

PROGRAMA DE LA ASIGNATURA. SEMIPRESENCIAL (PASARELA)

Curso académico: 2014-2015

Identificación y características de la asignatura			
Código	502286		Créditos ECTS 6
Denominación	PROCESAMIENTO AVANZADO DE LA INFORMACIÓN		
Titulaciones	GRADO EN INFORMACIÓN Y DOCUMENTACIÓN		
Centro	FACULTAD DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN		
Semestre	2º	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
VICENTE P. GUERRERO BOTE	303	guerrero@unex.es	
Área de conocimiento	Biblioteconomía y Documentación		
Departamento	Información y Comunicación		
Profesor coordinador (si hay más de uno)			
Competencias/Objetivos			
<ol style="list-style-type: none"> 1. (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. 2. (CG3) Conocimiento de las tecnologías de la información que se emplean en las unidades y servicios de información. 3. (CE6) Utilizar y aplicar herramientas informáticas para la implantación, desarrollo y explotación de sistemas de información. 4. (CG4) Habilidades en el manejo de las tecnologías como medio indispensable en los procesos de tratamiento y transferencia de la información. 5. (CT3) Habilidades en el uso de Internet y software genérico (ofimática). 6. (CT10) Capacidad para el aprendizaje autónomo. 			
Breve descripción del contenido			
Algoritmos de clustering, clasificación, condensación documental. Enriquecimiento con la información de enlaces, pagerank.			
Temario de la asignatura			
<p>1. Tema 1: Recuperación de Información.</p> <p>1. Historia</p>			

2. Modelo de Recuperación de Información.
 - i. Modelo Booleano.
 - ii. Modelos de emparejamiento parcial.
 1. Modelo del Espacio Vectorial.
 2. Modelo Probabilístico.
3. Medidas del éxito.

2. Tema 2: Clustering, Clasificación y Condensación Documental.

1. Clustering y Clasificación.
2. Hipótesis del Clustering.
3. Algoritmos de Clustering.
 - i. Distancia/Similitud.
 - ii. Tipos de clustering.
 1. Algoritmos de clustering jerárquicos.
 2. Algoritmos particional de la k-medias.
 3. Algoritmos de clustering espectral.
 4. Mapas Auto-Organizativos de Kohonen.
4. Mecanismos de condensación documental automatizadas.
 - i. Técnicas de Etiquetado.
 - ii. Técnicas de Resumen.

3. Tema 3: Recuperación de Información Web.

1. Marco de la Recuperación de Información Web.
 - i. La Web como Colección Documental
 1. Tamaño.
 2. Dinamismo.
 3. Auto-organización.
 4. Hiperenlazado.
 - ii. Eficacia en un buscador web.
 - iii. Elementos de un buscador web.
 - iv. Situación en 1998.

2. Algoritmo de PageRank.
3. Algoritmo de HITS.
 - i. Conjunto de resultados.
 - ii. Algoritmo.
 - iii. Algebra lineal subyacente.
 - iv. Pros y contras del algoritmo HITS.
4. Otros mecanismos de ponderación.
5. Fraudes para mejorar de la posición en los buscadores.
 - i. Anteriores a 1998.
 - ii. Granjas de Enlaces (link farming).
 - iii. Google Bombing.
6. Retos actuales de la Recuperación de la Información Web.

Nota: Todos los temas llevan su parte teórica y su parte de problemas práctica de usos de distintos productos documentales.

Actividades formativas y Metodología

Horas de trabajo del alumno por tema		Presencial		Actividad de seguimiento	No presencial
Tema	Total	GG	SL	TP	EP
1	43	1	1	0	41
2	53	1	2	0	50
3	54,5	1	2	1,5	50
Evaluación del conjunto	150	3	5	1,5	141

GG: Grupo Grande (100 estudiantes).

