

M2.851. Tipología y ciclo de vida de los datos

Guía de estudio: Bloque 1. Preliminares

Presentación

Esta guía de aprendizaje presenta una breve exposición de los objetivos, contenidos más relevantes, recursos de aprendizaje y actividades de evaluación del bloque “Preliminares” para orientar en su estudio.

Objetivos

1. Conocer qué se entiende por Data Science y otros términos usados actualmente relacionados con el mismo.
2. Conocer el ciclo de vida de los datos, desde su captura hasta su publicación.
3. Conocer el significado y los potenciales beneficios del web scraping.
4. Explicar los desafíos legales del web scraping.
5. Realizar una investigación a fondo de nuestro sitio web objetivo para conocer la dificultad de realizar web scraping.

Contenidos

El primer bloque de preliminares está orientado a conocer el ciclo de vida de los datos, sus etapas y qué son las características de los datos. Además, se hará una introducción al web scraping indicando sus objetivos y los principales conceptos a tener en cuenta.

En la asignatura pondremos en práctica los conceptos presentados en el material de “Introducción al ciclo de vida de los datos” utilizando la técnica de web scraping,

extraemos un conjunto de diferentes tipos de datos de Internet, y posteriormente lo procesaremos con técnicas de data mining.

Recursos de aprendizaje

- M1: Subirats, L., Pérez, D., Calvo, M. (2019). Introducción al ciclo de vida de los datos. Editorial UOC.
- M2: Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC. Capítulos 1 y 6.

Material complementario

- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 1. Introduction to Web Scraping.

Conceptos más importantes

En este módulo didáctico se ha explicado qué es un dato y el ciclo de vida de los datos. También se han revisado los aspectos fundamentales relacionados con web scraping. En primer lugar, se ha presentado la utilidad y potencial de esta herramienta, particularmente útil cuando se requiere obtener información de un sitio web que no dispone de una API para tal efecto, o la API no satisface del todo las necesidades de nuestro proyecto de ciencia de datos.

A continuación, se han planteado algunos retos legales así como una lista de acciones para analizar la dificultad de realizar web scraping en un sitio web.

Evaluación

La evaluación de este bloque se compone de una actividad teórica-práctica:

1. PEC 1: Se trabajan los conceptos generales del tipo de datos, el ciclo de vida de los datos y también los conceptos principales de web scraping.