



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA

GUIA DOCENTE



CURSO 2015/2016

Máster Oficial en Ingeniería Informática

DATOS DE LA ASIGNATURA

Nombre:

Procesamiento de Información Textual en Redes Sociales

Denominación en inglés:

Natural Language Processing over Social Networks Information

Código:

1140214

Carácter:

Optativo

Horas:

	Totales	Presenciales	No presenciales
Trabajo estimado:	150	60	90

Créditos:

Grupos reducidos				
Grupos grandes	Aula estándar	Laboratorio	Prácticas de campo	Aula de informática
3	0	3	0	0

Departamentos:**Áreas de Conocimiento:**

Tecnologías de la Información	Lenguaje y Sistemas Informáticos
-------------------------------	----------------------------------

Curso:**Cuatrimestre:**

2º - Segundo	Primer cuatrimestre
--------------	---------------------

DATOS DE LOS PROFESORES

Nombre:**E-Mail:****Teléfono:****Despacho:**

Vélez Vélez, Miguel Ángel	velez@uhu.es	87674	TUPB-24 (Torreumbría)
Manuel de la Villa Cordero	manuel.villa@dti.uhu.es	959217639	Edif. Torreumbría, Dcho. 41
*Manuel J. Maña López	manuel.mana@dti.uhu.es	959217389	Edif. Torreumbría, Dcho. 47

*Profesor coordinador de la asignatura

Consultar los horarios de la asignatura

1. Descripción de contenidos**1.1. Breve descripción (en castellano):**

Procesamiento del Lenguaje Natural (PLN). Análisis léxico, morfológico y sintáctico. Reconocimiento de entidades nombradas. Detección de la negación y la especulación. Recursos léxico-semánticos. PLN en la Web y las redes sociales. Análisis de la información textual. Búsqueda en la Web. Resumen automático. Agrupamiento. Categorización. Minería de opiniones. Análisis de sentimientos. Integración en sistemas de negocio para el soporte a la toma de decisiones. Analítica Web y social media.

1.2. Breve descripción (en inglés):

Natural Language Processing (NLP). Lexical and morphological analysis. Parsing. Named entity recognition. Negation and speculation detection. Lexical-semantic resources. NLP on the Web and social networks. Analysis of the textual information. Web search. Automatic summarization. Clustering. Categorization. Opinion mining. Sentiment analysis. Integration in business systems for decisions support. Web analytics and social media.

2. Situación de la asignatura**2.1. Contexto dentro de la titulación:**

Las redes sociales y, en general, la Web acumulan un gran volumen de información que puede ser de gran relevancia para las organizaciones. La mayor parte de esta información se encuentra en forma de texto. A las dificultades habituales del tratamiento de cualquier texto, como por ejemplo la ambigüedad presente en todos los niveles del lenguaje, se suman características que dificultan mucho más la tarea. Entre estas dificultades adicionales podemos destacar la carencia de una estructura de discurso, el uso de jergas, abreviaturas, emoticonos y etiquetas, la utilización de frases agramaticales y la carencia de contexto, sobre todo en el caso de los microblogs, que incrementa la ambigüedad. Sin embargo, cada vez es más relevante analizar todo lo que se está diciendo en cada momento en Internet sobre una persona, una marca, una institución o un producto, y detectar automáticamente las entidades, conceptos, polaridad de opiniones o reputación. A pesar de la relevancia que tiene el texto y del auge de las redes sociales y la Web, el procesamiento del texto no es una materia de la que se ocupe ninguna otra asignatura de este Máster, con la profundidad que requiere el tema. Tan solo la asignatura Inteligencia Computacional (asignatura obligatoria, primer curso, primer cuatrimestre) incluye un tema entre nueve dedicado al procesamiento del lenguaje natural. De esta forma, esta asignatura opcional viene a cubrir esa carencia.

2.2. Recomendaciones:

Las competencias básicas que deberían poseer los alumnos antes de comenzar la asignatura son:

- Tener destreza en el lenguaje de programación Java.
- Poseer destreza para buscar información útil en la Red.
- Poseer conocimientos suficientes para leer documentos escritos en inglés.
- Saber manejar fuentes bibliográficas.
- Tener capacidad de lectura comprensiva.
- Saber expresarse correctamente de forma oral y escrita.

3. Objetivos (Expresados como resultados del aprendizaje):

El objetivo general de la asignatura es dotar al alumno de las técnicas necesarias para organizar, resumir y analizar grandes volúmenes de información no estructurada en forma de texto con especial atención a las redes sociales y la Web.

Este objetivo general se puede concretar en los siguientes objetivos específicos:

- Comprender la dificultad de procesar y analizar la información textual.
- Comprender las técnicas para el procesamiento léxico, sintáctico y semántico del texto.
- Comprender la importancia que tiene para las organizaciones la información textual almacenada en la Web y en las redes sociales.
- Comprender alguna de las técnicas fundamentales para el tratamiento y extracción de conocimiento de la información textual almacenada en la Web y en las redes sociales.
- Ser capaz de implementar un sistema básico para el tratamiento de información textual.

Con la realización de esta asignatura, el estudiante será capaz de:

1. Conocer y aplicar las técnicas básicas para el procesamiento del texto y, en particular, cuando se trata de texto procedente de redes sociales, y
2. Diseñar sistemas que permitan acceder a la información textual y analizar y extraer conocimiento de la misma, especialmente cuando se trata de grandes volúmenes de texto.