SL: Seminario/Laboratorio (prácticas clínicas hospitalarias = 7 estudiantes; prácticas laboratorio o campo = 15; prácticas sala ordenador o laboratorio de idiomas = 30, clases problemas o seminarios o casos prácticos = 40).

TP: Tutorías Programadas (seguimiento docente, tipo tutorías ECTS).

EP: Estudio personal, trabajos individuales o en grupo, y lectura de bibliografía.

Sistemas de evaluación

1. Evaluación Continua a través de la red. Consistirá en una serie de actividades con las condiciones y plazos que se establezcan a lo largo del curso. La nota de estas actividades constituye el 60% de la nota final de la asignatura.
2. Evaluación Final de carácter presencial. La prueba constará de cinco preguntas de respuesta libre, de las cuales tres serán teóricas y dos versarán sobre las prácticas propuestas en clase. La nota del examen final constituye el 40% de la nota final de la asignatura, aunque deberá aprobarse en todo caso para superar la asignatura.
3. En las recuperaciones que se hagan, las tareas que se proponen a lo largo del curso no serán recuperables.

Bibliografía y otros recursos virtuales

Brandes, U. y Erlebach, T. (Editores). Network Analysis: Methodological Foundations. Springer Berlin Heidelberg: NewYork, 2005.

Chen, C., Ibekwe-SanJuan, F., y Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. Journal of the American Society for Information Science and Technology, 61(7), 1386-1409. doi: 10.1002/asi.21309.

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

Guerrero-Bote, V.P, Moya-Anegón, F. y Herrero-Solana, V. Document organization using Kohonen's algorithm. Information Processing & Management 38, no. 1 (January): 79-89, 2002.

Langville, A. N., Meyer, C. D. Google's PageRank and Beyond. Princeton University Press, 2006.

Moya Anegón, F. Los Sistemas Integrados de Gestión Bibliotecarias. ANABAD, 1994.

Nooy, W., Mrvar, A., Batagelj, V. Exploratory Social Network Analysis with Pajek. Cambridge University Press, 2005.

Page, L. Method for node ranking in a linked database. United States Patent Application 2001/6285999, Kind Code B 1, 2001.

Page, L., Brin, S., Motwani, R. & Winograd, T. The PageRank citation ranking: Bringing order to the Web. Technical report, Stanford University, Stanford, CA, 1998.

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

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

Horario de tutorías

Tutorías Programadas:
1,5 h. en la primera semana de mayo

Tutorías de libre acceso:

El horario de tutorías de libre acceso se establecerá para cada semestre dentro de los plazos previstos por la Universidad y podrá ser consultado en la web de la Facultad.

Recomendaciones

La Recuperación de Información siempre ha sido una de las materias centrales de la Información y la Documentación. De hecho el almacenamiento y recuperación de la información frente a una necesidad de información es una de las tareas centrales de la disciplina.

Esta asignatura comienza con un tema dedicado a lo que se conoce como recuperación de información tradicional donde se muestran los fundamentos del modelo booleano y vectorial, así como los problemas de la misma.

El siguiente tema está dedicado a las técnicas de clustering y condensación documental como indican los descriptores, también utilizados como técnicas de recuperación de la información.

Y la última parte está dedicada a la recuperación de información web, que se diferencia de la tradicional principalmente por el tamaño y localización de la conexión y por la red de enlaces existente entre los documentos.

En el grado en Información y Documentación de la Uex existen otras asignaturas que tienen contenidos relacionados. La falta de esos contenidos ha sido tenida en cuenta en el diseño de esta asignatura, de modo que la misma trata de cubrir esos otros temas, además de adaptar el nivel a tal circunstancia. Así, la asignatura ha sido especialmente diseñada para la pasarela en modalidad virtual de la Uex no siendo apta para su impartición en el grado de Información y Documentación.

