26-27

GUÍA DE ESTUDIO PÚBLICA

ANÁLISIS Y GESTIÓN DE LOS DATOS EN LA INVESTIGACIÓN EN HD

CÓDIGO 27040039

UNED

26-27

ANÁLISIS Y GESTIÓN DE LOS DATOS EN LA INVESTIGACIÓN EN HD

CÓDIGO 27040039

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
IGUALDAD DE GÉNERO

Nombre de la asignatura	ANÁLISIS Y GESTIÓN DE LOS DATOS EN LA INVESTIGACIÓN EN HD
Código	27040039
Curso académico	2026/2027
Título en que se imparte	MÁSTER UNIVERSITARIO EN HUMANIDADES DIGITALES: MÉTODOS Y BUENAS PRÁCTICAS
Tipo	CONTENIDOS
Nº ETCS	5
Horas	125
Periodo	SEMESTRE 1
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

La asignatura se imparte en el primer semestre del Máster Universitario en Humanidades Digitales: Métodos y Buenas Prácticas, y tiene carácter obligatorio. Ocurre en los proyectos que una vez recabados los tipos datos de interés en sus diferentes formas y formatos, es necesario extraer información de estos a partir de su análisis y gestión para la investigación en primer lugar y posteriormente para su uso profesional, si es el caso.
Una vez que se dispone de datos almacenados ¿cómo se pueden analizar y gestionar para que aporten información de interés? Se presentarán diferentes técnicas y herramientas para análisis de datos e información, principalmente textuales, así como ejemplos de interés en investigación y en el entorno profesional, tomados de proyectos nacionales e internacionales de especial relevancia.
Esta asignatura se presenta en paralelo a la asignatura de “Acceso y Extracción de Datos para la Investigación en Humanidades Digitales” donde se estudiarán los principios de las bases de datos relacionales y está relacionada tangencialmente con las asignaturas “Bases de Datos y Big Data”, “Recuperación de Información y Minería de Datos”, ¨Web Semántica y Sistemas Inteligentes para las HD” y “Datos Espaciales: Manejo y Publicación” (optativas en el segundo semestre).
Respecto a las actividades prácticas y los ejercicios de evaluación prácticos, los estudiantes trabajarán con bases de datos existentes y herramientas (principalmente de uso abierto) para familiarizarse con la gestión de los datos textuales, su análisi y así generar diferentes tipos de información.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Aunque no existen requisitos obligatorios para cursar esta asignatura, es recomendable el conocimiento del inglés para poder acceder a un mayor número de fuentes de recursos (datos, artículos, libros…) que no siempre están traducidos al castellano, o se traducen muy posteriormente a su publicación.
El tener soltura con herramientas informáticas básicas ayudará al éxito en esta materia. Se fomentará el uso de software libre y de aplicaciones web para la realización de las actividades y las practicas propuestas.

EQUIPO DOCENTE

Nombre y Apellidos	ANA Mª GARCIA SERRANO (Coordinador/a de asignatura)
Correo Electrónico	agarcia@lsi.uned.es
Teléfono	91398-7993
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS

Nombre y Apellidos	ALBA GARCIA SECO DE HERRERA
Correo Electrónico	alba.garcia@lsi.uned.es
Teléfono	91398-8407
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS

HORARIO DE ATENCIÓN AL ESTUDIANTE

El seguimiento del aprendizaje se realizará revisando la participación de los alumnos en los distintos foros de debate y las aportaciones de material nuevo, así como supervisando la entrega en fecha de los diferentes trabajos prácticos que se han planificado durante la evolución del curso.
En todo tipo de comunicación con el profesorado se deberá indicar la asignatura a la que se refiere y utilizar el correo de la UNED. En caso de necesitar contactar con el Equipo Docente por medios distintos al curso virtual, se utilizará preferentemente el correo electrónico. Adicionalmente, está también previsto, para temas personales que no afecten al resto de los estudiantes, atender consultas en remoto. Si se desea una entrevista personal, debe concertarse previamente.

