# Sistemas de recuperación de información

**Curso Académico: ( 2021 / 2022 )** **Fecha de revisión: 04-06-2021**

**Departamento asignado a la asignatura: Departamento de Biblioteconomía y Documentación**

**Coordinador/a: SERRANO LOPEZ, ANTONIO ELEAZAR**

**Tipo: Obligatoria Créditos ECTS : 6.0**

**Curso : 2 Cuatrimestre : 1**

## REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Ninguna

## OBJETIVOS

En general, buscar y recuperar información, gracias a métodos y herramientas informáticos y manuales, que permitan dar respuesta a las demandas de los usuarios en condiciones óptimas de costes y plazos, y evaluar la adecuación entre la demanda y la respuesta proporcionada, así como determinar y evaluar las necesidades tecnológicas relacionadas con la gestión de bases de datos documentales que puedan ser de interés y utilidad en el momento actual o en el próximo futuro para los servicios y unidades de información. Y específicamente:

1. Comprender y conocer la definición de términos relacionados con la Recuperación de Información (RI).
2. Comprender y aplicar los principios y las técnicas para la RI y su evaluación.
3. Conocer los modelos teóricos de recuperación de información.
4. Manejar con soltura los diferentes lenguajes de interrogación e interfaces de los sistemas de recuperación de información.
5. Interactuar con los sistemas de recuperación de información para resolver las posibles necesidades de información que se planteen.
6. Distinguir los diferentes modelos teóricos de RI y reconocerlos en los Sistemas de Recuperación de Información (SRI) reales.
7. Convertir una solicitud de información en una estrategia de búsqueda adecuada al sistema y transcribir y transmitir los resultados de una búsqueda.
8. Manejar con soltura, comparar y evaluar diversos lenguajes de interrogación e interfaces que permiten interactuar con un SRI local o con motores, metabuscadores y otras herramientas de búsqueda en la red.
9. Dominar al menos un software de recuperación de información, de características avanzadas, instalación extendida y calidad reconocida, que sirva como base para el análisis y evaluación de cualquier otro.
10. Evaluar los resultados de una búsqueda en términos de fiabilidad y relevancia, en cualquier entorno de uso de un SRI.

## DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

CONTENIDOS TEÓRICOS
UNIDAD DIDÁCTICA 1: Introducción a los Sistemas de Recuperación de Información (SRI).
- Tema T0: Recuperación de Información (RI) en bases de datos documentales
- Tema T1: El marco teórico de la RI: Relación con la Indización; diferencias con la Recuperación de Datos
UNIDAD DIDÁCTICA 2: Principales modelos formales de RI (según D. Blair).
- Tema T2: Modelos elementales (Modelos 1-4): un descriptor, varios descriptores, valor de corte y salida ordenada
- Tema T3: Modelos con descriptores ponderados (Modelos 5-8): sólo búsqueda ponderada, sólo indización ponderada, búsqueda e indización ponderada y búsqueda vectorial
- Tema T4: Modelos con búsqueda booleana (Modelos 9 y 10): Búsqueda booleana y particularidades sobre texto libre
- Tema T5: Modelos basados en tesauros (Modelos 11 y 12): Búsqueda con tesauros binarios y con tesauros ponderados
UNIDAD DIDÁCTICA 3: Evaluación de los Sistemas de Recuperación de Información.
- Tema T6: Principios de evaluación de la eficacia de la recuperación: Relevancia; Exhaustividad y Precisión y sus complementos; interrelación entre P y R; otras medidas

- Tema T7: Dificultades para la obtención de indicadores: El problema del Silencio; Relevancia, como Afinidad y Utilidad; Relevancia, como binaria o ponderada

CONTENIDOS PRÁCTICOS
UNIDAD DIDÁCTICA 4: Análisis de software de recuperación de información: BKM (basado en BRS).
- Tema P1: BKM (1). Organización de las bases de datos; interfaces y funciones básicas; operadores booleanos; criterios de búsqueda y fraccionamiento de la estrategia; referencias a conjuntos previos; operadores de proximidad, deseada y evitada; prioridades de los operadores
- Tema P2: BKM (2). Limitación por campos, buscados e ignorados; truncamientos y comodines; uso de los índices; histórico y archivo de estrategias; formatos de visualización y exportación
- Tema P3: BKM (3). Selección del universo de trabajo; búsquedas numéricas; ordenación y exportación de resultados; búsqueda a través del tesauro integrado, activo y pasivo; preferencias del usuario
UNIDAD DIDÁCTICA 5: RI en el entorno de las bases de datos distribuidas en línea.
- Tema P4: Entrez (1). Particularidades de la RI en este entorno: preprocesamiento de la búsqueda; importancia del control terminológico
- Tema P5: Entrez (2). Uso de filtros. Formatos de exportación.
- Tema P6: Solr/Lucene (1). Particularidades de la RI en este entorno: Búsqueda (o navegación) facetada. Relevancia flexible
- Tema P7: Solr/Lucene (2). agrupamiento ("clustering") de resultados de búsqueda. Anexo: Comparación de estos entornos con los motores de búsqueda en Internet (Google y otros). Particularidades de la RI en este entorno: Escasa estructura y control terminológico; la importancia de la ordenación de resultados; Precisión vs. Exhaustividad

## ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

CLASES TEÓRICO-PRÁCTICAS. En ellas se presentarán los conocimientos que deben adquirir los alumnos. Estos recibirán las notas de clase y tendrán textos básicos de referencia para facilitar el seguimiento de las clases y el desarrollo del trabajo posterior. Se resolverán ejercicios, prácticas problemas por parte del alumno y se realizarán talleres para adquirir las capacidades necesarias. Se dedicarán 42 horas con un 100% de presencialidad.

TRABAJO INDIVIDUAL O EN GRUPO DEL ESTUDIANTE. Se decicarán 78 horas 0% presencialidad.

- Adquisición de conocimientos teóricos (total 3 ECTS) a través de las clases teóricas, de materiales docentes elaborados por el profesor, tutoriales en línea, lecturas especializadas y debates (1,2 ECTS), así como del estudio y trabajo personal de los estudiantes (1,8 ECTS). Se relaciona con las competencias 1 a 3.
- Adquisición de habilidades y destrezas (total 3 ECTS) a través de supuestos prácticos diversos de recuperación de información en entornos diversos (sistemas locales, en línea y motores web), con los que puedan adquirir y desarrollar las competencias 4 a 10.

- Tutorías: Los días y horario de las tutorías se podrán consultar en el espacio destinado a la asignatura en Aula Global. Además de las horas y lugares fijadas oficialmente para la asignatura, los alumnos pueden solicitar y concertar con los profesores otras tutorías fuera de ese horario y no necesariamente presenciales. Se decicarán 28 horas con un 25% de presencialidad.

## SISTEMA DE EVALUACIÓN

La calificación final de la asignatura tendrá la siguiente distribución general atendiendo a los distintos aspectos evaluables:

1. Evaluación continua [= 60% de la calificación final, 6 puntos sobre 10]
- Cuestionarios: 30% de la calificación final, 3 puntos sobre 10
- Prácticas: 30% de la calificación final, 3 puntos sobre 10

2. Examen final [= 40% de la calificación final, 4 puntos sobre 10]

De acuerdo con la normativa de la Universidad, en la convocatoria ordinaria el alumno que no haya seguido la evaluación continua tendrá derecho a la realización de un examen que le permita obtener el 60% de la calificación final.
En la convocatoria extraordinaria, si no ha seguido la evaluación continua, tendrá derecho a la realización de un examen que le permita obtener el 100% de la calificación final, y si ha seguido la evaluación continua será evaluado de la forma que más le beneficie, considerando un peso del examen del 40% más la calificación de la evaluación continua, o un peso del examen del 100% descartando la puntuación obtenida en la evaluación continua.

Peso porcentual del Examen Final 40
Peso porcentual del resto de la evaluación 60

| | |
|---|---|
| **Peso porcentual del Examen Final:** | 60 |
| **Peso porcentual del resto de la evaluación:** | 40 |

BIBLIOGRAFÍA BÁSICA


- BLAIR, D.C. Language and Representation in Information Retrieval., Elsevier Science Publishers, 1990
- Baeza-Yates, R.; Ribeiro-Neto, B Modern information retrieval, Addison-Wesley, 1999
- CHOWDHURY, G.G. Introduction to modern information retrieval (3ª ed.), Library Association, 2010
- LANCASTER, F.W El control del vocabulario en la recuperación de la información (2ª ed. corr.), Universitat de València, 2002
- MEADOW, CH.T.; BOYCE, B.R.; KRAFT, D.H Text information retrieval systems (3ª ed.), San Diego, Academic Press, 2007


BIBLIOGRAFÍA COMPLEMENTARIA


- Buckland, M.K. Information and Information Systems, Greenwood Pres, 1991
- Chamis, A.Y. Vocabulary Control and Search Strategies in Online Searching, Greenwood Press, 1991
- Manning, C.D.; Raghavan, P.; Schütze, H. Introduction to Information Retrieval., Cambridge University Press, 2008
- Meadow, Ch.T Text information retrieval systems, Academic Press, 2000
- Salton, G Automatic text processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison Wesl, 1989