Semántica, datos conectados y minería de datos textual

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Másteres Universitarios

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.

DATOS GENERALES

Breve descripción

El principal objetivo de esta asignatura es que el estudiante sepa aplicar los conocimientos básicos sobre las distintas metodologías y técnicas de aprendizaje automático de forma crítica en problemas reales, incluyendo la minería de textos y la minería Web.

Un segundo objetivo, de tipo práctico, es proporcionar al alumno las capacidades y herramientas estándar necesarias para poder llevar a cabo de manera autónoma proyectos de analítica de datos.

Título asignatura

Semántica, datos conectados y minería de datos textual

Código asignatura

102270

Curso académico

2022-23

Planes donde se imparte

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE

Créditos ECTS

4

Carácter de la asignatura

OPTATIVA

Duración

Cuatrimestral

Idioma

Castellano e Inglés

CONTENIDOS

Contenidos

1. Redes semánticas.
2. Ontologías y aprendizaje de ontologías.
3. Linked data.
4. Análisis de redes complejas.
5. Minería de textos y minería Web.

COMPETENCIAS

Generales

CG1 - Capacidad para integrarse eficazmente en un grupo de trabajo y trabajar en equipo, compartir la información disponible e integrar su actividad en la actividad del grupo colaborando de forma activa en la consecución de objetivos comunes

CG2 - Capacidad de estudio, síntesis y autonomía suficientes para, una vez finalizado este programa formativo, iniciar una Tesis Doctoral

CG4 – Saber preparar y conducir presentaciones, ante públicos especializado, sobre una investigación o proyecto científico

CG7 - Conocer las herramientas metodológicas necesarias para desarrollar proyectos avanzados

CG8 - Capacidad de actualización de los conocimientos expuestos en el ámbito de la comunidad científica

Transversales

CT1 - Capacidad para buscar, obtener, seleccionar, tratar, analizar y comunicar información utilizando diferentes fuentes

CT2 - Capacidad para proyectar los conocimientos, habilidades y destrezas adquiridos para promover una sociedad basada en los valores de la libertad, la justicia, la igualdad y el pluralismo

CT3- Dominio de la gestión del tiempo

CT4- Capacidad para afrontar tareas y situaciones críticas

CT5- Capacidad de trabajo autónomo y toma de decisiones

CT6- Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, saber escuchar

Específicas

DSDA01 - Utilizar el análisis predictivo para analizar grandes volúmenes de datos y descubrir nuevas relaciones

DSDA02 - Utilizar técnicas estadísticas apropiadas sobre los datos disponibles para lograr una visión adecuada de los mismos

DSENG02 - Desarrollar y aplicar soluciones computacionales para problemas en un cierto

dominio de aplicación, usando una amplia gama de plataformas de análisis de datos

DSRM06 - Aplicar el ingenio propio para resolver problemas complejos y desarrollar ideas innovadoras

PLAN DE APRENDIZAJE

Actividades formativas

AF1 - Participación y asistencia a lecciones magistrales y seminarios (10 horas)

AF2 - Realización de prácticas de computación y análisis de datos (20 horas)

AF6 - Tutorías (presenciales o por medio de recursos telemáticos) (7 horas)

AF7 - Elaboración de informes de laboratorio y trabajos (15 horas)

AF8 - Estudio individual de contenidos de la asignatura (15 horas)

AF9 - Trabajo en grupo (30 horas)

A10 - Pruebas de evaluación (3 horas)

Resultados de aprendizaje

- Entender métodos complementarios para el análisis de grandes masas de datos no estructurados, introduciéndose en el campos de la minería de textos (y minería Web).
- Entender los fundamentos de la representación y análisis de datos con redes complejas.

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

SE1 - Examen (30%)

SE2 - Valoración de informes y trabajos escritos (50%)

SE4 - Seguimiento de actividades presenciales (20%)

PROFESORADO

Profesor responsable

Gómez Pérez, Domingo

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

Profesorado

Lloret Iglesias, Lara

Científica Titular del CSIC
Instituto de Física de Cantabria (IFCA), Consejo Superior de Investigaciones Científicas

Villar Fernández, Alejandro

Director/Consultor IT
Tecnologías de la Información y la Comunicación del Norte SL, Ticnor

Aguilar Gómez, Fernando

Investigador Postdoctoral
Instituto de Física de Cantabria (IFCA), CSIC-UC

Ramasco Sukia, José Javier

Investigador Postdoctoral
Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), CSIC-UIB

HORARIO

Horario

31/01/2023

17:30 - 19:30

Teoría de Ontologias

Alejandro Villar Fernández

Director/Consultor IT
Tecnologías de la Información y la Comunicación del Norte SL, Ticnor

02/02/2023

17:30 - 18:30

Prácticas de Ontologías

Alejandro Villar Fernández

Director/Consultor IT
Tecnologías de la Información y la Comunicación del Norte SL, Ticnor

18:30 - 19:30

Prácticas de Ontologías

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

07/02/2023

17:30 - 19:30

Redes semánticas

Alejandro Villar Fernández

Director/Consultor IT
Tecnologías de la Información y la Comunicación del Norte SL, Ticnor

09/02/2023

17:30 - 18:30

Prácticas de Redes Semánticas

Alejandro Villar Fernández

Director/Consultor IT
Tecnologías de la Información y la Comunicación del Norte SL, Ticnor

18:30 - 19:30

Prácticas de Redes Semánticas

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

14/02/2023

17:30 - 19:30

Teoría de redes semánticas

Alejandro Villar Fernández

Director/Consultor IT
Tecnologías de la Información y la Comunicación del Norte SL, Ticnor

16/02/2023

17:30 - 19:30

Linked Data

Fernando Aguilar Gómez

Investigador Postdoctoral
Instituto de Física de Cantabria (IFCA), CSIC-UC

21/02/2023

17:30 - 19:30

SPARQL

Fernando Aguilar Gómez

Investigador Postdoctoral
Instituto de Física de Cantabria (IFCA), CSIC-UC

23/02/2023

17:30 - 18:30

Prácticas SPARQL

Fernando Aguilar Gómez

Investigador Postdoctoral
Instituto de Física de Cantabria (IFCA), CSIC-UC

18:30 - 19:30

Prácticas SPARQL

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

28/02/2023

17:30 - 19:30

Minería de textos

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

07/03/2023

17:30 - 19:30

Minería de textos

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

14/03/2023

17:30 - 19:30

Práctica de minería de textos

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

17/03/2023

17:30 - 19:30

Redes Complejas

José Javier Ramasco Sukia

Investigador Postdoctoral
Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), CSIC-UIB

21/03/2023

17:30 - 19:30

Redes Complejas

José Javier Ramasco Sukia

Investigador Postdoctoral
Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), CSIC-UIB

23/03/2023

17:30 - 19:30

Prácticas de Minería Web

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

28/03/2023

17:30 - 19:30

Examen

Domingo Gómez Pérez

Doctor en Ciencias Matemáticas
Profesor Titular de Universidad del área de Lenguajes y Sistemas informáticos.
Universidad de Cantabria (UC).

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Bibliografía básica

Sholom M. Weiss, Nitin Indurkhya, Tong Zhang, Fred Damerau. *Text mining: predictive methods for analyzing unstructured information*. Springer Science+Business Media (2005)

Juan Antonio Pastor Sánchez. *Tecnologías de la web semántica*. UOC (2012)

Toby Segaran, Colin Evans, Jamie Taylor. *Programming the semantic web*. O'Reilly (2009)

Bibliografía complementaria

Jiawei Han, Micheline Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann Publishers (2006)