Contenidos en el Campus Virtual

A través del campus virtual de la Uex se facilitan:

1. *Contenidos Básicos.* Como su nombre indica son los contenidos esenciales de la asignatura.
2. *Contenidos de Ampliación Evaluables.* Son contenidos también evaluables, pero serán menos

prioritarios.

3. *Otros Contenidos y Recursos de Ampliación no evaluables directamente.* Estos contenidos se pueden necesitar manejar para la realización de alguna práctica.

Evaluación Continua

La evaluación continua se realizará a través de cuatro tipos de actividades:

1. *Cuestionarios de Teoría*¹. Estos cuestionarios versarán sobre los Contenidos Básicos y sobre los Contenidos de Ampliación evaluables.
2. *Problemas Prácticos*¹. En ellos se da un enunciado genérico, pero se tiene que realizar un cuestionario en cada uno de ellos que dará la puntuación de la evaluación.
3. *Resolución de una pregunta típica de examen.* Los alumnos tendrán que responder una pregunta de una lista reducida de preguntas típicas de examen. El profesor la valorará de 0 a 10 después del examen.
4. *Foros de dudas*². Estos foros están pensados para que los alumnos intercambien dudas y respuestas. Habrá uno en cada tema. La solución de las dudas por parte de los alumnos se valorará en centésimas en función de su dificultad, exactitud y completitud. Para ello el profesor dejará un tiempo prudencial antes de resolver la duda para permitir al resto de alumnos solucionar las dudas. La puntuación total obtenida irá de 0 a 100 puntos y será la suma de las puntuaciones obtenidas por la resolución de dudas a compañeros.

Aunque la plataforma informática lo permite, se recomienda al alumno no dejar para el final la realización de estas actividades.

Examen Presencial

El examen presencial constará de 5 preguntas de dos puntos cada una:

- 3 preguntas serán de teoría y versarán sobre los contenidos básicos y contenidos de ampliación evaluables (al menos 2 serán de los contenidos básicos).
- 2 son problemas prácticos y serán similares a los realizados en el desarrollo de la asignatura. Para estas dos preguntas se podrán utilizar los ordenadores de las aulas de informática. En el examen no se facilitará ninguna URL ni forma de acceso al producto en cuestión.

Se ha configurado el módulo de calificación de la plataforma virtual, de modo que el alumno podrá consultar todas las puntuaciones obtenidas y medias a través de dicha plataforma.

Fraude y difusión ilegítima de los contenidos del curso.

Los contenidos del curso, tanto los temas, como los cuestionarios, problemas prácticos, y todos los ficheros proporcionados en el curso han sido registrados en el registro de la propiedad intelectual. La difusión por cualquier medio de los mismos atentará contra la propiedad intelectual del curso.

Por otro lado, en el curso se fomenta la ayuda de unos estudiantes a otros. Se entiende por ayuda, explicación de los procedimientos para su aprendizaje, pero en ningún caso facilitar las respuestas directas a los cuestionarios que será considerado fraude. Si se detecta algún fraude de este tipo tendrá las mismas consecuencias que la copia en el examen, esto es:

- Para los implicados la reducción a cero de toda la evaluación continua (que supone el 60% de la calificación).
- Si estos resultados se mostraran de forma pública se anularán las actividades implicadas, con el

¹ Solamente podrán realizarse una vez. Al evaluarse automáticamente por la plataforma del campus virtual, se podrán realizar hasta el examen de la convocatoria de julio. En cada convocatoria solamente se garantiza que se tendrán en cuenta aquellas puntuaciones obtenidas antes del examen.

² Solamente se evaluará durante el periodo de clases del semestre correspondiente. El contenido del mismo será evaluable en el examen presencial.

consecuente perjuicio para los que ya las hubieran realizado. Si llegaran a anularse todas las actividades, se programarán seminarios prácticos evaluables en la sala de laboratorio donde se realizarán actividades sustitutivas.

Los foros de dudas de los temas correspondientes son herramientas dispuestas entre otras cosas para fomentar dicha ayuda.