

**Guía docente de la asignatura**

Asignatura	MÉTODOS MATEMÁTICOS APLICADOS AL DESARROLLO DE SISTEMAS Y SERVICIOS DE INTERNET		
Materia	SISTEMAS, APLICACIONES Y SERVICIOS DE INTERNET		
Módulo	TECNOLOGÍAS INFORMÁTICAS		
Titulación	MASTER EN INGENIERIA INFORMATICA		
Plan	510	Código	53191
Periodo de impartición	2º CUATRIMESTRE	Tipo/Carácter	OPTATIVA
Nivel/Ciclo	MASTER	Curso	PRIMERO
Créditos ECTS	3 ECTS		
Lengua en que se imparte	CASTELLANO		
Profesor/es responsable/s	CARLOS MARIJUÁN LÓPEZ		
Datos de contacto (E-mail, teléfono...)	DESPACHO: 2D041 TELÉFONO: 983 423731 E-MAIL: marijuan@mat.uva.es		
Horario de tutorías	Véase www.uva.es → Centros → Campus de Valladolid → Escuela Técnica Superior de Ingeniería Informática → Tutorías		
Departamento	MATEMÁTICA APLICADA		



1. Situación / Sentido de la Asignatura

1.1 Contextualización

En esta asignatura nos interesamos por dos problemas relativos a la información en la red de Internet: la ordenación de los resultados obtenidos por los motores de búsqueda y la clasificación de grandes volúmenes de información interrelacionada. Estudiaremos técnicas matemáticas que soportan el funcionamiento de los algoritmos con que actualmente se abordan estos problemas y las utilizaremos para mostrar el interés de su conocimiento aplicándolas a casos reales.

1.2 Relación con otras materias

Las técnicas espectrales estudiadas en esta asignatura son útiles en materias relacionadas con *Spectral Clustering*, *Data Mining* o *Big Data*.

1.3 Prerrequisitos

Nociones básicas de teoría de grafos, álgebra lineal y cálculo infinitesimal generalmente estudiados en los títulos de grado que permiten el acceso a este Máster.

2. Competencias

2.1 Generales

Código	Descripción
CG4	Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática.
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 multidisciplinares, siendo capaces de integrar estos conocimientos.

2.2 Específicas

Código	Descripción
CET2	Capacidad de comprender y saber aplicar el funcionamiento y organización de Internet, las tecnologías y protocolos de redes de nueva generación, los modelos de componentes, software intermediario y servicios.
CET9	Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

3. Objetivos

Código	Descripción
CET2.1	Conocer los fundamentos de la Teoría de Grafos que permiten comprender la organización de grandes redes y servicios de Internet.
CET9.1	Conocer los fundamentos de la Teoría Espectral de Grafos en que se basan los más relevantes servicios de Internet.
CET2.2	Ser capaz de utilizar estos recursos matemáticos para comprender y mejorar algunos servicios de Internet.
CET9.2	Ser capaz de aplicar técnicas de particionado espectral para la clasificación y comprensión de información de grandes redes en Internet.

**4. Tabla de dedicación del estudiante a la asignatura**

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teórico-prácticas (T/M)	14	Estudio y trabajo individual	20
Clases prácticas de aula (A)		Estudio y trabajo en grupo	25
Laboratorios (L)	10		
Prácticas externas, clínicas o de campo			
Seminarios (S)	4		
Tutorías grupales (TG)			
Evaluación (fuera del periodo oficial de exámenes)	2		
Total presencial	30	Total no presencial	45

5. Bloques temáticos**Bloque 1: Ordenación de búsquedas en Internet**Carga de trabajo en créditos ECTS: **a. Contextualización y justificación**

