

Seminario de investigación – 07 de noviembre de 2025

Curso avanzado de web scraping para la obtención de datos turísticos

**Campus de Vicálvaro, seminario 062 departamental
15:30 – 18:30**

Impartido por:

Sofía Blanco Moreno, Profesora Ayudante Doctora, Universidad de León

Organizado por:

Miriam Menchero Sánchez, Profesor Permanente Laboral, Universidad Rey Juan Carlos

Antecedentes y Justificación

Tras la adquisición de conocimientos fundamentales sobre técnicas de web scraping y análisis de datos en el primer taller, este segundo seminario se orienta hacia la aplicación práctica y avanzada de dichas herramientas en contextos reales de investigación y desarrollo.

El web scraping avanzado no solo permite recopilar información, sino también automatizar la extracción, limpieza y análisis de grandes volúmenes de datos de manera eficiente. Combinado con técnicas de machine learning, topic modeling y análisis de sentimiento, este enfoque posibilita la detección de patrones, percepciones y tendencias relevantes en los datos digitales, con aplicaciones directas en el ámbito del turismo, el marketing y la sostenibilidad.

De este modo, el seminario busca consolidar competencias técnicas y metodológicas en el uso de herramientas libres y entornos programables, fortaleciendo las capacidades de los investigadores para el desarrollo de proyectos orientados análisis inteligente de la huella digital de los destinos turísticos.

Objetivos del Seminario

Explicación en el aula por un lado del manejo de herramientas de web scraping para la descarga de datos web, y por otro, del análisis de sentimiento de reviews y de topics a los datos obtenidos de internet.

- Aplicar técnicas avanzadas de web scraping mediante herramientas programables (Python, Selenium, BeautifulSoup, Apify, Octoparse, etc.).
- Implementar procesos automatizados de limpieza, estructuración y traducción multilingüe de datos.
- Desarrollar análisis de sentimiento y modelado temático (LDA) con herramientas libres o APIs de IA.

- Visualizar los resultados mediante herramientas de análisis y representación gráfica (Power BI, Tableau o Python).
- Fomentar la colaboración interdisciplinar y el intercambio de experiencias entre investigadores interesados en la aplicación de la IA al turismo y la sostenibilidad.

Resultados esperados y networking para los miembros del Centro

- Competencia práctica en el uso de herramientas de web scraping y análisis avanzado de texto.
- Capacidad para integrar datos heterogéneos (textuales, visuales, geográficos) en un mismo flujo de trabajo analítico.
- Generación de una base de datos propia y procesada con información real de plataformas turísticas.
- Elaboración de un informe analítico final, con métricas de sentimiento, tópicos y visualizaciones interpretables.
- Creación de una red colaborativa de investigadores interesados en la minería de datos turísticos y la inteligencia de destino.

El seminario se inscribe dentro del Proyecto Doctores Emergentes (Comunidad de Madrid – URJC) denominado “De la Huella Digital a la Física: Turismo, Innovación y ODS en el Patrimonio Mundial de Madrid (TIODS – PMMAD)” en el que participan varios miembros del Centro de Investigación de Estudios Turísticos de la URJC (CIETUR).

Programa de Trabajo

1. Ejercicio práctico de scraping avanzado (1h)

- Extracción de datos en tiempo real desde plataformas como TripAdvisor, Booking, Google Reviews o Civitatis.
- Implementación de loops, XPath selectors, y gestión de paginación y bloques dinámicos.

2. Limpieza, normalización y traducción de datos multilingües (1h)

- Uso de librerías como pandas, regex, Google Sheet y deep_translator.
- Creación de datasets estructurados y listos para análisis textual.

3. Análisis de sentimiento y topic modeling (1h)

- Introducción a técnicas de Natural Language Processing (NLP).
- Aplicación de herramientas como Azure Cognitive Services, Sentiment Analyzer, o Python (VADER, LDA, BERTopic).
- Interpretación de resultados: métricas, visualizaciones y conclusiones.

Profesorado

Impartido por Sofía Blanco Moreno, profesora experta en Marketing y Turismo con especialización en Marketing Digital, y tecnologías como web scraping e inteligencia artificial (machine learning y deep learning) aplicadas en Turismo y Hotelería. Es profesora en la Universidad de León, tras haber desarrollado su carrera profesional como experta en Marketing en empresas internacionales como Meliá Hoteles y Grupo Oblanca. Ha recibido varios premios de investigación, como el de mejor comunicación en los congresos AIRSI y AIM 2023, y los premios de transferencia de conocimiento I UAM-ASSECO Business Case Award 2022 y Prueba de Concepto dentro del Plan de Transferencia de Conocimiento Universidad-Empresa para la plataforma Photo Data Tour Analytics (<https://mktingphotodatatour.unileon.es/>).