

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 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 la librería Selenium, especialmente útil a la hora de trabajar con páginas web dinámicas como a la que nos hemos enfrentado.

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 y jugadoras recorren el mundo compitiendo y proporcionando un espectáculo único.

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 los mismos.

Cabe mencionar que, en el ámbito de este deporte, la analítica de datos aún no he irrumpido con la misma magnitud que en otros deportes como el fútbol o el tenis, por lo que la publicación del conjunto de datos resultante supondrá uno de las primeras contribuciones.

2. Título del conjunto de datos

Top 100 World Padel Tour Dataset

3. Descripción del conjunto de datos

Como resultado del ejercicio de *web scraping*, se ha obtenido un conjunto de datos...

4. Representación gráfica

5. Contenido

A continuación se detallan cada una de las características que han sido recogidas para cada uno de los profesionales.

- **Nombre:** Nombre del jugador en el formato recogido por la página de World Padel Tour.
- **Posición:** Posición en el ránking World Padel Tour.

- **Puntos:** Puntuación actual.
- **Compañero:** Compañero profesional actual del jugador.
- **Pista:** Posición habitual en pista del jugador: izquierda o derecha.
- **Lugar de nacimiento:** Lugar de nacimiento del jugador.
- **Edad:** Edad del jugador en años.
- **Altura:** Altura del jugador.
- **Residencia:** Lugar de residencia del jugador.
- **Partidos jugados:** Número total de partidos jugados desde que se tienen registros en World Padel Tour.
- **Partidos ganados:** Número total de partidos ganados.
- **Partidos perdidos:** Número total de partidos perdidos.
- **Rendimiento:** Porcentaje total de victorias.
- **Racha:** Mayor racha de partidos ganados de manera consecutiva.

Por otra parte, para cada uno de los años disponibles (entre 2013 y 2020), se recogieron algunas características adicionales:

- **Partidos jugados**
- **Partidos ganados**
- **Partidos perdidos**
- **Rendimiento**
- **Torneos:** Número de torneos ganados.
- **Finales:** Número de veces en los que el jugador llegó a la final y la perdió.

6. Agradecimientos

7. Inspiración

8. Licencia