

Películas de Tom Hanks

Jiaqian Lin

2020/10/31

Descripción

El conjunto de datos generado en esta práctica recoge la información de las películas que participó el actor **Tom Hanks** tanto como protagonista, actor secundario o doblaje.

Imagen identificativa



Figure 1: Forograma de la película "Forrest Gump"

Contexto

El conjunto de datos contiene toda la información de las películas participadas del actor americano **Tom Hanks**. Hay películas en las que el actor no era el protagonista sino, solo un actor secundario y otras que actuó como doblaje.

Los datos son extraídos de **Wikipedia**. Más concretamente en la pagina de "*Filmografía de Tom Hanks*" donde listó todas las actuaciones del Tom Hanks.

Contenido

Para cada accidente, se ha recogido las siguientes características:

- **Title:** el título de la película.
- **Directed by:** director de la película.
- **Produced by:** productor cinematográfico de la película.
- **Written by:** guionista de la película.
- **Starring:** actores principales.
- **Music by:** banda sonora.
- **Based on:** Si la película es basado en alguna obra o un hecho real.
- **Cinematography:** director de fotografía.
- **Screenplay by:** guionista.
- **Edited by:** director de montaje.
- **Production companies:** Lista de las compañías productoras.
- **Distributed by:** Compañía distribuidora.
- **Story by/Narrated by:** narrador de la película.
- **Release date:** fecha de estreno.
- **Running time:** tiempo de duración.
- **Country:** país de producción.
- **Language:** lenguaje original de la película.
- **Budget:** el presupuesto de la película.
- **Box office:** la recaudación obtenida.

Agradecimientos

Los datos han sido recolectados desde la página web **Wikipedia**. Para realizar esta práctica se ha hecho uso del lenguaje de programación Python y de técnicas de *Web Scraping* para extraer la información alojada en la web.

Inspiración

Este conjunto de datos es más del tipo informativo que para trabajar con ello (construcción de modelos,...). Se podría ser utilizado para una web de fans de Tom Hanks donde se lista las actuaciones del dicho actor.

Licencia

Como había comentado en el apartado anterior, este conjunto de datos es del tipo informativo para que todo el mundo pueda compartir. Además, al ser un conjunto no completo (puede ir añadiendo más películas), estaría libre de que otros usuarios pueda añadir o modificar el dicho conjunto de datos. Por dichos motivos, la licencia escogida para la publicación será **CC BY-SA 4.0 Licence**.

Código fuente y dataset

El código fuente y el dataset se puede encontrar en el repositorio de Github: <https://github.com/jiaqian94/Practica-1-Web-Scraping>

Recursos

1. Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC.
2. Masip, D. (2010). El lenguaje Python. Editorial UOC.
3. Tutorial de Github <https://guides.github.com/activities/hello-world>.