Dra. Ana García Serrano
Dirección postal: ETSI Informática –UNED C/ Juan del Rosal 16 (Despacho 2.07) 28040 MADRID
Correo electrónico: agarcia@lsi.uned.es
Horario de tutorías: Jueves de 12h a 14h y de 16h a 18h

Dra. Alba Garcia Seco de Herrera
Dirección postal: ETSI Informática –UNED C/ Juan del Rosal 16 (Despacho 2.07) 28040 MADRID
Correo electrónico: alba.garcia@lsi.uned.es
Horario de tutorías: Martes y Jueves de 12h a 14h

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

COMPETENCIAS BÁSICAS Y GENERALES
CG1 - Administrar el trabajo en equipos multidisciplinares dedicados al ámbito de las Humanidades Digitales de forma eficiente, abordando los posibles conflictos de manera constructiva.
CG2 - Conocer e identificar las nuevas técnicas y herramientas digitales para su empleo en la práctica profesional e investigadora en el ámbito de las Humanidades Digitales.

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.
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 ESPECÍFICAS

CE3 - Analizar y formalizar la información con herramientas digitales en el ámbito de las Humanidades Digitales.
CE8 - Conocer y saber aplicar diferentes técnicas y tipos de representación de datos digitales y del resultado de su análisis, en el ámbito de las Humanidades Digitales.
CE9 - Utilizar bases de datos, archivos y centros documentales en línea para su consulta y aplicación a un análisis original y propio en el ámbito de las Humanidades Digitales.

RESULTADOS DE APRENDIZAJE

Los resultados que se pretenden alcanzar con el estudio de esta asignatura son los siguientes:

- Conocer las principales aproximaciones para la gestión y el análisis de datos en general y en la práctica en Humanidades Digitales.
- Familiarizarse con los problemas relacionados con el análisis de datos como el preprocesamiento de datos, anonimización, validación, etc.
- Tener criterios para seleccionar entre los modelos de bases de datos actuales, y familiarizarse con algunas de ellos de utilidad para tareas en HD.
- Saber qué se entiende por recuperación de información y descubrimiento de información en textos con ejemplos en HD.
- Conocer aspectos relacionados con la accesibilidad y la usabilidad en la web en proyectos de HD.
- Saber qué son los datos enlazados y en abierto.

CONTENIDOS

1. Conceptos básicos de análisis de datos y su gestión en las HD.

Se presentarán conceptos y recursos básicos y se mostrará cómo es necesario preparar los datos antes de poder gestionarlos. Se utilizarán ejemplos de proyectos relevantes en la investigación en HD.

2. Análisis de datos en sistemas no relacionales.

En este máster, hay una asignatura dedicada a las bases de datos relacionales y otra optativa dedicada a la Recuperación de información y la minería de datos, por lo que en esta se presentarán brevemente las denominadas bases de datos "no relacionales" así como formas de indexación de contenido textual cuando los datos lo son.
Se hace un recorrido práctico en aplicaciones de HD con ejemplos de proyectos y casos de éxito en la investigación.

3. Recuperación de información y descubrimiento de información en textos de HD

Se afrontarán los conceptos y las diferencias entre recuperación, extracción y descubrimiento de información textual. Se utilizarán ejemplos de proyectos relevantes en la investigación en HD.
Se presentarán formas de anotación, anonimización y procesado lingüístico de textos semiestructurados con ejercicios prácticos en aplicaciones web, así como algunos ejercicios con aplicaciones de IA actuales.

4. Accesibilidad y usabilidad de repositorios en abierto.

Se utilizarán ejemplos de recursos y repositorios en abierto de proyectos relevantes en la investigación en HD. Además, se pretende familiarizar al estudiante con aspectos relacionados con la accesibilidad y usabilidad de los recursos y las aplicaciones web. Se presentarán casos de estudio relacionados con museos.

METODOLOGÍA

La materia está planteada para su realización a través de la metodología general de la UNED, en la que se combinan distintos recursos y los medios impresos con los audiovisuales y virtuales. La metodología estará basada en los siguientes elementos:

1. Materiales de estudio: guía de estudio y web; textos obligatorios; materiales audiovisuales; bibliografía, etc.
2. Participación y utilización de las distintas herramientas del Entorno Virtual de Aprendizaje.
3. Tutorías en línea y telefónica: participación en los foros; comunicación e interacción con el profesorado.
4. Evaluación continua y sumativa: actividades prácticas de evaluación continua; pruebas presenciales; ejercicios de autoevaluación.
5. Trabajo individual: lectura analítica de cada tema; elaboración de esquemas; realización de las actividades de aprendizaje propuestas.

