24-25

GUÍA DE ESTUDIO PÚBLICA

INTRODUCCIÓN A LA INVESTIGACIÓN EN PROCESAMIENTO DEL LENGUAJE NATURAL

CÓDIGO 3107003-

UNED

24-25

INTRODUCCIÓN A LA INVESTIGACIÓN EN PROCESAMIENTO DEL LENGUAJE NATURAL

CÓDIGO 3107003-

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA
EQUIPO DOCENTE
HORARIO DE ATENCIÓN AL ESTUDIANTE
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE
RESULTADOS DE APRENDIZAJE
CONTENIDOS
METODOLOGÍA
SISTEMA DE EVALUACIÓN
BIBLIOGRAFÍA BÁSICA
BIBLIOGRAFÍA COMPLEMENTARIA
RECURSOS DE APOYO Y WEBGRAFÍA
IGUALDAD DE GÉNERO

Nombre de la asignatura	INTRODUCCIÓN A LA INVESTIGACIÓN EN PROCESAMIENTO DEL LENGUAJE NATURAL
Código	3107003-
Curso académico	2024/2025
Título en que se imparte	MÁSTER UNIVERSITARIO EN TECNOLOGÍAS DEL LENGUAJE
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150
Periodo	ANUAL
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

El objetivo principal de la asignatura es que el estudiante se familiarice con la metodología de investigación en Procesamiento del Lenguaje Natural, de forma que pueda plantear, planificar, desarrollar y comunicar un trabajo de investigación de forma autónoma y efectiva. Es una asignatura optativa de carácter metodológico recomendada para todos aquellos estudiantes que se planteen una carrera investigadora o de innovación relacionada con el Procesamiento del Lenguaje Natural o la Inteligencia Artificial en general, ya sea en la industria o en la academia.

EQUIPO DOCENTE

GONZALO ARRAYO, JULIO:

Julio Gonzalo es director del centro de investigación en Procesamiento del Lenguaje Natural y Recuperación de Información de la UNED (nlp.uned.es). Ha investigado en muchos aspectos de la Inteligencia Artificial relacionados con el lenguaje. Algunos de los temas más recientes son: la creatividad de los modelos de lenguaje, el análisis automático de las técnicas de propaganda política en las estrategias narrativas de las potencias mundiales y las métricas de evaluación para aplicaciones de Inteligencia Artificial. Ha sido co-chair de iniciativas internacionales de evaluación competitiva de sistemas inteligentes como CLEF, IberEval e IberLEF.

CENTENO SÁNCHEZ, ROBERTO:

Roberto Centeno forma parte del grupo NLP&IR de la UNED. Su actividad investigadora se centra en la inteligencia artificial distribuida y la detección de noticias falsas y opiniones en las redes sociales. En particular, mis principales intereses se centran en la regulación de Sistemas Multiagente Abiertos y el estudio de aspectos sociales como la confianza y la reputación y su aplicación tanto en redes sociales como en teorías de eLearning.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Conocimientos previos recomendables:
- Conocimientos básicos de matemáticas, estadística y programación.
- Conocimientos básicos de aprendizaje automático
- Lectura fluida en inglés

Esta asignatura puede ser cursada de forma aislada del resto de las asignaturas del máster. Se recomienda para cualquier estudiante que tenga intención de desarrollar una carrera investigadora en la academia o en la industria.

EQUIPO DOCENTE

Nombre y Apellidos	JULIO ANTONIO GONZALO ARROYO (Coordinador/a de asignatura)
Correo Electrónico	julio@lsi.uned.es
Teléfono	
Facultad	
Departamento	

Nombre y Apellidos	ROBERTO CENTENO SANCHEZ
Correo Electrónico	rcenteno@lsi.uned.es
Teléfono	
Facultad	
Departamento	

HORARIO DE ATENCIÓN AL ESTUDIANTE

Se realizará mediante la plataforma de posgrados de la UNED.

•Julio Gonzalo Arroyo
email: julio@lsi.uned.es
Tfno: 913987922
Horario: Jueves de 16:00 a 20:00 horas

•Roberto Centeno Sánchez
email: rcenteno@lsi.uned.es
Tfno: 913989696
Horario: Jueves de 11:00 a 13:00 horas y de 15:00 a 17:00 horas

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

COMPETENCIAS

C1 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.
C2 Abstracción, análisis, síntesis y relación de ideas.
C3 Capacidad crítica y de decisión.
C4 Capacidad de estudio y autoaprendizaje
C5 Capacidad creativa y de investigación.
C7 Capacidad de estudio de los sistemas y aproximaciones existentes y para distinguir las aproximaciones más efectivas.
C8 Capacidad para detectar carencias en el estado actual de la ciencia y la tecnología.
C9 Capacidad para proponer nuevas aproximaciones que de solución a las carencias detectadas.

