20-21

MÁSTER UNIVERSITARIO EN
TECNOLOGÍAS DEL LENGUAJE

GUÍA DE ESTUDIO PÚBLICA

ACCESO INTELIGENTE A LA INFORMACIÓN

CÓDIGO 31101019

UNED

20-21

ACCESO INTELIGENTE A LA INFORMACIÓN

CÓDIGO 31101019

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

Nombre de la asignatura	ACCESO INTELIGENTE A LA INFORMACIÓN
Código	31101019
Curso académico	2020/2021
Título en que se imparte	MÁSTER UNIVERSITARIO EN TECNOLOGÍAS DEL LENGUAJE
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	ANUAL
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

Tipo	Optativa
Cuatrimestre	Anual
Créditos/horas totales	6/150
Horas de estudio teórico	100
Horas de prácticas	50
Horas complementarias	0

Esta asignatura se encuadra en el módulo "ESP-LSI-1 Tecnologías del Lenguaje en la Web" dentro de la especialidad con el mismo nombre de la titulación de posgrado "Master en Lenguajes y Sistemas Informáticos". Dentro de esta especialidad, "Acceso Inteligente a la Información" revisa las áreas de investigación más activas dentro de las tecnologías de la lengua y la recuperación de información.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Lectura fluida del inglés y conexión a Internet.

EQUIPO DOCENTE

Nombre y Apellidos	FERNANDO LOPEZ OSTENERO (Coordinador de asignatura)
Correo Electrónico	flopez@lsi.uned.es
Teléfono	91398-7793
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS

Nombre y Apellidos	ALVARO RODRIGO YUSTE
Correo Electrónico	alvarory@lsi.uned.es
Teléfono	91398-9693
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA

Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS

HORARIO DE ATENCIÓN AL ESTUDIANTE

La tutorización de los estudiantes se llevará a cabo a través de la plataforma de enseñanza virtual de posgrado de la UNED.
D. Fernando López Ostenero: Asistencia al estudiante Martes y Miércoles de 11:00 a 13:00 horas.
D. Álvaro Rodrigo Yuste: Asistencia al estudiante Jueves de 11:00 a 13:00 horas y de 15:00 a 17:00 horas.

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

Competencias Básicas:
CB6 - 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
CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
CB8 - Que los estudiantes sean capaces 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
CB9 - Que los 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
CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Competencias Generales:
CPG1 - Adquirir capacidad de abstracción, análisis, síntesis y relación de ideas.
CPG2 - Adquirir capacidad crítica y de decisión
CPG3 - Adquirir capacidad de estudio y autoaprendizaje
CPG4 - Adquirir capacidad creativa y de investigación
CPG5 - Adquirir habilidades sociales para el trabajo en equipo

Competencias Específicas:
CE1 - Adquirir capacidad de comprender y manejar de forma básica los aspectos más importantes relacionados con los lenguajes y sistemas informáticos en general y, de manera especial, en los siguientes ámbitos: Tecnologías del lenguaje y de acceso a la información en web
CE3 - Adquirir capacidad de estudio de los sistemas y aproximaciones existentes y para distinguir las aproximaciones más efectivas.

CE4 - Adquirir capacidad para detectar carencias en el estado actual de la ciencia y la tecnología
CE5 - Adquirir capacidad para proponer nuevas aproximaciones que den solución a las carencias detectadas.
CE6 - Adquirir capacidad de especificar, diseñar, implementar y evaluar tanto cualitativa como cuantitativamente los modelos y sistemas propuestos.
CE7 - Adquirir 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.

RESULTADOS DE APRENDIZAJE

El curso pretende introducir a los estudiantes en los temas de investigación más relevantes relacionados con el Procesamiento del Lenguaje Natural aplicado a la exploración y manipulación de grandes volúmenes de información textual (como la disponible en la Web). El temario inicial cubre la Recuperación de Información Multilingüe, la extracción de resúmenes, la búsqueda automática de respuestas y la extracción de información, todos ellos temas en los que se puede iniciar, con posterioridad al curso, un trabajo de iniciación a la investigación en la especialidad de Tecnologías del Lenguaje en la Web. El temario se actualizará anualmente para adecuarse a las nuevas aplicaciones que surjan en este campo.
El estudiante adquirirá las siguientes destrezas y competencias:

- Debe tener una visión de conjunto de la investigación en sistemas de acceso a la información, como punto de confluencia entre los campos de Recuperación de Información y Procesamiento del Lenguaje Natural.
- Debe ser capaz de realizar una lectura crítica de artículos científicos sobre el tema, de localizar y discriminar información bibliográfica relevante, y de sintetizar información de distintas fuentes.
- Debe ser capaz de redactar con rigor científico y de comunicar y debatir con pares (en este caso, sus compañeros) sus análisis y opiniones en torno a los temas de la asignatura.
- Debe ser capaz de diagnosticar los puntos débiles del estado del arte actual en el campo y de esbozar temas y líneas de investigación potenciales.

