normalización de textos
  - 3.3. Tema 3.3. Distancia entre textos
4. Unidad 4. Minería de textos
  - 4.1. Tema 4.1. Análisis de sentimientos
  - 4.2. Tema 4.2. Clasificación de textos
  - 4.3. Tema 4.3. Creación de resúmenes
5. Unidad 5. Recuperación y extracción de información
  - 5.1. Tema 5.1. Introducción a la recuperación y extracción
  - 5.2. Tema 5.2. Extracción de relaciones
  - 5.3. Tema 5.3. Extracción de eventos y fechas
  - 5.4. Tema 5.4. Lenguajes para consultas

## 5. Cronograma

### 5.1. Cronograma de la asignatura \*

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<p><b>Explicación de los contenidos teóricos de la Unidad 1</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 1</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
2	<p><b>Explicación de los contenidos teóricos de la Unidad 2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 2</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
3	<p><b>Explicación de los contenidos teóricos de la Unidad 2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 2</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
4	<p><b>Explicación de los contenidos teóricos de la Unidad 2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 2</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
5	<p><b>Explicación de los contenidos teóricos de la Unidad 3</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 3</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
6	<p><b>Explicación de los contenidos teóricos de la Unidad 3</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 3</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			

7	<p><b>Explicación de los contenidos teóricos de la Unidad 3</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 3</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
8				<p><b>Examen unidades 1-3</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 02:00</p>
9	<p><b>Explicación de los contenidos teóricos de la Unidad 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
10	<p><b>Explicación de los contenidos teóricos de la Unidad 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
11	<p><b>Explicación de los contenidos teóricos de la Unidad 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
12	<p><b>Explicación de los contenidos teóricos de la Unidad 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
13	<p><b>Explicación de los contenidos teóricos de la Unidad 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
14	<p><b>Explicación de los contenidos teóricos de la Unidad 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			

15	<p><b>Explicación de los contenidos teóricos de la Unidad 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Resolución de ejercicios de la Unidad 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			
16				<p><b>Entrega de la práctica</b> EP: Técnica del tipo Examen de Prácticas Evaluación continua Presencial Duración: 00:00</p>
17				<p><b>Examen prueba global.</b> EX: Técnica del tipo Examen Escrito Evaluación sólo prueba final No presencial Duración: 02:00</p> <p><b>Examen de las Unidades 4 y 5</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 02:00</p> <p><b>Examen de recuperación unidades 1-3</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 02:00</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.

## 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
8	Examen unidades 1-3	EX: Técnica del tipo Examen Escrito	Presencial	02:00	50%	3 / 10	CB03 CB05 CE17
16	Entrega de la práctica	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	%	5 / 10	CG02 CG01
17	Examen de las Unidades 4 y 5	EX: Técnica del tipo Examen Escrito	Presencial	02:00	50%	3 / 10	CB03 CB05 CE17
17	Examen de recuperación unidades 1-3	EX: Técnica del tipo Examen Escrito	Presencial	02:00	0%	3 / 10	CB03 CB05 CE17

#### 6.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Examen prueba global.	EX: Técnica del tipo Examen Escrito	No Presencial	02:00	100%	5 / 10	CG02 CB03 CB05 CG01 CE17

#### 6.1.3. Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.

## 6.2. Criterios de evaluación

Los criterios de evaluación en la Evaluación Progresiva son:

1. La evaluación consta de dos partes, una a través de la realización de dos exámenes y otra a través de la realización de una práctica.
2. Para aprobar es necesario que en cada examen se tenga como mínimo un 3, la media sea 5 o superior, y la práctica tenerla APTA.
3. En todos los exámenes de la semana 17 aparecerán preguntas sobre la práctica.
4. Si en el examen de las unidades 1-3 de la evaluación la nota es superior o igual a 3, pero inferior a 5, el alumno puede optar a presentarse al examen sobre las unidades 3-4 o al examen de recuperación de la semana 17. Si es inferior a 3 tendrá que presentarse obligatoriamente al examen de recuperación de la semana 17.
5. No se guarda ninguna nota de una convocatoria para otra.

Los criterios de evaluación en la Evaluación por Prueba Global son:

1. La evaluación constará de un examen, en el cual debe sacar un 5 o más para aprobar, y de una práctica, la cual debe estar calificada como APTA para aprobar. Si la nota del examen fuera inferior a un 5 o la práctica fuera NO APTA, la nota final en la asignatura sería suspenso.
2. La nota final en la asignatura sería la obtenida en el examen, siempre y cuando la calificación de la práctica sea APTA.
3. En el examen habrá preguntas sobre la práctica.

## 7. Recursos didácticos

### 7.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Jurafsky y Martin. Speech and Language Processing (El procesamiento del lenguaje y el discurso)	Bibliografía	En inglés. Borrador de la tercera edición (en inglés): <a href="https://web.stanford.edu/~jurafsky/slp3/">https://web.stanford.edu/~jurafsky/slp3/</a> Hay libro de la segunda edición (un clásico de 2008).
Desagulier. Corpus Linguistics and Statistics with R	Bibliografía	En inglés.
Allen. Natural Language Understanding (Comprensión del lenguaje natural)	Bibliografía	En inglés. Otro clásico (de 1994).
Indurkha y Damerau. Handbook of Natural Language Processing (Manual sobre procesamiento del lenguaje natural)	Bibliografía	En inglés.
Gelbukh y Sidorov. Procesamiento automático del español	Bibliografía	En ESPAÑOL (2010). Con pdf gratis en <a href="https://www.gelbukh.com/libro-procesamiento-2/">https://www.gelbukh.com/libro-procesamiento-2/</a>
Text mining in R for the social sciences and digital humanities	Bibliografía	Libro online de 2020 de Andreas Niekler y Gregor Wiedemann <a href="https://tm4ss.github.io/docs/index.html">https://tm4ss.github.io/docs/index.html</a>
An Introduction to Text Processing and Analysis with R	Bibliografía	libro online de 2018 en Michael Clark <a href="https://m-clark.github.io/text-analysis-with-R/">https://m-clark.github.io/text-analysis-with-R/</a>
RStudio y/o RStudio Connect	Otros	Servidor para RStudio online y servidor para publicación (entrega) de trabajos en R o Rmd.

## 8. Otra información

---

### 8.1. Otra información sobre la asignatura