

30233 - Recuperación de información

Información del Plan Docente

Año académico: 2022/23

Asignatura: 30233 - Recuperación de información

Centro académico: 110 - Escuela de Ingeniería y Arquitectura

Titulación: 439 - Graduado en Ingeniería Informática

Créditos: 6.0

Curso: 4

Periodo de impartición: Primer semestre

Clase de asignatura:

Materia:

1. Información Básica

1.1. Objetivos de la asignatura

Tras haber cursado un conjunto de asignaturas de programación, bases de datos y sistemas de información, el alumno es competente para aplicar estrategias de recuperación de datos. La recuperación de datos consiste principalmente en identificar aquellos registros de un repositorio (fichero, base de datos, etc) que contienen o cumplen exactamente con los términos especificados en las consultas del usuario. Sin embargo, cuando se trabaja con fuentes de datos heterogéneas y poco estructuradas (por ejemplo, la Web o grandes repositorios de textos o datos multimedia) este tipo de búsquedas exactas no es suficiente para satisfacer las necesidades de información del usuario. El objetivo de esta asignatura es aprender a aplicar un conjunto de técnicas de recuperación de información que están más enfocadas a recuperar información acerca de una temática o tópico y no tanto en la recuperación de los datos que satisfacen exactamente una consulta.

La asignatura tiene un marcado carácter aplicado. El alumno aprenderá las técnicas de recuperación de información aplicándolos sobre un conjunto de ejemplos de sistemas de recuperación de información integrados tanto en bibliotecas digitales como en otro tipo de repositorios de documentos para proporcionar funcionalidades de búsqueda. Estos sistemas serán presentados tanto en las clases de problemas, como en las distintas sesiones de las prácticas de laboratorio.

Estos planteamientos y objetivos están alineados con algunos de los Objetivos de Desarrollo Sostenible, ODS, de la Agenda 2030 (<https://www.un.org/sustainabledevelopment/es/>) y determinadas metas concretas, de tal manera que la adquisición de los resultados de aprendizaje de la asignatura proporciona capacitación y competencia al estudiante para contribuir en cierta medida a su logro:

- Objetivo 9: Construir infraestructuras resilientes, promover la industrialización inclusiva y sostenible y fomentar la innovación.

Meta 9.c: Aumentar significativamente el acceso a la tecnología de la información y las comunicaciones y esforzarse por proporcionar acceso universal y asequible a Internet en los países menos adelantados.

1.2. Contexto y sentido de la asignatura en la titulación

La asignatura de Recuperación de Información es una asignatura obligatoria englobada en la materia de tecnología específica en "Aprendizaje y Recuperación de Información" dentro de la intensificación de Computación. Aunque en la asignatura obligatoria de "Inteligencia Artificial" del tercer curso se presenta la Recuperación de Información como ejemplo de aplicación de las técnicas de inteligencia artificial, esta asignatura de Recuperación de Información permite profundizar en esta disciplina de la informática que permite desarrollar sistemas de recuperación de información sobre distintos tipos de fuentes de datos de grandes volúmenes y poco estructuradas.

1.3. Recomendaciones para cursar la asignatura

El alumno que curse esta asignatura ha de contar con una formación en métodos y técnicas de inteligencia artificial del nivel de la asignatura de Ingeniería Artificial.

2. Competencias y resultados de aprendizaje

2.1. Competencias

- Conocer y aplicar los procedimientos algorítmicos básicos de las tecnologías informáticas para diseñar soluciones a problemas, analizando la idoneidad y complejidad de los algoritmos propuestos.
- Conocer y aplicar las herramientas necesarias para el almacenamiento, procesamiento y acceso a los Sistemas de información, incluidos los basados en web.
- 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.
- Desarrollar y evaluar sistemas interactivos y de presentación de información compleja y su aplicación a la resolución de problemas de diseño de interacción persona computadora.
- Conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.

2.2. Resultados de aprendizaje

El estudiante, para superar esta asignatura, deberá demostrar los siguientes resultados...

- Domina las técnicas de recuperación de información sobre colecciones de datos almacenados en distintos repositorios (incluyendo repositorios hipermedia y multimedia).
- Sabe aplicar las técnicas de recuperación de información ante nuevos problemas que se le plantean.
- Domina las técnicas basadas en ontologías para representar la información disponible en un dominio específico.
- Sabe aplicar las técnicas de recuperación semántica para desarrollar aplicaciones de búsqueda.

