

17-18

MÁSTER UNIVERSITARIO EN I.A.  
AVANZADA: FUNDAMENTOS, MÉTODOS  
Y APLICACIONES

# GUÍA DE ESTUDIO PÚBLICA



## DESCUBRIMIENTO DE INFORMACIÓN EN TEXTOS

CÓDIGO 31101254



Ámbito: GUJ - La autenticidad, validez e integridad de este documento puede ser verificada mediante el "Código Seguro de Verificación (CSV)" en la dirección <https://sed3.uned.es/valida/>



DA3513B2B2EDD3C4CFC5742AF4AA89B243

17-18

DESCUBRIMIENTO DE INFORMACIÓN EN  
TEXTOS

CÓDIGO 31101254

# ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN  
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA  
ASIGNATURA  
EQUIPO DOCENTE  
HORARIO DE ATENCIÓN AL ESTUDIANTE  
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE  
RESULTADOS DE APRENDIZAJE  
CONTENIDOS  
METODOLOGÍA  
SISTEMA DE EVALUACIÓN  
BIBLIOGRAFÍA BÁSICA  
BIBLIOGRAFÍA COMPLEMENTARIA  
RECURSOS DE APOYO Y WEBGRAFÍA



Nombre de la asignatura	DESCUBRIMIENTO DE INFORMACIÓN EN TEXTOS
Código	31101254
Curso académico	2017/2018
Títulos en que se imparte	MÁSTER UNIVERSITARIO EN I.A. AVANZADA: FUNDAMENTOS, MÉTODOS Y APLICACIONES
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	ANUAL
Idiomas en que se imparte	CASTELLANO

## PRESENTACIÓN Y CONTEXTUALIZACIÓN

Ficha técnica:

- Tipo: Optativa
- Duración: Anual
- Créditos Totales y Horas: 6 / 150
- Horas de estudio teórico: 70
- Horas de trabajo práctico: 70
- Horas de actividades complementarias: 10

Reseña del Profesorado:

### **MARTÍNEZ UNANUE, RAQUEL:**

Ha realizado la mayor parte de su actividad docente en el campo de la programación, la algoritmia, la documentación electrónica y la minería de textos. Su actividad investigadora reciente se centra en la minería de textos, especialmente en clustering de documentos tanto monolingües como multilingües, aplicada a diversos tipos de textos (páginas web, noticias, redes sociales, ...) y dominios, en particular el dominio médico.

Desde el año 2000 hasta la actualidad ha colaborado en programas de doctorado de tres universidades: la Universidad Complutense de Madrid, la Universidad Rey Juan Carlos y la UNED.

e.mail: raquel@lsi.uned.es

### **ARAUJO SERNA, LOURDES:**

Desde 1990 ha desarrollado en universidades públicas diversa actividad docente relacionada con los lenguajes de programación y la algoritmia. Desde 1994 hasta la actualidad ha colaborado de forma continua en programa de doctorado de la Universidad Complutense de Madrid y de la UNED. Su tema de tesis fue el estudio del paralelismo de Prolog y



posteriormente ha trabajado en programación con restricciones sobre arquitecturas paralelas, ámbito en el que comenzó a aplicar técnicas de programación evolutiva. En la actualidad investiga en procesamiento del lenguaje natural, recuperación de información y en la aplicación de programación evolutiva a dichas áreas.

e.mail: lurdes@lsi.uned.es

### FRESNO FERNÁNDEZ, VÍCTOR

Su actividad docente se ha centrado principalmente en el campo de la documentación electrónica y su investigación en el campo de la representación automática de textos, en especial de páginas web, así como en la clasificación y clustering de documentos HTML. Desde el año 2000 hasta la actualidad ha trabajado en el Instituto de Automática industrial (CSIC), la Universidad Rey Juan Carlos (URJC) y la Universidad Nacional de Educación a Distancia (UNED), colaborando en los programas de doctorado de dichas universidades.

e.mail: vfresno@lsi.uned.es

## REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Ninguno diferente de los generales de acceso a este programa de posgrado orientado a la investigación.

## EQUIPO DOCENTE

Nombre y Apellidos  
Correo Electrónico  
Teléfono  
Facultad  
Departamento

M. LOURDES ARAUJO SERNA  
lurdes@lsi.uned.es  
91398-7318  
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA  
LENGUAJES Y SISTEMAS INFORMÁTICOS

Nombre y Apellidos  
Correo Electrónico  
Teléfono  
Facultad  
Departamento

RAQUEL MARTINEZ UNANUE  
raquel@lsi.uned.es  
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA  
LENGUAJES Y SISTEMAS INFORMÁTICOS

Nombre y Apellidos  
Correo Electrónico  
Teléfono  
Facultad  
Departamento

VICTOR DIEGO FRESNO FERNANDEZ  
vfresno@lsi.uned.es  
91398-8217  
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA  
LENGUAJES Y SISTEMAS INFORMÁTICOS



## HORARIO DE ATENCIÓN AL ESTUDIANTE

La tutorización de los alumnos se llevará a cabo a través de la plataforma de e-Learning Alf, por teléfono y por correo electrónico:

- Raquel Martínez (coordinadora)

email: raquel@lsi.uned.es

Tfno: 913988725

Horario guardias: Jueves de 11:30 a 13.30 y de 14.30 a 16:30

- Lourdes Araujo

email: lurdes@lsi.uned.es

Tfno: 913987318

Horario de guardias: Jueves de 11 a 13.30 y de 14.30 a 16:00

- Víctor Fresno

email: vfresno@lsi.uned.es

Tfno: 913988217

Horario guardias: Jueves de 15:00 a 19:00

## COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

### RESULTADOS DE APRENDIZAJE

El objetivo del curso es proporcionar al alumno una visión global de las técnicas y tecnologías involucradas en el descubrimiento de información en textos.

