

17-18

MÁSTER UNIVERSITARIO EN  
LENGUAJES Y SISTEMAS  
INFORMÁTICOS

# GUÍA DE ESTUDIO PÚBLICA



## MOTORES DE BÚSQUEDA WEB

CÓDIGO 31101042



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



9CA1BDF9AF15834BF272FB25B27D6699

17-18

MOTORES DE BÚSQUEDA WEB  
CÓDIGO 31101042

# 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	MOTORES DE BÚSQUEDA WEB
Código	31101042
Curso académico	2017/2018
Títulos en que se imparte	MÁSTER UNIVERSITARIO EN LENGUAJES Y SISTEMAS INFORMÁTICOS
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	ANUAL
Idiomas en que se imparte	CASTELLANO

## PRESENTACIÓN Y CONTEXTUALIZACIÓN

Tipo	Optativa
	Cuatrimestre
Primero	Créditos/horas totales
6/150	Horas de estudio teórico
100	Horas de prácticas
50	Horas complementarias

Esta es la guía de la asignatura "Motores de búsqueda Web" que se imparte dentro del máster en Lenguajes y Sistemas Informáticos de la UNED. En esta guía se contextualiza la asignatura dentro del máster, se especifican los conocimientos previos necesarios para cursarla con éxito, sus objetivos de aprendizaje y contenidos, y la metodología con la que se estudiará.

## REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Lectura fluida del inglés y conexión a Internet, además de los requisitos propios del máster.

## EQUIPO DOCENTE

Nombre y Apellidos	JULIO ANTONIO GONZALO ARROYO
Correo Electrónico	julio@lsi.uned.es
Teléfono	91398-7922
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS

Nombre y Apellidos	JUAN MARTINEZ ROMO
Correo Electrónico	juaner@lsi.uned.es
Teléfono	91398-9378
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS



## HORARIO DE ATENCIÓN AL ESTUDIANTE

Se realizará mediante la plataforma de posgrados de la UNED.

## COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

## RESULTADOS DE APRENDIZAJE

Objetivos generales de la materia

En este curso se estudian los aspectos esenciales para la recuperación de información en la Web: desde la naturaleza del problema (topología de la Web y características de los usuarios) hasta los retos tecnológicos planteados en la nueva generación de buscadores, pasando por los sistemas clásicos de recuperación de información, la arquitectura básica de un buscador Web, y los sistemas de recuperación basados en notoriedad, de los que Google es el ejemplo canónico.

Al finalizar el curso, el alumno debe ser capaz de plantear la arquitectura completa de un buscador Web, y debe ser capaz de diagnosticar las limitaciones de los sistemas actuales y proponer soluciones novedosas para superarlas.

Destrezas y competencias

El alumno adquirirá las siguientes destrezas y competencias:

Debe tener una visión de conjunto de las tecnologías relacionadas con la búsqueda Web, comprendiendo su evolución temporal y los retos de investigación que se plantean en la actualidad.

Debe ser capaz de realizar una lectura crítica de artículos científicos sobre el tema, de localizar y discriminar información bibliográfica relevante, y de sintetizar información de distintas fuentes.

Debe ser capaz de redactar con rigor científico y de comunicar y debatir con pares (en este caso, sus compañeros) sus análisis y opiniones en torno a los temas de la asignatura.

Debe ser capaz de diagnosticar las limitaciones del campo de investigación en motores de búsqueda Web y apuntar caminos para superarlas.

## CONTENIDOS

## METODOLOGÍA

La general del programa de posgrado. En particular, el alumno realiza dos tipos de actividades en esta asignatura: las relacionadas con la consulta bibliográfica y las de implementación y experimentación. Las primeras son comunes a todos los alumnos y están fijadas dentro del material de estudio correspondiente a cada tema. En una segunda parte de la asignatura, cada alumno realiza un trabajo individual sobre un tema acordado con el equipo docente. Todo el material de estudio está disponible en el entorno virtual del



posgrado, y toda la interacción entre profesores y alumnos se puede llevar a cabo en este entorno.

## SISTEMA DE EVALUACIÓN

### BIBLIOGRAFÍA BÁSICA

Arvind Arasu, Junghoo Cho, Hector García-Molina, Andreas Paepcke and Sriram Raghavan. Searching the Web. ACM Transactions on Internet Technology, vol. 1, n. 1, August 2001, pages 2-43.

