



M2.851 - Tipología y ciclo de vida de los datos aula 1

Práctica 1: Web scraping

Inicio:	Entrega:	Solución:	Calificación:	Dedicación:
09/03/18	16/04/18	23/04/18	27/04/18	35 %

Enunciado de la práctica 1

Fichero PDF a descargar, con el enunciado de la Práctica 1.

M2.851_20172_Practica1.pdf

Descripción

En esta práctica se elabora un caso práctico orientado a aprender a identificar los datos relevantes para un proyecto analítico y usar las herramientas de extracción de datos. En la realización de la práctica deberéis trabajar en grupo de 3 ó 2 personas, o si preferís, también podréis hacerlo de manera individual.

La entrega de la práctica deberéis realizarla en un único fichero en el enlace Github (<https://github.com>) con las soluciones incluyendo los nombres de los componentes del equipo. Podéis utilizar la Wiki de Github para describir vuestro equipo y los diferentes archivos de vuestra entrega. Cada miembro del equipo deberá contribuir con usuario de Github.

Podéis consultar los siguientes ejemplos como guía:

- Ejemplo: <https://github.com/rafoelhonrado/foodPriceScraper>
- Ejemplo complejo: <https://github.com/tteguayco/Web-scraping>

Competencias y objetivos

Competencias

En esta Práctica se desarrollan las siguientes competencias del Master de Data Science:

- Capacidad de analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para abordarlo y resolverlo.
- Capacidad para aplicar las técnicas específicas de web scraping.

Objetivos

Los objetivos concretos de esta práctica son:

- Aprender a aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios o multidisciplinares.
- Saber identificar los datos relevantes que su tratamiento aportan valor a una empresa y la identificación de nuevos proyectos analíticos.
- Saber identificar los datos relevantes para realizar un proyecto analítico.
- Capturar datos de diferentes fuentes de datos (tales como redes sociales, web de datos o repositorios) y mediante diferentes mecanismos (tales como queries, API y scraping).
- Actuar con los principios éticos y legales relacionados con la manipulación de datos en función del ámbito de aplicación.
- Desarrollar la capacidad de búsqueda, gestión y uso de la información y de los recursos.

Recursos

Los siguientes recursos son de utilidad para la realización de la práctica:

- Masip, D. (2010). El lenguaje Python. Editorial UOC
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Simon Munzert, Christian Rubba, Peter Meißner, Dominic Nyhuis. (2015). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.
- Tutorial de Github <https://guides.github.com/activities/hello-world> .

Materiales y fuentes de información

Materiales

El lenguaje Python

 [PDF](#)

Fuentes de información
