

Desarrollo  
profesional y personal

Curso académico 2017-2018

## Plataforma docente

Análisis de textos y estilometría usando R (ACTUALIZACIÓN PROFESIONAL)

del 16 de enero al 31 de mayo de 2018

**CERTIFICADO DE ENSEÑANZA ABIERTA**

6 créditos

**Características:** material impreso, material multimedia, actividades presenciales optativas, página web, curso virtual y guía didáctica.

Departamento

*Literatura Española y Teoría de la Literatura*

Facultad de Filología

Departamento

*Sistemas de Comunicación y Control*

E.t.s. de Ingeniería Informática

## PROGRAMA DE DESARROLLO PROFESIONAL Y PERSONAL Curso 2017/2018

El Programa de Desarrollo Profesional y Personal acoge cursos que dan derecho a la obtención de un Título Propio por la UNED. Cada curso se impartirá en una de las siguientes categorías: Experto Profesional, Enseñanza Abierta, Actualización Profesional (\*) y atienden una demanda de formación en respuesta a las necesidades e intereses del mundo del trabajo. Sus cursos estarán promovidos por la UNED o entidades sociales y económicas que busquen el apoyo académico de la Universidad para la mejor calidad de sus acciones de formación y promoción de sus colectivos laborales y profesionales.

Requisitos de acceso: no hay requisitos mínimos de acceso, salvo los específicos de cada curso establecidos por su director.

(\*) En los cursos que se ofertan en Enseñanza Abierta que en su denominación contengan la especificación de (ACTUALIZACIÓN PROFESIONAL), la superación del curso se acreditará mediante un "Certificado de Actualización Profesional" (Reunión del Consejo de Gobierno de la UNED, celebrada el día dieciséis de diciembre de dos mil catorce, por la que se aprueba la implantación de cursos para cubrir la formación profesional y empresarial con la denominación de Certificado de Actualización Profesional).

## Destinatarios

Para la realización de este curso no es necesario cumplir con ningún requisito académico específico.

Está destinado principalmente a:

- Estudiantes con formación en distintas áreas de Humanidades que deseen adquirir conocimientos tecnológicos para enfrentarse a los nuevos retos que ofrece el análisis digital automatizado del texto, abriendo nuevas perspectivas.

- Investigadores actualmente trabajando en proyectos de diferentes disciplinas humanísticas (filología, arte, historia, filosofía...) que deseen ponerse al día sobre las herramientas y perspectivas tecnológicas existentes con el fin de poder aplicarlas a sus proyectos reales de investigación.
- Profesionales que trabajen en el campo de las humanidades (principalmente GLAM: Galerías, Bibliotecas, Archivos y Museos) y que deseen aumentar sus competencias digitales en análisis textual para poder aplicarlas en su propio trabajo.
- Personas pertenecientes a otras disciplinas que estén interesados por conocer las Humanidades Digitales y acercarse a un tipo de trabajo interdisciplinar y muy actual.

En especial, este curso está diseñado para aquellas personas que se dedican a la investigación y el trabajo con textos y quieran ir más allá de las posibilidades que ofrecen los programas del estilo de WordSmith, AntConc y Excel.

Además, el curso está dirigido a cualquier licenciado o graduado de humanidades, ciencias sociales o cualquier otra disciplina, con unos conocimientos básicos de informática que desee profundizar en los conocimientos propuestos.

