



## Grado en Ingeniería Informática 30233 - Recuperación de información

Guía docente para el curso 2015 - 2016

Curso: , Semestre: , Créditos: 6.0

---

### Información básica

---

#### Profesores

- **Javier Lacasta Miguel** jlacasta@unizar.es
- **Francisco Javier Nogueras Iso** jnog@unizar.es

#### Recomendaciones para cursar esta 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.

#### Actividades y fechas clave de la asignatura

El calendario de exámenes y las fechas de entrega de trabajos se anunciará con suficiente antelación.

---

### Inicio

---

#### Resultados de aprendizaje que definen la asignatura

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

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

# Introducción

## Breve presentación de la asignatura

Esta asignatura persigue que el alumno se familiarice con las técnicas y algoritmos que faciliten la recuperación de información. La recuperación de información (Information Retrieval) es una disciplina dentro de las ciencias de la computación sobre la que se fundamenta el desarrollo de las herramientas de búsqueda basadas en computador. Esta disciplina proporciona los modelos y algoritmos que permiten abordar facetas tan diversas como la representación, el almacenamiento, la organización y el acceso a elementos de información.

---

## Contexto y competencias

---

### Sentido, contexto, relevancia y objetivos generales de la asignatura

#### La asignatura y sus resultados previstos responden a los siguientes planteamientos y objetivos:

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.

#### 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.

#### Al superar la asignatura, el estudiante será más competente para...

- 1:** 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.
- 2:** Conocer y aplicar de las herramientas necesarias para el almacenamiento, procesamiento y acceso a los Sistemas de información, incluidos los basados en web.
- 3:** 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.

**4:**

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.

- 5:** 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.

### **Importancia de los resultados de aprendizaje que se obtienen en la asignatura:**

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.

---

## **Evaluación**

---

### **Actividades de evaluación**

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

- 1:** **EN LA ESCUELA DE INGENIERÍA Y ARQUITECTURA DE ZARAGOZA**

**Convocatoria de junio.** La evaluación global 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 práctico realizado en equipos. 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. El trabajo se entregará de forma electrónica con anterioridad a la fecha que la dirección de la EINA establezca para la realización de la prueba global de la asignatura. Con anterioridad, a lo largo del cuatrimestre, se habrán tenido que entregar algunos de los elementos que forman parte del trabajo, acompañados de presentaciones en clase, que servirán para que los equipos reciban realimentación sobre el trabajo por parte de los profesores de la asignatura. Si no se han realizado estas entregas y presentaciones parciales, además de enviar todos los entregables se deberá realizar un examen sobre el trabajo.

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.

**Convocatoria de septiembre.** La evaluación global de la asignatura se realiza a través de dos pruebas análogas a las de la convocatoria de junio, con las mismas ponderaciones y exigencia de notas mínimas. La prueba P2 en septiembre consistirá en la entrega del trabajo y, en el caso de que no se hubieran realizado las entregas y presentaciones parciales durante el periodo de impartición de la asignatura, en la realización de un examen sobre el trabajo. Las calificaciones del alumno obtenidas en la convocatoria de junio en cualquier de las pruebas (P1 y P2) se mantienen en septiembre, salvo que el alumno opte por presentarse a la prueba correspondiente en esta nueva convocatoria, en cuyo caso prevalecerá la nueva calificación.

---

## Actividades y recursos

---

### 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.

### Actividades de aprendizaje programadas (Se incluye programa)

**El programa que se ofrece al estudiante para ayudarle a lograr los resultados previstos comprende las siguientes actividades...**

- 1:**
- 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 de 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.

### Planificación y calendario

#### Calendario de sesiones presenciales y presentación de trabajos

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 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.

Presentación de trabajos objeto de evaluación:

- El trabajo práctico realizado en equipos será entregado con anterioridad a la fecha límite para su entrega, que será la misma que la dirección de la EINA establezca para la realización de la prueba P1. Las entregas y presentaciones parciales relativas al trabajo dependerán del calendario académico y se anunciarán oportunamente tanto en clase el día de la presentación de la asignatura, como en la plataforma Moodle, a través del enunciado del trabajo.

# 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
- Evaluación de motores de búsqueda
- El modelo probabilístico
- Realimentación de relevancia y expansión de consultas

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

## 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, aproximadamente, de actividades presenciales (sesiones en el aula teóricas -30-, de problemas -15- y sesiones en el laboratorio -15-)
- 50 horas de trabajo en grupo
- 35 horas de trabajo y estudio individual efectivo (estudio de apuntes y textos, resolución de problemas, preparación de clases y prácticas, desarrollo de programas, etc.)
- 5 horas dedicadas a distintas pruebas de evaluación

## Referencias bibliográficas de la bibliografía recomendada

1. Manning, C.D. Introduction to Information Retrieval / C.D. Manning, P. Raghavan, H. Schuetze Cambridge University Press, 2008
2. Baeza-Yates, Ricardo. Modern information Retrieval / Ricardo Baeza-Yates, Berthier Ribeiro-Neto . - [1st. ed.] New York : ACM Press ; Harlow : Addison-Wesley, 1999
3. Witten, Ian H.. Managing gigabytes : compressing and indexing documents and images / Ian H. Witten, Alistair Moffat, Timothy C. Bell . - 2nd ed. San Francisco, Calif. : Morgan Kaufmann Publishers, 1999
4. Marti, A. Search User Interfaces / Marti A. Hearst Cambridge University Press, 2009
5. Gómez-Pérez, A.. Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web / A. Gómez-Pérez, M. Fernández-López, O. Corcho Springer Verlag, 2003
6. Antoniou, G. Semantic Web Primer (Cooperative Information Systems) / G. Antoniou, F. van Harmelen. 3rd ed. The MIT Press, 2012