

PROGRAMA DE LA ASIGNATURA

Curso académico: 2022-2023

Identificación y características de la asignatura			
Código	502285		Créditos ECTS
			6
Denominación (español)	RECUPERACIÓN DE LA INFORMACIÓN		
Denominación (inglés)	INFORMATION RETRIEVAL		
Titulaciones	GRADO EN INFORMACIÓN Y DOCUMENTACIÓN PCEO: Información y Documentación/Comunicación Audiovisual PCEO: Información y Documentación/Periodismo		
Centro	FACULTAD DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN		
Semestre	1º	Carácter	OBLIGATORIO
Módulo	TECNOLOGÍAS Y APLICACIÓN PRÁCTICA DE LA INFORMACIÓN Y DOCUMENTACIÓN		
Materia	REPRESENTACIÓN Y RECUPERACIÓN DE LA INFORMACIÓN		
Profesor/es			
Nombre	Despacho	Correo-e	Página Web
Cristina López Pujalte	22A	cloppuj@unex.es	
Área de conocimiento	Biblioteconomía y Documentación		
Departamento	Información y Comunicación		
Profesor coordinador (si hay más de uno)			
Competencias *			
<i>Competencias básicas</i>			
<p>CB1 - Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.</p> <p>CB2 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.</p> <p>CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.</p> <p>CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.</p> <p>CB5 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.</p>			

*Los apartados relativos a competencias, breve descripción del contenido, actividades formativas, metodologías docentes, resultados de aprendizaje y sistemas de evaluación deben ajustarse a lo recogido en la memoria verificada del título.

Competencias generales

CG2 - Conocimiento de los principios teóricos y metodológicos para la reunión, selección, organización, representación, preservación, recuperación, acceso, difusión e intercambio de la información.

CG3 - Conocimiento de las tecnologías de la información que se emplean en las unidades y servicios de información.

CG4 - Habilidades en el manejo de las tecnologías como medio indispensable en los procesos de tratamiento y transferencia de la información.

Competencias específicas

CE1 - Conocimiento de los principios teóricos y metodológicos para la planificación, organización y evaluación de sistemas, unidades y servicios de información.

CE4 - Comprender y aplicar los principios y las técnicas para la planificación, organización y evaluación de sistemas, unidades y de servicios de información.

CE5 - Comprender y aplicar los principios y las técnicas para la reunión, selección, organización, representación, preservación, recuperación, acceso, difusión e intercambio de la información.

CE10 - Capacidad de usar y aplicar las técnicas, las normativas y otros instrumentos utilizados en la reunión, selección, organización, representación, preservación, recuperación, acceso, difusión e intercambio de la información.

Competencias transversales

CT1 - Capacidad de análisis y de síntesis aplicadas a la gestión y organización de la información.

CT3 - Habilidades en el uso de Internet y software genérico (ofimática).

CT5 - Capacidad de organización y planificación del trabajo propio.

CT6 - Capacidad de trabajar en equipo y de integración en equipos multidisciplinares.

CT8 - Razonamiento crítico en el análisis y la valoración de alternativas.

CT10 - Capacidad para el aprendizaje autónomo.

Contenidos

Breve descripción del contenido*

Descriptores del Título:

Historia, Evaluación y Leyes aplicadas a la Recuperación y Representación Documental. Estrategias Booleanas de Búsqueda.

Temario de la asignatura

Denominación del tema 1: *Introducción. Conceptos básicos.*

Contenidos del tema 1:

- 1.1. Introducción. Acotación del ámbito abarcado por la asignatura
- 1.2. Concepto de Recuperación de Información (RI). Su papel dentro del Proceso Documental
- 1.3. Recuperación de información e indización
- 1.4. Sistemas de Recuperación de Información
- 1.5. Breve historia de la RI: *Astia, Cranfield, SMART...*

Denominación del tema 2: *Evaluación de la Recuperación de Información*

Contenidos del tema 2:

- 2.1. Criterios de evaluación. Eficacia. Concepto de relevancia
- 2.2. Medidas de evaluación
- 2.3. Evaluación según la técnica de recuperación empleada
- 2.4. Test de diferencias de medias

Descripción de las actividades prácticas del tema 2: Batería de ejercicios prácticos correspondiente a los contenidos del tema 2.

Denominación del tema 3: *Modelos de RI y Representación documental*

Contenidos del tema 3:

- 3.1. Sistemas basados en índices invertidos
- 3.2. Leyes de Zipf
- 3.3. Valor de información
- 3.4. Modelos de RI
- 3.5. Representación documental. Modelo del Espacio Vectorial
- 3.6. Técnicas de recuperación. Medidas de similitud
- 3.7. Selección de las componentes y ponderación
- 3.8. Componentes de un buscador

Descripción de las actividades prácticas del tema 3: Batería I y II de ejercicios prácticos correspondiente a los contenidos del tema 3.

Denominación del tema 4: *Mejoras de la RI*

Contenidos del tema 4:

- 4.1. Concepto de Unitérminos. Limitaciones
- 4.2. Mejoras a los unitérminos
- 4.3. La importancia de las consultas. Estrategias de búsqueda. Técnicas de modificación automática de consultas: *Relevance Feedback*
- 4.4. Tendencias de investigación relacionadas

Descripción de las actividades prácticas del tema 4: Batería de ejercicios prácticos correspondiente a los contenidos del tema 4.

Actividades formativas*								
Horas de trabajo del estudiante por tema		Horas Gran grupo	Actividades prácticas				Actividad de seguimiento	No presencial
Tema	Total	GG	PCH	LAB	ORD	SEM	TP	EP
1	13	4						9
2	40,5	15				5	0,5	20
3	40,5	15				5	0,5	20
4	34,5	12				2	0,5	20
Evaluación**	21,5	1,5						20
TOTAL	150	47,5				12	1,5	89

GG: Grupo Grande (100 estudiantes).
PCH: prácticas clínicas hospitalarias (7 estudiantes)
LAB: prácticas laboratorio o campo (15 estudiantes)
ORD: prácticas sala ordenador o laboratorio de idiomas (30 estudiantes)
SEM: clases problemas o seminarios o casos prácticos (40 estudiantes).
TP: Tutorías Programadas (seguimiento docente, tipo tutorías ECTS).
EP: Estudio personal, trabajos individuales o en grupo, y lectura de bibliografía.

** Indicar el número total de horas de evaluación de esta asignatura.

Metodologías docentes*

- Explicación en clase de los temas programados
- Utilización de material docente en diferentes tipos y formatos.
- Discusión de los contenidos.
- Aplicación práctica de los conocimientos teóricos a través de los laboratorios, talleres, etc.
- Análisis y resolución de problemas prácticos propuestos.
- Actividades de seguimiento del aprendizaje.
- Trabajo en grupo.

Resultados de aprendizaje*

- Encuadrar el concepto de recuperación de información en el proceso documental.
- Comprender la importancia de la representación documental (indización) y su relación con la recuperación de información.
- Conocer la historia de los sistemas de recuperación de información y cómo han evolucionado a lo largo del tiempo.
- Comprender la importancia de una correcta evaluación de la recuperación de la información, y aplicar las técnicas necesarias para llevarla a cabo.
- Conocer y emplear los distintos modelos de recuperación de información.
- Utilizar las metodologías de automatización de la representación documental y determinar su repercusión en la posterior recuperación a partir de sus ventajas e inconvenientes.
- Aplicar las técnicas básicas de recuperación y de representación documental, tanto estadísticas como lingüísticas.
- Comprender la importancia de la indización de las consultas en la mejora de los resultados de la recuperación, así como de las estrategias de búsquedas.

Sistemas de evaluación*

Criterios de evaluación

Según la normativa de evaluación vigente, el estudiante comunicará al profesor por el Campus Virtual de la Universidad (CVUEx) el tipo de evaluación elegido (modalidad de evaluación continua o modalidad de evaluación global) al principio de cada semestre. Cuando un estudiante no realice esta comunicación, se entenderá que opta por la modalidad de evaluación continua.

El sistema de evaluación por defecto, es decir, la **modalidad de evaluación continua** se realizará según los siguientes parámetros:

La evaluación de esta asignatura se realizará sobre los conocimientos teóricos y prácticos adquiridos por el alumno, la asistencia y participación en las clases magistrales y a las prácticas realizadas. En todo caso, el examen teórico/práctico deberá aprobarse para superar la asignatura.

