**Отчет по скрипту Python для веб-скрейпинга**

**URL сайта**

**URL**: [https://www.wildberries.ru](https://www.wildberries.ru/)

**Описание**

Цель этого проекта заключалась в извлечении информации о книгах Фридриха Дюрренматта с сайта Wildberries. Извлеченные данные включали название книги, цену, бренд (издательство), артикул, автора, жанр, язык, год выпуска и обложку. Эти данные можно использовать для анализа доступности и цен на книги этого автора на платформе.

**Подход**

Для навигации по сайту и извлечения необходимых данных был использован веб-скрапинг с помощью библиотеки Selenium. Скрипт выполнял следующие действия:

1. **Инициализация браузера**: создание экземпляра браузера Chrome с нужными параметрами.
2. **Поиск книг**: откройте главную страницу Wildberries, введите запрос «книги Фридриха Дюрренматта» в поле поиска и нажмите «Enter».
3. **Сбор ссылок на книги**: прокрутка страницы для загрузки всех доступных карточек книг и сбор их URL-адресов.
4. **Извлечение сведений о книгах**: для каждой ссылки на книгу открывался новый экземпляр браузера, из которого извлекались характеристики книги, такие как название, цена и другие атрибуты, хранящиеся в таблице на странице.
5. **Сохранение данных**: собранные данные записывались в файлы CSV и JSON для дальнейшего анализа.

**Трудности**

Во время реализации проекта возникли несколько проблем:

* **Динамическая загрузка контента**: некоторые элементы на странице загружались динамически, что требовало использования явных ожиданий (WebDriverWait) для гарантии их доступности перед доступом к ним.
* **Ошибка при поиске элементов**: иногда скрипт не мог найти элементы из-за изменений в структуре HTML на сайте. Это потребовало адаптации XPath-запросов и классов, используемых для поиска.
* **Ограничения на количество запросов**: при слишком частом скрейпинге сайт мог временно блокировать доступ. Для решения этой проблемы были добавлены задержки между запросами, чтобы имитировать поведение пользователя.

**Результаты**

В результате выполнения скрипта была собрана информация о количестве книг, а также их характеристиках. Пример извлечённых данных в формате JSON:

[

{

"name": "Книга 1",

"price": 500.0,

"brend": "Издательство 1",

"url": "https://www.wildberries.ru/catalog/123456",

"article": "123456", "author": "Фридрих Дюрренматт",

"genre": "Драма",

"language": "Русский",

"year": "2019",

"cover": "Твердый"

},

{ "name": "Книга 2",

"price": 750.0,

"brend": "Издательство 2",

"url": "https://www.wildberries.ru/catalog/654321",

"article": "654321", "author": "Фридрих Дюрренматт",

"genre": "Философия",

"language": "Русский",

"year": "2020",

"cover": "Мягкий"

}

]

Данные могут быть использованы для анализа рынка книг Фридриха Дюрренматта в России. Скрипт учитывает условия обслуживания сайта и избегает чрезмерного скрейпинга, что обеспечивает корректную работу платформы.