-----

Se aplicarán los siguientes descuentos específicos sobre el precio de matrícula de este curso a:

- Los miembros del Laboratorio de Innovación en Humanidades Digitales de la UNED (30%)
- Los alumnos matriculados en este curso o en años anteriores en el Experto Profesional en Humanidades Digitales o en el Experto Profesional en Edición Digital Académica (20%)
- Los miembros de la Asociación de Humanidades Digitales Hispánicas que estén al corriente del pago de la cuota (10%)

**Estos incentivos son incompatibles con otro tipo de Ayudas y no podrán acumularse en la misma matrícula becas, ayudas, bonificaciones / incentivos.**

**SOLICITUD DEL ALUMNO :ver enlace en apartado "Matriculación"**

## 1. Objetivos

El objetivo del curso es dotar a los investigadores de humanidades de los conocimientos básicos para realizar investigaciones de minería de textos (*Text mining*), problemas de atribución de autoría, lingüística forense, procesamiento de lenguaje natural y modelado de temas (*Topic Modelling*), por medio de un lenguaje de programación libre y gratuito ampliamente empleado por la comunidad científica en diversos campos y materias, desde la estadística a la lingüística forense pasando por la filogénesis, la bioestadística, el análisis político, o el marketing.

Este sistema -basado en la utilización del software R, procedente del campo de la estadística, sobre el que se han desarrollado paquetes específicos para el tratamiento y el análisis de textos (como Stylo)- permite el procesamiento automatizado de grandes cantidades de texto por ordenador, arrojando interesantes datos, que pueden representarse visualmente sobre cuestiones tan relevantes como las distintas manos o autores que intervienen sobre una obra, o la frecuencia de un determinado tipo de palabras o recursos estilísticos para tratar determinados temas.

Se trata, sin duda, de una revolucionaria aproximación con enormes perspectivas dentro de la investigación en el ámbito hispánico de las humanidades que no se había planteado hasta hace poco tiempo como reto.

## Otra Información

Será responsabilidad exclusiva del Equipo Docente la información facilitada en la siguiente relación de hipervínculos. En caso de detectarse alguna contradicción, prevalecerá la oferta formativa aprobada por el Consejo de Gobierno para cada convocatoria, así como del Reglamento de Formación Permanente y del resto de la legislación Universitaria vigente.

[Más Información](#)

## 2. Contenidos

### TEMARIO

- 1.- Introducción a la estilometría y a R (instalación y primeros pasos)
- 2.- Calcular la frecuencia de palabras de un texto
- 3.- Análisis de la distribución de *tokens* en un texto
- 4.- Análisis de correlación
- 5.- Medir la variedad y riqueza léxica de un texto
- 6.- Cálculo de *hapax*
- 7.- Calidad del texto, variedad de textos y el XML
- 8.- Análisis de grupos (*clustering*)
- 9.- Clasificación
- 10.- Modelado de temas (*Topic modelling*)

### ACADEMIC PROGRAM

- 1.-Introduction to stylometry and R (setup and fist steps)
- 2.-Calculation of word frequencies in a text
- 3.-Analysis of *token* distribution in a text
- 4.-Correlation analysis
- 5.-Measurement of variety and lexical richness in a text
- 6.-Hapax calculation
- 7.-Text quality and variety with XML
- 8.-Group analysis (*clustering*)
- 9.-Classification
- 10.-Topic modelling

### 3. Metodología y actividades

El curso será completamente online. La metodología de enseñanza será a distancia, siguiendo el sistema de enseñanza virtual de la UNED, a través de la plataforma moodle.

Se ofrecerán materiales digitales específicos, unos elaborados por el equipo docente y otros materiales académicos de calidad disponibles en abierto en formato digital, materiales audiovisuales, foros y chats como herramientas principales de comunicación entre los alumnos y el equipo docente.

Se trabajará de una forma práctica sobre corpora digitales de texto y se irá proponiendo la realización de diferentes experimentos (minería, análisis de datos textuales, visualizaciones en distintos formatos, extracción de conclusiones, etc.).

El objetivo final es sentar las bases para la realización de análisis textuales cuantitativos en áreas de estudio que hasta ahora solo se han preocupado por los métodos cualitativos.

Se utilizará la instalación del software R local y también se facilitará acceso a la nube de Bluemix de IBM para trabajar con el software en la nube.

Los alumnos realizarán un trabajo final sobre un corpus de textos en el que aplicarán todos los conocimientos aprendidos en el curso, y podrán extraer conclusiones como la atribución de autoría, la similitud entre textos de diferente origen o la visualización de gráficos de las entidades y personajes que en ellos aparecen.