4. Competencias a adquirir por los estudiantes

4.1. Competencias específicas:

4.2. Competencias básicas, generales o transversales:

- **CB6:** Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- **CB7:** Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- **CG8:** Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinarios, siendo capaces de integrar estos conocimientos
- **CT3:** Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, responsabilidad
- **CT4:** Capacidad para el aprendizaje autónomo y toma de decisiones

5. Actividades Formativas y Metodologías Docentes

5.1. Actividades formativas:

- Sesiones de Teoría sobre los contenidos del Programa.
- Sesiones de Resolución de Problemas.
- Sesiones Prácticas en Laboratorios Especializados o en Aulas de Informática.
- Actividades Académicamente Dirigidas por el Profesorado: seminarios, conferencias, desarrollo de trabajos, debates, tutorías colectivas, actividades de evaluación y autoevaluación.

5.2. Metodologías docentes:

- Clase Magistral Participativa.
- Desarrollo de Prácticas en Laboratorios Especializados o Aulas de Informática en grupos reducidos.
- Resolución de Problemas y Ejercicios Prácticos.
- Tutorías Individuales o Colectivas. Interacción directa profesorado-estudiantes.
- Planteamiento, Realización, Tutorización y Presentación de Trabajos.
- Evaluaciones y Exámenes.

5.3. Desarrollo y justificación:

Las clases teóricas permitirán presentar los conceptos teóricos de forma clara y concisa, ayudándose de las herramientas docentes adecuadas al alcance del profesor. En ocasiones se proporcionará al alumno material para que pueda trabajar ciertos contenidos de forma personal, estimulando, así, el aprendizaje autónomo.

En las prácticas de laboratorio los alumnos tendrán que desarrollar ejercicios prácticos que permitirán afianzar los conocimientos adquiridos en teoría.

Se realizarán actividades académicamente dirigidas con el objetivo de profundizar en aspectos más avanzados y concretos de algunos contenidos teóricos y prácticos.

6. Temario desarrollado:

Bloque I. Procesamiento del Lenguaje Natural.

- Tema 1. Introducción.
 - Motivación. Conceptos básicos. Características del texto. Formatos de documento.
- Tema 2. Análisis del texto
 - Análisis léxico. Eliminación de palabras vacías. Reducción a la raíz. Análisis morfológico. Análisis sintáctico. Reconocimiento de entidades nombradas. Índice invertido: construcción y búsqueda. Detección de la negación y la especulación.
- Tema 3. Recursos léxico-semánticos.

Bloque II. Análisis de la información textual en Redes Sociales.

- Tema 4. PLN en la Web y las redes sociales.
- Tema 5. Búsqueda en la Web y Posicionamiento.
 - Características de la Web. Arquitectura de un buscador. Analítica web y social media. SEO y SEM. Ayuda para toma de decisiones.
- Tema 6. Otras tareas de análisis textual.
 - Resumen automático. Agrupamiento. Categorización. Minería de opiniones. Análisis de sentimiento.

7. Bibliografía

7.1. Bibliografía básica:

Daniel Jurafsky y James H. Martin. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Second Edition. Pearson Education International. 2009

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.

7.2. Bibliografía complementaria:

Christopher D. Manning y Hinrich Schütze. Foundations of statistical natural language processing. MIT Press. 1999.

David A. Grossman y Ophir Frieder. Information Retrieval: Algorithms and Heuristics. Springer, 2004.

P. Jackson y I. Moulinier. Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization. John Benjamins Publishing Company, 2002.

I.H. Witten, A. Moffat y T.C. Bell. Managing Gigabytes: Compressing and Indexing Documents and Images. Second Edition. Morgan Kaufmann Publishers, 1999.

8. Sistemas y criterios de evaluación.

8.1. Sistemas de evaluación:

- Examen de teoría/problemas
- Defensa de Prácticas
- Defensa de Trabajos e Informes Escritos

8.2. Criterios de evaluación y calificación:

La asignatura se califica utilizando diferentes instrumentos:

- I1. Examen objetivo. Se corresponde con el sistema de evaluación "Examen de teoría/problemas". 40% de la nota.
- I2. Trabajos de laboratorio. Se corresponde con el sistema de evaluación "Defensa de prácticas". 40% de la nota.
- I3. Trabajos/problemas individuales y en grupo. Se corresponde con el sistema de evaluación "Defensa de trabajos e informes escritos". 20% de la nota.

La asistencia a clase es obligatoria. Si algún alumno no pudiera hacerlo por motivos justificados debe comunicárselo a los profesores de la asignatura lo antes posible. En ese caso, se diseñará un plan de evaluación y calificación alternativo y adaptado a las necesidades del alumno.

Para aprobar la asignatura es necesario aprobar por separado I1 e I2 y sumar 5 entre las 3 calificaciones.

9. Organización docente semanal orientativa:

Semanas	Grupos Grandes	Grupos Reducidos	Aula Estándar	Grupos Reducidos	Aula de Informática	Grupos Reducidos	Laboratorio	Grupos Reducidos	prácticas de campo	Pruebas y/o actividades evaluables	Contenido desarrollado
#1	2	0	0	2	0						Tema 1.
#2	2	0	0	2	0						Tema 1 y 2.
#3	2	0	0	2	0						Tema 2.
#4	2	0	0	2	0						Tema 2.
#5	2	0	0	2	0						Tema 2.
#6	2	0	0	2	0						Tema 2.
#7	2	0	0	2	0						Tema 2 y 3.
#8	2	0	0	2	0	Entrega y defensa Trabajo 1.					Tema 3 y 4.
#9	2	0	0	2	0						Tema 4 y 5.
#10	2	0	0	2	0						Tema 5.
#11	2	0	0	2	0						Tema 5.
#12	2	0	0	2	0						Tema 5.
#13	2	0	0	2	0						Tema 6.
#14	2	0	0	2	0						Tema 6.
#15	2	0	0	2	0	Entrega y defensa Trabajo 2.					Tema 6.
	30	0	0	30	0						