

Curso Académico: 2023/24

31373 - Corpus y Herramientas Informáticas

Información de la Guía Docente

Curso Académico: 2023/24

Centro académico: 803 - Centro Másteres del Departamento de Traducción i Ciencias del Lenguaje

Estudio: 8037 - Máster Universitario en Lingüística Teórica y Aplicada

Asignatura: 31373 - Corpus y Herramientas Informáticas

Créditos: 5.0

Curso: 1

Idiomas de docencia:

Teoría: Grupo 1: Inglés

Seminario: Grupo 101: Inglés

Grupo 102: Inglés

Profesorado: Maria Nuria Bel Rafecas

Periodo de Impartición: Primer trimestre

Presentación

La asignatura Corpus y herramientas de software presenta la metodología para realizar investigaciones empíricas basadas en corpus sobre lingüística y lingüística aplicada. En particular, se trata de utilizar programas como herramientas básicas para gestionar grandes cantidades de datos de texto.

El objetivo de la asignatura es que el estudiante conozca la razón de los programas y herramientas que ayudan a los investigadores a la hora de trabajar con datos lingüísticos procedentes de un objeto de estudio llamado corpus, que es una colección de textos.

Competencias asociadas

CGS1. Creatividad para la investigación de posgrado y la práctica profesional

CT1.6. Capacidad para diseñar y revisar procesos de forma sistemática

CE. Desarrollar con la metodología del análisis lingüístico argumentativo y empírico.

Resultados del aprendizaje

- Aplicación de criterios de para el diseño de corpus y uso de herramientas última generación para compilar un corpus con fines específicos.
- Uso de conceptos como representatividad y significación para la investigación empírica en lingüística
- Definición de los requisitos necesarios para encontrar herramientas (y fuentes de información) y funcionalidades para utilizarlas.
- Descubrimiento, instalación y uso de herramientas que realizan funciones típicas de la lingüística de corpus, incluida la comprensión de la búsqueda de patrones con expresiones regulares y herramientas de anotación de corpus.
- Conocimiento de la terminología de PNL y procesamiento de textos.

Contenidos

Sección 1

- ¿Qué es un corpus? ¿Por qué utilizar ordenadores?

- Herramientas para funciones básicas. Palabra clave en contexto, KWIC y concordancias. Contar frecuencias de palabras. Importancia de la frecuencia relacionada con los contextos. Contar las frecuencias de secuencias de palabras. Contar las frecuencias de secuencias de palabras especialmente relacionadas, es decir, colocaciones. Evaluación de la fuerza de una relación, es decir, información mutua. Búsqueda de patrones y expresiones regulares.

Sección 2

- Representatividad, equilibrio y muestra. Corpus de referencia. Corpus de referencia más conocidos y otras fuentes de textos.

- Tipo de corpus. Corpus generales. Corpus especializados. Corpus escritos. Corpus hablados. Corpus sincrónicos. Corpus diacrónicos. Corpus del aprendiz. Monitor de corpus. Derechos de autor y otras cuestiones legales.

Sección 3

- Mark-up y corpus. Desde la codificación de caracteres hasta los lenguajes de marcación Corpus. Metadatos para describir el corpus.

- Anotación del corpus. Niveles de anotación lingüística. Herramientas para la anotación de corpus.

Sección 4

- Corpus paralelos y herramientas específicas. Herramientas para encontrar textos paralelos. Alineación de textos paralelos. Explotación de corpus paralelos.

Objetivos de Desarrollo Sostenible

ODS4 Educación de calidad

4.3 Para 2030, garantizar la igualdad de acceso de todos los hombres y mujeres a la enseñanza técnica, profesional y superior, incluida la universitaria, accesible y de alta calidad.

4.4 De aquí a 2030, aumentar sustancialmente el número de jóvenes y adultos que poseen las competencias necesarias, en particular técnicas y profesionales, para acceder al empleo, al trabajo decente y al espíritu empresarial.

ODS5 Igualdad de género

5.b Mejorar el uso de la tecnología instrumental, en particular la tecnología de la información y la comunicación, para promover el empoderamiento de las mujeres

ODS9 Industria, innovación e infraestructuras

9.5 Aumentar la investigación científica y mejorar la capacidad tecnológica de los sectores industriales de todos los países, en particular de los países en desarrollo, entre otras cosas fomentando, para 2030, la innovación y un aumento sustancial del número de personas que trabajan en el ámbito de la investigación y el desarrollo por millón de habitantes, así como aumentando el gasto en investigación y desarrollo en los sectores público y privado

Sistema de evaluación y calificación

La evaluación principal se basará en la obtención de evidencias de la adquisición de las competencias antes mencionadas y la nota final se valorará a partir de los siguientes ratios:

a. Prácticas: 45%

b. Proyecto final, el artículo: 45%

c. Participación en el proceso de evaluación por iguales del proyecto final: 10%

En caso de suspenso después de la evaluación principal, el estudiante deberá entregar un trabajo final (revisado) en un plazo de dos meses.