

22-23

MÁSTER UNIVERSITARIO EN
INGENIERÍA INFORMÁTICA

GUÍA DE ESTUDIO PÚBLICA



SISTEMAS DE INFORMACIÓN NO ESTRUCTURADA

CÓDIGO 31106116

UNED

22-23

SISTEMAS DE INFORMACIÓN NO
ESTRUCTURADA
CÓDIGO 31106116

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	SISTEMAS DE INFORMACIÓN NO ESTRUCTURADA
Código	31106116
Curso académico	2022/2023
Título en que se imparte	MÁSTER UNIVERSITARIO EN INGENIERÍA INFORMÁTICA
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	SEMESTRE 2
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

Esta asignatura de carácter optativo pertenece al Máster Universitario en "Ingeniería Informática" que se imparte por profesores del Departamento de "Lenguajes y Sistemas Informáticos" de la ETSI Informática de la UNED. Esta asignatura es de carácter anual con una carga de 6 ECTS. El objetivo fundamental es dar una visión integradora de técnicas relacionadas con los sistemas de acceso a la información y en concreto en el caso de información no estructurada.

En la actualidad, la cantidad de información online es prácticamente ilimitada, lo que ha supuesto una de las transformaciones sociales más importantes en este comienzo de siglo. Las aplicaciones prácticas son muy amplias, incluyendo estudios de mercado, buscadores especializados, sistemas de recomendación, análisis de redes sociales, etc.

En este contexto, uno de los principales retos es el acceso a información no estructurada, es decir, a información textual o multimedia, debido a la variabilidad del lenguaje y a la dificultad de interpretar y organizar contenidos. En otras palabras, cómo enlazar, agrupar o clasificar toda esta información según las necesidades del usuario. Dentro del acceso a información no estructurada se incluyen las tecnologías de búsqueda, los clasificadores automáticos, las técnicas de agrupación (clustering), la extracción de información y la minería de datos textuales, etc.

Un sistema de acceso a la información cubre aspectos tan diversos como la eficiencia de acceso y escalabilidad de bases de datos, la compresión, el diseño de interfaces de usuario, reconocimiento de patrones en imágenes, evaluación de componentes o la interacción con el usuario. No es posible cubrir todos estos aspectos en un curso. Sin embargo, a través de esta asignatura el alumno adquirirá las capacidades básicas para el desarrollo de sistemas, centrándose en los desafíos específicos dados por el uso de información no estructurada.

Las competencias de esta asignatura se pueden consultar en la guía del máster. La asignatura conecta con la asignatura de "Aprendizaje Automático", por ejemplo, en el uso de dichas técnicas para clasificación de documentos o ranking supervisado. Conecta también, de forma colateral con las asignaturas de "Temas Avanzados en Redes e Internet", "Gestión de la Información en la Web" y "Métodos de Desarrollo y Análisis de Entornos Colaborativos y Redes Sociales".

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Los propios del máster. Además son necesarios conocimientos de inglés técnico (lectura y escritura) para manejar las fuentes bibliográficas.

EQUIPO DOCENTE

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

ENRIQUE AMIGO CABRERA (Coordinador de asignatura)
enrique@lsi.uned.es

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

JORGE AMANDO CARRILLO DE ALBORNOZ CUADRADO
jcalbornoz@lsi.uned.es

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

JULIO ANTONIO GONZALO ARROYO
julio@lsi.uned.es

HORARIO DE ATENCIÓN AL ESTUDIANTE

Los alumnos serán tutorizados vía los cursos virtuales en los días lectivos, o vía telefónica en horario de tutorías, atendiendo a las consultas de carácter teórico y dudas relativas a los trabajos prácticos. El horario de tutorías es de 3 a 7 de la tarde, todos los jueves del periodo lectivo, pudiendo contactar con el equipo docente en los teléfonos 913988651 y 913987922, además de los correos electrónicos indicados en la sección del equipo docente y la dirección postal C/Juan del Rosal 16, 28040, Madrid. Asimismo, se establecerán foros de discusión para temas concretos.

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

Competencias Básicas:

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

CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación

de sus conocimientos y juicios

CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Competencias Generales:

G1 - Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la ingeniería informática.

G2 - Capacidad para la dirección de obras e instalaciones de sistemas informáticos, cumpliendo la normativa vigente y asegurando la calidad del servicio.

