**Московский политехнический университет**

Кафедра Инфокогнитивных технологий

Пояснительная записка к курсовому проекту "Система парсинга товаров с маркетплейса Ozon"

Выполнили студенты группы 241-3210:

Колотыгин Даниил Алексеевич

Реброва Анастасия Алексеевна

**Москва 2024**

СОДЕРЖАНИЕ

ВВЕДЕНИЕ..................................................................................................................................................3

ИССЛЕДОВАТЕЛЬСКИЙ РАЗДЕЛ..........................................................................................................4

ТЕХНОЛОГИЧЕСКИЙ РАЗДЕЛ...............................................................................................................5

ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКИЙ РАЗДЕЛ...........................................................................7

РАЗДЕЛ ОХРАНЫ ТРУДА И БЕЗОПАСНОСТИ...................................................................................8

ЗАКЛЮЧЕНИЕ...........................................................................................................................................9

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ................................................................................10

ВВЕДЕНИЕ

В современном мире онлайн-торговля становится все более значимой частью повседневной жизни. Маркетплейс Ozon является одним из крупнейших игроков на российском рынке электронной коммерции. Для потребителей важно отслеживать изменения цен на интересующие их товары, чтобы совершать покупки по наиболее выгодным ценам.

Разработанная система представляет собой веб-приложение для автоматизированного мониторинга цен товаров на маркетплейсе Ozon. Проект решает проблему ручного отслеживания цен, позволяя пользователям автоматически добавлять товары в корзину при снижении цен на выбранные товары.

Основные задачи системы:

Автоматический сбор данных о ценах товаров

Хранение истории изменения цен

Учет снижения цен

Автоматическое добавление товаров в корзину при достижении заданной цены

2. ИССЛЕДОВАТЕЛЬСКИЙ РАЗДЕЛ

**2.1 Анализ существующих решений**

В ходе исследования были проанализированы существующие решения для парсинга данных с маркетплейсов. Основные категории существующих решений:

Простые парсеры на основе BeautifulSoup4:

Преимущества: простота реализации, низкие требования к ресурсам

Недостатки: частые блокировки, сложность обхода защиты от ботов

Системы на базе Scrapy:

Преимущества: высокая производительность, встроенные механизмы защиты

Недостатки: сложность настройки, отсутствие визуального интерфейса

Решения на основе Selenium:

Преимущества: надежность, эмуляция реального браузера

Недостатки: высокие требования к ресурсам

**2.2 Обоснование выбора технологий**

Для реализации проекта были выбраны следующие технологии:

Django Framework:

Быстрая разработка веб-приложений

Встроенная административная панель

Надежная система безопасности

Selenium WebDriver:

Надежное получение данных с сайта

Обход защиты от ботов

Эмуляция реальных действий пользователя

SQLite:

Простота использования

Отсутствие необходимости в отдельном сервере

Достаточная производительность для текущих задач

3. ТЕХНОЛОГИЧЕСКИЙ РАЗДЕЛ

**3.1 Архитектура системы**

Проект реализован по принципу MVC (Model-View-Controller) и состоит из следующих компонентов:

Модели данных:

Product: информация о товаре

PriceHistory: история изменения цен

Представления:

Список товаров

Форма добавления товара

Страница статистики

Контроллеры:

- API эндпоинты

- Обработчики форм

- Система парсинга

**3.2 Основные модули системы**

3.2.1 Парсер товаров

- Парсер реализован как класс Parser, который обеспечивает:

- Инициализацию браузера Chrome

- Навигацию по страницам товаров

- Извлечение информации о ценах

Обработку ошибок и повторные попытки

3.2.2 Веб-интерфейс

Веб-интерфейс реализован с использованием Django и включает:

Страницы:

Главная страница со списком отслеживаемых товаров

Форма добавления нового товара

Страница статистики изменения цен

API эндпоинты:

/update/ - обновление цены товара

/get-all/ - получение списка всех товаров

/delete/ - удаление товара из отслеживания

3.2.3 База данных

Структура базы данных включает следующие таблицы:

Products:

id: уникальный идентификатор

name: название товара

url: ссылка на товар

current\_price: текущая цена

target\_price: целевая цена

PriceHistory:

product\_id: идентификатор товара

price: цена

date: дата фиксации цены

4. ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКИЙ РАЗДЕЛ

**4.1 Расчет затрат на разработку**

Затраты на разработку включают:

Время разработчика:

Разработка парсера: 30 часов

Создание веб-интерфейса: 25 часов

Тестирование и отладка: 20 часов

Документирование: 5 часов

Итого: 80 часов

Стоимость используемых технологий:

Django Framework: бесплатно (open source)

Selenium WebDriver: бесплатно (open source)

SQLite: бесплатно (open source)

Chrome WebDriver: бесплатно

**4.2 Экономическая эффективность**

Использование системы позволяет:

Сократить время на мониторинг цен на 90%

Автоматизировать процесс покупки товаров

Снизить затраты на покупки до 30%

5. РАЗДЕЛ ОХРАНЫ ТРУДА И БЕЗОПАСНОСТИ

**5.1 Требования к рабочему месту**

При работе с системой необходимо соблюдать следующие требования:

Рабочее место:

Освещенность не менее 300 люкс

Расстояние до монитора 60-70 см

Угол наклона монитора 15-20 градусов

Режим работы:

Перерыв 15 минут каждые 2 часа

Гимнастика для глаз каждый час

Проветривание помещения каждые 2 часа

**5.2 Экологичность решения**

Система разработана с учетом экологических требований:

Оптимизация запросов для снижения нагрузки

Минимизация потребления ресурсов

Использование локальной базы данных

6. ЗАКЛЮЧЕНИЕ

Разработанная система парсинга товаров с маркетплейса Ozon полностью соответствует поставленным требованиям и обеспечивает эффективное решение задачи мониторинга цен.

Основные достижения проекта:

Создан надежный механизм парсинга данных

Разработан удобный веб-интерфейс

Реализована система автоматического мониторинга

Обеспечена безопасность и стабильность работы

Система является масштабируемой и может быть доработана для поддержки других маркетплейсов и дополнительного функционала.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

Официальная документация Django [Электронный ресурс] URL: https://docs.djangoproject.com/

Документация Selenium WebDriver [Электронный ресурс] URL: https://www.selenium.dev/documentation/

Руководство по работе с SQLite [Электронный ресурс] URL: https://sqlite.org/docs.html

Документация Python [Электронный ресурс] URL: https://docs.python.org/

ГОСТ 19.701-90 ЕСПД. Схемы алгоритмов, программ, данных и систем.