CONTENIDOS

Tema 1: Recuperación de Información Multilingüe.

1. Necesidad del acceso multilingüe a la información
- Diferencia entre IR monolingüe e IR multilingüe
2. Recursos para la Recuperación de Información dependientes del idioma:
- Lematizadores.

•Stemmers.
•Segmentación de compuestos y palabras.
3. Traducción de Consultas, el problema de la Fusión Documental:
•Diccionarios bilingües, diccionarios con información gramatical.
•Estructuración de la consulta según la traducción.
•Idiomas pivote.
•Utilización de corpora paralelos/comparables.
•Programas de traducción automática.
•Tesauros.
•Múltiples idiomas: el problema de la fusión documental.
4. Traducción de Documentos, otros enfoques al problema:
•Traducción de los documentos vs. traducción de las consultas.
•Traducciones bidireccionales.
•Representación conceptual.
5. Sistemas interactivos de búsqueda de información multilingüe.
6. Evaluación
•CLEF y otros foros de evaluación
•Consideraciones de evaluación (medidas, colecciones, etc)

Tema 2: Extracción de Información.

1. Definición:
•Perspectiva histórica y objetivos
•Reconocimiento de información relevante
2. Arquitectura de los sistemas de Extracción de Información:
•Preprocesado y análisis de los documentos.
•Reconocimiento de patrones y entidades
•Resolución de co-referencias.
•Generación de la salida esperada.
3. Aprendizaje Máquina aplicado a la Extracción de Información:
•Aprendizaje de reglas para la extracción de información.
•Aprendizaje estadístico.
4. Extracción de Información Multilingüe:
•Reconocedores del idioma.
•Traducción de la información extraída vs. traducción de los documentos:
•Sistemas de extracción de información translingües.
5. Ejemplos de Sistemas de Extracción de Información:
•Freeling

•Annie
•Stanford NER
6. Evaluación:
•foros de evaluación (MUC, ACE, TAC)
•consideraciones de evaluación (medidas, colecciones, etc)

Tema 3: Extracción Automática de Resúmenes y Síntesis de Información.

1. Definición.
2. Tipos de resumen: consideraciones sobre el texto a procesar y los objetivos del resumen:
•Resumen mono/multi-documento
•Resumen genérico vs. orientado a consulta
•Resumen informativo vs. indicativo
•Resumen multi-evento vs. mono-evento
3. Caracterización de fragmentos relevantes:
•Localización y longitud del fragmento.
•Presencia de términos relevantes
•Expresiones indicativas de relevancia
•Nombres propios.
4. Técnicas de resumen basadas en coherencia y cohesión:
•Conceptos de cohesión y coherencia.
•Aplicación de la cohesión y la coherencia en la generación automática de resúmenes.
•Aplicación combinada de cohesión y coherencia.
5. Resumen multidocumento y síntesis de información:
•Características del resumen multi-documento frente a mono-documento.
•Síntesis de Información frente a resumen multi-documento.
6. Resúmenes multilingües:
•Diversas aproximaciones al problema.
7. Evaluación de resúmenes
•Evaluación basada en la coherencia o en la información contenida.
•Evaluación mediante resúmenes de referencia.
•Evaluación en relación a los documentos de partida.
•Consideraciones de evaluación (medidas, colecciones, etc)

Tema 4: Sistemas de Búsqueda Automática de Respuestas.

1. Búsqueda de Respuestas vs. Recuperación de Información:
•Evolución de la Recuperación de Información hacia la Búsqueda de Respuestas.
2. Arquitectura básica de un sistema de Búsqueda de Respuestas:

•Análisis de la pregunta.
•Selección de documentos.
•Extracción de respuestas.
•Validación de respuestas.
3. Clasificación de Sistemas de búsqueda automática de respuestas:
•Nivel de utilización de técnicas de PLN.
•Taxonomía de Moldovan.
•Situación actual de la investigación en este campo.
4. Tipos de preguntas y respuestas:
•Clasificación de los diferentes tipos y subtipos de preguntas.
5. La barrera del idioma en la Búsqueda de Respuestas:
•El track QA@CLEF: búsqueda translingüe de respuestas.
•Diferentes enfoques al problema.
6. Interacción con el usuario:
•Sistemas de ayuda para la búsqueda de respuestas.
•Comparación entre sistemas automáticos de búsqueda de respuestas y asistentes interactivos de ayuda a la búsqueda de respuestas.
7. Ejemplos de Sistemas de Búsqueda de Respuestas.
•Watson de IBM en el desafío Jeopardy!
8. Evaluación:
•Conferencias TREC, CLEF y NTCIR.
•Consideraciones de evaluación (medidas, colecciones, etc)