En el **Curso Virtual** de la asignatura el estudiante encontrará:

- Una **guía de la asignatura** en la que se hace una descripción detallada del plan de trabajo propuesto.
- Un **calendario** con la distribución temporal de los temas propuesta por el Equipo Docente y con las fechas de entrega de las actividades teórico-prácticas que el estudiante tiene que realizar para su evaluación.
- Enunciado de las **actividades teórico-prácticas** propuestas y zona donde depositar los entregables asociados a dichas actividades.
- Los **foros** por medio de los cuales el Equipo Docente aclarará las dudas de carácter general y que se usarán también para actividades de tipo colaborativo (comentar aportaciones de estudiantes y otros). También se comunican todas aquellas novedades que surjan a lo largo del curso ya que será el principal medio de comunicación entre los distintos participantes en la asignatura.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen: No hay prueba presencial

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad: No

Descripción

No hay prueba presencial ni trabajos, solamente una PEC y una Práctica Final.

Criterios de evaluación

Ponderación de la prueba presencial y/o los trabajos en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?: Si,PEC no presencial

Descripción

PEC de actividades en la plataforma virtual: Consiste en varias prácticas sencillas que deben ser discutidas individualmente por el estudiante y reportadas en un único fichero. Son un trabajo de revisión de casos de éxito de análisis de datos en proyectos relevantes. Se califica con un máximo de 10 puntos.

Criterios de evaluación

Se valora la calidad de las respuestas y la presentación por escrito de las actividades realizadas.

Ponderación de la PEC en la nota final: 50%

Fecha aproximada de entrega: 30/11/2026

Comentarios y observaciones

Si se usa IA generativa durante la realización de los trabajos escritos se debe indicar al igual que cuando se extrae contenido de un libro o un artículo se cita en el texto y se referencia correctamente en el apartado de bibliografía.
Es una declaración de autoría indicando cómo se ha utilizado una IAGen o alguna herramienta de IA. Por ejemplo: "Se utilizó NOMBRE 1 para mejorar la expresión gramatical y refinar la estructura de las oraciones. La figura XX también se mejoró mediante la herramienta de inteligencia artificial NOMBRE 2".
Se pueden utilizar el tipo de indicaciones de la UNED para el caso de los Trabajo Fin de Título: https://www.uned.es/universidad/dam/inicio/unidad/cpri/Portal-antiplagio/Guia-de-integridad-en-I.A.-generativa/Gu%C3%ADa-para-asegurar-integridad-acad%C3%A9mica-con-IA.pdf

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? Si,no presencial

Descripción

Práctica Final: La segunda actividad evaluable es un conjunto de ejercicios prácticos con herramientas disponibles en la web o el aula virtual que conforman una práctica informática. Deben ser realizada individualmente por el estudiante y reportada en un único fichero. Se califica con un máximo de 10 puntos.

Criterios de evaluación

Se valora la calidad de práctica realizada y de su presentación por escrito.

Ponderación en la nota final 50%

Fecha aproximada de entrega 29/1/2027

Comentarios y observaciones

Si se usa IA generativa durante la realización de los trabajos escritos se debe indicar al igual que cuando se extrae contenido de un libro o un artículo se cita en el texto y se referencia correctamente en el apartado de bibliografía.
Es una declaración de autoría indicando cómo se ha utilizado una IAGen o alguna herramienta de IA. Por ejemplo: "Se utilizó NOMBRE 1 para mejorar la expresión gramatical y refinar la estructura de las oraciones. La figura XX también se mejoró mediante la herramienta de inteligencia artificial NOMBRE 2".
Se pueden utilizar el tipo de indicaciones de la UNED para el caso de los Trabajo Fin de Título: https://www.uned.es/universidad/dam/inicio/unidad/cpri/Portal-antiplagio/Guia-de-integridad-en-I.A.-generativa/Gu%C3%ADa-para-asegurar-integridad-acad%C3%A9mica-con-IA.pdf

¿CÓMO SE OBTIENE LA NOTA FINAL?

