**Nikita Iashin**

**04.09.2023**

**Task report**

When I started task – I immediately met a problem. I couldn’t open site without VPN. And I couldn’t find any good free PC-based VPN, and only one browser-based VPN worked with this site. Second problem – site has “antibot” system that was I wasn’t been able to pass it even “by hands”, like it wasn’t working at all. And only way to pass it was to close browser and open it again.

With this, I started to parse first page. It needed to “open” all blocks of items before getting urls, but after tapping on button with script – I started to receive 403, cause “human verification”. New data wasn’t loading at all. With that, I did it half-manually – loaded whole page and saved it as .html file. Then with script I parsed all links on items with BeautifulSoup and created json file with all needed links to items pages. Also, I filtered some pages so there were two items with same names but different images on main page, I put them into one list in dictionary. Code for parsing is in “parse\_main\_page.py”

After parsing main page and creating list of all item page urls, I started to parse pages for getting image urls. In “parse\_item\_page.py” I made function to load two images by url. After manual research I found that ~95% pages have first image as person image and second as cloth. Other pages have two cloth images. With that I decided to parse both images’ urls in order and just add them to list.

But, because of problems with VPN and anti-bot, I could load data from only one page at one script run. With that, process was not faster than ‘manual’ one, so I loaded data only about few pages. It could be fixed with using good PC-based VPN, or running script on PC located in other country.

After loading urls of images, I made script to load images (that I was able to open without VPN) and put them into corresponding folders, using script in “load\_images.py”. Example of related images for “The RL Fleece Hoodie” item for cloth and person folders:

