

PROGRAMA DE LA ASIGNATURA

Curso académico 2008/2009

Identificación y características de la asignatura			
Denominación	Indización y Condensación Documental Automatizadas	Código	102883
Créditos (T+P)	4.5 (3.5 T + 1 P)		
Titulación	Licenciatura en Documentación		
Centro	Facultad de Biblioteconomía y Documentación		
Curso	1º	Temporalidad	1º cuatrimestre
Carácter	Troncal		
Descriptor (BOE)	Sistemas de Recuperación de Información. Compatibilidad de idiomas y sistemas. Condensación de contenidos documentales.		
Profesor/es	Nombre	Cristina López Pujalte	
	Despacho	309	
	Correo-e	clopez@alcazaba.unex.es	
	Página web		
Área de conocimiento	Biblioteconomía y Documentación		
Departamento	Información y Comunicación		
Profesor coordinador (si hay más de uno)			

Objetivos y/o competencias
<ol style="list-style-type: none"> 1. Recordar el concepto de indización y condensación documental, y las funciones que realizan dentro del Proceso Documental, así como el resto de conceptos básicos (documentales, estadísticos y tecnológicos) necesarios para poder acometer la docencia de la asignatura. 2. Comprender la importancia de la indización y cómo está íntimamente relacionada con la Recuperación de Información. Asimismo, asimilar el papel fundamental del resumen en estos tiempos de sobrecarga de información. 3. Conocer las distintas metodologías de automatización de la indización y condensación documental de manera general, cómo surgen, sus orígenes, características, y las bases científicas en las que se apoyan. 4. Profundizar posteriormente en cada una de ellas, analizando sus características específicas, sus técnicas básicas, sus principales ventajas y limitaciones, así como las principales investigaciones en torno a ellas.

5. Ahondar en una de estas metodologías, el Procesamiento del Lenguaje Natural, al ser ésta la que encierra mayor complejidad y a la vez mayores expectativas.
6. Comprender la importancia de la indización de las consultas, así como de las técnicas automáticas de formulación de las mismas.
7. Manejar con soltura las técnicas básicas de automatización, tanto estadísticas como lingüísticas
8. Conocer las últimas tendencias en sistemas automáticos de indización y resumen, y sus perspectivas de futuro en la sociedad de la información.

Temas y contenidos
(especificar prácticas, teoría y seminarios, en su caso)

METODOLOGÍA:

Para el desarrollo de esta asignatura se emplearán tanto clases teóricas donde el profesor expondrá los temas correspondientes con ayuda de presentaciones gráficas, como clases prácticas donde los alumnos con ayuda del profesor tendrán la oportunidad de practicar las metodologías expuestas o hacer ejercicios prácticos que ayuden a la comprensión de la teoría, que se irán alternando en función de las necesidades.

Por otro lado, el alumno deberá realizar un trabajo que se entregará durante el cuatrimestre y antes de realizar el examen.

Además del horario de clases, existe un horario de tutoría así como una cuenta de correo electrónico donde se podrán resolver todo tipo de dudas.

TEMARIO:

Tema 1: Introducción. Conceptos básicos.

- 1.1. Introducción. Acotación del ámbito abarcado por la asignatura
- 1.2. Conceptos de indización y condensación documental: su papel dentro de la Cadena Documental
- 1.3. Recuperación de información (RI). Relación con la indización y el resumen
- 1.4. Breve historia de la RI: *Astia, Cranfield, SMART...*
- 1.5. Necesidad y ventajas de la automatización de la indización y condensación documental
- 1.6. Ventajas del lenguaje natural en la sociedad de la información
- 1.7. Sistemas basados en índices invertidos
- 1.8. Modelos de representación documental. Modelo del espacio vectorial
- 1.9. Evaluación de los sistemas de recuperación de información

Tema 2: Metodología estadística

- 2.1. Introducción
- 2.2. Leyes de Zipf
- 2.3. Crecimiento del vocabulario
- 2.4. Teoría matemática de la comunicación. Concepto de entropía
- 2.5. Primeras investigaciones de Luhn
- 2.6. Esquemas de ponderación de términos
- 2.7. Modelo probabilístico
- 2.8. *Stemming*
- 2.9. Proceso de indización automática básico
- 2.10. Limitaciones de los métodos estadísticos. Conclusiones

Tema 3: Metodología pseudolingüística: a medio camino entre los métodos estadísticos y los lingüísticos

- 3.1. Problemas del empleo exclusivo de unitérminos
- 3.2. Relaciones entre términos: tesauros y formación de frases
- 3.3. Indización mediante citas
- 3.4. Ejemplos de aplicaciones
- 3.5. Limitaciones de los métodos pseudolingüísticos. Conclusiones

Tema 4: Procesamiento del Lenguaje Natural (PLN)

- 4.1. Introducción: Definición de PLN y contextualización
- 4.2. Clasificación del conocimiento de la Lengua
- 4.3. Principales aplicaciones del PLN
- 4.4. Modelos computacionales del lenguaje
- 4.5. Fases o niveles de análisis del PLN desde la perspectiva del análisis textual
- 4.6. Ejemplos de aplicaciones
- 4.7. Conclusiones

Criterios de evaluación

CRITERIOS DE EVALUACIÓN

- Explicar razonadamente los conceptos básicos y las técnicas más utilizadas en la automatización de la indización y resumen
- Analizar de manera crítica los resultados obtenidos con estas técnicas, ventajas, inconvenientes y limitaciones
- Aplicar las metodologías de automatización estudiadas a casos prácticos y evaluar los rendimientos obtenidos por ellas aplicando las técnicas de diagnóstico explicadas
- Participar activamente en la resolución de problemas realizados en clase
- Participar activamente en los comentarios/discusiones de artículos realizados en clase, analizándolos con rigor y desde el punto de vista del método científico
- Realizar un análisis crítico o una revisión sobre investigaciones recientes de indización/resumen automático (siguiendo las normas de

elaboración de trabajos para revistas científicas)

ACTIVIDADES E INSTRUMENTOS DE EVALUACIÓN

- Valoración de la participación del alumno en clase mediante evaluación continua, tanto en la resolución de problemas como en la discusión de artículos científicos.
- Será necesario haber realizado positivamente un trabajo de tipo práctico para poder presentarse al examen.
- Dado que dentro de los objetivos de la asignatura están tanto la adquisición de unos conocimientos como el desarrollo de una serie de habilidades, la evaluación final consistirá en una prueba escrita que constará de cinco preguntas de respuesta libre, que serán teóricas y ejercicios teórico-prácticos del tipo de los realizados en clase. Todas ellas tendrán la misma puntuación, y será necesario conseguir al menos cinco puntos en total para superar la asignatura.

Bibliografía

Ellis, D. Progress and problems in information retrieval. The Library Association. London, 1996.

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

Keyser, P. de. Indexing From Thesauri to the Semantic Web. Chandos Publishing. 2007.

Mani, I.; Maybury, M.T.(Eds.). Advances in Automatic Text Summarization. MIT Press. Massachusetts, USA, 2001.

Moya Anegón, F. de. Los Sistemas Integrados de Gestión Bibliotecarias. ANABAD. Madrid, 1995.

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

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

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

Tutorías

	Horario	Lugar
Lunes	17:00 – 18:00 h.	Despacho nº 309
Martes	10:00 – 13:00	Despacho nº 309
Miércoles	17:00 – 18:00 20:00 – 21:00	Despacho nº 309
Jueves		
Viernes		