Отчет по «Сбор и разметка данных» семинару №7.

**Исполнитель**: Григорьев Н.А.

**Цель работы**: Провести автоматический сбор данных с сайта, используя Python с библиотеками Selenium и BeautifulSoup.

**URL**: https://www.foroffice.ru/

**Описание сайта:**

Был выбран сайт поставщика промышленного и офисного оборудования и на сайте можно найти обширную номенклатуру различных видов товаров (принтеры, столы, резаки, прессы) с указанием цены.

**Анализ сайта.**

Скрипт в начале своей работы запрашивает требуемую для поиска информацию, проводит сбор результатов поиска на сайте в соответствии с этим запросом. Скрипт записывает корректный URL каждого товара, его название и стоимость в отдельную переменную.

Собранная информация со страницы передавалась в словарь result\_dict (ключ – URL товара, значение – название товара и его стоимость).

**Навигация по сайту.**

Python Selenium открывает страницу в Chrome и вводит в строку поиска запрос, после чего автоматически нажимает на кнопку поиска и открывается первая страница результатов поиска.

Переключение страниц результатов поиска реализовано через изменение самого URL страницы и открытие страницы по этой URL. Для скорости обработки, сбор данных осуществляется через BeautifulSoup (при необходимости можно отключить открывание новых страниц поиска). В коде реализовано только поиск по 3 страницам, хотя код легко модернизируется для поиска по всем результатам поиска.

**Проблемны и трудности кода.**

1. Переключение страниц реализовано не по клику на элемент “next-page” внизу страницы, а при помощи модернизации URL страницы результата поиска. Это исполнено так как реализация клика на “next-page” в Selenium систематически вызывает ошибку, решение которой не было найдено.
2. Проводились тесты с использованием браузера FireFox, который хоть и показал меньшую стабильность работы чем Chrome, но скорость загрузки у него значительно выше.