ASIGNATURA DE MÁSTER:

SISTEMAS DE INFORMACIÓN NO ESTRUCTURADA

Curso 2016/2017

(Código:31106116)

1.PRESENTACIÓN

Actualmente, con el crecimiento constante de páginas web y redes sociales, existe accesible en la red una cantidad de información prácticamente ilimitada. Esto ha supuesto una de las transformaciones sociales más importantes en las últimas dos décadas. Las aplicaciones prácticas son muy amplias, incluyendo estudios de mercado, buscadores especializados, sistemas de recomendación, análisis de redes sociales, etc.

La mayor barrera a la que se enfrenta esta línea de innovación es el acceso a información no estructurada, es decir, texto o documentos multimedia, debido fundamentalmente a la variabilidad del lenguaje y a la dificultad de interpretar contenidos. Es decir, cómo enlazar, agrupar o clasificar toda esta información según las necesidades de un sistema.

Un sistema de acceso a la información cubre aspectos tan diversos como la eficiencia de acceso y escalabilidad de bases de datos, la compresión, el diseño de interfaces de usuario, reconocimiento de patrones en imágenes, evaluación de componentes o la interacción con el usuario. No es posible cubrir todos estos aspectos en un curso. Sin embargo, a través de esta asignatura el alumno adquirirá las capacidades básicas para el desarrollo de sistemas de acceso a información, centrándose en los desafíos específicos dados por el uso de información no estructurada.

2.CONTEXTUALIZACIÓN

Esta asignatura es una asignatura de 6 créditos y obligatoria en el segundo cuatrimestre. Se incluye en el módulo de "Tecnologías informáticas". El objetivo fundamental es dar una visión integradora de técnicas relacionadas con los sistemas de acceso a la información y en concreto en el caso de información no estructurada.

Las competencias de esta asignatura se pueden consultar en la guía del máster.

3.REQUISITOS PREVIOS RECOMENDABLES

Los propios del máster y además es necesario disponer de un dominio de inglés técnico (leer y escribir) para manejar con facilidad las fuentes bibliográficas.

4.RESULTADOS DE APRENDIZAJE

Los resultados del aprendizaje esperados en este curso son:

- Entender los principales retos a los que se enfrenta un sistema de acceso a información no estructurada, así como las posibilidades que ofrece el desarrollo de este tipo de sistemas en un contexto industrial.
- Ser capaz de descomponer el desarrollo de un sistema de acceso a la información, con el fin de poder abordar cada uno de los aspectos implicados de forma independiente.
- Conocer las técnicas básicas más empleadas en las distintas fases que componen un sistema de acceso a la información, desde el procesamiento de texto, tratamiento de la estructura de la WEB y redes sociales, o los mecanismos de clasificación, ranking o agrupación de contenidos.

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección https://sede.uned.es/valida/
D5115621?0F47A3D7A75E583B48C6809

- Experiencia en implementación de algunos mecanismos básicos de tratamiento de la información no estructurada.
- Capacidad de recopilar y sintetizar conocimiento de aplicación industrial a partir de publicaciones de carácter científico.

5.CONTENIDOS DE LA ASIGNATURA

El temario de la asignatura cubrirá los siguientes contenidos:

1. Introducción: Fundamentos y arquitectura de un sistema de acceso a la información no estructurada. Adquisición, transformación e indexación.
2. Adquisición y procesamiento de contenidos en internet: *Crawling* y tratamiento de texto.
3. Técnicas de organización de la información: ranking, agrupación y clasificación de documentos.
4. WEB.2.0: Redes sociales, recomendación y búsqueda colaborativa.
5. Otros objetos de búsqueda: XML, entidades y contenidos multimedia.

La inmensa mayoría de ingenieros informáticos dedicados al acceso a la información no estructurada, desarrollan sistemas sobre conjuntos de datos locales o centrados en un dominio. Por ello, con el fin de evitar exceso de contenidos, no se abordarán técnicas específicas de indexación y compresión para búsquedas sobre cantidades masivas de datos. Estas técnicas requieren un estudio en profundidad que no es compatible en tiempo con el estudio de aspectos más generales como son las técnicas de organización de documentos o más novedosos como son la búsqueda en entornos sociales.

Otro aspecto que queda fuera de temario por los mismos motivos es la evaluación de sistemas de acceso a la información. La evaluación es particular de cada problema concreto independientemente del tipo de información que se accede, mientras que este curso se centra en los aspectos característicos de la información no estructurada. Por ello, más que en la evaluación global de un sistema de información, el curso cubrirá evaluación de componentes específicos del tratamiento de información no estructurada, como son el ranking, agrupación o clasificación de documentos.