Los motores de búsqueda de información en Internet utilizan algoritmos para ordenar los resultados de las búsquedas. Estos algoritmos valoran la importancia de cada página web por procedimientos que dependen fundamentalmente de los vínculos entre las páginas de Internet. Conocer el funcionamiento de estos algoritmos y los fundamentos matemáticos que los soportan puede ser útiles para comprender qué modificaciones pueden hacerse en una red local para mejorar la importancia de sus páginas principales y, de este modo, mejorar su visibilidad en Internet. Se estudiará especialmente el algoritmo PageRank de Google.

b. Objetivos de aprendizaje

Código	Descripción
CET2.1	Conocer los fundamentos de la Teoría de Grafos que permiten comprender la organización de grandes redes y servicios de Internet.
CET9.1	Conocer los fundamentos de la Teoría Espectral de Grafos en que se basan los más relevantes servicios de Internet.
CET2.2	Ser capaz de utilizar estos recursos matemáticos para comprender y mejorar algunos servicios de Internet.

c. Contenidos**TEMA 1: Conectividad en grafos**

- 1.1 Estructuras jerárquicas. Árboles dirigidos
- 1.2 Digrafos acíclicos
- 1.3 Conexión fuerte en digrafos
- 1.4 Digrafo condensación



TEMA 2: Teoría de Perron-Frobenius

- 2.1 Estructura espectral de matrices no negativas
- 2.2 Matrices positivas, primitivas e irreducibles
- 2.3 Matrices reducibles
- 2.4 Análogos diigráficos

TEMA 3: La importancia de una página web

- 3.1 Fórmulas para valorar la importancia de una página web
- 3.2 Aplicación de los temas 1 y 2 al diseño de algoritmos de ordenación de búsquedas
- 3.3 Modificaciones en una web local jerárquica
- 3.4 Clasificación de redes jerárquicas por su importancia
- 3.5 Optimización de la estructura de vínculos

TEMA 4: Topologías más complejas

- 4.1 Árboles bidireccionales
- 4.2 Árboles cíclicos
- 4.3 Modificaciones esenciales
- 4.4 Estructuras más generales
- 4.5 Reglas para la mejora de visibilidad

d. Métodos docentes

Ver 8. Anexo: Métodos docentes

e. Plan de trabajo

Ver 9. Anexo: Cronograma de actividades previstas

f. Evaluación

Ver 7. Sistema de calificaciones – Tabla resumen.

g. Bibliografía básica

A. Arratia and C. Marijuán, *Ranking pages and the topology of the web*.
<http://arxiv.org/abs/1105.1595v2>

S. Brin, R. Motwami, L. Page and T. Winograd, *The PageRank citation ranking: Bringing order to the web*. Technical Report, Comp. Sci. Dept., Stanford University, 1998.

M. Brinkmeier, *PageRank revisited*, ACM Transactions on Internet Technologies, 2006, **6** (3).

N. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.



h. Bibliografía complementaria

A. Arasu, J. Novak, A. Tomkins and J. Tomlin, *PageRank computation and the structure of the Web: Experiments and algorithms*. The Eleventh International World Wide Web Conference, Posters 2002.

M. Bianchini, M. Gori and F. Scarselli, *Inside PageRank*. ACM Transactions on Internet Technologies, 2005 **4**(4).

S. Brin and L. Page, *The anatomy of a large scale hypertextual web search engine*. Computer Networks and ISDN Systems, 1998, **33**, 107-117.

M. Brinkmeier. *Distributed calculation of PageRank using strongly connected components*, Proceedings of the I2CS'05 in Paris (LNCS), 2005.

T. Cormen, C. Leiserson, R. Rivest and C. Stein. *Introduction to Algorithms*, The MIT Press, 2nd. Edition, 2001.

F. Harary, *Graph Theory*, Addison--Wesley, 1972.

D. E. Knuth, *The Art of Computer Programming*, Addison-Wesley, volumes 1, 2, and 3, 3rd edition, 1998.

S. D. Kamvar, T. H. Haveliwala, C. D. Manning, and G. H. Golub. *Exploiting the block structure of the Web for computing PageRank*. Stanford University Technical Report, 2003.

