

**FICHA IDENTIFICATIVA****Datos de la Asignatura**

Código	46578
Nombre	Procesado del lenguaje natural
Ciclo	Máster
Créditos ECTS	6.0
Curso académico	2023 - 2024

Titulación(es)

Titulación	Centro	Curso	Periodo
2262 - M.U. en Ciencia de Datos	Escuela Técnica Superior de Ingeniería	1	Segundo cuatrimestre

Materias

Titulación	Materia	Caracter
2262 - M.U. en Ciencia de Datos	10 - Procesado del lenguaje natural	Obligatoria

Coordinación

Nombre	Departamento
VIVES GILABERT, YOLANDA	242 - Ingeniería Electrónica

RESUMEN

Actualmente, gran parte de los datos disponibles para el análisis están formados por información no estructurada en forma de textos en lenguaje natural. Entre esta información se encuentran páginas web (Wikipedia, periódicos digitales, blogs) o redes sociales (Facebook, Twitter). Poder analizar estos textos, mediante algoritmos de Procesamiento de Lenguaje Natural (PLN), resulta muy útil para que las organizaciones puedan tomar mejores decisiones.

Los algoritmos de aprendizaje automático no son capaces de entender texto o caracteres, por lo que el PLN realiza todo el preprocesamiento necesario para convertir estos datos en forma de texto en un formato comprensible para las máquinas (números) y así poder realizar todo tipo de análisis posterior para obtener la información relevante. Entre las aplicaciones más comunes del PLN se encuentran la clasificación de textos, búsqueda y extracción de información, resumen de datos, traducción automática o sistemas de respuesta automática, entre otros.



Se trata de una asignatura teórico-práctica, donde habrá sesiones de teoría, sesiones prácticas y sesiones mixtas. Las clases de teoría se impartirán en castellano.

CONOCIMIENTOS PREVIOS

Relación con otras asignaturas de la misma titulación

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

Otros tipos de requisitos

Se recomienda haber superado la asignatura de Aprendizaje Máquina (primer cuatrimestre).

COMPETENCIAS

2262 - M.U. en Ciencia de Datos

- Que los/las estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- Que los/las estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo
- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- Ser capaces de valorar la necesidad de completar su formación técnica, científica, en lenguas, en informática, en literatura, en ética, social y humana en general, y de organizar su propio autoaprendizaje con un alto grado de autonomía
- Capacidad para trabajar en equipo para llegar a soluciones de problemas interdisciplinarios usando técnicas de análisis de datos.
- Ser capaces de acceder a herramientas de información (bibliográficas y de empleo) y utilizarlas apropiadamente.
- Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio, aplicando los conocimientos adquiridos en la identificación de salidas profesionales y yacimientos de empleo.
- Extraer conocimiento de conjuntos de datos en diferentes formatos.
- Modelar la dependencia entre una variable respuesta y varias variables explicativas, en conjuntos de datos complejos, mediante técnicas de aprendizaje máquina, interpretando los resultados obtenidos.
- Usar las técnicas de procesado de lenguaje natural para analizar textos extrayendo conocimiento útil de ellos.



RESULTADOS DE APRENDIZAJE

Aprender las diferentes técnicas de extracción y procesado de la información sobre datos no estructurados de texto.

Saber transformar texto en vectores numéricos (embeddings) para su uso e algoritmos de ML. Conocer y saber aplicar modelos de DL sobre texto, especialmente los modelos basados en atención (transformers).

Conocer las diferentes aplicaciones del NLP en clasificación, extracción de conocimiento y generación de texto.

DESCRIPCIÓN DE CONTENIDOS

1. Introducción al Procesamiento del Lenguaje Natural

- 1.1. Qué es el PLN
- 1.2. La importancia del texto
- 1.3. Aproximaciones históricas al PLN
- 1.4. Aplicaciones y flujo de trabajo

2. Uso de texto en Python

- 2.1. Cadenas de texto en Python
- 2.2. Expresiones regulares
- 2.3. Carga de texto
- 2.4. Captura de contenido web (web scraping)
- 2.5. Llibrerías PLN en Python

3. Preprocesamiento de texto

- 3.1. División de texto
- 3.2. Limpieza y normalización de texto
- 3.3. Análisis morfológico
- 3.4. Análisis semántico
- 3.5. Análisis gramatical

4. Extracción de características

- 4.1. Características simples
- 4.2. Modelo Bag of Words
- 4.3. Modelo TF-IDF
- 4.4. Vectores de palabra (word embeddings)
- 4.5. Vectores de document



5. Aplicaciones de PLN

- 5.1. Clasificación
- 5.2. Extracción de la información
- 5.3. Minería de texto
- 5.4. Búsqueda de información
- 5.5. Modelos secuenciales

6. Aprendizaje profundo en PLN

- 6.1 Redes neuronales recurrentes en PLN
- 6.2 Atención en aprendizaje profundo
- 6.3 Transformers
- 6.4 Aplicaciones

VOLUMEN DE TRABAJO

ACTIVIDAD	Horas	% Presencial
Prácticas en laboratorio	28,00	100
Clases de teoría	28,00	100
Clases teórico-prácticas	4,00	100
Asistencia a eventos y actividades externas	2,00	0
Elaboración de trabajos en grupo	15,00	0
Elaboración de trabajos individuales	3,00	0
Estudio y trabajo autónomo	15,00	0
Lecturas de material complementario	5,00	0
Preparación de actividades de evaluación	10,00	0
Preparación de clases de teoría	10,00	0
Preparación de clases prácticas y de problemas	10,00	0
Resolución de casos prácticos	15,00	0
Resolución de cuestionarios on-line	5,00	0
TOTAL	150,00	

METODOLOGÍA DOCENTE

Actividades teóricas.



- Desarrollo expositivo de la materia con la participación del estudiante en la resolución de cuestiones puntuales.
- Realización de cuestionarios individuales de evaluación.

Actividades prácticas.

Aprendizaje mediante resolución de problemas, ejercicios y casos de estudio a través de los cuales se adquieren competencias sobre los diferentes aspectos de la materia.

Trabajos en laboratorio y/o aula ordenador.

Aprendizaje mediante la realización de actividades desarrolladas de forma individual o en grupos reducidos y llevadas a cabo en aulas de ordenador.

EVALUACIÓN

Prueba objetiva, consistente en uno o varios exámenes que constarán tanto de cuestiones teórico-prácticas como de problemas (entre el 20% y el 80%).

Evaluación de las actividades prácticas a partir de la elaboración de trabajos/memorias, exposiciones orales y herramientas de e-learning de la Universitat (entre 20% y el 80%).

Evaluación basada en la participación y grado de implicación del alumno en el proceso de enseñanza-aprendizaje, teniendo en cuenta la asistencia regular a las actividades presenciales previstas y la resolución de cuestiones y problemas propuestos periódicamente. (entre el 0 y el 20%)

REFERENCIAS

Básicas

- Sohom Ghosh, Dwight Gunning. Natural Language Processing Fundamentals. Packt Publishing, 2019
- Akshay Kulkarni, Adarsha Shivananda. Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python. Apress, 2019 (disponible e-libro)
- Dipanjan Sarkar. Text Analytics with Python: A Practitioner's Guide to Natural Language Processing. Apress 2019 (disponible e-libro)
- Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python. O'Really Media, 2009

Complementarias



- Jacob Eisentein. Natural Language Processing. 2018 (disponible bajo licencia CC-BY-NC-ND)
- Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana. Practical Natural Language Processing. O'Really Media, 2020