G4 - 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.

G8 - 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.

Competencias Transversales:

CT1 - Capacidad para emprender y liderar proyectos innovadores en entornos científicos, tecnológicos y multidisciplinarios.

CT2 - Capacidad para tomar decisiones y formular juicios basados en criterios objetivos (datos experimentales, científicos o de simulación disponibles).

Competencias Específicas:

TI1 - Capacidad para modelar, diseñar, definir la arquitectura, implantar, gestionar, operar, administrar y mantener aplicaciones, redes, sistemas, servicios y contenidos informáticos.

TI2 - 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.

TI5 - Capacidad para analizar las necesidades de información que se plantean en un entorno y llevar a cabo en todas sus etapas el proceso de construcción de un sistema de información.

TI9 - 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.

TI11 - Capacidad para conceptualizar, diseñar, desarrollar y evaluar la interacción persona-ordenador de productos, sistemas, aplicaciones y servicios informáticos.

TI12 - Capacidad para la creación y explotación de entornos virtuales, y para la creación, gestión y distribución de contenidos multimedia.

RESULTADOS DE APRENDIZAJE

Los resultados del aprendizaje esperados en este curso son:

1. Conocimiento sobre las oportunidades de desarrollo de sistemas que ofrece la información no estructurada disponible en Internet y otros entornos locales.
2. Capacidad para realizar un diseño global de un sistema de acceso a información no estructurada en base a las necesidades del usuario.
3. Capacidad para identificar las barreras y limitaciones a las que se enfrenta un sistema de información por el hecho de tratar información no estructurada.
4. Comprensión del funcionamiento de Internet, tanto en cuanto a páginas estáticas como a redes sociales (WEB 2.0).
5. Conocimiento y experiencia en el manejo de las técnicas más extendidas para el tratamiento de este tipo de información.
6. Capacidad para diseñar e implementar un sistema básico de acceso a información en Internet desde el proceso de descarga de datos a la organización de la información.

CONTENIDOS

1. Introducción

El objetivo de este tema consiste en dar una visión global de un sistema de acceso a información no estructurada, con todos sus componentes principales. Como caso paradigmático, siguiendo los primeros capítulos del libro base, nos centraremos en el planteamiento general de un sistema genérico de recuperación de información orientado a texto, que incluya los componentes básicos de este tipo de sistemas como son la adquisición de texto, procesamiento, interacción con el usuario, y el ranking de documentos.

Los resultados específicos para este tema son:

- Entender qué es la información no estructurada y su relevancia hoy en día.
- Conocer la estructura y los elementos básicos de un sistema de acceso a información no estructurada tomando como referencia los sistemas de recuperación de información.
- Ser capaz de identificar qué papel jugarían los componentes básicos de un sistema de recuperación de información en un escenario concreto.

La motivación y arquitectura de un sistema de información no estructurada guarda cierta relación con los conocimientos adquiridos en la asignatura del mismo master *Métodos de Desarrollo y Análisis de Entornos Colaborativos y Redes Sociales*.

Teniendo en cuenta las limitaciones en cuanto a horas de dedicación en el presente curso, se evitará el esfuerzo de síntesis de fuentes por medio de la focalización en el libro base. Dado que el libro se centra en sistemas de recuperación de documentos de texto, se aportará en el curso material introductorio describiendo la importancia de la información no estructurada en la actualidad, además de referencias a artículos de interés. Éstos estarán disponibles en el curso virtual.

2. Adquisición y procesamiento de textos

La información no estructurada no se ciñe necesariamente a documentos de textos extraídos de internet. Existen bases de datos documentales en formato digital que no forman parte de la WEB, y existe también otros tipos de información como imágenes, audios, etc. Sin embargo, la inmensa mayoría de la información que puede ser explotada en un contexto comercial mediante un sistema de información no estructurada es la que se encuentra accesible en Internet y en formato textual. En este tema se describen las herramientas básicas para obtener y procesar estos datos.

Los resultados de aprendizaje esperados para este tema son fundamentalmente:

- Conocer los retos a los que se enfrenta el proceso de adquisición dentro de un sistema de acceso a información no estructurada.
- Conocer las técnicas básicas de tratamiento de texto que se aplican en este tipo de sistemas.
- Ser capaz de identificar las necesidades de adquisición y procesamiento básico de texto en un problema concreto.