A. N. Langville and C. D. Meyer, *Deeper Inside PageRank*, Internet Mathematics, **2005 1** (3): 335--380.

i. Recursos necesarios

El profesor de la asignatura proporcionará los recursos necesarios para realizar este bloque.

Bloque 2: Clasificación de información interrelacionada en Internet

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

En este bloque pretendemos clasificar grandes volúmenes de información en grupos cuyos elementos están densamente relacionados entre sí dentro de cada grupo, pero escasamente relacionados si pertenecen a grupos distintos. Desarrollaremos la Teoría de Particionado Espectral por ser especialmente eficiente cuando se trata de clasificar volúmenes de información de inmensa magnitud. La dotaremos de una implementación real en un sistema informático y trataremos con herramientas apropiadas para su aplicación sobre distintas áreas de trabajo en Internet.

b. Objetivos de aprendizaje

Código	Descripción
CET2.1	Conocer los fundamentos de la Teoría de Grafos que permiten comprender la organización de grandes redes y servicios de Internet.
CET9.1	Conocer los fundamentos de la Teoría Espectral de Grafos en que se basan los más relevantes servicios de Internet.
CET9.2	Ser capaz de aplicar técnicas de particionado espectral para la clasificación y comprensión de información de grandes redes en Internet.



c. Contenidos

TEMA 5: Fundamentos Matemáticos

- 5.1 Formas cuadráticas
- 5.2 Normas matriciales. Ortogonalidad.
- 5.3 Matrices asociadas a grafos.
- 5.4 Particionado de conjuntos. Algoritmo *k-means*

TEMA 6: Teoría de Particionado Espectral

- 6.1 Grafos de similitud
- 6.2 Matrices laplacianas
- 6.3 Algoritmos de Particionado Espectral
- 6.4 Análisis y justificación de los algoritmos

TEMA 7: Consideraciones prácticas

- 7.1 Computación de autovectores. Método de Arnoldi
- 7.2 Acondicionamiento de matrices
- 7.3 Elección del número de núcleos, matriz laplaciana y función de similitud

TEMA 8: Exploración en la web

- 8.1 *Web crawlers*
- 8.2 La araña WIRE
- 8.3 Aplicación a un problema real

d. Métodos docentes

Ver 8. Anexo: Métodos docentes

e. Plan de trabajo

Ver 9. Anexo: Cronograma de actividades previstas

f. Evaluación

Ver 7. Sistema de calificaciones – Tabla resumen.



g. Bibliografía básica

U. von Luxburg, *A Tutorial on Spectral Clustering*. Technical Report No. TR-149. Max-Planck-Intitut fur biologische Kybernetik. Department for Empirical Inference, 2006.

A. Ng, M. Jordan and Y. Weiss. *On spectral clustering: analysis and an algorithm*. En T. Dietterich, S. Becker and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14*. MIT Press, 2002.

J. Shi and J. Malik. *Normalized cuts and image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, **22** (8), 888-905.

C. Castillo. *Effective Web Crawling* (Ph. D. in Computer Science), 2004.

h. Bibliografía complementaria

F. Chung. *Spectral graph theory*, Washington: Conference board of the Mathematical Sciences, 1997.

D. Arthur and S. Vassilvitskii, *How Slow is the k-means Method?* Proceedings of the 2006 Symposium on Computational Geometry (SoCG), 2006.

D. Arthur and S. Vassilvitskii, *k-means++ The Advantages of Careful Seeding*. Symposium on Discrete Algorithms (SODA), 2007.

Z. Markov and D.T. Larose, *Data Mining the Web: uncovering patterns in web content, structure and usage*. Wiley Series on methods and Applications in Data Mining. Wiley-Interscience. A John Wiley & Sons, Inc. Publication, 2007.

