Facultad de Ciencias

GUÍA DOCENTE DE LA ASIGNATURA

G687 - Natural Language Processing

Grado en Ingeniería Informática
Optativa. Curso 4

Curso Académico 2023-2024

1. DATOS IDENTIFICATIVOS

Título/s	Grado en Ingeniería Informática	Tipología y Curso	Optativa. Curso 4
Centro	Facultad de Ciencias		
Módulo / materia	MATERIA DE COMPUTACIÓN MENCIÓN EN COMPUTACIÓN		
Código y denominación	G687 - Natural Language Processing		
Créditos ECTS	6	Cuatrimestre	Cuatrimestral (2)
Web			
Idioma de impartición	Inglés	Forma de impartición	Presencial

Departamento	DPTO. MATEMATICAS, ESTADISTICA Y COMPUTACION
Profesor responsable	CRISTINA TIRNAUCA
E-mail	cristina.tirnauca@unican.es
Número despacho	Facultad de Ciencias. Planta: + 1. DESPACHO CRISTINA TIRNAUCA (1046)
Otros profesores	

2. CONOCIMIENTOS PREVIOS

Se requieren unos conocimientos básicos de cálculo de probabilidades y cálculo vectorial (tal como se adquiere en las asignaturas iniciales de la materias 'Análisis Matemático y Métodos Numéricos' y 'Álgebra Lineal y Discreta'), y habilidades en la utilización de algún lenguaje de programación imperativa (C, Java, Python). También se recomienda haber cursado las asignaturas de 'Lenguajes Formales' y 'Aprendizaje Automático y Minería de Datos'.

3. COMPETENCIAS GENÉRICAS Y ESPECÍFICAS DEL PLAN DE ESTUDIOS TRABAJADAS
Competencias Genéricas
(Conocimiento) 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.
(Aplicación) 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.
(Análisis) Reunir e interpretar datos relevantes (dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
(Comunicación) Transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
(Aprendizaje) Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
Capacidad de análisis, síntesis y evaluación.
Capacidad de organización y planificación.
Capacidad de gestión de la información.
Capacidad de resolución de problemas aplicando técnicas de ingeniería.
Capacidad para argumentar y justificar lógicamente las decisiones tomadas y las opiniones.
Capacidad de trabajo en equipo.
Razonamiento crítico.
Aprendizaje autónomo.
Adaptación a nuevas situaciones.
Creatividad.
Capacidad de comprensión auditiva, lectura, interacción y expresión oral y escrita en Inglés
Competencias Específicas
Capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable para la resolución de problemas mediante un sistema informático en cualquier ámbito de aplicación, particularmente los relacionados con aspectos de computación, percepción y actuación en ambientes o entornos inteligentes.

3.1 RESULTADOS DE APRENDIZAJE
- Saber desarrollar los mecanismos adicionales de proceso de lenguaje necesarios para la construcción de sistemas inteligentes de interacción en lenguas humanas, incluyendo el análisis morfológico, el etiquetado “part-of-speech”, la desambiguación semántica, la extracción de información a partir de textos en lenguas humanas y los rudimentos de los sistemas inteligentes de traducción.

4. OBJETIVOS
Diferenciar entre lenguajes naturales y lenguajes artificiales.
Comprender la complejidad del lenguaje humano.
Familiarizarse con la terminología lingüística.
Conocer los algoritmos, técnicas y métodos más utilizados actualmente para el tratamiento automático de los fenómenos léxicos, sintácticos y semánticos del lenguaje humano.
Asumir que el procesamiento del lenguaje natural no se puede automatizar completamente, pero que se pueden desarrollar soluciones satisfactorias en la práctica.
Identificar los aspectos del lenguaje sobre los que debemos trabajar más para obtener sistemas de procesamiento del lenguaje natural útiles.
Reflexionar sobre los avances realizados en el campo y los errores cometidos a lo largo de las últimas décadas.

