

TALLER DE WEB SCRAPING

FECHA: 9 de junio de 2025

HORARIO: de 10 h a 14 h

LUGAR: Campus de Vegazana

Objetivos:

Este taller tiene como objetivo presentar las distintas técnicas de extracción de datos de sitios web para aplicarlas de forma ética a la investigación. Para ello, se analizará el funcionamiento de las páginas web y se describirán los distintos tipos de datos presentes, teniendo en consideración las implicaciones legales y éticas asociadas a la explotación de estos datos. Se presentarán los métodos y herramientas necesarios para la obtención, limpieza y almacenamiento de los datos extraídos. Estos métodos se detallarán en profundidad en cuatro ejercicios prácticos, con el fin de cubrir los casos de uso principales de extracción de datos.

Programa:

- 1. Introducción al Web Scraping**
 - a. ¿Qué es el Web Scraping?
 - b. Casos de uso
 - c. Aspectos legales y éticos
- 2. Fundamentos del Web Scraping**
 - a. Funcionamiento de las páginas web
 - b. Tipologías de datos y su tratamiento
- 3. Herramientas y librerías para Web Scraping**
 - a. Web Scraping con Python
 - b. Librerías para cada caso de uso
- 4. Primer caso de uso**
 - a. Web Scraping de páginas web estáticas
- 5. Segundo caso de uso**
 - a. Web Scraping de páginas web dinámicas

6. Tercer caso de uso

- a. Web Scraping de grandes volúmenes de datos

7. Cuarto caso de uso

- a. Web Scraping con modelos de lenguaje

8. Cierre

- a. Comentarios finales y discusión

Ponente: D. Miguel Molina Abad (Deloitte)