

MÁSTER EN ECONOMÍA, FINANZAS Y COMPUTACIÓN • 2021-2022

1. DATOS BÁSICOS DE LA ASIGNATURA

ASIGNATURA ESPAÑOL)	SISTEMAS DE PROCESADOS MASIVO DE DATOS Y BIG DATA				
SUBJECT	MASSIVE DATA PROCESSING SYSTEMS AND BIG DATA				
CÓDIGO	1210102	AÑO DE PLAN DE ESTUDIOS	2021		
TIPO	OBLIGATORIO	OPTATIVO	X		
MÓDULO	FUNDAMENTOS DE PROGRAMACIÓN INFORMÁTICA				
SEMESTRE	2º				
CRÉDITOS (ECTS)	6,0	TEORÍA (80%)	4,8	PRÁCTICAS (20%)	1,2

HORARIO DE CLASES

GRUPO	FECHA	Lunes	Martes	Miércoles	Jueves	Viernes
SESIÓN 1A	09/03/2021			9:15-11:15 (BD)		
SESIÓN 1B	09/03/2021			11:15-13:15 (MT)		
SESIÓN 2A	16/03/2021			9:15-11:15 (MT)		
SESIÓN 2B	16/03/2021			11:15-13:15 (BD)		
SESIÓN 3A	23/03/2021			9:15-11:15 (MT)		
SESIÓN 3B	23/03/2021			11:15-13:15 (BD)		
SESIÓN 4A	30/03/2021			9:15-11:15 (MT)		
SESIÓN 4B	30/03/2021			11:15-13:15 (BD)		
SESIÓN 5A	06/04/2021			9:15-11:15 (MT)		
SESIÓN 5B	06/04/2021			11:15-13:15 (BD)		
SESIÓN 6A	27/04/2021			9:15-11:15 (MT)		
SESIÓN 6B	27/04/2021			11:15-13:15 (BD)		
SESIÓN 7	06/05/2021					9:15-13:15 (DB)
SESIÓN 8	13/05/2021					9:15-13:15 (DB)
SESIÓN 9	20/05/2021					9:15-13:15 (DB)
SESIÓN 10	27/05/2021					9:15-13:15 (DB)
SESIÓN 11A	31/05/2021		9:15-11:15 (MT)			
SESIÓN 11B	31/05/2021		11:15-13:15 (BD)			
SESIÓN 12A	01/06/2021			9:15-11:15 (MT)		
SESIÓN 12B	01/06/2021			11:15-13:15 (BD)		

2. DOCENTES

RESPONSABLE DE LA ASIGNATURA

NOMBRE	DR. ANTONIO PEREGRIN RUBIO				
UNIVERSIDAD	HUELVA				
DEPARTAMENTO	TECNOLOGÍAS DE LA INFORMACIÓN				
ÁREA DE CONOCIMIENTO	CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL				
Nº DESPACHO	156	UBICACIÓN	ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA		
CORREO ELECTRÓNICO	peregrin@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. Se puede hacer uso de las tutorías asincrónicas a través de la plataforma y el correo electrónico.

OTROS DOCENTES

NOMBRE	DR. MANUEL MAÑA LÓPEZ				
UNIVERSIDAD	HUELVA				
DEPARTAMENTO	TECNOLOGÍAS DE LA INFORMACIÓN				
ÁREA DE CONOCIMIENTO	LENGUAJES Y SISTEMAS INFORMÁTICOS				
Nº DESPACHO	135	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. Se puede hacer uso de las tutorías asincrónicas a través de la plataforma y el correo electrónico.

OTROS DOCENTES

NOMBRE	DR. FRANCISCO ALFREDO MÁRQUEZ HERNANDEZ				
UNIVERSIDAD	HUELVA				
DEPARTAMENTO	TECNOLOGÍAS DE LA INFORMACIÓN				
ÁREA DE CONOCIMIENTO	CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL				