2.3. Importancia de los resultados de aprendizaje

Actualmente, la difusión y compartición de datos de muy diversos formatos a través de la Web permite que cualquier tipo de aplicación informática pueda procesar y explotar un conjunto de recursos de información que hasta hace unos pocos años era impensable.

Sin embargo, la heterogeneidad sintáctica y semántica de los datos que se pueden descargar exige también la aplicación de una serie de técnicas y procedimientos formales que permitan procesar y extraer adecuadamente información para aprovechar todos esos recursos. Por ello, la formación en los conceptos, técnicas y métodos presentados en esta asignatura serán muy importantes para abordar con solvencia cualquier tipo de problema de búsqueda de información.

3. Evaluación

3.1. Tipo de pruebas y su valor sobre la nota final y criterios de evaluación para cada prueba

El estudiante deberá demostrar que ha alcanzado los resultados de aprendizaje previstos mediante las siguientes actividades de evaluación

Primera convocatoria. La evaluación de la asignatura se realiza a través de dos pruebas:

P1. Prueba escrita sobre los conceptos básicos de la asignatura en la que el estudiante tendrá que responder cuestiones cortas y resolver pequeños ejercicios. Se requiere una nota mínima de 5,0 puntos en esta prueba para aprobar la asignatura. Si se obtiene esta nota mínima, entonces la prueba pondera un 50% en la nota de la asignatura. La fecha de realización de esta prueba será la que la dirección de la EINA determine para la realización de la prueba global de evaluación de la asignatura.

P2. Trabajo de la asignatura y resultados de prácticas de laboratorio realizados en equipo. Se requiere una nota mínima de 5,0 puntos en esta prueba para aprobar la asignatura. Si se obtiene esta nota mínima, entonces la prueba pondera un 50% en la nota de la asignatura. Cada equipo, formado por 2 alumnos (salvo excepciones justificadas), deberá asistir a cada sesión de prácticas y realizar las entregas que se indiquen. Asimismo, en coordinación con las sesiones prácticas de laboratorio, cada equipo deberá realizar un trabajo de la asignatura. A lo largo del cuatrimestre se tendrán que ir realizando entregas parciales del trabajo. Si no se ha asistido a las sesiones prácticas de laboratorio o no se han realizado las entregas exigidas para las prácticas y el trabajo, además de enviar todos los entregables se deberá realizar un examen de la parte práctica de la asignatura en la fecha que establezca la EINA para la realización de la prueba global de la asignatura.

Es obligatorio realizar y entregar ambas pruebas para poder aprobar la asignatura. Si en una de las pruebas, o en las dos, la nota obtenida fuera inferior a 5,0, la calificación final de la asignatura será la media ponderada de las dos calificaciones (50% P1 y 50% P2), con un máximo de 4,0.

Se considerarán como no presentados en esta convocatoria a los alumnos que no hayan realizado ninguna de las dos pruebas P1 y P2, así como a aquellos que se hayan presentado a una de las pruebas, aprobándola, pero no se hayan presentado a la otra.

Segunda convocatoria. La evaluación global de la asignatura se realiza a través de dos pruebas análogas a las de la primera convocatoria, con las mismas ponderaciones y exigencia de notas mínimas. La prueba P2 en segunda convocatoria consistirá en el envío de los entregables exigidos para las prácticas y el trabajo y en la realización de un examen de la parte práctica de la asignatura. Las calificaciones del alumno obtenidas en la primera convocatoria en cualquier de las pruebas (P1 y P2) se mantienen en segunda convocatoria, salvo que el alumno opte por presentarse a la prueba correspondiente en esta nueva convocatoria, en cuyo caso prevalecerá la nueva calificación.

