

PROGRAMA

27/02/2026

Yanco Amor Tortero Orta

Departamento de Lenguajes y Sistemas
Informáticos de la UNED

“Label Studio para OCR: construye tu
dataset de transcripciones”

23/03/2026

Olga Batiukova

Centro de Investigación en Lenguaje y Tecnologías
de la UAM (LINTEC)

“Procesamiento de corpus de prensa
histórica con Sketch Engine”

27/04/2026

Rocío Ortuño Casanova

Laboratorio de Innovación y Humanidades Digitales
de la UNED (LINHD)

“Exploración de prensa histórica con IA:
qué aporta al estudio”

Todos los talleres consistirán en una parte
teórica y otra práctica



**INSCRIPCIÓN GRATUITA
AFORO LIMITADO
SEGUIR QR PARA REGISTRO**

LUGAR Y HORARIO

DE 10:00 A 13:30H

AULA 1007 (laboratorio)

Edificio Multiusos, Universidad Complutense



CLARIAH-CM



II Ciclo CLARIAH-CM

Formación en Prensa Histórica

Herramientas y metodologías digitales

CLARIAH-CM

27/02/2026

Yanco Amor Torterolo Orta

Departamento de Lenguajes y Sistemas
Informáticos de la UNED

Label Studio para OCR: construye tu dataset de transcripciones

Este taller versa sobre cómo utilizar la herramienta open-source Label Studio para la transcripción de documentos. Se explicará en qué consiste la anotación para obtener un dataset que se pueda utilizar para *fine-tuning*, con énfasis en la tarea de OCR. Se enseñará anotación desde cero; además de introducir la preanotación con modelos de layout y/o OCR ya existentes, para su posterior revisión.

Yanco Amor Torterolo Orta es Personal Técnico de Apoyo a la investigación (PTA) en la UNED, en el departamento de Lenguajes y Sistemas Informáticos (LSI). Además, es doctorando en Lingüística Computacional por la UAM. Sus temas principales de investigación son el RAG y el OCR en documentos históricos.



Label Studio



23/03/2026

Olga Batiukova

Centro de Investigación en Lenguaje y Tecnologías
de la UAM (LINTEC)

Procesamiento de corpus de prensa histórica con Sketch Engine

En este taller práctico se introducirá a los participantes en el manejo del sistema de consulta y gestión de corpus Sketch Engine y se aplicarán algunas de sus herramientas para compilar corpus propios de prensa histórica y analizarlos.

Olga Batiukova es profesora de Lengua Española en el Departamento de Filología Española de la Universidad Autónoma de Madrid. Su investigación se centra en los modelos del léxico mental y la anotación semántica de corpus.

27/04/2026

Rocío Ortuño Casanova

Laboratorio de Innovación y Humanidades Digitales
de la UNED (LINHD)

Exploración de prensa histórica con IA: qué aporta al estudio

En este taller exploraremos la prensa histórica con herramientas digitales, contrastando métodos cuantitativos de las humanidades digitales clásicas con enfoques de inteligencia artificial para analizar, interpretar y descubrir patrones.

Rocío Ortuño Casanova es profesora de Literatura en la UNED, especialista en literatura hispanofilipina, directora del Laboratorio de Innovación en Humanidades Digitales y co-coordinadora del Grupo INHDIE de Investigación en Humanidades DigitalEs. Es también IP1 del proyecto "GRESEL-UNED: Narrativas poscoloniales en periódicos en lengua española de Asia, España y el Caribe hispánico"(PID2023-151280OB-C22) financiado por el Ministerio de Ciencia, Innovación y Universidades de España. Entre sus publicaciones destaca la coedición del libro Las humanidades digitales en la enseñanza de las literaturas hispánicas. Aplicaciones prácticas (Peter Lang 2023).