### 4. Nivel del curso

Iniciación

### 5. Duración y dedicación

Duración: del 16 de enero de 2018 al jueves 31 de mayo de 2018.

Dedicación: 150 horas.

### 6. Equipo docente

#### Director/a

Director - UNED

*ROBLES GOMEZ, ANTONIO*

#### Colaboradores UNED

Colaborador - UNED

*HERNANDEZ BERLINCHES, ROBERTO*

Colaborador - UNED

*ROS MUÑOZ, SALVADOR*

## Colaboradores externos

Colaborador - Externo

*FRADEJAS RUEDA, JOSE MANUEL*

Colaborador - Externo

*GARRIDO TEIXEIRA, PATRICIA*

Colaborador - Externo

*GONZALEZ-BLANCO GARCIA, ELENA*

## 7. Material didáctico para el seguimiento del curso

### 7.1 Material disponible en la plataforma virtual

El equipo docente colocará en la plataforma materiales elaborados por ellos mismos en formato digital.

Materiales académicos de calidad disponibles en abierto en formato digital.

Cuestionarios de evaluación.

## 8. Atención al estudiante

Cada profesor indicará en la web su horario y sistema preferente de tutorías.

Para cuestiones relativas al funcionamiento general del curso, pueden dirigirse a:

Elena González-Blanco García

Dpto. de Literatura Española y Teoría de la Literatura, Despacho 722

Facultad de Filología, UNED

Paseo Senda del Rey 7

28040 MADRID

tel. 91 3986873

Horario preferente de tutorías presenciales y de consultas telefónicas: martes 9:30-19:00 y miércoles 9:30-14:00, previa cita por email, a ser posible, a [egonzalezblanco@flog.uned.es](mailto:egonzalezblanco@flog.uned.es)

Salvador Ros Muñoz

Dpto. de Sistemas de Comunicación y Control

Escuela Técnica Superior de Ingeniería Informática, Despacho 5.05

C/ Juan del Rosal, 16

28040 MADRID

tel. 913987205

[sros@scscc.uned.es](mailto:sros@scscc.uned.es)

web informativa del curso: <http://linhd.uned.es/p/titulo-propio-analisis-del-texto-y-estilometria/>

## 9. Criterios de evaluación y calificación

La evaluación se realizará mediante trabajos planteados, tutelados y corregidos por el Equipo Docente, así como mediante cuestionarios a través de la plataforma. Se evaluarán los conocimientos adquiridos en la lectura de los materiales, pero los trabajos tendrán una orientación eminentemente práctica, de aplicación de las herramientas tecnológicas estudiadas a problemas concretos y reales de las distintas áreas de humanidades mediante el uso del análisis de textos con la metodología aprendida en el curso.

## 10. Precio público del curso

Precio público de matrícula: 408 €

## 11. Matriculación

Del 7 de septiembre de 2017 al 15 de enero de 2018.

Atención administrativa:

Cursos de Certificado de Enseñanza Abierta

[ea@adm.uned.es](mailto:ea@adm.uned.es)

Tf. 91 398 7711 / 7530

C/ Juan del Rosal, 14 - 1ª planta

28040 Madrid

Se aplicarán los siguientes descuentos específicos sobre el precio de matrícula de este curso a:

-Los miembros del Laboratorio de Innovación en Humanidades Digitales de la UNED (30%)

-Los alumnos matriculados en este curso o en años anteriores en el Experto Profesional en Humanidades Digitales o en el Experto Profesional en Edición Digital Académica (20%)

-Los miembros de la Asociación de Humanidades Digitales Hispánicas que estén al corriente del pago de la cuota (10%)

**Estos incentivos son incompatibles con otro tipo de Ayudas y no podrán acumularse en la misma matrícula becas, ayudas, bonificaciones/incentivos.**

[FORMULARIO DE SOLICITUD](#)