



Comparador de precios de videojuegos.

En este proyecto se ha tratado de extraer el precio de videojuegos de diferentes fuentes de datos mediante "Web Scraping", este proceso consiste en recoger la información que necesitamos directamente desde el código html de la pagina desde la que extraemos la información. Las paginas que se han utilizado son las siguientes:

<https://www.otogami.com/>

<https://www.xtralife.es/>

Para la obtención de los datos que requerimos, en este caso el nombre, precio, y link al producto se ha utilizado la librería JSOUP. La herramienta sobre la que se ha trabajado es Eclipse Neon 3.

La url desde la que se puede acceder al sistema es:

<https://gamecomparator-168709.appspot.com>

Una vez aclarado esto paso a explicar el funcionamiento del programa.

1. Clases

En este programa se han 2 paquetes diferentes: Uno que contiene el Servlet y otro que contiene todas las otras clases de programa. Comenzaremos por las clases de java:

- Core.java: Esta clase es la que, una vez obtenido un Array con los resultados de las búsquedas genera la pagina html y ordena los resultados obtenidos en el Array de menor a mayor.
- Game.java: En esta clase se ha definido el objeto Game, que contendra el nombre, link, sitio y precio de cada juego.
- JsoupMethods.java: Esta clase se ha utilizado para agrupar los dos métodos de JSOUP que se utilizan. Estos son GetHtmlDocument(url) y GetStatusConnection(url):
 - GetHtmlDocument(url): Este metodo descarga todo el contenido del HTMLde la url.
 - GetStatusConnection(url): Este método comprueba cual es el StatusCode de la respuesta que se recibe de la url al hacer la petición, si el codigo es 200 significa que todo esta bien.
- WebScanning: En esta clase se define la forma en la que se va a hacer el scraping en cada pagina. El método es diferente en cada una de ellas ya que la arquitectura de la pagina también cambia, pero podemos resumirlo en que primero coge la cadena a buscar y la introduce en la url de la pagina, mas tarde obtiene el html de la pagina y a partir de una etiqueta o una clase en el html obtenemos los datos que necesitamos. Ademas de esto, dado que es posible que el string que recuperemos de la pagina no tenga el formato que queremos como es el caso de string, hay que recorrer todo el string buscando solo caracteres que sean o numeros o un separador de decimales, una coma o un punto.

Ademas de estas clases hay un servlet, que es el que se encarga de la comunicación con el servidor. Este hace las llamadas a las clases mencionadas anteriormente para obtener, ordenar y mostrar los resultados de la búsqueda.

2. WEB

Este sistema tiene 2 webs o documentos html, el primero muestra una simple introducción con un botón para ir a la siguiente pagina a traves de la cual se realiza la busqueda de un videojuego.

La comunicación entre la web y el servlet se ha especificado a través del atributo “action” del formulario, en este atributo se ha añadido la dirección del servlet, enviándole así el texto de búsqueda. Por otro lado en el servlet se ha obtenido a través del método “doGet”.