5. MODALIDADES ORGANIZATIVAS Y MÉTODOS DOCENTES	
ACTIVIDADES	HORAS DE LA ASIGNATURA
ACTIVIDADES PRESENCIALES	
HORAS DE CLASE (A)	
- Teoría (TE)	21
- Prácticas en Aula (PA)	15
- Prácticas de Laboratorio Experimental(PLE)	
- Prácticas de Laboratorio en Ordenador (PLO)	24
- Prácticas Clínicas (CL)	
Subtotal horas de clase	60
ACTIVIDADES DE SEGUIMIENTO (B)	
- Tutorías (TU)	7,5
- Evaluación (EV)	7,5
Subtotal actividades de seguimiento	15
Total actividades presenciales (A+B)	75
ACTIVIDADES NO PRESENCIALES	
Trabajo en grupo (TG)	30
Trabajo autónomo (TA)	45
Tutorías No Presenciales (TU-NP)	
Evaluación No Presencial (EV-NP)	
Total actividades no presenciales	75
HORAS TOTALES	150

6. ORGANIZACIÓN DOCENTE

CONTENIDOS		TE	PA	PLE	PLO	CL	TU	EV	TG	TA	TU-NP	EV-NP	Semana
1	Clases de teoría, problemas o laboratorio	0,00	0,00	0,00	0,00	0,00	7,50	0,00	0,00	0,00	0,00	0,00	Todas
1.1	Presentación de la asignatura. Introducción al procesamiento de lenguaje natural.	4,00	2,00	0,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1-2
1.2	Procesamiento básico de texto.	1,00	0,00	0,00	2,00	0,00	0,00	0,00	5,00	0,00	0,00	0,00	3
1.3	Modelos lingüísticos y correctores ortográficos.	3,00	2,00	0,00	6,00	0,00	0,00	0,00	5,00	0,00	0,00	0,00	3-6
1.4	Clasificación de textos y análisis de sentimientos.	1,00	1,00	0,00	2,00	0,00	0,00	0,00	5,00	0,00	0,00	0,00	6-7
1.5	Recuperación de la información.	2,00	1,00	0,00	2,00	0,00	0,00	0,00	5,00	0,00	0,00	0,00	7-8
1.6	Análisis léxico.	1,00	0,00	0,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	8-9
1.7	Análisis sintáctico.	5,00	3,00	0,00	8,00	0,00	0,00	0,00	10,00	0,00	0,00	0,00	9-13
1.8	Análisis semántico.	4,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	13-14
1.9	Repaso	0,00	4,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	15
2	Proyecto final	0,00	0,00	0,00	0,00	0,00	0,00	4,50	0,00	45,00	0,00	0,00	Todas
3	Examen final	0,00	0,00	0,00	0,00	0,00	0,00	3,00	0,00	0,00	0,00	0,00	16
TOTAL DE HORAS		21,00	15,00	0,00	24,00	0,00	7,50	7,50	30,00	45,00	0,00	0,00	

Esta organización tiene carácter orientativo.

TE	Horas de teoría
PA	Horas de prácticas en aula
PLE	Horas de prácticas de laboratorio experimental
PLO	Horas de prácticas de laboratorio en ordenador
CL	Horas de prácticas clínicas
TU	Horas de tutoría
EV	Horas de evaluación
TG	Horas de trabajo en grupo
TA	Horas de trabajo autónomo
TU-NP	Tutorías No Presenciales
EV-NP	Evaluación No Presencial

7. MÉTODOS DE LA EVALUACIÓN

Descripción	Tipología	Eval. Final	Recuper.	%
Examen final	Examen escrito	Sí	Sí	50,00

Calif. mínima	4,00
Duración	3 horas
Fecha realización	En las fechas indicadas por la Facultad para la realización de exámenes finales
Condiciones recuperación	
Observaciones	Se realizará una prueba escrita en el periodo oficial de exámenes. No se podrán llevar apuntes o libros. Si un estudiante no tuviese la calificación mínima requerida para la superación de esta prueba, la calificación global de la asignatura será el menor valor entre 4,9 y la media ponderada de todas las pruebas de evaluación.

