

Documentación - Indeed web scraper for data scientist in Barcelona

Ricardo Martinez y Joan Prieto

11/04/21

Contents

1	Contexto	2
2	Título dataset	2
3	Contenido	2
4	Agradecimientos	2
5	Inspiración	2
6	Licencia	3
7	Codigo	3

1 Contexto

Este proyecto se ha creado para cubrir la necesidad de entrar de manera reiterada a las paginas webs que listas las ofertas de trabajo (indeed.com). En este caso lo hemos adptado a trabajos de data scientist ubicados en Barcelona.

2 Título dataset

Indeed web scraper for data scientist in Barcelona

3 Contenido

Nos encontraremos con los siguiente datos:

- **title:** Titulo de la oferta de trabajo
- **company:** Nombre de la compañía que ha creado la oferta
- **location:** Localización del puesto de trabajo
- **descripcion:** Detalles de la oferta.
- **URL:** Link para visualizar la oferta en su web.
- **Fecha de publicación:** Cuando se publicó
- **Profile:** Si es junior senior o indefinido.

Para recoger los datos hemos utilizado beautifulsoup4 y selenium. A través de bs4 hacemos la petición y transformamos el html a datos, mientras que con *selenium* podemos interactuar con la página, así entrar en cada una de las ofertas y seguir extrayendo mas información. A todo esto le hemos puesto un *timeout* para evitar sobrecargar el servidor y que nos bloqueen.

4 Agradecimientos

A Indeed.com, porque ha creado una plataforma simple y con los datos necesarios para poder resumir un puesto de trabajo de manera sencilla y clara. Evidentemente, con la modernización de las tecnologías es necesario la creación de portales que recogen las ofertas laborales, así poder consultar en cada momento los posibles trabajos.

5 Inspiración

Aunque ambos tenemos trabajo, es importante seguir al día con las ofertas de trabajo, así que se nos ocurrió crear un pequeño scraper para recoger la información de las nuevas ofertas de data-science con mas facilidad. Para ello recurrimos a indeed.com, una web reconocida mundialmente por listar ofertas de trabajo de muchos paises.

6 Licencia

La licencia que hemos decidido usar es la *Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)*, es una licencia que permite copiar,redistribuir y transformar para cualquier propósito (incluido comercial).

Bajo los términos:

- *Attribution*: Deberán dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. De esta manera se reconocerá el trabajo del autor y el que haga uso y aporte sobre el proyecto.
- *ShareAlike*: Deberán compartir el código de cualquier modificación con la misma licencia, con esto beneficiamos a la creación de un mejor proyecto y que sea de dominio publico.

Hemos escogido esta licencia ya que la simpleza del script y su fácil adaptación (es tan sencillo como cambiar la url para cambiar el trabajo o la localización, o añadir un parámetro para dinamizar los parámetros de la url) permiten ser modificados con facilidad para cualquier otro fin.

7 Codigo

Se puede acceder al código mediante el enlace: [Indeed Web Scraper](#)

Contribuciones	Firma
Investigación previa	Joan Prieto y Ricardo Martinez
Redacción de las respuestas	Joan Prieto y Ricardo Martinez
Desarrollo código	Joan Prieto y Ricardo Martinez