Nº DESPACHO	129	UBICACIÓN	SEDE SANTA MARÍA DE LA RÁBIDA
CORREO ELECTRÓNICO	alfredo.marquez@dti.uhu.es		TELÉFONO 959 217 641
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. Se puede hacer uso de las tutorías asincrónicas a través de la plataforma y el correo electrónico.			
3. DESCRIPTOR			
ESPAÑOL	Sistemas de almacenamiento masivo y arquitecturas, tecnologías de los sistemas de procesado masivo de datos, principios de la minería de texto, principios de la lingüística computacional, introducción a la ciencia de los datos y el big data, framework Spark, procesamiento con Python en entornos Spark, conectores con bases de datos, uso del aprendizaje automático en entornos de big data, y aplicaciones de la tecnología de big data con datos de economía		
ENGLISH	Mass storage systems and architectures, technologies of mass data processing systems, principles of text mining, principles of computational linguistics, introduction to data science and big data, Spark framework, Python processing in Spark environments, connectors with databases, use of machine learning in big data environments, and applications of Big Data technology with economics data		
4. SITUACIÓN			
PRERREQUISITOS Conocimientos básicos de programación en Python (cursados en la asignatura “Programación“)			
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			
<ul style="list-style-type: none"> • Poseer destreza para buscar información útil en la Red • Poseer conocimientos básicos de inglés • Saber manejar fuentes bibliográficas 			
5. COMPETENCIAS			
BÁSICAS Y GENERALES			
<ul style="list-style-type: none"> • CG1 - Capacidad para organizar, planificar y desarrollar trabajos y proyectos propios de su ámbito científico o profesional. • CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo. 			
TRANSVERSALES			
<ul style="list-style-type: none"> • CT1 - Gestionar adecuadamente la información adquirida expresando conocimientos avanzados y demostrando, en un contexto de investigación científica y tecnológica o altamente especializado, una comprensión detallada y fundamentada de los aspectos teóricos y prácticos y de la metodología de trabajo en el campo de estudio. • CT2 - Dominar el proyecto académico y profesional, habiendo desarrollado la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro de su ámbito temático, en contextos interdisciplinares y, en su caso, con un alto componente de transferencia del conocimiento. • CT3 - Desarrollar una actitud y una aptitud de búsqueda permanente de la excelencia en el quehacer académico y en el ejercicio profesional futuro. 			
ESPECÍFICAS			
<ul style="list-style-type: none"> • CE5 - Conocer y saber utilizar principios de programación informática y de gestión de bases de datos. • CE6 - Conocer las fuentes y opciones que permite el big data en los ámbitos de la administración de empresas, del marketing, de la economía y de las finanzas. • CE8 - Saber diseñar y manejar los Sistemas de Información en la empresa 			
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 los 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	51
	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)	93
	Trabajo autónomo en grupo	Resolución de problemas y casos prácticos	
Tutorías	Individuales	Actividades académicas dirigidas	6
		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 de teleformación. Para las clases teóricas, los recursos que se utilizarán son la pizarra (tradicional y en su versión electrónica), las proyecciones de presentaciones con la ayuda del ordenador y material suplementario suministrado por el profesorado (fotocopias, archivos electrónicos, etc.). En las clases prácticas se aplicarán los contenidos abordados en las clases teóricas, se hará hincapié en los mecanismos de resolución, sus limitaciones y ventajas, así como un análisis crítico de los resultados alcanzados. Estas clases prácticas serán interactivas y la participación del/la alumno/a será tenida en cuenta a la hora de valorar su adaptación al grado de aprendizaje.

8. BLOQUES TEMÁTICOS Y CONTENIDOS

BREVE DESCRIPCIÓN DE LOS CONTENIDOS

Módulo I: BIG DATA (BD) (Profesores Dr. A. Peregrín y Dr. F.A. Márquez)

- Introducción al Big Data
- Ciencia de los Datos
- Tecnologías para Almacenamiento y Procesamiento de Big Data I y II.

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.

D) ORGANIZACIÓN DEL TRABAJO DURANTE EL CURSO

11. NÚMERO DE HORAS DE TRABAJO DEL ESTUDIANTE

Nº de Horas: 150 (6 créditos ECTS)

- Actividades presenciales: 51 horas
 - Clases de aula teóricas: Método expositivo. 36 horas
 - Clases de aula de problemas: Método expositivo. 8 horas
 - Sesiones de evaluación: 6 horas
- Trabajo autónomo tutelado: 93 horas
 - Trabajo autónomo individual: 73 horas
 - Trabajo autónomo en grupo: 20 horas
- Tutorías docentes: 6 horas

E) TEMARIO DESARROLLADO

MÓDULO I: BIG DATA

Tema 1.1. 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 1.2. Minería de Datos Computacional

Definiciones, proceso, algoritmos, validación, etc.

Tema 1.3. Tecnologías para Big Data I

Paradigma MapReduce y Apache Hadoop: Modelos de procesamiento y almacenamiento.

Tema 1.4. Ecosistema Hadoop

Elementos del ecosistema, Hive, Pig, Sqoop, Flume, etc.

Tema 1.5. Tecnologías para Big Data II

Apache Spark, DataFrames, Machine Learning con Spark.

Tema 1.6. Ecosistema Spark

Esquema del ecosistema Spark, SparqSQL, GraphX, etc.

Tema 1.7. Tecnologías Emergentes en Big Data

Apache Flink. Tendencias.

MÓDULO II: PROCESAMIENTO DEL TEXTO**Tema 2.1. 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 2.2. 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 2.3. Minería de textos

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

F) MECANISMOS DE CONTROL Y SEGUIMIENTO DE LA ASIGNATURA

Mientras se desarrollan los módulos del programa se propondrán ejercicios y prácticas diversas para establecer el nivel de captación de competencias y contenidos de la asignatura.