### BIBLIOGRAFÍA COMPLEMENTARIA

#### **Tema 1.** Características de la búsqueda de información en la WWW

Sobre estructura de la WWW:

- Kleinberg, JM. Hubs, authorities, and communities, ACM computing surveys 1999.

<http://www.cs.brown.edu/memex/ACMCSHT/10/10.html>

- A Borodin, GO Roberts, JS Rosenthal, P. Tsaparas. Finding authorities and hubs from link structures on the World Wide Web. Proc. WWW 2001.

<http://www10.org/cdrom/papers/314/>

Sobre tipología de búsquedas web:

- Rose, D. y Levinson, D. Understanding User Goals in Web Search. WWW 2004.

<http://wwwconf.ecs.soton.ac.uk/archive/00000537/01/p13-rose.pdf>

Sobre navegación versus consulta:

- Marti A. Hearst. Next Generation Web Search: Setting Our Sites In IEEE Data Engineering Bulletin, 2002.

<http://www.sims.berkeley.edu/hearst/papers/data-engineering>

- A. Peñas, F. Verdejo, J. Gonzalo, 2002. Terminology Retrieval: towards a synergy between thesaurus and free text searching. Advances in Artificial Intelligence - IBERAMIA 2002, LNAI 2527.

<http://nlp.uned.es/pergamus/pubs/iberamia2002.pdf>

#### **Tema 2.** Arquitectura básica de un motor de búsqueda.

Sobre crawling:

- J Cho, H Garcia-Molina, L Page. Efficient Crawling Through URL Ordering, WWW 1998.

- Allan Heydon and Marc Najork. Mercator: A Scalable, Extensible Web Crawler. In Proceedings of World Wide Web Conference, 1999, pages 219-229.

Sobre soporte hardware:



- L. A. Barroso, J. Dean, U. Hoelzle. Web search for a planet: the Google cluster architecture. IEEE 2003.

**Tema 3.** Motores de búsqueda pre-Google: recuperación basada en contenidos.

- D Hiemstra. Using Language Models for Information Retrieval. CTIT Ph.D. Thesis, 2001.
- G Salton, A Wong, CS Yang. A Vector Space Model for Automatic Indexing. Comm. ACM, 1975.
- N Fuhr. Probabilistic Models in Information Retrieval. The Computer Journal, 1992.

**Tema 4.** Motores de búsqueda actuales (generalistas): recuperación basada en autoridad.

Referencias:

- M Hollander. Google's PageRank Algorithm to Better Internet Searching. TR UMN.
- Brin, S. y Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. WWW 1998.
- CHQ Ding, X He, P Husbands, H Zha, HD Simon. PageRank, HITS and a unified framework for link analysis. SIGIR 2002.
- TH Haveliwala. Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search. IEEE T. on Knowledge and data engineering, 2003.

**Tema 5.** Temas avanzados.

- Guha, R. y Garg, A. Disambiguating People in Search. Proc. WWW 2004.
  - S Lawrence, NJ Princeton. Context in Web Search, IEEE data engineering bulletin, 2000.
  - J Sivic, A Zisserman. Video google: A text retrieval approach to object matching in videos, ICCV 2003.
  - SK Bhavnani, CK Bichakjian, TM Johnson, RJ Little. Strategy Hubs: Next-Generation Domain Portals with Search Procedures. Proc. ACM Conference on Human Factors in Computing Systems, 2003, ACM Press NY, USA.
  - T Berners-Lee, J Hendler, O Lassila. The semantic Web. Scientific American, 2001.
  - J Heflin, J Hendler. A Portrait of the Semantic Web in Action. IEEE Intelligent Systems, 2001.
  - S Eissen, B Stein. Analysis of Clustering Algorithms for Web-Based Search. Springer-Verlag, 2002.
  - J. Cigarrán, A. Peñas, J. Gonzalo, F. Verdejo, 2005. Automatic selection of noun phrases as document descriptors in an FCA-based Information Retrieval system. ICFA 2005. Springer LNCS 3403.
- Search Engines: Technology, Society, and Business. Materiales online del curso:  
<http://www.sims.berkeley.edu/courses/is141/f05/schedule.html>



## RECURSOS DE APOYO Y WEBGRAFÍA

La plataforma de enseñanza virtual de posgrados de la UNED será la interfaz de interacción entre el alumno y sus profesores. Esta plataforma permite impartir y recibir formación, gestionar y compartir documentos, crear y participar en comunidades temáticas, así como realizar proyectos online.

---

## 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.

