

Tipología y ciclo de vida de los datos

Práctica 1 - Web Scraping

Jose Maria Bernet

1. Titulo

Ofertas de trabajo del sector TIC en España

2. Descripción - Subtitulo

El presente dataset contiene las últimas ofertas de trabajo publicadas por la web www.tecnoempleo.com en España pertenecientes el sector de las TIC.

3. Imagen



BÚSQUEDA DE EMPLEO POR ANUNCIOS CLASIFICADOS

4. Contexto

La materia del conjunto de datos corresponde con las últimas ofertas de empleo publicadas en la web www.tecnoempleo.com a fecha de **12/11/2018**. La información que contiene el dataset incluye información cómo el puesto solicitado, el rango de salarios a ofrecer, la localidad de la oferta etc.

5. Contenido

Cada oferta de trabajo corresponde con un registro del dataset, las columnas que se recogen en el dataset son las siguientes:

- **name:** Nombre descriptivo de la oferta
- **id:** Identificador de la oferta
- **url:** Enlace a la oferta

- **puesto:** Puesto solicitado en la oferta
- **contrato:** Tipo de contrato ofertado
- **jornada:** Tipo de jornada de trabajo
- **localización:** Localidad o ciudad de la oferta
- **date:** Fecha de publicación o de actualización de la oferta
- **company:** Empresa que publica la oferta
- **min_salario:** Umbral mínimo de salario ofertado
- **max_salario:** Umbral máximo de salario ofertado

Los registros del dataset son una porción de ofertas recogidas a fecha **12/11/2018** de la web www.tecnoempleo.com más concretamente de su listado principal con el único filtro de que sean españolas, la web página las ofertas cada 30 y se ha ejecutado el script para que recoja las 20 primeras páginas, por lo que resulta un dataset de 600 ofertas.

6. Agradecimientos

La información ha sido recogida del portal de empleo **TecnoEmpleo**, siendo este el mayor portal de ofertas de trabajo del sector TIC en España. Para la recogida de los datos se ha usado el lenguaje **Python** y las librerías de **Request** y **BeautifulSoup** así como la librería **Pandas** para la creación del conjunto de datos y posterior exportación a CSV.

7. Inspiración

Creo que el uso de este dataset puede usarse para fines estadísticos, poder hacer minería de datos y extraer información acerca de los empleos más demandados, de las ciudades que más ofertas publican o incluso cuáles son las localidades que más ofrecen por qué tipo de puesto.

La información que se recoge es una mínima parte de toda la que se puede recoger, además si la recogida de información se hiciera de manera continua en el tiempo, podríamos sacar modelos predictivos pudiendo identificar incluso que fechas del año son en las que se produce un aumento de la demanda de empleo.

8. Licencia

La licencia sobre la cual publico el dataset y el código de script que he usado para realizar el web scrapping es **CC0: "Public Domain License"** ya que la información recogida en el mismo es de dominio publico y a la que todos pueden acceder sin necesidad si quiera de estar registrado en la web en cuestión (www.tecnoempleo.com).

Así mismo invito a otros desarrolladores a mejorar las técnicas usadas en mi script de scrapping o incluso que pueda servir de ayuda a otros desarrolladores en sus propios proyectos personales.

9. Código y Dataset

Tanto el código del script cómo el dataset resultante se encuentran alojados en el repositorio <https://github.com/JBF91/tecnoScrapping>