

Сбор и разметка данных (семинары)

Урок 7. Selenium в Python

Краткий отчет:

URL сайта: <https://book24.ru/catalog/business-1671/>

Описание: Извлекаемая информация включает названия книг, авторов, цены и о товаре из раздела бизнес-литературы.

Подход: Используется Selenium для навигации по сайту и BeautifulSoup для парсинга HTML-контента. Информация извлекается из элементов, идентифицированных по классам.

Трудности: Основной проблемой было управление динамической навигацией и обработка исключений, таких как StaleElementReferenceException. Проблемы были решены путем переопределения элементов и использования JavaScript для кликов.

Результаты: Данные сохранены в файл books.csv в формате CSV с колонками "Title", "Author", "Price", "About".