El aprendizaje está diseñado para permitir que el alumno adquiera una serie de *destrezas* y *competencias* que se enumeran a continuación:

- Saber lo que es un corpus y conocer los criterios por los que se clasifican, los tipos de anotaciones más comunes y los estándares utilizados.
- Conocer los modelos de representación comúnmente utilizados, así como los métodos de selección y reducción del número de rasgos.
- Saber distinguir los diversos niveles de información lingüística que se pueden utilizar en la representación de textos.
- Saber qué se entiende por minería de textos y conocer las principales técnicas y tecnologías implicadas.
- Saber qué es la clasificación automática de textos y sus características y tipos.
- Conocer diversos tipos de técnicas de aprendizaje automático que se pueden utilizar en la clasificación automática de textos.



- Conocer los modelos estadísticos más utilizados en el procesamiento del lenguaje.
- Saber utilizar las herramientas disponibles de clasificación automática de textos y tener criterios para seleccionar las más adecuadas.
- Saber qué es el clustering de textos y sus características y tipos.
- Conocer diversos tipos de algoritmos de clustering.
- Saber utilizar las herramientas disponibles de clustering de textos y tener criterios para seleccionar las más adecuadas.
- Conocer algoritmos de etiquetado léxico y análisis sintáctico.

Como actividades formativas se tienen:

**1. Actividades teóricas interaccionando con equipos docentes, tutores y compañeros.**

Resolución de dudas de contenido teórico de forma presencial, vía telefónica o en línea sobre la metodología, los contenidos o las actividades a realizar. Intercambio de información a través de un foro virtual.

**2. Actividades prácticas interaccionando con equipos docentes, tutores y compañeros.**

Resolución de dudas de contenido práctico de forma presencial, vía telefónica o en línea sobre la metodología, los contenidos o las actividades a realizar. Intercambio de información a través de un foro virtual.

**3. Actividades teóricas desempeñadas autónomamente.**

Lectura reflexiva y crítica de las orientaciones metodológicas de la asignatura. Estudio de los materiales didácticos.

**4. Actividades prácticas desempeñadas.**

Elaboración de trabajos individuales originales.

## CONTENIDOS

## METODOLOGÍA

La metodología es la general del programa de postgrado; junto a las actividades y enlaces con fuentes de información externas, existe material didáctico propio preparado por el equipo docente. Se trata de una metodología adaptada a las directrices del EEES, de acuerdo con el documento del IUED.

La asignatura no tiene clases presenciales. Los contenidos teóricos se impartirán a distancia,



de acuerdo con las normas y estructuras de soporte telemático de la enseñanza en la UNED. El material docente incluye un resumen de los contenidos de cada tema y distintos tipos de actividades relacionadas con la consulta bibliográfica y la utilización de herramientas asociadas a las tecnologías y técnicas presentadas en el curso. Tratándose de un master orientado a la investigación, las actividades de aprendizaje se estructuran en torno al estado del arte en cada una de las materias del curso y a los problemas en los que se va a focalizar la práctica que el alumno deberá realizar.

## SISTEMA DE EVALUACIÓN

## BIBLIOGRAFÍA BÁSICA

Como bibliografía de la asignatura se deberán estudiar capítulos seleccionados de las siguientes referencias:

- Gordon, A.D. Classification. 2nd Edition. Chapman &Hall/CRC, 1999.
- McEnery, T., Wilson, A. Corpus Linguistics. Edinburgh University Press, 1996.
- Manning, C.D., and Schütze, H. Foundations of Statistical Natural Language Processing. The MIT Press (2000).
- Mitchell, T. Machine Learning. McGraw Hill, 1997. (Nuevos capítulos creados en 2006 y disponibles en <http://www.cs.cmu.edu/%7Etom/mlbook.html> )
- S. Weiss; N. Indurkha; T. Zhang; F. Damerou. Text Mining: Predictive Methods for Analyzing Unstructured Information, 2004.

## BIBLIOGRAFÍA COMPLEMENTARIA

## RECURSOS DE APOYO Y WEBGRAFÍA

La plataforma de e-Learning Alf proporcionará el adecuado interfaz de interacción entre el alumno y sus profesores. Alf es una plataforma de e-Learning y colaboración que permite impartir y recibir formación, gestionar y compartir documentos, crear y participar en comunidades temáticas, así como realizar proyectos online.

Se ofrecerán las herramientas necesarias para que, tanto el equipo docente como el alumnado, encuentren la manera de compaginar tanto el trabajo individual como el aprendizaje cooperativo.

---

## IGUALDAD DE GÉNERO



En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no hayan sido sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.