La calificación final será la suma de la evaluación continua y del examen atendiendo a los siguientes criterios:

- Evaluación continua (30%): realización de prácticas tuteladas relacionadas con la materia desarrolladas durante la impartición de la asignatura que se expondrán en clase. La evaluación continua **no es recuperable en ningún caso**, y deberá realizarse en los plazos establecidos por el profesor durante el semestre correspondiente (1º semestre). La calificación obtenida en la evaluación continua solo estará vigente para las convocatorias del curso académico correspondiente.

- Examen (70%): realización de una prueba escrita sobre conocimientos teóricos y prácticos de la materia. En todo caso **el examen deberá aprobarse para poder superar la asignatura**, y solo en este caso se sumará a la nota obtenida en el examen la nota de la evaluación continua.

En cuanto a la **modalidad de evaluación global**, para aquellos alumnos que lo hayan solicitado, consistirá en una prueba final de carácter global, de manera que la superación de ésta suponga la superación de la asignatura.

En ambas modalidades, la calificación final mínima para superar la asignatura será de Aprobado (5).

Para los aspectos no contemplados en estos criterios se seguirá lo dispuesto en la normativa de evaluación vigente en la Universidad de Extremadura.

Bibliografía (básica y complementaria)

Bibliografía básica:

Baeza-Yates, R.; Ribeiro-Neto, B. (1999). *Modern information retrieval*. Essex, UK: Addison-Wesley.

Büttcher, S.; Clarke, C.L.A.; Cormack, G.V. (2016). *Information Retrieval: Implementing and Evaluating Search Engines*. Cambridge: MIT Press.

Cacheda Sijo, F.; Manuel Fernández Luna, J.; Huete Guadix, J.F. (coord.) (2011). *Recuperación de información: un enfoque práctico y multidisciplinar*. Madrid: RA-MA, cop.

Chu, H. (2007). *Information Representation and Retrieval in the Digital Age*. Medford, New Jersey: Information Today.

Ellis, D. (1996). *Progress and problems in information retrieval*. London: The Library Association.

Frakes, W. B. & Baeza-Yates, R. (1992). *Information retrieval: data structures & algorithms*. New Jersey: Upper Saddle River.

Harman, D.K. (1992). Relevance feedback and other query modification techniques. In: Frakes, W.B.; Baeza-Yates, R. (Eds.) *Information retrieval: data structures and algorithms*. NJ: Prentice Hall, 241-263.

Manning, C.D.; Raghavan, P.; Schütze, H. (2008). *Introduction to Information Retrieval*. UK: Cambridge University Press.

Moya Anegón, F. de. (1995). *Los Sistemas Integrados de Gestión Bibliotecarias*. Madrid: ANABAD.

Salton, G. (1989). *Automatic text procesing: the transformation, analysis, and retrieval of information by computer*. Reading, Massachusetts: Addison-Wesley.

Salton, G. & McGill, M.J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. New York: Association for Computing Machinery and Morgan & Claypool Publishers.

Bibliografía complementaria:

Cheng, L., Yang, Y., Zhao, K., & Gao, Z. (2020). Research and improvement of tf-idf algorithm based on information theory. *Advances in Intelligent Systems and Computing*, 905, 608-616.

Keyser, P. de. (2007). *Indexing From Thesauri to the Semantic Web*. Oxford, UK: Chandos Publishing.

Langville, A.N. & Meyer, C.D. (2011). *Google's PageRank and Beyond: The Cience of Search Engine Rankings*. Princeton, N.J.: Princeton University Press.

Mayer-Schönberger, V. & Cukier, K. (2013). *Big data. La revolución de los datos masivos*. Madrid: Turner Publicaciones S.L.

Pestana Caldes, A.I.; López-Pujalte, C. (2009). Algoritmos Genéticos aplicados a la Recuperación de Información. Em: Borges, M.M. & Sanz Casado, E. (ed.). *A Ciência da Informação Criadora de Conhecimento*. Vol. II, 405-408.

Coimbra: Coimbra University Press.

Strzalkowski, T. (ed.) (1999). *Natural Language Information Retrieval*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Otros recursos y materiales docentes complementarios

- Espacio virtual de la asignatura en el campus virtual de la UEx "*Recuperación de información*"
<http://campusvirtual.unex.es>
- Se utilizarán colecciones de test específicas de evaluación de recuperación de información (*Cranfield, CISI, Medline*, etc.) accesibles on-line:
http://ir.dcs.gla.ac.uk/resources/test_collections/