

23-24

MÁSTER UNIVERSITARIO EN
HUMANIDADES DIGITALES: MÉTODOS Y
BUENAS PRÁCTICAS

GUÍA DE ESTUDIO PÚBLICA



ANÁLISIS ESTILOMÉTRICO Y DE TEXTO

CÓDIGO 27040200

UNED

23-24

ANÁLISIS ESTILOMÉTRICO Y DE TEXTO

CÓDIGO 27040200

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 ESTILOMÉTRICO Y DE TEXTO
Código	27040200
Curso académico	2023/2024
Título en que se imparte	MÁSTER UNIVERSITARIO EN HUMANIDADES DIGITALES: MÉTODOS Y BUENAS PRÁCTICAS
Tipo	CONTENIDOS
Nº ETCS	4
Horas	100.0
Periodo	SEMESTRE 2
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

En esta asignatura se impartirán los conceptos básicos para que los investigadores de humanidades puedan realizar investigaciones sobre sus corpora de texto. Estos trabajos de investigación pueden estar relacionados con la minería de textos (*text mining*), la atribución de autoría y otras cuestiones de interés para los estudios literarios computacionales, así como el procesamiento del lenguaje natural y el modelado de temas (*topic modelling*). Para ello utilizaremos R o Python (cuando sea necesario desde un punto de vista formativo), ambos lenguajes son de programación libres y gratuitos ampliamente empleados por la comunidad científica en múltiples campos y materias, desde la estadística a la lingüística forense, pasando por la filogénesis, la bioestadística, el análisis político, el marketing, el análisis de ideas, la autoría de fuentes, etc.

La estilometría nos permite el procesamiento automatizado de grandes cantidades de textos con un ordenador y acercarnos a nuevos paradigmas de investigación dentro del *close reading* o en el *distant reading*, ofreciéndonos un gran volumen de datos que pueden representarse visualmente, y que pueden ayudar a resolver cuestiones tan interesantes como las distintas manos o autores que intervienen en una obra, descubrir el autor de un texto anónimo, la frecuencia de un determinado tipo de palabras o recursos estilísticos para tratar determinados temas y tópicos, o cuál es la forma que sigue el relato. Es, sin duda, una asignatura con enormes perspectivas dentro de la investigación en el ámbito hispánico de las humanidades

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Se recomienda que el estudiante tenga conocimientos básicos de inglés y nociones básicas de programación en especial R o Python y su entorno de desarrollo.

EQUIPO DOCENTE

Nombre y Apellidos	SALVADOR ROS MUÑOZ (Coordinador de asignatura)
Correo Electrónico	sros@dia.uned.es
Teléfono	7205/7564
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL
Nombre y Apellidos	SALVADOR ROS MUÑOZ (Coordinador de asignatura)
Correo Electrónico	sros@scc.uned.es
Teléfono	7205/7564
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	SISTEMAS DE COMUNICACIÓN Y CONTROL
Nombre y Apellidos	JOSE GIOVANNI HIGUERA RUBIO
Correo Electrónico	jhiguerarubio@fsof.uned.es
Teléfono	
Facultad	FACULTAD DE FILOSOFÍA
Departamento	FILOSOFÍA

HORARIO DE ATENCIÓN AL ESTUDIANTE

Las/os estudiantes pueden contactar con las/os profesoras/es para resolver dudas sobre la asignatura en primer lugar y prioritariamente a través del foro de la asignatura en el campus virtual correspondiente y a través del correo electrónico. Si se desea una entrevista personal, debe concertarse previamente. 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.

Dr. Salvador Ros Muñoz

Horario de atención: Lunes y Martes de 12 a 14 horas

Correo electrónico: sros@scc.uned.es

Dra. Eugenia San Segundo Fernández

Horario de atención: Miércoles de 10 a 14 horas

Dirección postal: Facultad de Filología. Departamento de Lengua Española y Lingüística General

Paseo de la Senda del Rey, 7. Despacho 7.10 28040 MADRID

Teléfono: 91 398 6875

Correo electrónico: esansegundo@flog.uned.es

Dr. José Higuera Rubio

Horario de atención: Martés 10.00-13.00 y Jueves 14.00-17.00

Correo electrónico: jhiguerarubio@fsof.uned.es

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.

