

AGENTE DE RECOLECCIÓN

Rubén Giráldez de la Fuente

- ¿Qué información recoge el agente?

El agente que se ha programado en esta práctica de la asignatura recopila los datos relativos a los precios de la sesión diaria de las acciones de la Bolsa de Madrid. Estos datos son:

- Nombre de la compañía
- Último valor
- Diferencia
- Máximo valor
- Mínimo valor
- Volumen
- Efectivo
- Fecha de la sesión
- Hora

- ¿Cómo se han conseguido los datos?

Para poder recopilar los datos de las acciones de la Bolsa de Madrid, primero se ha hecho un análisis de cómo están organizados los elementos de la página web en el contenido HTML, en este caso cómo se organiza la tabla de acciones. Una vez que se han localizado los elementos que eran de interés, se han escrito expresiones regulares que discriminen a todos los elementos de la tabla, menos al que se pretende encontrar en cada caso particular. Como en algunos casos el contenido que se buscaba se encontraba rodeado de una estructura idéntica a otras que se encontraban inmediatamente después y antes, se ha procedido a filtrar discriminando por el formato del contenido en cuestión, como por ejemplo para poder encontrar los campos mínimo y volumen.

En el resto de casos, se ha filtrado en base a la clase de la etiqueta o a la manera en la que estaba alineado el contenido en la celda de la tabla.

- ¿Cómo se han ido guardando los datos?

Para guardar los datos, se ha creado una lista para cada campo de la tabla. Cuando se llama a la función que recoge los datos de cada campo, se realiza una operación append en la lista correspondiente a cada campo.

- ¿Cómo se han presentado los datos recopilados?

Una vez recogidos los datos de la tabla de acciones, primero se han convertido las listas de listas en listas, y después se ha creado un DataFrame con pandas para poder visualizar los datos de una mejor manera. Esto se ha conseguido gracias a utilización de la librería pandas, y a la creación de un diccionario que contiene todos los campos, distinguidos por su etiqueta que los define. En caso de que se quiera realizar un análisis más exhaustivo sobre los datos recopilados, se han volcado los datos en un archivo excel. Esto ha sido posible gracias a la utilización de la librería openpyxl.