Este tema en concreto guarda estrecha relación con los primeros temas de la asignatura de este master *Métodos de Desarrollo y Análisis de Entornos Colaborativos y Redes Sociales*, dado que mucha de la información no estructurada disponible en la red se encuentra en este tipo de entornos.

El estudio de este tema se apoyará fundamentalmente en el libro base, junto con la descripción del escenario a tratar y enlaces relacionados. Los enlaces relacionados propuestos por el equipo docente dependerán del problema propuesto, que estará disponible en el curso virtual.

3. Técnicas de organización de la información: ranking, agrupación y clasificación de documentos

En este tema nos centraremos en las técnicas de organización de la información no estructurada. Básicamente, los documentos o unidades de información pueden ordenarse (ranking), clasificarse en base a un conjunto predefinido de categorías o agruparse en base a algún criterio de similitud.

Los resultados de aprendizaje esperados para este tema son fundamentalmente:

- Conocer las técnicas básicas de ranking, clasificación y agrupación en información no estructurada.
- Ser capaz de identificar las necesidades de clasificación, agrupación o clasificación de texto en un problema concreto y la adecuación de las técnicas existentes.

Las técnicas de clasificación y agrupación tienen su fundamento en tecnologías básicas de inteligencia artificial, por lo que este tema se interrelaciona con las asignaturas de Aprendizaje Automático y Fundamentos de la Inteligencia Artificial de este master.

El estudio de este tema se apoyará fundamentalmente en el libro base, junto con la descripción del problema a tratar y enlaces relacionados. Los enlaces relacionados propuestos por el equipo docente dependerán del problema propuesto, que estará disponible en el curso virtual.

4. Redes sociales y otros objetos de búsqueda

Muchas de las técnicas empleadas para el acceso y manejo de información no estructurada no se apoyan directamente en el análisis de contenidos (textual o multimedia) sino en cómo interactúan los usuarios entre sí y con las piezas de información. Es decir, el proceso de acceso a la información es colaborativo. Bajo esta filosofía, se recomienda a usuarios productos que han sido consumidos por usuarios próximos, o se identifica la relevancia de un video o post en función del número de “likes”. A medida que las redes sociales y los procesos colaborativos cobran peso, este tipo de técnicas adquieren más relevancia en el ámbito comercial. Por otro lado, es importante tener en cuenta que los contenidos textuales son solo una parte de toda la información no estructurada. En este tema se abordarán también otros objetos de búsqueda distintos del documento tradicional. En particular, se estudiarán algunos conceptos básicos la búsqueda de entidades, en donde el objeto de búsqueda es una persona, organización o concepto en general, y la búsqueda de imágenes y audio.

Análogamente a los temas anteriores, los resultados de aprendizaje esperados para este tema son fundamentalmente:

- Conocer los mecanismos de acceso a información no estructurada en entornos sociales.
- Adquirir una visión global sobre las técnicas de acceso a información multimedia y búsqueda de entidades.
- Ser capaz de identificar el papel que pueden jugar dichos mecanismos en un escenario concreto.

Este tema está íntimamente ligado a la asignatura de Métodos de Desarrollo y Análisis de Entornos Colaborativos y Redes sociales.

METODOLOGÍA

Las diferentes asignaturas que integran este Master se imparten conforme a la metodología no presencial que caracteriza a la UNED, en la cual prima el autoaprendizaje del alumno, pero asistido por el profesor y articulado a través de diversos sistemas de comunicación docente-discente. El Máster en Ingeniería Informática se imparte con apoyo en una plataforma virtual interactiva de la UNED, donde el alumno encuentra tanto materiales didácticos básicos como complementarios, informaciones, noticias, ejercicios y también permite la evaluación correspondiente a las diferentes materias.

Las diferentes actividades formativas se organizan como sigue:

- Estudio de contenidos: 50 horas.
- Tutorías:10h
- Actividades en la plataforma virtual: 10 horas.
- Trabajos individuales: 40 horas.
- Trabajos en equipo: 10 horas.
- Práctica 30 horas.