La nota final se calcula con la siguiente fórmula:
Nota Final = 50% PEC + 50% PF
Aprobarán la asignatura los estudiantes que consigan al menos 5 puntos en la nota final calculada con las ponderaciones definidas más arriba.
Los estudiantes que decidan presentarse en la convocatoria de septiembre tendrán esta misma forma de evaluación.

BIBLIOGRAFÍA BÁSICA

Dadas las características de la asignatura, que abarca una diversidad de contenidos y enfoques, no existe un texto único que recoja de manera completa todos los temas del programa. Por ello, el equipo docente propone bibliografía seleccionada pertinente y en algunos casos, ha elaborado materiales específicos que cubren los aspectos esenciales de cada tema. La selección de lecturas y apuntes complementarios estarán disponibles en el curso virtual.

BIBLIOGRAFÍA COMPLEMENTARIA

- Alcaraz Martínez, Rubén*. Conceptos, materiales y métodos en humanidades digitales. Ultreia 2025.* https://ultreia.ucv.es/index.php/ultreia/catalog/book/35
- Aldama, N., Guerrero, M., Montoro, H., and Samy, D. (2022) 'Anotación de corpus lingüísticos: Metodología utilizada en el Instituto de Ingeniería del Conocimiento (IIC)', 17. httpss://www.iic.uam.es/whitepapers/anotacion-corpus-linguisticos-metodologia-utilizada-iic/
- Al-Thanyyan, S. S., &Azmi, A. M. (2021). Automated Text Simplification: A Survey. *ACM Computing Surveys*, *54*(2), 1-36. httpss://doi.org/10.1145/3442695
- Ayuso García, M. (2022) 'Las ediciones de Arnao Guillén de Brocar de BECLaR transcritas con ayuda de Transkribus y OCR4all: Creación de un modelo para la red neuronal y posible explotación de los resultados',*Historias Fingidas*, *0*, 151-173. httpss://doi.org/10.13136/2284-2667/1102
- Baptiste, B., Favre, B., Auguste, J., and Henriot, C. (2021) 'Transferring Modern Named Entity Recognition to the Historical Domain: How to Take the Step?',*Workshop on Natural Language Processing for Digital Humanities (NLP4DH)*.Available at:httpss://hal.archives-ouvertes.fr/hal-03550384
- Cámara, Alicia; Molina, Álvaro y Margarita A. Vázquez. (2020). Manassero (eds.). 'La ciudad de los saberes en la Edad Moderna', Gijón, Ediciones Trea, 296 pp. Available at:https://e-spacio.uned.es/fez/view/bibliuned:404-Amolina-1011.
- Campillos-Llanos, L., Terroba, A., Zakhir, S., Valverde, A., &Capllonch, A. (2022). Building a comparable corpus and a benchmark for Spanish medical text simplification. *Procesamiento*