4. Metodología, actividades de aprendizaje, programa y recursos

4.1. Presentación metodológica general

El proceso de aprendizaje que se ha diseñado para esta asignatura se basa en lo siguiente:

- El estudio y trabajo continuado desde el primer día de clase.
- El aprendizaje de conceptos y técnicas a través de las clases magistrales, en las que se favorecerá la participación de los alumnos.
- La aplicación de tales conocimientos en las clases de casos y problemas. En estas clases los alumnos desempeñarán un papel activo en la discusión de los casos y resolución de los problemas.
- Las clases de prácticas en laboratorio en las que el alumno aprenderá la tecnología necesaria para implementar los algoritmos y estrategias presentados en las clases magistrales y aplicados en las clases de casos y problemas.
- El trabajo en equipo, en el que cada equipo deberá desarrollar y evaluar dos sistemas de recuperación de información que faciliten las búsquedas sobre una colección de documentos descargable de la Web. El primer sistema será un sistema de recuperación de información tradicional donde se apliquen técnicas tradicionales de recuperación de información sobre texto. El segundo sistema será un sistema de recuperación semántico que transformará los documentos de la colección en descripciones semánticas de recursos (RDF) que se volcarán sobre un almacén de tripletas para facilitar después su búsqueda con un lenguaje de consulta semántico.

4.2. Actividades de aprendizaje

Las actividades de aprendizaje que se ofrecen al estudiante para ayudarle a lograr los resultados previstos son las siguientes:

- En las clases impartidas en el aula se desarrollará el temario de la asignatura a través de clases magistrales, del análisis de caso y de la resolución de problemas de aplicación de los conceptos y técnicas presentadas en el programa de la asignatura.
- Las sesiones de prácticas se desarrollan en un laboratorio informático. A lo largo de sus sesiones cada alumno deberá realizar, individualmente o en equipo, trabajos directamente relacionados con los temas estudiados en la asignatura.
- Adicionalmente, se realizará un trabajo en equipos bajo la tutela del profesor, en el que cada equipo deberá desarrollar y evaluar distintos tipos de sistemas de recuperación de información sobre una colección de documentos descargable de la Web.

4.3. Programa

Programa de la asignatura

Bloque I - Recuperación de información tradicional

- Introducción a la recuperación de información: el modelo booleano
- El proceso de indexación
- El modelo vectorial
- El modelo probabilístico
- Evaluación de motores de búsqueda
- Aspectos avanzados de recuperación de información

Bloque II - Sistemas hipermedia y multimedia

- Búsqueda en la Web
- Interfaz de usuario y visualización

Bloque III - Recuperación semántica

- Introducción a las ontologías y la Web Semántica
- El lenguaje de representación RDF
- El lenguaje de consulta SPARQL
- El lenguaje de representación OWL

4.4. Planificación de las actividades de aprendizaje y calendario de fechas clave

Calendario de sesiones presenciales

La organización docente de la asignatura prevista es la siguiente.

- Clases en el aula (magistrales y de casos y problemas) (3 horas en una semana ordinaria, de las cuales, en el cómputo global del curso, aproximadamente 2 horas semanales corresponderán a clases magistrales y aproximadamente una hora semanal a clases de casos y problemas, de acuerdo con el calendario académico que establezca la dirección de la EINA).
- Clases prácticas de laboratorio (una sesión de 2,5 horas cada dos semanas, en función del calendario académico que establezca la dirección de la EINA y de la reserva de laboratorios). Son sesiones de trabajo en la utilización de tecnologías, tuteladas por un profesor.
- Tutorías de los trabajos en equipo donde los alumnos desarrollan y evalúan distintos tipos de sistemas de recuperación de información sobre una colección de documentos descargable de la Web. Estas tutorías se realizarán al finalizar cada una de las sesiones prácticas de laboratorio de 2,5 horas, tendrán una duración de 20 minutos y se atenderá a los alumnos asignados a cada grupo de prácticas de la sesión previa.

Trabajo del estudiante

La dedicación del estudiante para alcanzar los resultados de aprendizaje en esta asignatura se estima en 150 horas distribuidas del siguiente modo:

- 60 horas de actividades presenciales (sesiones en el aula teóricas -30-, de problemas -15- y sesiones en el laboratorio -15-)
- 24 horas de trabajo en equipo
- 60 horas de trabajo y estudio (estudio de apuntes y textos, resolución de problemas, preparación de clases y prácticas, desarrollo de programas, etc.)
- 6 horas dedicadas a distintas pruebas de evaluación

El calendario de exámenes será determinado por la dirección de la EINA y las fechas de las entregas asociadas a las prácticas y el trabajo se anunciarán con suficiente antelación.

4.5. Bibliografía y recursos recomendados

<http://psfunizar10.unizar.es/br13/egAsignaturas.php?codigo=30233>