En cada caso se refieren a:

1. Estudios de contenidos teóricos: Lectura de las orientaciones, material multimedia y estudio del libro base.
2. Tutorías: Interacción presencial o virtual con los tutores y aclaración de cuestiones planteadas por los estudiantes.
3. Actividades en la plataforma virtual: Lectura de las orientaciones para la realización de las actividades prácticas planteadas, problemas y ejercicios, e interacción con los compañeros en el foro.
4. Trabajos individuales: Realización de trabajos y resúmenes guiados por cuestiones planteadas por el equipo docente. Todos los trabajos realizados se apoyarán en un escenario hipotético de acceso a información no estructurada definido al comienzo del curso.
5. Trabajos en equipo: Discusión y evaluación conjunta de trabajos.
6. Práctica: Uso de herramientas básicas de recuperación de información sobre conjuntos reducidos de documentos, o bien la implementación de un algoritmo en tareas más específicas como problemas de recomendación o agrupación de documentos. Los ejercicios prácticos se realizarán mediante software libre disponible y recursos de libre disposición.

El estudio de la asignatura se realiza a través de tres tipos de actividades:

- Estudio del libro base* (y materiales complementarios) guiado por análisis. En cada tema, el alumno deberá analizar la aplicabilidad de cada estudiada dentro del contexto de un problema específico que se planteará al comienzo del curso. El alumno partirá de cuestiones concretas para realizar un resumen guiado.
- Análisis comparativo con los trabajos del resto de alumnos*. Una vez entregada una actividad, el estudiante tendrá acceso a las respuestas de sus compañeros, y participará en

una discusión colectiva a través de un foro específico. La actividad consistirá en aportar en el foro las consideraciones oportunas tras el análisis de las respuestas de los compañeros.

Es decir, qué modificaría de su primer resumen tras leer el resto de trabajos.

- Práctica.** Consistirá en la implementación de un componente software dentro del problema planteado al principio del curso.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen	Examen de desarrollo
Preguntas desarrollo	4
Duración del examen	90 (minutos)
Material permitido en el examen	

Ninguno

Criterios de evaluación

Se valorará la capacidad de análisis y manejo de los conceptos fundamentales, más que el conocimiento en detalle de técnicas específicas.

% del examen sobre la nota final	50
Nota del examen para aprobar sin PEC	5
Nota máxima que aporta el examen a la calificación final sin PEC	5
Nota mínima en el examen para sumar la PEC	4
Comentarios y observaciones	

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad	Si
Descripción	

El examen es de carácter obligatorio y estará centrado en cuestiones teórico-prácticas relacionadas con los conceptos tratados en la realización de resúmenes guiados. Por ello, se recomienda seguir las guías de los resúmenes de evaluación continua para identificar aquellos conceptos de mayor peso en el examen.

Criterios de evaluación

Se valorará la capacidad de análisis y manejo de los conceptos fundamentales, más que el conocimiento en detalle de técnicas específicas.

Ponderación de la prueba presencial y/o los trabajos en la nota final	El examen aportará un máximo de cinco puntos a la nota final. El examen es de carácter obligatorio, y es necesario obtener un mínimo de dos puntos para poder aprobar la asignatura.
---	--

Fecha aproximada de entrega	
Comentarios y observaciones	

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?

Si, PEC no presencial

Descripción

Realización de la práctica obligatoria. El objetivo de la práctica es interiorizar las limitaciones a las que están sujetos un sistemas de acceso a información no estructurada mediante la implementación de una solución a un determinado problema. Se recomienda el uso de los foros en esta fase para compartir todas las dudas o problemas que surjan.

Criterios de evaluación

Realización de la práctica obligatoria. Se valorará ante todo el análisis de resultados, más que la efectividad del sistema implementado. Se recomienda el uso de los foros en esta fase para compartir todas las dudas o problemas que surjan.

Ponderación de la PEC en la nota final

Aportará un máximo de 3 puntos a la nota final. Es indispensable aprobar la práctica para aprobar la asignatura en cualquiera de las convocatorias.

Fecha aproximada de entrega

Las fechas de realización de la práctica serán las semanas anteriores a la realización del examen.

Comentarios y observaciones

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s?

