

MÁSTER EN ECONOMÍA, FINANZAS Y COMPUTACIÓN • 2019-2020

1. DATOS BÁSICOS DE LA ASIGNATURA

ASIGNATURA (ESPAÑOL)		SISTEMAS DE PROCESADO MASIVO DE DATOS			
SUBJECT		MASSIVE DATA PROCESSING SYSTEMS			
CÓDIGO		AÑO DE PLAN DE ESTUDIOS		2015	
TIPO	OBLIGATORIO		OPTATIVO	X	
MÓDULO		PROGRAMACION			
SEMESTRE		2º			
CRÉDITOS (ECTS)	3,0	TEORÍA (80%)	2.4	PRÁCTICAS (20%)	0.6

HORARIO DE CLASES

GRUPO	FECHA	Lunes	Martes	Miércoles	Jueves	Viernes
SESIÓN 1	06/03/2020			9:15-13:15 (BD)		
SESIÓN 2	10/03/2020			9:15-13:15 (PT)		
SESIÓN 3	13/03/2020			9:15-11:15 (BD)		
SESIÓN 4	17/03/2020			9:15-13:15 (PT)		
SESIÓN 5	20/03/2020			9:15-11:15 (BD)		
SESIÓN 6	24/03/2020			9:15-13:15 (PT)		
SESIÓN 7	27/03/2020			9:15-11:15 (BD)		
SESIÓN 8	31/03/2020			9:15-13:15 (PT)		
SESIÓN 9	02/04/2020			9:15-11:15 (BD)		
SESIÓN 10	14/04/2020			9:15-13:15 (PT)		
SESIÓN 11	17/04/2020			9:15-11:15 (BD)		
SESIÓN 12	21/04/2020			9:15-13:15 (PT)		

2. DOCENTES

RESPONSABLE DE LA ASIGNATURA	Dr. Antonio Peregrín Rubio				
NOMBRE	Universidad de Huelva				
UNIVERSIDAD	Tecnologías de la Información				
DEPARTAMENTO	Ciencias de la Computación e Inteligencia Artificial				
ÁREA DE CONOCIMIENTO	156	UBICACIÓN	Escuela Técnica Superior de Ingeniería (El Carmen)		
Nº DESPACHO	peregrin@dti.uhu.es			TELÉFONO	959 217 653
CORREO ELECTRÓNICO		CAMPUS VIRTUAL	Moodle		
URL WEB					

HORARIO DE TUTORÍAS (*)

La celebración de las tutorías se comunicará a los alumnos a través de la plataforma virtual de la asignatura.

	Dr. Manuel J. Maña López				
NOMBRE	Tecnologías de la Información				
DEPARTAMENTO	Lenguajes y Sistemas Informáticos				
ÁREA DE CONOCIMIENTO	135	UBICACIÓN	Escuela Técnica Superior de Ingeniería (El Carmen)		
Nº DESPACHO	manuel.mana@dti.uhu.es			TELÉFONO	959 217 389
CORREO ELECTRÓNICO		CAMPUS VIRTUAL	Moodle		
URL WEB					

HORARIO DE TUTORÍAS (*)

La celebración de las tutorías se comunicará a los alumnos a través de la plataforma virtual de la asignatura.

3. DESCRIPTOR (ESPAÑOL)	<ul style="list-style-type: none"> - Sistemas de almacenamiento masivo. Características generales - Arquitectura de un SAM - Tecnologías de los sistemas de procesamiento de datos - Principios de Minería de Texto - Aplicaciones a la extracción de datos en la Web - Principios de Lingüística Computacional
-------------------------	---

(ENGLISH)	<ul style="list-style-type: none"> - Massive data processing systems. Main features - Architecture of a massive data processing system - Technologies of massive data processing systems - Foundations of Text Mining - Applications to data extraction on the Web - Foundations of Computational Linguistics
------------------	---

4. SITUACIÓN

PRERREQUISITOS

Ninguno.

