

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Волгоградский государственный технический университет»

Факультет Электроники и вычислительной техники

Кафедра Программное обеспечение автоматизированных систем

Согласовано

Утверждаю
Зав. кафедрой

(должность гл. специалиста предприятия)

О. А. Сычев

(подпись)

(инициалы, фамилия)

(подпись)

(инициалы, фамилия)

« ____ » _____ 20 ____ г.

« ____ » _____ 20 ____ г.

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к выпускной квалификационной работе бакалавра на тему
(наименование вида работы)

Совершенствование веб-сервиса для автоматизированного сбора информации
о турецкой недвижимости

Автор _____ Кузьмин Данила Григорьевич
(подпись и дата подписания) (фамилия, имя, отчество)

Обозначение ВКРБ-09.03.04-10.19-16-23
(код документа)

Группа ПрИн-467
(шифр группы)

Направление 09.03.04 – Программная инженерия,
Разработка программно-информационных систем
(код и наименование направления, наименование программы (профиля))

Руководитель работы _____ Терехов Г.В.
(подпись и дата подписания) (инициалы и фамилия)

Консультанты по разделам:

(краткое наименование раздела) _____
(подпись и дата подписания) _____
(инициалы и фамилия)

(краткое наименование раздела) _____
(подпись и дата подписания) _____
(инициалы и фамилия)

Нормоконтролер: _____ Кузнецова А.С.
(подпись и дата подписания) (инициалы и фамилия)

Волгоград 2025 г.

Кафедра Программное обеспечение автоматизированных систем

Зав. кафедрой

(подпись)

(инициалы, фамилия)

« » 20 Г.

на выпускную квалификационную работу бакалавра

(наименование вида работы)

Студент

Кузьмин Данила Григорьевич

(фамилия, имя, отчество)

Код кафедры

10.19

Группа

ПрИН-467

Тема Совершенствование веб-сервиса для автоматизированного сбора информации о турецкой недвижимости

Утверждена приказом по университету « 06 » сентября 20 22 г. № 1235-ст

Срок представления готовой работы (проекта)

(дата, подпись студента)

Исходные данные для выполнения работы (проекта)

Задание, выданное научным руководителем кафедры «ПОАС»

Содержание основной части пояснительной записки

[illegible]

Перечень графического материала

- 1) _____

- 2) _____

- 3) _____

- 4) _____

- 5) _____

- 6) _____

- 7) _____

- 8) _____

- 9) _____

- 10) _____

- 11) _____

- 12) _____

Руководитель работы (проекта)	_____	Терехов Г.В.
	(подпись и дата подписания)	(инициалы и фамилия)

Консультанты по разделам:

_____	_____	_____
(краткое наименование раздела)	(подпись и дата подписания)	(инициалы и фамилия)
_____	_____	_____

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Волгоградский государственный технический университет»

Кафедра «Программное обеспечение автоматизированных систем»

УТВЕРЖДАЮ:

Зав. кафедрой ПОАС

_____ О. А. Сычев

«___» _____ 20 ____ г.

Разработка мобильного приложения для заказа и доставки еды

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

ВКРБ–09.03.04–10.19–16–23–81

Листов 65

Руководитель работы

_____ Терехов Г.В.

«___» _____ 20 ____ г.

Нормоконтролер

Исполнитель

студент группы ПриИ-467

_____ Кузнецова А.С.

_____ Кузьмин Данила

«___» _____ 20 ____ г.

Григорьевич «___» _____ 20 ____ г.

Волгоград, 2025 г.

Аннотация

Настоящий документ является пояснительной запиской к выпускной квалификационной работе бакалавра на тему: «Совершенствование веб-сервиса для автоматизированного сбора информации о турецкой недвижимости».

В работе обосновывается актуальность выбранной темы, проводится анализ проблемы поиска и сбора информации.

Документ включает в себя страниц -..., рисунков -, приложений -...

