

## “Introducción al procesamiento de datos textuales con el sistema de consulta y gestión de corpus Sketch Engine”

### INFORMACIÓN GENERAL

Responsable-s de la actividad: Olga Batiukova.

Persona de contacto con los alumnos (email): Olga Batiukova (volha.batsiukova@uam.es)

Correo contacto cuestiones administrativas: doctorado.gestion@uam.es

Duración: 10 horas

Fechas de realización: 8 de enero de 2024 de 9:30 a 13:00 - 9 de enero de 2024 de 9:30 a 13:00 - 10 de enero de 2024 de 9:30 a 12:30.

Lugar: Facultad de Filosofía y Letras, aula 201 del módulo V.

Plazas: 25.

ECTS: 1 crédito (10 horas presenciales + 20 horas de trabajo no presencial de cada doctorando/a).

### OBJETIVOS

El objetivo general de la actividad consiste en proporcionar formación teórico-práctica que permita a los doctorandos trabajar con corpus digitales a través de las distintas funcionalidades del sistema de consulta y gestión de corpus Sketch Engine.

Las competencias genéricas que han de adquirir los doctorandos a través de esta actividad están relacionadas con las siguientes habilidades transversales de Eurodoc Report (2018) “Identifying Transferable Skills and Competences to Enhance Early-Career Researchers’ Employability and Competitiveness”:

- Competencias digitales: búsqueda y procesamiento de la información; intercambio de información; uso de *software*
- Competencias relacionadas con la investigación: análisis de datos; conocimiento de disciplinas específicas y su terminología; gestión de datos en abierto; interdisciplinaridad.

Las competencias específicas son las siguientes:

- Hacer búsquedas básicas y avanzadas en corpus digitales.
- Generar resúmenes de las propiedades distribucionales de las unidades léxicas.
- Ser capaz de ajustar las reglas de Word Sketch Grammar a las necesidades específicas del usuario.
- Comparar propiedades contextuales de palabras diferentes.
- Identificar términos especializados en un corpus digital.
- Compilar corpus propios a partir de documentos preseleccionados y de documentos disponibles en internet

### CRITERIOS DE ADMISIÓN

Por centrarse en un *software* de procesamiento de datos textuales, esta actividad es altamente transversal. Será especialmente recomendable para alumnos matriculados en los programas de doctorado asociados a las Ciencias Sociales y Humanidades (de la Facultad de Filosofía y Letras, Formación de Profesorado, Psicología, etc.). Es recomendable en cualquier fase de la formación doctoral, pero

preferentemente en los primeros dos años.

No hay requisitos de admisión previos. No serán necesarios conocimientos previos sobre los corpus y su tratamiento informático. Es imprescindible que los asistentes traigan su ordenador portátil a clase. Se dará prioridad a los estudiantes de la Facultad de Filosofía y Letras.

## PROCEDIMIENTO DE INSCRIPCIÓN

Se realiza a través de [SIGMA](#) > Otra Formación UAM > Doctorandos – Actividades Formativas – Preinscripción.

- Fechas de Preinscripción: 20/11/2022 a 17/12/2023
- Fechas de Admisión: 18/12/2023 a 19/12/2023
- Fechas de Matrícula: 20/12/2023 a 27/12/2023
- Fechas de Impartición y horario de impartición: 8 de enero (09:30 a 13:00), 9 de enero (09:30 a 13:00), 10 de enero de 2024 (09.30 a 12.30).

## CERTIFICADOS

Una vez recibido el listado de alumnos que resulten "aptos" o "no aptos" desde el punto de vista académico por parte del director de la actividad transversal, la Escuela de Doctorado generará un certificado que será remitido por correo electrónico a la dirección facilitada por el alumno.

## CONTENIDOS

1. Tipos de corpus y lenguas incluidas en Sketch Engine.
2. Tipos de búsqueda: búsquedas básicas y búsquedas avanzadas con Corpus Query Language. La sintaxis de las expresiones regulares.
3. Word Sketch: resumen de las propiedades contextuales de las unidades léxicas. Definición de relaciones gramaticales en las gramáticas de Word Sketch existentes y propias.
4. Thesaurus y Word Sketch Differences: comparación de las propiedades contextuales.
5. Extracción terminológica con Keywords and Terms.
6. Obtención de ejemplos representativos con GDEX (Good Dictionary Examples).
7. Compilación de corpus propios a través de documentos preseleccionados y la web.

## ENFOQUE METODOLÓGICO

El curso se impartirá en formato de taller y las sesiones tendrán un enfoque eminentemente práctico. Tras la presentación de conceptos y herramientas nuevos por parte de la docente, se profundizará en su aplicación y se harán ejercicios prácticos para aprender a desenvolverse en el entorno de Sketch Engine.

Sesiones presenciales: 10 horas distribuidas en tres sesiones.

Trabajo autónomo de los alumnos (estudio de los materiales de clase y elaboración de ejercicios): 10 horas.

**Sistema de evaluación:** Para aprobar este módulo será necesario obtener un mínimo de cinco puntos en la suma de los siguientes apartados:

- Asistencia (será necesario asistir a un mínimo del 80% de las horas lectivas): 50%.
- Elaboración de ejercicios propuestos durante las sesiones y después de las sesiones:

50%.

**Número de horas exigidas:** Será necesario asistir a un mínimo del 80% de las horas lectivas

**Idioma de impartición:** Español.