**Сбор и разметка данных (семинары)**

**Урок 7. Selenium в Python**

**Краткий отчет:**

**URL сайта:** <https://books.toscrape.com>

**Описание:** Извлекаемая информация включает названия книг, ссылки, цены и наличие товара из раздела Travel.

**Подход:** Используется Selenium, файл «Selenium\_HW\_7.py», для навигации по сайту и с помощью библиотеки BeautifulSoup для парсинга HTML-контента.

На сайте с помощью функции «Просмотр кода элемента» получили все необходимые данные и теги.

Далее, в файле «BeautifulSoup\_HW\_7.py», сформировали словарь book\_data, который записали в json файл «travel\_books\_data.json», используя метод try-except.

**Трудности:** Основной проблемой было нахождение нужных тегов, так как очень много дублируемых тегов не только на исследуемом сайте, но и на многих других. Поиск осуществлялся методом подбора и тестирования через Jupyter Notebook.

**Результаты:** Данные сохранены в файл «travel\_books\_data.json» в формате json с данными "Title", "URL", "Price", "Availability".