CG3 - Describir y aplicar las tecnologías para la gestión y organización de la información y la documentación 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.

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 ESPECÍFICAS

CE3 - Analizar y formalizar la información con herramientas digitales en el ámbito de las Humanidades Digitales.

CE5 - Desarrollar aplicaciones o proyectos originales e innovadores en el campo de las Humanidades Digitales de carácter profesional e investigador.

CE7 - Aplicar las tecnologías digitales en el tratamiento y la preservación de datos de diferente tipología 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

- Introducir al estudiante en el software R y Python para el análisis de textos de manera cualitativa y cuantitativa.
- Dotar de conocimientos básicos para realizar investigaciones de minería de textos.
- Introducirse al procesamiento de textos.
- Aprender a modelar tópicos de interés.
- Tratamiento de análisis de textos para el procesamiento automatizado de textos.
- Descubrir características que identifican a los texto como sentimientos en los textos.
- Detectar autorías de textos.
- Visualización y representación de la información obtenida, una vez procesados los textos.

CONTENIDOS

1.- Lectura y curación de datos

Se comenzará aprendiendo a cómo preparar los datos para su futura explotación haciendo uso de los ordenadores.

2.- Análisis estadístico descriptivo de textos

Se introducirá los conceptos básicos de estadística exploratoria aplicado a textos.

3.- Introducción a la estilometría

Primeros pasos en las técnicas estilométricas

4.- Otras técnicas de utilidad en el análisis de texto.

En este tema se completará el estudio de las técnicas de estilometría y se incluirán otras como *topic modelling* o análisis de sentimientos.

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 o en grupo: lectura analítica de cada tema; elaboración de esquemas; realización de las actividades de aprendizaje propuestas.

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

Memoria o proyecto con desarrollo informático sobre un tema abordable con los contenidos explicados en la asignatura.

Criterios de evaluación

- Se valorará la estructuración del trabajo
- **La resolución de las preguntas de investigación**
- **La aplicación de los conocimientos de estilometría a la solución presentada**

Ponderación de la prueba presencial y/o los trabajos en la nota final 80

Fecha aproximada de entrega

Comentarios y observaciones

Para aprobar la asignatura es obligatorio aprobar el trabajo final con al menos un 5 sobre la nota máxima de esta prueba que será 10.

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?

Si, PEC no presencial

Descripción

Pruebas de evaluación continua consistirá en un ejercicio práctico o un test. La información sobre esta prueba se dará en el curso virtual.

Criterios de evaluación

- Se valorará la estructuración del trabajo
- **La resolución de las preguntas de investigación**
- **La aplicación de los conocimientos de estilometría a la solución presentada**

Ponderación de la PEC en la nota final 20

Fecha aproximada de entrega

Comentarios y observaciones

Para aprobar la asignatura es obligatorio aprobar la PEC con al menos un 5 sobre la nota máxima de esta prueba que será 10.

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? No

Descripción

Criterios de evaluación

Ponderación en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

La nota se obtiene como suma ponderada de la nota de las PEC y el Trabajo Final. La ponderación es la siguiente 80% el trabajo final y el 20 % la PEC. Para incluir las notas de estas pruebas en la ponderación final es necesario haberlas aprobado por separado. Es decir, sacar al menos un 5 en cada uno de ellas. Si no se obtiene esa nota media el peso de esa prueba en la ponderación será cero.