Proyecto final	Trabajo	No	Sí	20,00

Calif. mínima	0,00
Duración	4,5 horas
Fecha realización	durante el cuatrimestre
Condiciones recuperación	
Observaciones	El alumno entregará un informe sobre un tema relacionado con la asignatura. El ingrediente básico es la iniciativa personal. Al final del cuatrimestre, los alumnos tendrán que defender oralmente su trabajo.

Prácticas en grupo	Trabajo	No	Sí	20,00

Calif. mínima	0,00
Duración	14 semanas
Fecha realización	durante el cuatrimestre
Condiciones recuperación	
Observaciones	El enunciado de esas prácticas es el mismo para todos (se trata de implementar varios algoritmos de procesamiento de lenguaje natural en grupos de 2 o 3 personas). Los ficheros de las prácticas se tienen que entregar a través de la plataforma Moodle. Todas las prácticas tienen el mismo peso.

Problemas	Actividad de evaluación con soporte virtual	No	Sí	10,00

Calif. mínima	0,00
Duración	14 semanas
Fecha realización	durante el cuatrimestre
Condiciones recuperación	
Observaciones	Se realizará una evaluación de problemas y cuestiones, con una periodicidad semanal, mediante cuestionarios informatizados. En ocasiones, también habrá cuestionarios informatizados acompañando las sesiones prácticas que tendrán que responderse durante el horario de clase.

TOTAL	100,00
Observaciones	

Cuando un estudiante que no se presenta al examen final en el periodo ordinario de exámenes no haya realizado actividades de evaluación cuyo peso supere el 50% de la calificación de la asignatura, figurará en su acta como no presentado. Cuando el estudiante haya realizado pruebas que supongan el referido 50% o más, en el acta figurará la calificación correspondiente. En el periodo extraordinario, un estudiante que no se presenta al examen final figuraría como no presentado en cualquiera de los dos casos mencionados anteriormente.

El estudiante tendrá derecho a realizar un examen en la convocatoria extraordinaria con un valor del 100% de la calificación total de las actividades recuperables de la asignatura (es decir, el peso del examen final en la convocatoria extraordinaria es del 100% de la calificación final de la asignatura). Dicho examen se desarrollará de la misma manera que en el periodo ordinario (presencial si las condiciones sanitarias lo permiten y a través de Moodle en caso contrario).

Cualquier alumno que disponga o se valga de medios ilícitos en la celebración de un examen, o que se atribuya indebidamente la autoría de trabajos académicos requeridos para la evaluación, tendrá la calificación de 'suspenso' o de '0', según se trate de calificaciones literales o numéricas, respectivamente. Cuando se dé esta circunstancia, el profesor podrá elevar un informe al Centro que, en el plazo máximo de dos meses, y previa audiencia al alumno, procederá a decidir sobre la eventual inclusión de este hecho en el expediente del alumno.

Criterios de evaluación para estudiantes a tiempo parcial

Para los alumnos que por motivo justificado (estudiantes a tiempo parcial) no hayan seguido la evaluación continua, el examen tiene un peso de 100% (tanto en periodo ordinario como en el extraordinario).

8. BIBLIOGRAFÍA Y MATERIALES DIDÁCTICOS

BÁSICA

Speech and Language Processing (2nd edition). Daniel Jurafsky, James H. Martin. 2014

Complementaria

Natural Language Processing. H.M. Noble. Blackwell Scientific Publications. 1988

Introduction to Information Retrieval. Christopher D. Manning. 2008

Natural Language Understanding (2nd edition). James Allen. 1995

Foundations of Statistical Natural Language Processing. Christopher D. Manning, Hinrich Schütze. MIT Press. Cambridge, MA.1999.

Natural Language Processing with Python. Steven Bird, Ewan Klein, Edward Loper. 2009

Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition (2nd edition). Daniel Jurafsky, James H. Martin. 2009

9. SOFTWARE

PROGRAMA / APLICACIÓN	CENTRO	PLANTA	SALA	HORARIO
Python				

10. COMPETENCIAS LINGÜÍSTICAS

- ☑ Comprensión escrita
- ☑ Comprensión oral
- ☑ Expresión escrita
- ☑ Expresión oral
- ☑ Asignatura íntegramente desarrollada en inglés

Observaciones