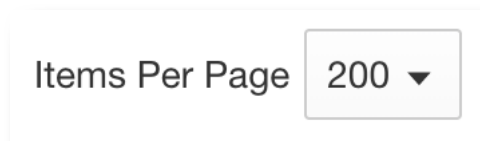**Student**: Agustin Luques

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] 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.

```
▼<a class="s-item__link" href="https://www.ebay.com/itm/Military-Leather-Stainless-Steel-Quartz-Analo…rist-Watches/
222682747892?hash=item33d8ed13f4:m:m_SYb82acvQH3zGPd5-6SIA" data-track="{"eventFamily":"LST","eventAction":
"ACTN","actionKind":"NAVSRC","actionKinds":["NAVSRC"],"operationId":"2351460","flushImmediately":false,"eventProperty":
{"moduledtl":"mi:1686|iid:1|li:7400|luid:1|scen:Listings","parentrq":"6d55be411690aaf854f3acf7ffab688e","pageci":"f8e58ea1-
4410-11e9-b37d-1a4ec0efa356"}}" _sp="p2351460.m1686.l7400">
    <h3 class="s-item__title" role="text">Military Leather Stainless Steel Quartz Analog Army Men's Quartz Wrist Watches
    </h3>
</a>
```

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.

```
▼<a href="https://www.ebay.com/usr/uktoper?_trksid=p2047675.l2559" aria-label="Member ID: uktoper" id="mbgLink">
    <span class="mbg-nw">uktoper</span>
</a>
```

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

```
▼<div id="user_info" class="usrinfo">
 ▶<div class="mbg">…</div>
   <div class="perctg">
           98.5% positive feedback</div>
 </div>
```