Si, no presencial

Descripción

Realización de resúmenes guiados. El objetivo de estas actividades es la asimilación de contenidos por medio del análisis. Por ello, se recomienda ir generando dichos resúmenes a medida que se leen los contenidos. Los resúmenes vendrán guiados por preguntas concretas en relación a la aplicación de las técnicas descritas en el libro sobre un escenario planteado al principio de curso. Los resúmenes deben de ser por tanto breves. Es importante tener en cuenta que una vez pasada la fecha de entrega, los resúmenes generados serán accesibles para el resto de alumnos. Se recomienda hacer uso del curso virtual para resolver cualquier duda que pudiera surgir.

Actividades de discusión. El objetivo de estas actividades es la auto-evaluación y discusión de conceptos tratados en la asignatura, asimilando diferentes perspectivas del análisis mediante el estudio comparativo de las aportaciones del resto de los alumnos. Una vez pasada la fecha de entrega de cada resumen, todos ellos serán accesibles para el resto de alumnos. Esta actividad consistirá en aportar en los foros al menos un comentario (entre 200 y 300 palabras) en relación a lo que la lectura del resto de trabajos le haya podido aportar.

Criterios de evaluación

Realización de resúmenes guiados. Se valorará la concreción y fundamentación de los argumentos.

Actividades de discusión. El objetivo no es calibrar la calidad del trabajo individual, sino enriquecerlo con otras aportaciones. No es necesario cubrir todos los aspectos. Se valorará ante todo el trabajo de análisis y capacidad colaborativa.

Ponderación en la nota final

Aportará un máximo de tres puntos a la nota final.

Fecha aproximada de entrega

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

La calificación de la asignatura se obtendrá sumando, hasta un máximo de 10 puntos, de la siguiente manera:

Examen final, aportando a la nota final hasta un máximo de **cinco** puntos. Es de carácter obligatorio, y es necesario obtener un mínimo de cuatro puntos sobre 10 en el examen para poder aprobar la asignatura.

Práctica, aportando a la nota final hasta un máximo de **tres** puntos. Es de carácter obligatorio, y es necesario aprobarla para poder aprobar la asignatura.

Evaluación continua: la participación del estudiante en las tareas colaborativas propuestas a lo largo del curso se valorará con un máximo de **tres** puntos. Estas tareas son de carácter voluntario.

Nótese que este procedimiento de evaluación permite al estudiante adaptarse a diferentes metodologías.

Sin realizar las pruebas de evaluación continua y obteniendo la máxima nota en el examen y la práctica, se obtendría un 8 como nota final.

Realizando todas las pruebas de evaluación continua con la máxima calificación, y obteniendo la mínima puntuación necesaria en el examen y la práctica, se obtendría un 8 en la nota final.

BIBLIOGRAFÍA BÁSICA

ISBN(13):9780136072249

Título:SEARCH ENGINES: INFORMATION RETRIEVAL IN PRACTICE (Primera)

Autor/es:Bruce Croft ; Trevor Strohman ; Donald Metzler ;

Editorial:PEARSON

El libro se encuentra disponible de forma gratuita en

<https://ciir.cs.umass.edu/downloads/SEIRiP.pdf>

BIBLIOGRAFÍA COMPLEMENTARIA

El acceso a la información es una línea de investigación en constante desarrollo. Por ello, el texto base se complementará con artículos académicos publicados en los principales foros del área, como los congresos:

- SIGIR: Association for Computing Machinery's Special Interest Group on Information Retrieval. Since 1963, we have promoted research, development and education in the area of search and other information access technologies.
- CLEF: Conference and Labs of the Evaluation Forum. Information Access Evaluation meets Multilinguality, Multimodality and Interaction.
- CERI: Spanish Conference in Information Retrieval.
- SPIRE: rnational Symposium on String Processing and Information Retrieval.

Y revistas especializadas como:

- Information Processing and Management (IP&M), Elsevier
- Information Retrieval, Springer
- Transactions on the Web (TWEB), ACM

RECURSOS DE APOYO Y WEBGRAFÍA

Para las actividades prácticas se utilizarán lenguajes de programación estándar y herramientas de libre distribución. Se proporcionarán a los estudiantes los dataset o conjuntos de prueba (colecciones de documentos y anotaciones manuales) necesarios. Para la realización de trabajos de carácter teórico se partirá de guías elaboradas por el equipo docente sobre un escenario hipotético de acceso a la información.

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 se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.