CONTEXTO DENTRO DE LA TITULACIÓN

Esta asignatura proporciona el marco formal para entender el proceso de almacenamiento, manipulación y extracción de conocimiento a partir de grandes volúmenes de datos estructurados y no estructurados, y de texto. Se introducirá a los alumnos en las tecnologías del Big Data, los paradigmas entorno a él, sus retos y las tecnologías adyacentes y relacionadas con el mismo como la Computación en la Nube, el Internet de las Cosas o las Bases de Datos No Estructuradas. Asimismo, se buscará la conexión entre las tecnologías del Big Data con la Analítica de Datos, la Inteligencia de Negocio, y la Minería de Datos, y las herramientas disponibles para llevarlas a cabo, las infraestructuras necesarias, y las alternativas para las empresas. Se abordarán las técnicas fundamentales de procesamiento de texto y las aplicaciones más relevantes de la minería de textos. Los conceptos estudiados en esta asignatura son de carácter transversal para las asignaturas del Máster en las que se tenga acceso a grandes volúmenes de información.

RECOMENDACIONES

- Poseer destreza para buscar información útil en la Red
- Poseer conocimientos básicos de inglés
- Saber manejar fuentes bibliográficas
- Poseer destrezas básicas de programación

5. COMPETENCIAS

COMPETENCIAS

Al finalizar el curso, el alumno deberá ser capaz de:

- Capacidad para identificar los entornos de Big Data, y participar en los equipos multidisciplinares que lo integren en una empresa u organismo.
- Capacidad para identificar aplicaciones del procesamiento del lenguaje natural y de la minería de textos
- Capacidad para identificar la técnica más adecuada para resolver un problema de minería de textos

6. OBJETIVOS DE LA ASIGNATURA

El objetivo general es dotar al alumno de los conocimientos fundamentales necesarios para comprender y utilizar sistemas de procesado y almacenamiento masivo de datos. Este objetivo general se puede refinar en los siguientes objetivos específicos:

- Conocer las tecnologías del Big Data, el alcance del problema, y los recursos necesarios para utilizarlo.
- Comprender la relación entre el tratamiento masivo de datos, y otras tecnologías de la información emergentes directamente relacionadas.
- Conocer los recursos disponibles que integran Big Data con Analítica de Datos, Inteligencia de Negocio, etc.
- Comprender la dificultad de analizar grandes bases de datos de información textual y extraer conocimiento de las mismas
- Conocer lo retos más relevantes y las áreas de aplicación de la minería de textos
- Conocer las herramientas más importantes que existen actualmente y que pueden aplicarse en algunos de los problemas de minería de textos
- Comprender las técnicas fundamentales relacionadas con el agrupamiento, clasificación, análisis de sentimientos y minería de opiniones

7. ACTIVIDADES FORMATIVAS, METODOLOGÍA DE ENSEÑANZA-APRENDIZAJE Y SU RELACIÓN CON COMPETENCIAS QUE DEBE ADQUIRIR EL ESTUDIANTE

Metodología	Actividad	Descripción	Horas
Actividades presenciales	Clases teóricas	Clases teóricas. Fundamentos y planteamientos teóricos	25,5
	Clases prácticas	Problemas y casos prácticos: planteamiento y resolución de problemas concretos relacionados con la materia	
	Evaluación	Realización de exámenes parciales y finales, escritos u orales	
Trabajo autónomo tutelado	Trabajo autónomo individual	Uso de manuales, monografías y artículos (científicos, didácticos y divulgativos)	46,5
	Trabajo autónomo en grupo	Resolución de problemas y casos prácticos	
Tutorías	Individuales	Actividades académicas dirigidas	3
		Presenciales o virtuales (Campus Virtual, correo electrónico)	

Todo el material oportuno para el seguimiento de las clases teóricas y prácticas estará disponible en la plataforma Moodle de la UHU. Para las clases teóricas, los recursos que se utilizarán son la pizarra, las proyecciones de presentaciones con la ayuda del ordenador y material suplementario suministrado por el profesorado. En las clases prácticas se aplicarán los contenidos abordados en las clases teóricas, se hará hincapié en los mecanismos de resolución. Estas clases prácticas serán interactivas.