BIBLIOGRAFÍA BÁSICA

Fradejas Rueda, J. M. (2020). *Cuentapalabras. Estilometría y análisis de texto con R para filólogos*. Página web: <http://www.aic.uva.es/cuentapalabras/>

BIBLIOGRAFÍA COMPLEMENTARIA

- McMenamin, G. (2002). *Forensic linguistics: Advances in forensic stylistics*. CRC press.
- Olsson, J. (2004). *Forensic Linguistics*. London, New York: Continuum International Publishing Group.
- M.L.,Jokers (2014)Text Analysis with R for Students of Literature. Springer Verlag
- MORETTI F.: Distant reading. Verso, 2013. 2
- Archer, Jodie &Matthew L. Jockers (2016). The Bestseller Code. Penguin Books, 2017.
- Blatt, Ben (2017). Nabokov's Favorite Word is Mauve: What the Numbers Reveal about the Classics, Bestsellers, and Our Own Writing. Londres: Simon &Schuster.
- Booker, Christopher (2004). The Seven Basic Plots. Why we tell stories. Londres: Bloomsbury, 2017
- Burrows, John (2002). «'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship». *Literary and Linguistic Computing*, 17:3, 267 –287.
- Calvo Tello, José &Juan Cerezo Soler (2018). «La conquista de Jerusalén ¿de Cervantes? Análisis estilométrico sobre autoría en el teatro del Siglo de Oro español». *Digital Humanities Quarterly*, 12: 1, 1–10. Accesible en línea
- Eder, Maciej (2013). «Mind your corpus: systematic errors in authorship attribution». *Literary and Linguistic Computing*, 28:4, 603–614.

- Eder, Maciej (2015). «Does size matter? Authorship attribution, small samples, big problem». *Digital Scholarship in the Humanities*, 30:2, 167–182.
- Eder, Maciej (2017a). «Short samples in authorship attribution: a new approach». *Digital Humanities 2017. Conference abstracts*. McGill University & Université de Montréal. August 8-11, 2017.
- Eder, Maciej (2017b). «Visualization in stylometry: Cluster analysis using networks». *Digital Scholarship in the Humanities* 32:1, 50–64.
- Eder, Maciej & Jan Rybicki (2013). «Do birds of a feather really flock together, or how to choose training samples for authorship attribution». *Digital Scholarship in the Humanities*, 28: 2, 229–236.
- Eder, Maciej et al. (2016). «Stylometry with R: A Package for Computational Text Analysis». *The R Journal*, 8:1, 107-121
- Fradejas Rueda, José Manuel (2016). «Lingüística forense y crítica textual. El caso Ayala–Cervantes», *Signa*, 25, 193-220. DOI: <https://doi.org/10.5944/signa.vol25.2016.16927>
- Fradejas Rueda, José Manuel (2016). «El análisis estilométrico aplicado a la literatura española: las novelas policiacas e históricas», *Caracteres* 5:2, 196-245. Accesible en línea.
- Fradejas Rueda, José Manuel (2019). «Estilometría y la Edad Media castellana», en Nanette Rissler-Pipka (ed.), *Theorien von Autorschaft und Stil in Bewegung: Stilistik und Stilometrie in der Romania*, (Romanische Studien, Beiheft 6). Múnich: AVM.edition, 49-74. Accesible en línea.
- Fradejas Rueda, José Manuel (2020). *Cuentapalabras*. Valladolid: Universidad. Accesible en línea.
- Hernández Lorenzo, Laura (2019a). «Fernando de Herrera y la autoría de Versos. Un primer acercamiento al drama textual desde la Estilometría», en Nanette Rissler-Pipka (ed.), *Theorien von Autorschaft und Stil in Bewegung: Stilistik und Stilometrie in der Romania*, (Romanische Studien, Beiheft 6). Múnich: AVM.edition, 75–90. Accesible en línea
- Hernández-Lorenzo, Laura (2019b). «Poesía áurea, estilometría y fiabilidad: métodos supervisados de atribución de autoría atendiendo al tamaño de las muestras». *Caracteres. Estudios culturales y críticos de la esfera digital*, 8:1, 189-228. Accesible en línea
- Hernández-Lorenzo, Laura (2022). «Stylistic Change in Early Modern Spanish Poetry Through Network Analysis (with an Especial Focus on Fernando de Herrera's Role) ». *Neophilologus*, 106, 397-417. <https://doi.org/10.1007/s11061-021-09717-2>

RECURSOS DE APOYO Y WEBGRAFÍA

<https://www.dariah.eu/>

<https://www.clarin.eu/>

<https://programminghistorian.org/>

<https://eadh.org/>

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.