Ключевые слова:

Содержание

Введение	8
1 Анализ проблемы автоматизированного сбора информации об объектах турецкой недвижимости	11
1.1 Базы данных в работе аналитиков	11
1.2 Актуальные направления зарубежной недвижимости в России	12
1.3 Актуальность турецкой недвижимости в России	13
1.4 Источники данных о турецкой недвижимости	16
1.5 Веб-сайты с турецкой недвижимостью	17
1.6 Способы формирования баз данных	19
1.7 Автоматизированный сбор информации (Парсинг)	21
1.8 Требования к системе автоматизированного сбора информации	22
1.9 Обзор систем для автоматизированного сбора информации с веб-сайтов	24
1.10 Требования к логированию информации о процессе парсинга	31
1.11 Выбор и обоснование методов	31
Выводы	34
3 Реализация виртуального тура	34
3.1 Требования к функциональным характеристикам	34
Выводы	34
4 Тестирование виртуального тура	34
4.1 Mind Map карта областей тестирования	34
Выводы	34
Заключение	34
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	36
Приложение А	37
Справка о результатах проверки выпускной квалификационной работы на наличие заимствований	37
Приложение Б	38

Техническое задание	38
Приложение В	39
Руководство системного программиста	39

Введение

С каждым годом объем информации, создаваемой и накапливаемой человечеством, стремительно растет благодаря непрерывному технологическому прогрессу. В прошлом возможности людей были значительно ограничены из-за отсутствия современных, передовых технологий и недостатка накопленных знаний. Способы хранения и передачи информации в основном сводились к устной или письменной форме, передаваемой из поколения в поколение. Эти методы хранения информации могли быть ненадежны: они могли быть искажены или вовсе забыты со временем. Более того, самые ценные и важные сведения старались сохранить в узком кругу доверенных лиц, и они могли легко утратиться вместе с носителями этих знаний.

Благодаря вкладу человечества в развитие технологий, появилась возможность обеспечивать сохранность информации в более надежной форме, чем бумажной или устной. Информация стала храниться более массово, и ее распространение стало проще и быстрее. Компьютеры и интернет открыли новые возможности для накопления, передачи и хранения больших объемов данных. Сегодня каждый человек, может беспрепятственно сохранять информацию, практически в неограниченных объемах. Оцифровка данных позволила их структурировать, обрабатывать и в дальнейшем анализировать, анализ которой может быть пригоден для извлечения ценных выводов для научных исследований, бизнеса, медицины или других сфер.

Одним из видов хранения информации выступают базы данных, предназначенные для обработки и хранения информации практически любых объемов в различных предметных областях, таких как машинное обучение, бизнес-аналитика, медицина и другие. Одной из востребованных областей применения баз данных, является анализ постоянно расширяющегося рынка недвижимости. В настоящее время все больше людей проявляют интерес к

приобретению зарубежной недвижимости, что стимулирует рост предложений на рынке. В данной сфере базы данных позволяют обрабатывать информацию о недвижимости, включая данные о цене, агентстве, типу здания, жилой площади и другие. Анализ этих баз данных помогает агентствам, застройщикам и аналитическим компаниям делать прогнозы, а также отслеживать тенденции и реагировать на различные изменения.

Одним из распространенных способов заполнить базу данных является автоматизированный сбор данных(парсинг). Этот процесс может проходить без прямого участия человека. Парсинг значительно облегчает получение данных, особенно если речь идет о постоянно обновляющейся информации, что обеспечивает ее актуальность, также преимуществом автоматизации является высокая скорость сбора данных.

Целью данной бакалаврской работы является совершенствование веб-сервиса для автоматизированного сбора данных о зарубежной недвижимости.

Для достижения поставленной цели были определены следующие задачи:

- анализ актуальности баз данных с зарубежной недвижимостью, обзор парсеров веб-сайтов с зарубежной недвижимостью и библиотек для парсинга html на языке ruby;
- доработка системы автоматизированного сбора информации о зарубежной недвижимости;
- доработка админ-панели для отображения информации о парсерах;
- разработка системы автоматизированного сбора информации о зарубежной недвижимости;
- тестирование разработанного системы, проверка эффективности работы;

Объектом исследования в работе является методы автоматизированного сбора информации для формирования базы данных о зарубежной недвижимости.

Предметом исследования является база данных о зарубежной недвижимости.

Для решения поставленных задач были использованы методы математического моделирования, системного анализа, проектирования баз данных, программной инженерии, объектно-ориентированного программирования, технологии проектирования человеко-машинного взаимодействия.