RESULTADOS DE APRENDIZAJE

CONOCIMIENTOS O CONTENIDOS

CO1 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.

HABILIDADES O DESTREZAS

H1 Capacidad de aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios relacionados con su área de estudio.
H2 Capacidad de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
H3 Poseer las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
H4 Capacidad de especificar, diseñar, implementar y evaluar tanto cualitativa como cuantitativamente los modelos y sistemas propuestos.
H5 Capacidad para proponer y llevar a cabo experimentos con la metodología adecuada como para poder extraer conclusiones y determinar nuevas líneas de actuación e investigación.

COMPETENCIAS

C1 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.
C2 Abstracción, análisis, síntesis y relación de ideas.
C3 Capacidad crítica y de decisión.
C4 Capacidad de estudio y autoaprendizaje.
C5 Capacidad creativa y de investigación.

C7 Capacidad de estudio de los sistemas y aproximaciones existentes y para distinguir las aproximaciones más efectivas.
C8 Capacidad para detectar carencias en el estado actual de la ciencia y la tecnología.
C9 Capacidad para proponer nuevas aproximaciones que de solución a las carencias detectadas.

CONTENIDOS

Tema 1: Metodología de investigación en Procesamiento del Lenguaje Natural: introducción.

Introducción y motivación a la asignatura.

Tema 2: Planteamiento de problemas de investigación

En este tema veremos cómo elegir y plantear un tema de investigación en el campo del Procesamiento del Lenguaje Natural.
- elección del tema: interés científico, interés aplicado, viabilidad.
- Escenarios y tareas
- cuestiones de investigación ("research questions")
- objetivos
- determinación del estado del arte

Tema 3: Abstracción del problema y métricas de evaluación.

En este tema discutiremos cómo abstraer un problema de interés para desarrollar una investigación con metodología científica. Haremos especial énfasis en la determinación adecuada de las métricas de evaluación adecuadas a partir de un problema y un escenario de aplicación concretos, como punto de partida para una abstracción adecuada.

Tema 4: Diseño experimental

En este tema estudiaremos a fondo el diseño experimental conocido como A/B testing, en el que se comparan un sistema de referencia (normalmente el estado del arte) y un sistema contrastivo (que implementa nuevas soluciones), para establecer si las mejoras propuestas producen una ventaja significativa sobre el estado del arte. También veremos otros mecanismos de comparación habituales como los experimentos de ablación (en los que se van cambiando componentes de un sistema complejo para establecer cómo influyen las partes en el resultado global de un sistema) y los leaderboard (aplicaciones web donde se

compara en tiempo real el estado del arte respecto a un dataset específico).

Tema 5: Recolección y anotación de datasets

En este tema estudiaremos los mecanismos más habituales para recolectar datos anotados que nos permitan evaluar sistemas de Procesamiento del Lenguaje Natural. Trataremos:
- Anotación de expertos vs crowdsourcing
- Identificación y prevención de sesgos
- Tratamiento del desacuerdo entre anotadores

Tema 6: Análisis de resultados

En este tema estudiaremos las herramientas analíticas más habituales para extraer conclusiones de los resultados de nuestra experimentación. Hablaremos de análisis cuantitativo y cualitativo, uso de múltiples métricas, tests de relevancia estadística, análisis por casos, identificación de alcance de los resultados y análisis de limitaciones.

Tema 7: Comunicación científica

En este tema veremos cómo preparar artículos de investigación en Procesamiento del Lenguaje Natural. Estudiaremos los procesos de peer review, las revistas académicas y congresos científicos más relevantes del área, y medidas de impacto científico. También discutiremos la publicación de código, datasets y otros recursos; los aspectos de reproducibilidad, y los repositorios de investigación. Finalmente, estudiaremos pautas para la presentación oral de resultados de investigación en ámbitos académicos y en eventos de divulgación.