METODOLOGÍA

La general del programa de posgrado. En particular, los estudiantes realizan dos tipos de actividades en esta asignatura: una serie de tareas para cada uno de los temas que componen la asignatura. Éstas son comunes a todos los estudiantes y están fijadas dentro del material de estudio correspondiente a cada tema. En una segunda parte de la asignatura, cada estudiante realiza un trabajo individual sobre un tema acordado con el equipo docente. Todo el material de estudio está disponible en el entorno virtual del posgrado, y toda la interacción entre profesores y estudiantes se puede llevar a cabo en este entorno.

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

No existen exámenes en esta asignatura.

Criterios de evaluación

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

La evaluación de la asignatura se basa en los trabajos realizados por los estudiantes como respuesta a las actividades propuestas en cada tema y al trabajo final de la asignatura.

Estos trabajos deberán realizarse con la estructura de los artículos de investigación, con una extensión mínima de entre cinco y siete páginas, donde se deberán incluir (al menos) las siguientes secciones:

Resumen

Introducción

Conclusiones

El estudiante podrá añadir cualesquiera otras secciones que considere oportunas para poder completar la tarea o aquellas que se indiquen en la descripción de cada tarea concreta.

Además se deberá incluir la bibliografía utilizada, con especial atención a aquella no proporcionada por el Equipo Docente que haya sido encontrada en la fase de búsqueda bibliográfica de cada tema. Es importante que TODA la bibliografía se referencie desde el texto, en aquellos puntos donde se comenten temas relacionados con la fuente referenciada.

Criterios de evaluación

Para cada tarea se valorará:

Completitud: se tratan todos los puntos importantes de la tarea.

Corrección: el análisis realizado no contiene incorrecciones.

Bibliografía: ha de cubrir la temática de la tarea y deberá ser referenciada correctamente desde el texto.

Cuando se haga una referencia a una fuente hay que abstenerse de copiar literalmente fragmentos de la misma, se deberá optar por realizar una descripción de los conceptos relevantes que se quieren citar. Si un trabajo incluye copias literales, se calificará como suspenso.

Ponderación de la PEC en la nota final	100%, la asignatura se evalúa únicamente mediante estos trabajos.
Fecha aproximada de entrega	Las fechas aproximadas de entrega de cada tarea se encuentran indicadas en el plan de trabajo de la asignatura.
Comentarios y observaciones	

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s?	No
Descripción	
Criterios de evaluación	
Ponderación en la nota final	
Fecha aproximada de entrega	
Comentarios y observaciones	

¿CÓMO SE OBTIENE LA NOTA FINAL?

Para aprobar la asignatura se deberán entregar todas las tareas y el trabajo final. Los pesos de las tareas de los temas y el trabajo final serán los siguientes:
Tema 1: 15%
Tema 2: 15%
Tema 3: 15%
Tema 4: 15%
Trabajo Final: 40%
Los pesos de las tareas individuales de cada tema se encontrarán detallados en el curso virtual.

BIBLIOGRAFÍA BÁSICA

Bibliografia general de consulta

De la Revista Iberoamericana de Inteligencia artificial, n. 22, vol. 8, 2004 (número especial sobre Acceso a Información Multilingüe), los siguientes textos de referencia:

•López-Ostenero, F., Gonzalo, J. y Verdejo, M. F. Búsqueda de información multilingüe: estado del arte.
•Vicedo, José Luis. La búsqueda de respuestas: estado actual y perspectivas de futuro.
•Turmo, Jordi. Information Extraction: Multilinguality and Portability.
•Alonso, Laura, Castellon, Irene, Climent, Salvador, Fuentes, María, Padró, Lluís y Rodríguez, Horacio. Approaches to Text Summarization: Questions and Answers.

BIBLIOGRAFÍA COMPLEMENTARIA

En el curso virtual se propondrá una serie de artículos de investigación complementarios para cada una de las tareas a realizar por los estudiantes.

RECURSOS DE APOYO Y WEBGRAFÍA

El campus virtual de posgrados de la UNED, proporcionará interfaz de interacción entre el estudiante y sus profesores. La plataforma UNED (Alf) permite impartir y recibir formación, gestionar y compartir documentos, crear y participar en comunidades temáticas, así como realizar proyectos online.

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.