

Волгоградский Государственный Технический Университет
Факультет «Электроники и вычислительной техники»
Кафедра «Программное обеспечение автоматизированных систем»

Совершенствование веб-сервиса для автоматизированного сбора информации о турецкой недвижимости

Исполнитель: Студент группы ПриИ-467 – Кузьмин Д. Г.

Научный руководитель: Терехов Г. В., старший преподаватель

Волгоград, 2025

Актуальность баз данных с недвижимостью

База данных(БД) – хранилище данных, содержащее информацию в структурированном виде по определенной предметной области.

- Рынок недвижимости растет. **Турция** входит в ТОП-10, но соответствующей БД нет.
- Базы данных по недвижимости необходимы для пользователей, для бизнеса и для науки.
- Составление БД о турецкой недвижимости является актуальной задачей.

Проблема поиска готовой базы знаний

Существует сервис «Зарубежная недвижимость», которому нужна база данных по турецкой недвижимости, но в открытом доступе такой нет. В этой системе также есть админ-панель для просмотра парсеров, которая требует доработок.

Ручной сбор информации требует значительных временных затрат и подвержен ошибкам и опечаткам из-за влияния человеческого фактора. Поэтому процесс сбора данных рекомендуется автоматизировать, чтобы повысить эффективность и точность.

Далеко не все сайты предоставляют доступ к своему API. Веб-сайт www.hepsiemlak.com является популярным в Турции, но открытого API не имеет.

Автоматизированный сбор информации (Парсинг) с HTML наиболее подходящий способ, за счет своей простоты, отсутствия ограничений по доступу и надежному относительно ручного метода сбора.

Цель и задачи работы

Цель: совершенствование системы для автоматизированного сбора информации о турецкой недвижимости, а также сокращение времени на сбор данных о турецкой недвижимости.

Задачи:

- анализ актуальности баз данных с зарубежной недвижимостью, обзор парсеров веб-сайтов с зарубежной недвижимостью;
- доработка системы автоматизированного сбора информации о турецкой недвижимости;
- доработка админ-панели для отображения информации о парсерах;
- тестирование разработанного системы, проверка эффективности работы.

Объект и предмет исследования

Объектом исследования в работе выступают методы автоматизированного сбора информации для формирования базы данных о турецкой недвижимости. Процесс сбора и обработки данных о недвижимости на рынке Турции, включая автоматизацию операций по извлечению, хранению и анализу информации. Удобство админ-панели для просмотра информации о парсерах.

Предметом исследования является база данных о турецкой недвижимости, админ-панель для просмотра информации о парсерах.

Существующее решение и процесс решения задачи

Получить информацию о турецкой недвижимости можно разными способами. Через буклеты, брошюры или каталоги, видеоролики или веб-сайты.

Наполнить БД информацией о веб-сайте можно разными способами, например, ручной сбор информации, API, веб-скрейпинг, готовые БД, парсинг.

Наилучшим является парсинг, который состоит из следующих этапов: получение HTML-страницы, предобработка, формирование необходимой информации, передача данных на выход (сохранение в БД).

Анализ аналогов систем сбора информации

	AviParser	Octoparse	«Зарубежная недвижимость»
Пагинация в админ-панели	-	-	-
Парсинг сайта hepsiemlak.com	-	-	-
Гибкость в настройке	-	-	+
Сбор данных об объектах	+	+	+
GUI с информацией о парсерах	+	+	+
Сохранение изображений и их сопоставление к БД	-	-	+

Наиболее подходящим является «Зарубежная система»

Предлагаемое решение

Предлагается совершенствование веб-сервиса для парсинга турецкой недвижимости. Решение состоит из двух модулей:

- модуль парсинга.
- модуль админ-панели.

Добавление пагинации на админ-панель, создание парсера для сайта hepsiemlak.com.

Практическая значимость

Результаты исследования будут применяться для автоматизации сбора данных о турецкой недвижимости, что значительно упростит анализ рынка и принятие решений. Также будет улучшена админ-панель, для более эффективного управления парсерами.

Пользователи смогут оперативно находить актуальную информацию об объектах недвижимости, экономя время и снижая вероятность ошибок.

Бизнес сможет прогнозировать рыночные тенденции, оптимизируя работу агентств недвижимости.

Наука сможет использовать базы данных для проведения исследований в области урбанизации, экономики а также аналитики.

Спасибо за внимание!

Программный средства используемые в разработке

Используемые в «Зарубежной недвижимости»:

- Ruby 2.7.5;
- Ruby on rails 6.1.4.4;
- PostgreSQL 14.1;
- HTML 5; CSS 3;
- JavaScript ES6;
- Mechanize 2.8.4;
- Bootstrap 3.3.7.
- Nokogiri 1.13.1