

FICHA DE ASIGNATURA

Título: M1. Sistemas y tecnologías para la gestión de la información en entornos Big Data

Descripción: Esta materia, tras introducir el Big Data y el mercado de profesionales especializados en sus distintas tecnologías, proporciona a los alumnos la formación que precisan para controlar las tecnologías de rastreo, extracción, indexación y recuperación a fin de gestionar repositorios documentales masivos que acumulan datos en una gran variedad de formatos (por ejemplo, texto, imágenes, vídeos, XML, web), así como los sistemas de procesamiento de lenguaje natural y minería de textos útiles clasificación y el agrupamiento documental. La asignatura profundiza en las fases iniciales del proceso de gestión de la información masiva, que son la captura y el almacenamiento de la información. La elección de los modelos correctos en estas dos fases en gran manera la eficiencia del proceso de análisis. De ahí la importancia que se concede a las técnicas para identificar posibles fuentes de datos y para el almacenamiento de la información.

Carácter: Obligatoria

Créditos ECTS: 6

Contextualización: El programa de esta asignatura pretende proporcionar al alumno una serie de conocimientos en el campo de Big Data al proporcionarles una visión global de los entornos de Big Data y los conocimientos básicos para el desarrollo de la infraestructura, almacenamiento y procesamiento de información.

Modalidad: Online

Temario:

1. Unidad 1: Introducción a Big Data

- 1.1. Introducción al Big Data.
- 1.2. Áreas del Big Data: integración, infraestructura, preservación, análisis, explotación y visualización.
- 1.3. Definiciones relacionadas con Big Data: Data Science, Data Analytics, Data Mining, Data Science, Machine Learning, Business Intelligence.
- 1.4. Profesionales relacionados con Big Data. Chief Data Officer, Data Manager, y Data Scientist.

2. Unidad 2: Fuentes de datos en entornos Big data

- 2.1. Diferencias respecto a las tecnologías de datos tradicionales.
- 2.2. Tipos de datos, flujo de datos, volumen de datos, calidad de datos.
- 2.3. Las V's del Big Data.

3. Unidad 3: Estructuras de datos en entornos Big data

- 3.1. Estructura de datos y tecnologías para selección de datos útiles;
- 3.2. El lenguaje de consulta SPARQL.

4. Unidad 4: Métodos de integración y motores de búsqueda

- 4.1. Métodos de integración de fuentes de información
 - 4.1.1. Servicios web, APIs, crawlers.
- 4.2. Motores de búsqueda y de información.

5. Unidad 5: Almacenamiento y modelado de información

- 5.1. Almacenamiento y modelado de información eficiente, distribuida y tolerante a fallos.
- 5.2. Tecnologías de almacenamiento de información con formato de ficheros planos (CSV, XML, JSON).

6. Unidad 6: Bases de datos en entornos de Big Data

- 6.1. Bases de datos relacionales (Amazon RDS para datos masivos)
- 6.2. Bases de datos no relacionales (NoSQL) para datos no estructurados o semiestructurados.
- 6.3. Tecnologías de rastreo, procesamiento, indexación y recuperación para acceso eficiente a contenidos multi-formato (texto, web, XML, imágenes).
- 6.4. Tecnologías de almacenamiento escalable: modelos de agregación, modelos en grafo y sistemas distribuidos.
- 6.5. Gestión de datos con propuestas comerciales: Cassandra, CouchDB, MongoDB, Neo4.

Competencias:

CE1. Gestionar y procesar datos en un entorno de captación y almacenamiento masivos.

CE2. Utilizar diferentes modelos de almacenamiento de datos, así como sistemas de gestión de bases de datos mediante la implementación de código en diferentes lenguajes de programación especializados en el procesamiento de datos masivos.

CE3. Resolver problemas reales en la clasificación, modelización e interpretación de un conjunto de datos haciendo uso de los recursos técnicos disponibles y apropiados en cada caso en particular.

CE4. Entender las implicaciones legales, morales y éticas, en lo referente al uso de datos personales.

Actividades Formativas:

Actividad Formativa	Horas	Presencialidad
Clases expositivas	30	0
Clases prácticas sobre laboratorio informático	30	0
Seminarios y conferencias online	10	0
Tutorías online	10	0
Trabajo autónomo del estudiante	70	0

Metodologías docentes:

- Explicación del contenido temático, presentación de los conceptos fundamentales y desarrollo del contenido teórico.
- Clases prácticas consistentes en actividades síncronas y asíncronas a través de las cuales se pretende mostrar al alumnado cómo debe actuar a partir de la aplicación de los conocimientos adquiridos. Realización de diferentes pruebas para la verificación de la adquisición tanto de conocimientos teóricos como prácticos y la adquisición de competencias.

- Sesiones periódicas entre el profesorado y el alumnado para la resolución de dudas, orientación, supervisión, etc.
- Lectura crítica de la bibliografía, estudio sistemático de los temas, reflexión sobre problemas planteados, resolución de actividades propuestas, búsqueda, análisis y elaboración de información, investigación e indagación, elaboración de memorias, informes, y trabajos etc.
- Defensa de los trabajos, individual: Grado de adecuación del trabajo a las indicaciones dadas por el tutor. Capacidad de argumentar.
- Dominio de la terminología propia de la disciplina. Claridad expositiva, tanto escrita como verbal

Sistema de Evaluación:

Sistemas de evaluación	Ponderación mínima	Ponderación máxima
Evaluación continua	60.0	60.0
Evaluación final	40.0	40.0

Bibliografía:

- Abu-Mostafa, Y.S., Magdon-Ismail, M., Lin, H.T (2012) Learning from Data. Online: AML Book
- Baesens, B. (2014) Analytics in a Big Data World: The Essential Guide to Data Science and its Applications. New Jersey: John Wiley & Sons, Inc.
- Davenport, T. H. (2014). Big Data at Work: Dispelling the Myths, Uncovering the Opportunities. USA: Harvard Business School Publishing Corporation
- Foreman, J.W. (2014) Data Smart: Using Data Science to Transform Information into Insight. Indiana: John Wiley & Sons, Inc.
- Maheshwari, A. (2014). Data Analytics Made Accessible (2017 edition) Online: Amazon.
- Simon, P. (2013). Too Big to Ignore. The Business Case for Big Data. New Jersey: John Wiley & Sons, Inc.
- White, T. (2015) Hadoop: The Definitive Guide (4th Edition). USA: O'Really Media, Inc.