

# MÁSTER EN ECONOMÍA, FINANZAS Y COMPUTACIÓN • 2015-2016

## 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		1º			
Créditos totales (ECTS)		3,0	TEORÍA (80%)	2.4	PRÁCTICAS (20%) 0.6

## HORARIO DE CLASES

GRUPO	FECHA	Lunes	Martes	Miércoles	Jueves	Viernes
SESIÓN 1	17/06/2016					X
SESIÓN 2	22/06/2016			X		
SESIÓN 3	23/06/2016				X	
SESIÓN 4	30/06/2016				X	
SESIÓN 5	01/07/2016					X
SESIÓN 6	07/07/2016				X	

## 2. DOCENTES

### RESPONSABLE DE LA ASIGNATURA

NOMBRE	Dr. Antonio Peregrín Rubio				
UNIVERSIDAD	Universidad de Huelva				
DEPARTAMENTO	Tecnologías de la Información				
ÁREA DE CONOCIMIENTO	Ciencias de la Computación e Inteligencia Artificial				
Nº DESPACHO	TUPB-23	UBICACIÓN	Escuela Técnica Superior de Ingeniería		
CORREO ELECTRÓNICO	peregrin@dti.uhu.es			TELÉFONO	959 217 653
URL WEB		CAMPUS VIRTUAL	Moodle		

### HORARIO DE TUTORÍAS (\*)

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

NOMBRE	Dr. Manuel J. Maña López				
DEPARTAMENTO	Tecnologías de la Información				
ÁREA DE CONOCIMIENTO	Lenguajes y Sistemas Informáticos				
Nº DESPACHO	TUPB-47	UBICACIÓN	Escuela Técnica Superior de Ingeniería		
CORREO ELECTRÓNICO	manuel.mana@dti.uhu.es			TELÉFONO	959 217 389
URL WEB		CAMPUS VIRTUAL	Moodle		

### 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"> <li>- Sistemas de almacenamiento masivo. Características generales</li> <li>- Arquitectura de un SAM</li> <li>- Tecnologías de los sistemas de procesamiento de datos</li> <li>- Principios de Minería de Texto</li> <li>- Aplicaciones a la extracción de datos en la Web</li> <li>- Principios de Lingüística Computacional</li> </ul>
ENGLISH	<ul style="list-style-type: none"> <li>- Massive data processing systems. Main features</li> <li>- Architecture of a massive data processing system</li> <li>- Technologies of massive data processing systems</li> <li>- Foundations of Text Mining</li> <li>- Applications to data extraction on the Web</li> <li>- Foundations of Computational Linguistics</li> </ul>

#### 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 el 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
Clases presenciales	Clases teóricas	Clases teóricas. Fundamentos y planteamientos teóricos	22.5
	Clases prácticas	Laboratorio: realización de ejercicios relacionados con los contenidos teóricos	
Trabajo autónomo tutelado	Preparación de trabajos teórico-práctico individuales (prácticas planteadas y trabajo de curso)	Documentación Lectura Empleo de tecnologías de la información Resolución de problemas	52.5
	Estudio y preparación de contenidos	Contenidos teóricos (conceptos y fundamentos) y prácticos de la materia	
	Lectura/s complementaria/s	Uso de manuales, monografías y artículos	
	Actividades Académicas Dirigidas	Temas no explicados en las clases teóricas, que desarrolla de forma autónoma el alumnado	
Tutorías	Individuales	Presenciales o no (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

- Introducción al Big Data
- Tecnologías para Big Data

#### Módulo II: PROCESAMIENTO DEL TEXTO

- 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

Técnica empleada	Descripción	Criterio	Valor sobre el total de la nota
Prueba objetiva	Prueba individual teórico-práctica, en la que se evaluará la adquisición conocimientos teórico-prácticos	<ul style="list-style-type: none"> <li>• Grado de capacidad de resolución de problemas y aplicación de los contenidos teóricos a la práctica</li> <li>• Grado de conocimiento y comprensión</li> </ul>	100%

Los criterios de evaluación y calificación serán los que marca la normativa de exámenes y evaluaciones de la Universidad de Huelva y esta guía.

El conjunto de las actividades de evaluación estarán sujetas a la Normativa de Evaluación para las Titulaciones de Grado de la Universidad de Huelva (Consejo de Gobierno de 16 de julio de 2009):

[http://www.uhu.es/sec.general/Normativa/Texto\\_Normativa/Normativa\\_de\\_Evaluacion\\_grados.pdf](http://www.uhu.es/sec.general/Normativa/Texto_Normativa/Normativa_de_Evaluacion_grados.pdf)

## 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 y del acceso a la información. 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: aplicaciones de la minería de textos. 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.