B. Mohar, *Some applications of Laplace eigenvalues of graphs*. En G. Hahn and G. Sabidussi (Eds.), *Graph simmetry: Algebraic Methods and applications* (Vol. NATO ASI Ser. C 497, p. 225-275). Kluwer 1991.

D.S. Watkins. *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*. Siam, 2007.

i. Recursos necesarios

El profesor de la asignatura proporcionará los recursos necesarios para realizar este bloque.

6. Temporalización (por bloques temáticos)

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Bloque 1: Ordenación de búsquedas en Internet	1.3 ECTS	Semanas 1 a 4
Bloque 2: Clasificación de información interrelacionada en Internet	1.7 ECTS	Semanas 4 a 8

**7. Sistema de calificaciones – Tabla resumen**

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Proyecto Bloque 1	30%	Entrega en la semana 4
Examen Temas 1 a 7	40%	En la primera sesión de la semana 7
Proyecto Bloque 2	30%	Entrega en la semana 8
Seminarios	adicional	En las semanas indicadas en el Anexo 9
Examen final		Periodo de exámenes

CRITERIOS DE CALIFICACIÓN

- **Evaluación continua:**
Se puede aprobar esta asignatura mediante la realización del examen y de los dos proyectos indicados en la tabla anterior. Se requiere obtener al menos el 40% en cada una de las tres pruebas y que la suma de las tres no sea inferior al 50% de la nota total.
La participación activa y creativa en los Seminarios será evaluada positivamente y la correspondiente calificación se añadirá a la nota total obtenida mediante las otras tres pruebas.
- **Convocatoria ordinaria:**
Cada alumno se examinará, al menos, de las pruebas consideradas en la evaluación continua en las que no haya obtenido el 40% de la calificación. También podrá examinarse, a su elección, de las pruebas en las que haya obtenido al menos el 40% de la calificación. Cuando haya dos calificaciones de una misma prueba se considerará la superior de las dos.
- **Convocatoria extraordinaria:**
Mismos criterios aplicados ahora a la evaluación continua y/o a la convocatoria ordinaria.

8. Anexo: Métodos docentes

Actividad	Metodología
Clases teórico-prácticas	<ul style="list-style-type: none">• Clase magistral participativa• Estudio de casos• Resolución de problemas
Laboratorio	<ul style="list-style-type: none">• Clase magistral participativa• Resolución de casos prácticos con apoyo informático• Realización de un proyecto en grupo (2/3 alumnos) guiado por el profesor, siguiendo un enfoque colaborativo.
Seminarios	<ul style="list-style-type: none">• Sesiones de debate entre alumnos y profesor sobre su aprendizaje, las técnicas estudiadas y su aplicación práctica a casos reales.

**9. Anexo: Cronograma de actividades previstas**

Semana	Sesión	Contenido	Actividad	Estudio individual	Estudio en grupo
1	1 2h.	Tema 1	Clase T/M	2h	
	2 2h.	Tema 2	Clase T/M	2h	
2	1 2h.	Tema 3	Clase T/M	2h	
	2 2h.	Tema 3	Laboratorio		5h
3	1 1h.	Temas 1, 2 y 3	Seminario	2h	
	1 1h.	Tema 4	Clase T/M	2h	
	2 2h.	Tema 4	Laboratorio		5h
4	1 1h.	Temas 1 a 4	Seminario	1h	
	1 1h.	Tema 5	Clase T/M	1h	
	2 2h.	Tema 5	Clase T/M	2h	
5	1 2h.	Tema 6	Clase T/M	2h	
	2 2h.	Tema 6	Laboratorio		5h
6	1 2h.	Tema 7.1, 7.2	Clase T/M	2h	
	2 2h.	Temas 5, 6 y 7	Seminario	2h	
7	1 2h.	Temas 1 a 7	Examen	discrecional	
	2 2h.	Tema 8	Laboratorio		5h
8	1 2h.	Tema 8	Laboratorio		5h