8. BLOQUES TEMÁTICOS Y CONTENIDOS

BREVE DESCRIPCIÓN DE LOS CONTENIDOS

Módulo I: BIG DATA (BD) (Profesor Dr. A. Peregrín)

- Introducción al Big Data
- Tecnologías para Big Data

Módulo II: PROCESAMIENTO DEL TEXTO (PT) (Profesor Dr. M.J. Maña)

- Acceso a la información textual
- Análisis textual

Minería de textos

9. BIBLIOGRAFÍA

BIBLIOGRAFÍA GENERAL

Módulo I

- Big Data. La revolución de los datos masivos. Viktor Mayer-Schönberger, Kenneth Cukier, HoughtonMifflin Harcourt, 2013
- Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. IBM. Paul Zikopoulos, Chris Eaton. McGraw Hill Professional, 29/05/2015.
- A. Fernandez, S. Río, V. López, A. Bawakid, M.J. del Jesus, J.M. Benítez, F. Herrera, Big Data with Cloud Computing: An Insight on the Computing Environment, MapReduce and Programming Frameworks. WIREs Data Mining and Knowledge Discovery 4:5 (2014) 380-409.

Módulo II

- R. Baeza-Yates y B. Ribeiro-Neto. Modern Information Retrieval: The Concepts and Technology behind Search. Second Edition. Pearson, 2011.
- Christopher D. Manning, Prabhakar Raghavan y Hinrich Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.

BIBLIOGRAFÍA COMPLEMENTARIA

Módulo I

- Hadoop, Soluciones Big Data,, Boris Lublinsky, Kevin Smith, Alex Yakubovich, Anaya Multimedia 2014
- Del Cloud Computing al Big Data: Vision introductoria para jóvenes emprendedores. Jordi Torres i Vinnals. Editorial UOC – PID_00194204. Primera edición: septiembre 2012. FUOC, 2012. v.3.0 España de Creative Commons.
- Big Data, Técnicas, Herramientas y Aplicaciones, María Pérez Marqués, 2015.

Módulo II

- Bo Pang y Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. Journal Foundations and Trends in Information Retrieval, vol. 2: 1-2, pp. 1-135.

10. SISTEMA DE EVALUACIÓN

En la convocatoria ordinaria I, II y III (Junio, Septiembre y Diciembre), y salvo que el alumno lo solicite (mediante escrito firmado y entregado a través del Registro General de la Universidad (presencial o telemático) dirigido al Departamento de Tecnologías de la Información y a la atención del profesor coordinador de la materia, Antonio Peregrín Rubio (recomendándose, enviar también copia de dicho escrito al profesor por correo electrónico (peregrin@dti.uhu.es) sólo como medida complementaria, y por agilidad administrativa), se empleará la Evaluación Continua, la cual consiste en:

- Los conocimientos teóricos y prácticos (grado de capacidad de resolución de problemas y aplicación de los contenidos teóricos a la práctica, y grado de conocimiento y comprensión) de la materia se evaluarán mediante una prueba o examen (de tipo test en la convocatoria I, escrito de preguntas cortas en las demás convocatorias) presencial e individual en el aula de informática a través de la plataforma de apoyo a la enseñanza presencial Moodle, según la convocatoria oficial para esta materia. En dicho examen, además del ordenador del aula, no se permitirá el uso de ningún otro dispositivo electrónico, y no se empleará ningún elemento documental externo. El peso de este examen será de un 50%, y el alumno debe obtener al menos un 4 sobre 10 puntos para el cómputo de la nota final considerando esta parte superada. La duración del examen será de 1h.
- La participación activa del estudiante en las actividades propuestas (seguimiento del estudiante), es decir, su contribución a los debates, foros de ideas, respuesta a ejercicios, actividades académicas, etc., será un elemento que se valorará con un peso de un 50%, a criterio de los profesores de la asignatura. El alumno debe obtener al menos un 4 sobre 10 puntos para el cómputo de la nota final considerando esta parte superada, y mantenerla en un mismo curso académico entre convocatorias ordinarias.