del Lenguaje Natural, *69*(189-196). httpss://doi.org/10.26342/2022-69-16

- De Toni, F., Akiki, C., De La Rosa, J., Fourrier, C., Manjavacas, E., Schweter, S., and Van Strien, D. (2022) 'Entities, Dates, and Languages: Zero-Shot on Historical Texts with T0.' *Proceedings of BigScience Episode #5 -- Workshop on Challenges and Perspectives in Creating Large Language Models*, 75-83. Available at: httpss://doi.org/10.18653/v1/2022.bigscience-1.7
- Devlin J., Ming-Wei Chang, Kenton Lee, Kristina Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, V 1.
- Ehrmann, M., Romanello, M., Fluckiger, A., and Clematide, S. (2020). 'Extended Overview of CLEF HIPE 2020: Named Entity Processing on Historical Newspapers.' 38.
- García-Marco, Francisco-Javier. "La pirámide de la información revisitada: enriqueciendo el modelo desde la ciencia cognitiva". *El profesional de la información*, 2011, enero-febrero, v. 20, n. 1, pp. 11-24. DOI: 10.3145/epi.2011.ene.02
- García-Serrano, A. and Castellanos, A. (2016) 'Representación y organización dedocumentos digitales: detalles y práctica sobre la ontología DIMH'. *Revista de Humanidades Digitales*, v.1, 314-344, ISSN 2531-1786. Available at: httpss://doi.org/10.5944/rhd.vol.1.2017.17155
- García-Serrano, A., and Menta-Garuz, A (2022). 'La inteligencia artificial en las Humanidades Digitales: dos experiencias con corpus digitales.'*Revista de Humanidades Digitales*, 7, 19-39. ISSN 2531-1786. Disponible en (Fecha de acceso: 21 jun. 2022) httpss://doi.org/10.5944/rhd.vol.7.2022.30928
- García-Serrano, Ana; Menta-Garuz, Antonio (2021) Orientaciones y evaluaciones de técnicas en humanidades digitales: de la estadística al *Deep-Learning*, "La vía de la implementación: Herramientas e infraestructuras ", V congreso de la Sociedad Internacional de Humanidades Digitales Hispánicas - HDH 2021 Libro de resúmenes. Santiago de Compostela, 4-8 octubre, pag: 58-64 (7).
- García-Serrano, Ana; Fco Paños Merino; César Férnandez Sánchez (2023) Una mirada digital a los datos: el trabajo de un estudioso de la Antigua Grecia VI congreso de la Sociedad Internacional de Humanidades Digitales Hispánicas - HDH 2023 "Encuentros y transformaciones" Logroño, 18 y 19 octubre.
- Li, J., Sun, A., Han, J., and Li, C. (2022) 'A Survey on Deep Learning for Named Entity Recognition.'*IEEE Transactions on Knowledge and Data Engineering*, *34*(1), 50-70. Available at:httpss://doi.org/10.1109/TKDE.2020.2981314
- Martínez Cantón, C., I., Ortuño Casanova, R. (2023). *Las humanidades digitales en la enseñanza de las literaturas hispánicas*. *Aplicaciones prácticas*. Berlin, Germany: Peter Lang Verlag. https://www.peterlang.com/document/1366954

- Menta, A., Sánchez-Salido, E., and García-Serrano, A. (2022) 'Transcripción de periódicos históricos: Aproximación CLARA-HD',*Proceedings of the Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations (SEPLN-PD 2022)*. Available at: httpss://ceur-ws.org/Vol-3224/paper17.pdf
- Molina Martín, Á. (2021) 'Cartografías del adorno en las residencias nobiliarias de la corte de Carlos IV: redes y modelos de buen gusto y distinción'*Magallanica. Revista de Historia Moderna*, 7(14), 205-235.
- Moreno Sandoval, Antonio (2019) "Lenguas y computación" ISBN: 9788491712732. Páginas 288. ISBN Digital: 9788491718246. Editorial Síntesis.
- Moreno-Sandoval, A., Gisbert, A., &Montoro, H. (2020). FinT-esp: A corpus of financial reports in Spanish. *Multiperspectives in analysis and corpus design*, *January*, 89-102.
- Rubio, A., &Martínez, P. (2022). *HULAT-UC3M at SimpleText@CLEF-2022: Scientific text simplification using BART*. Conference and Labs of the Evaluation Forum.
- Sanchez-Salido, A. Menta and A. García-Serrano (2023) Seeking Information in Spanish Historical Newspapers: The Case of Diario de Madrid (18th and 19th Centuries). V17 N 4. DHQ journal. ISSN: 1938-4122. https://digitalhumanities.org/dhq/vol/17/4/000735/000735.html
- Tasovac, Toma; Sally Chambers; Erzsébet Tóth-Czifra. (2020) Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper. hal-02961317
- Waters, D.J. The emerging digital infrastructure for research in the humanities. *Int J Digit Libr* 24, 87–102 (2023). https://doi.org/10.1007/s00799-022-00332-3

RECURSOS DE APOYO Y WEBGRAFÍA

La asignatura se encontrará virtualizada en la plataforma que para ello disponga la Universidad. En ella el estudiante encontrará las últimas novedades relativas a la asignatura y los materiales necesarios para su estudio, así como una serie de foros para comunicarse con el equipo docente y con sus compañeros.

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.