METODOLOGÍA

La asignatura no tiene clases presenciales. Los contenidos teóricos se imparten a distancia y de acuerdo con las normas y estructuras de soporte telemático de la enseñanza en la UNED. El material docente incluye un resumen de los contenidos de cada tema y distintos tipos de actividades relacionadas con la consulta bibliográfica, la realización de ejercicios prácticos y, en ocasiones, la utilización de herramientas asociadas a las tecnologías y técnicas presentadas en el curso. Junto a las actividades y enlaces con fuentes de información externas, existe material didáctico propio preparado por el equipo docente con ejercicios que organizan las actividades prácticas del alumno.
Tratándose de un máster orientado a la investigación, las actividades de aprendizaje se estructuran alrededor del estado del arte en cada uno de los temas del curso y a la parte práctica que el alumno deberá realizar en cada sección. Se utilizarán los recursos de la UNED para las actividades complementarias que se propondrán a lo largo del curso, actividades de asistencia o visualización (on-line, off-line) de seminarios de temas avanzados, impartidos por investigadores de reconocido prestigio.
El plan de trabajo marcará unas pautas adecuadas para que el alumno alcance los objetivos al final del curso. Incluye una descripción ordenada y secuenciada de los contenidos de cada tema y orientaciones sobre las actividades a realizar, tanto de estudio teórico como práctico. Los ejercicios prácticos correspondientes a cada tema son para apoyo y refuerzo del aprendizaje, y su realización es obligatoria para superar la asignatura.

SISTEMA DE EVALUACIÓN

TIPO DE PRIMERA PRUEBA PRESENCIAL

Tipo de examen: No hay prueba presencial

TIPO DE SEGUNDA PRUEBA PRESENCIAL

Tipo de examen2: No hay prueba presencial

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad: No

Descripción

El trabajo por el que se evalúa la asignatura son los ejercicios prácticos propuestos en cada tema (que se discuten en el apartado de PECs), modulados en su caso por una entrevista personal con el estudiante por videoconferencia.

Criterios de evaluación

Los ejercicios prácticos se evalúan en función de la relevancia y corrección de las respuestas, de su completitud, de las aportaciones originales del estudiante y de su interacción con el equipo docente y el resto de los estudiantes donde sea apropiado. Además, la nota puede estar modulada por la discusión de las respuestas en una entrevista personal con el estudiante mediante videoconferencia.

Ponderación de la prueba presencial y/o los trabajos en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC? Si,PEC no presencial

Descripción

Para cada tema se proponen ejercicios prácticos.

Criterios de evaluación

La evaluación de los ejercicios prácticos dependerá de la relevancia y corrección de las respuestas, de su completitud, de las aportaciones originales del estudiante y de su interacción con el equipo docente y el resto de los estudiantes donde sea apropiado.

Ponderación de la PEC en la nota final: La resolución de los ejercicios prácticos supondrá el 100% de la calificación de la asignatura. En los casos en los que el equipo docente lo estime oportuno se complementará con una entrevista sobre el trabajo realizado, por medios telemáticos.

Fecha aproximada de entrega

Comentarios y observaciones

En los casos en los que el equipo docente lo estime oportuno, las pruebas de evaluación continua se complementarán con una entrevista sobre el trabajo realizado, por medios telemáticos.

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? Si,no presencial

Descripción

En los casos en los que el equipo docente lo estime oportuno, las actividades prácticas realizadas por el estudiante se complementarán con una entrevista con el equipo docente mediante videoconferencia.

Criterios de evaluación

En la entrevista se comprobarán los conocimientos adquiridos y se discutirán aspectos de los ejercicios prácticos entregados por el estudiante.

Ponderación en la nota final: la entrevista no tiene un porcentaje de calificación en la nota final, pero puede influir en la calificación de cada una de las PEC.

Fecha aproximada de entrega

Comentarios y observaciones

En caso de que el equipo docente requiera la realización de una entrevista, los resultados pueden influir en la calificación de cada una de las PEC de la asignatura.

¿CÓMO SE OBTIENE LA NOTA FINAL?

Hay dos situaciones posibles:

- En los casos en los que no se realice una entrevista personal, la calificación será la media ponderada de las calificaciones obtenidas en cada una de las PEC. La ponderación de cada tarea se comunicará a los estudiantes al inicio del curso.

- En los casos en los que se realice una entrevista personal, la entrevista podrá modificar (tanto al alza como a la baja) las calificaciones obtenidas en cada una de las PEC. Una vez realizadas las modificaciones oportunas, la calificación final, como en el caso anterior, será la media ponderada de las calificaciones consolidadas obtenidas en cada una de las PEC.

BIBLIOGRAFÍA BÁSICA

Como bibliografía se aportarán referencias a trabajos de investigación dentro del curso virtual.

BIBLIOGRAFÍA COMPLEMENTARIA

RECURSOS DE APOYO Y WEBGRAFÍA

Se pondrá a disposición de los estudiantes material de estudio para cada tema en el campus virtual.

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.