En la convocatoria extraordinaria (Noviembre), (así como si el alumno lo solicitó expresamente como se indicó en el párrafo inicial para las convocatorias ordinarias), se realizará una Evaluación Única, la cual consiste en:

- Un examen presencial e individual (en el que se medirá el grado de capacidad de resolución de problemas y aplicación de los contenidos teóricos a la práctica, y grado de conocimiento y comprensión de la materia) escrito en papel (que se le suministrará) y se rellenará a bolígrafo (azul o negro), consistente en preguntas cortas, largas y problemas combinados, en la fecha, hora y lugar en la que se convoque oficialmente para esta materia. No se permitirá el uso de dispositivos electrónicos (teléfonos móviles, tabletas, auriculares, ordenadores, relojes inteligentes, *wearables*, etc.) ni cualquier otro material didáctico que no haya sido expresamente autorizado por el profesor. El peso del examen es de un 100% en la calificación. La duración de este examen será de entre 2h y 2h y 30m.

Nota sobre la calificación “Matrícula de Honor”: si existieran más alumnos con una calificación que les permita aspirar a la matrícula de honor (10 sobre 10 puntos en la media ponderada), es decir, en caso de equidad, la calificación de Matrícula de Honor se asignará basándose en la mayor participación en clase e implicación del alumno en la asignatura (en todo tipo de sesiones), a juicio de los profesores de la asignatura.

MEDIDAS PREVISTAS PARA RESPONDER A NECESIDADES EDUCATIVAS ESPECIALES

Se adoptarán, las medidas adecuadas a cada caso para que aquellos alumnos que presenten necesidades especiales puedan adquirir los conocimientos y capacidades necesarias para la superación de la materia.

11) ORGANIZACIÓN DEL TRABAJO DURANTE EL CURSO

Nº de Horas: 75 (3 créditos ECTS)

- Clases: 22,5 horas
 - Clases Teóricas: 18 h.
 - Clases Prácticas: 4,5 h.
- Trabajo autónomo tutelado: 52,5 horas
Sin presencia del profesor: el alumno debería invertir 30 horas en la resolución de los problemas propuestos.

12) TEMARIO DESARROLLADO

MÓDULO I: BIG DATA

Tema 1. Introducción al Big Data

Motivación y Objetivos. Tecnologías de la Información emergentes relacionadas con Big Data. La Ciencia de los Datos y el Big Data en relación con la Analítica de Datos, la Inteligencia de Negocio, el Aprendizaje Automático y la Minería de Datos.

Tema 2. Tecnologías para Big Data

Paradigma MapReduce. Arquitectura, implementación de algoritmos y ecosistema Apache Hadoop: Sistemas de Almacenamiento HDFS. Nuevos retos y tendencias: Apache Spark y su ecosistema. Herramientas.

MÓDULO II: PROCESAMIENTO DEL TEXTO

Tema 3. Acceso a la información textual

Motivación: la era del conocimiento, el acceso a la información y la minería de textos. Características del texto. La investigación en acceso a la información textual en la UHU.

Tema 4. Análisis textual

Introducción. Análisis léxico. Eliminación de palabras vacías. Reducción a la raíz. Análisis morfológico. Análisis sintáctico. Reconocimiento de entidades nombradas. Índice invertido: construcción y búsqueda. Recursos léxicos

Tema 5. Minería de textos

Motivación: definiciones y aplicaciones de la minería de textos. Evaluación. Clasificación. Agrupación. Análisis de sentimiento y minería de opiniones.

13) MECANISMOS DE CONTROL Y SEGUIMIENTO DE LA ASIGNATURA

Los previstos en los mecanismos de evaluación.