

**Student**: Agustin Luques

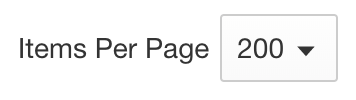
**Subject**: MIE-DDW

**Homework 1**

**Deadline**: 17.03.2019

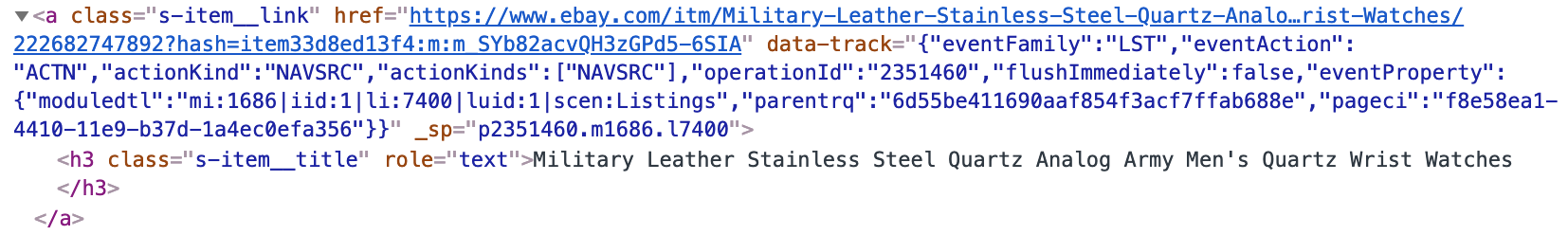
Firstly, the domain that was crawled id Ebay.com.

I searched for “watches” and changed the amount of iems per page in order to crawl as much websites as I could.



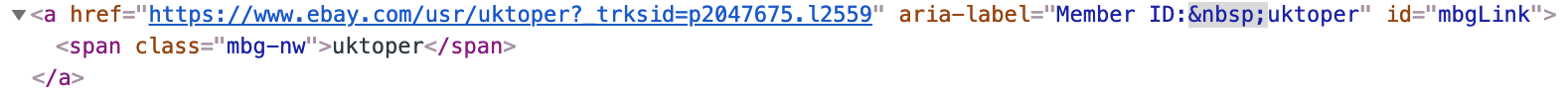
After that, I copied the link[[1]](#footnote-1) and set it as the “start url”.

I get all the products which are in the main page by searching for all the <a> tags which have the class “s-item\_\_link”. For each product, I extracted the href attribute and the text from the <h3> inside.



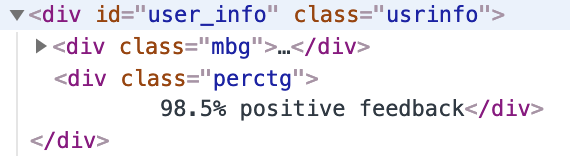
If the product link was not in the items I had crawled, I added it to the list and crawl it.

In the new type of website, I searched for the product’s owner information. To get it, I had to search by the id “mbgLink” where from I took the “href” pointing to the owner profile and the text from the <span> inside.



Before, I’d saved the product information in order to make a clear JSON output. Here, I did not have a list with the owners I had crawled because it could be that the same owner had many products.

Then, I crawled the owner’s profile and got the information of the feedback people gave to him/her. To do this, I simply took the text from the element in which that info was.



1. <https://www.ebay.com/sch/i.html?_from=R40&_nkw=watches&_sacat=0&_ipg=200> [↑](#footnote-ref-1)