

Tipología y ciclo de vida de los datos

PRA1 - Web Scraping

Alberto García Galindo y Federico Alejandro Floriano Pardal

26 de octubre de 2020

0. Introducción

En este documento se expone el informe final del caso práctico resuelto de *web scraping* acerca de la construcción de un conjunto de datos con información práctica de los mejores jugadores de padel del mundo. En concreto, se extrajo información del sitio web del World Padel Tour, y se recogió información tanto de la clasificación mundial masculina como femenina.

Para ello, se utilizó Python como herramienta principal, destacando el uso de algunas librerías útiles en el ámbito de *web scraping* como Selenium.

1. Contexto

A día de hoy, el pádel se sitúa como uno de los deportes más en auge de España. Cada vez son más las personas que disfrutan de él a todos los niveles y que lo practican de manera habitual. A nivel profesional, el circuito profesional del World Padel Tour se sitúa como el principal campeonato en el ámbito internacional, donde los mejores jugadores recorren el mundo y compiten por ser los mejores.

En este contexto, se enmarca la elaboración de esta práctica, que pretende realizar un ejercicio de *web scraping* para extraer de la página oficial de World Padel Tour la información tanto personal como deportiva de los mejores jugadores y jugadoras del mundo. Para ello, accederemos a la página web de la clasificación y exploraremos la manera de elaborar un conjunto de datos que permita la posterior explotación de dichos datos.

2. Título del conjunto de datos

Top 100 World Padel Tour Dataset

3.