6.EQUIPO DOCENTE

- ENRIQUE AMIGO CABRERA
- JULIO ANTONIO GONZALO ARROYO
- JORGE AMANDO CARRILLO DE ALBORNOZ CUADRADO

7.METODOLOGÍA

Las diferentes asignaturas que integran este Master, se impartirán todas ellas conforme a la metodología no presencial que caracteriza a la UNED, en la cual prima el autoaprendizaje del alumno, pero asistido por el profesor y articulado a través de diversos sistemas de comunicación docente-discente. Dentro de estos sistemas, cabe destacar que el Máster en Ingeniería Informática se imparte con apoyo en una plataforma virtual interactiva de la UNED donde el alumno encuentra tanto materiales didácticos básicos como materiales didácticos complementarios, informaciones, noticias, ejercicios y también permite la evaluación correspondiente a las diferentes materias.

Las diferentes actividades formativas se organizan como sigue:

- Estudio de contenidos: 50 horas.
- Tutorías:10h
- Actividades en la plataforma virtual: 10 horas.
- Trabajos individuales: 20 horas.
- Trabajos en equipo: 10 horas.

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección https://sede.uned.es/valida/
D5115621701F47A3D7A75E583B48C6809

- Práctica 50 horas.

En cada caso se refieren a:

1. Estudios de contenidos teóricos: Lectura de las orientaciones, materiales impresos y revisión de los exámenes con los docentes.

2. Tutorías: Interacción presencial o virtual con los tutores y aclaración de cuestiones planteadas por los estudiantes.

3. Actividades en la plataforma virtual: Lectura de las orientaciones para la realización de las actividades prácticas planteadas, realización de actividades de autoevaluación, problemas y ejercicios, e interacción con los compañeros en el foro.

4. Trabajos individuales: Comentarios de artículos de carácter científico.

5. Trabajos en equipo: Seminarios on-line y realización conjunta de tareas y evaluación de tareas entre pares.

6. Práctica: Uso de herramientas básicas de recuperación de información sobre conjuntos reducidos de documentos, o bien la implementación de un algoritmo en tareas más específicas como problemas de recomendación o agrupación de documentos. Los ejercicios prácticos se realizarán mediante software libre disponible y recursos de libre disposición.

8.BIBLIOGRAFÍA BÁSICA

ISBN(13): 9780136072249
Título: SEARCH ENGINES: INFORMATION RETRIEVAL IN PRACTICE (Primera)
Autor/es: Bruce Croft ; Trevor Strohman ; Donald Metzler ;
Editorial: PEARSON

Buscarlo en libreria virtual UNED

Buscarlo en bibliotecas UNED

Buscarlo en la Biblioteca de Educación

Buscarlo en Catálogo del Patrimonio Bibliográfico

9.BIBLIOGRAFÍA COMPLEMENTARIA

Comentarios y anexos:

El acceso a la información es en realidad una línea de investigación en constante desarrollo. Por ello, la bibliografía dada por el libro base se complementará con artículos de carácter científico publicados en congresos relacionados, como son:

- SIGIR: Association for Computing Machinery's Special Interest Group on Information Retrieval. Since 1963, we have promoted research, development and education in the area of search and other information access technologies.
- CLEF: Conference and Labs of the Evaluation Forum. Information Access Evaluation meets Multilinguality, Multimodality and Interaction.
- CERI: Spanish Conference in Information Retrieval.
- SPIRE: rnational Symposium on String Processing and Information Retrieval.

Y revistas especializadas, como son:

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección https://sede.uned.es/valida/
D5115621701F47A3D7A75E583B48C6809

- Information Processing and Management (IP&M), Elsevier
- Information Retrieval, Springer
- Transactions on the Web (TWEB), ACM

10.RECURSOS DE APOYO AL ESTUDIO

Para las actividades prácticas se utilizarán lenguajes estándar y herramientas de libre distribución para sistemas de búsqueda de información (Lucene). Se aportará a los alumnos corpora y anotaciones para la realización de dichas prácticas.

11.TUTORIZACIÓN Y SEGUIMIENTO

Los alumnos serán tutorizados vía los cursos virtuales o vía telefónica en horario de tutorías, atendiendo a las consultas de carácter teórico, dudas en cuanto a la realización de los trabajos prácticos. Asimismo, se establecerán foros de discusión para temas concretos.

12.EVALUACIÓN DE LOS APRENDIZAJES

La calificación se calcula sobre la base de:

- Examen Presencial: 50% de ponderación.
- Trabajos prácticos individuales y en equipo 30%.
- Pruebas Evaluación Continua 20% (Ejercicios y comentarios de artículos científicos).

La práctica consistirá en la utilización de una herramienta básica para el desarrollo de un motor de búsqueda sobre información textual, o bien la implementación de un algoritmo en tareas más específicas como problemas de recomendación o agrupación de documentos.

13.COLABORADORES DOCENTES

Véase equipo docente.

Ámbito: GUI - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección https://sede.uned.es/valida/
D511562170F47A3D7A75E583B48C6809