Практическая ценность работы заключается в том, что разработанная система позволит упростить и ускорить сбор информации об объектах недвижимости из объявлений о продаже недвижимости в зарубежных странах, а также облегчит управление парсерами через админ-панель, предоставляющую функционал для мониторинга их состояния, логов, а также запуска или прерывания выполнения работы.

1 Анализ проблемы автоматизированного сбора информации об объектах турецкой недвижимости

1.1 Базы данных в работе аналитиков

С развитием человечества данные стали играть ключевую роль, и превратились в ресурс, сопоставимый с нефтью или золотом, обработка которых стала одной из главных задач 21 века. Информационные системы, другими словами базы данных, стали ключевым аспектом в упрощении процедуры обработки информации. Базы данных могут хранить в себе не только информацию о какой-либо сфере, но и алгоритмы, процедуры и определения. Базы данных – универсальны, они могут составляться в зависимости от конкретной предметной области, например, научной, медицинской, производственной или любой другой, и применяться для любого рода задач, аналитика, хранение данных, обучение моделей искусственного интеллекта, моделирование графов и другие.

Так, базы данных находят широкое применение. Люди, не имеющие технического образования, или специально не изучавшие базы данных, могли использовать их в той или иной мере, например, при поиске информации на веб-сайтах агрегаторах или онлайн магазинах.

Ученные, университеты и разного рода исследователи хранят важные результаты экспериментов в базах данных, чтобы в дальнейшем иметь возможность обращаться к ним или редактировать.

Государственные структуры, такие как медицинские учреждения, налоговые службы, правоохранительные органы также используют базы данных о гражданах страны.

Также базы данных могут быть использованы как инструмент для сбора и хранения тестовых данных для обучения моделей искусственного интеллекта.

Аналитики также являются пользователями баз данных, так как они удобны для сбора больших объемов данных, которые в последующем используются для построения графиков, прогнозирования и анализа, применяя такие методы анализа как: регрессионный анализ, кластерный анализ, классификация данных и другие.

1.2 Актуальные направления зарубежной недвижимости в России

Спрос на зарубежную недвижимость обусловлен разными факторами, которые отражают разнообразные цели и потребности покупателей.

Недвижимость за рубежом часто рассматривается как способ диверсификации активов, из-за нестабильности локального рынка, и благодаря стабильному росту цен в других странах, приобретение недвижимости расценивается как долгосрочная инвестиция. И со временем такая инвестиция может значительно подорожать и принести существенную прибыль.

Зарубежную недвижимость также рассматривают как источник пассивного дохода, ведь ее можно сдавать в аренду. Туристически привлекательные страны, такие как Испания, Турция, ОАЭ или Таиланд, обеспечивают стабильный рост туристического рынка и спрос на аренду жилья.

Кто-то рассматривает покупку недвижимости, как способ защиты от инфляции, в условиях непредсказуемости рынка или экономической нестабильности.

Также есть люди, которые стремятся сменить место жительства, и для этого им необходимо получения ВНЖ или гражданства. В ряде стран получить вид на место жительства или гражданство можно через вклад в экономику страны, т.е. покупку недвижимости.

1.3 Актуальность турецкой недвижимости в России

Исходя из исследования международного брокера недвижимости Tranió для отчета РБК, Турция продолжает находиться в списке популярных стран для покупки недвижимости среди Россиян

№	Направление	Доля от всех заявок	Динамика спроса, год к году
1	Таиланд	14,2%	+64,2%
2	Грузия	13,5%	+5,1%
3	ОАЭ	7,6%	-22,3%
4	Греция	6,9%	-14,6%
5	Франция	6,6%	-5,9%
6	Турция	6,1%	-24,1%
7	Испания	5,8%	-25,3%
8	Кипр	5,5%	+17,1%
9	Индонезия	4,3%	+6,3%
10	Чехия	4,2%	+72,6%
	Всего	100%	-8,2%

Источник: данные Tranió за январь – сентябрь 2024 года

Рисунок 1 – Статистика популярных стран для покупки недвижимости среди россиян

Это явление обусловлено рядом факторов. Одним из этих факторов, является предоставление иностранцам гражданства за вклад в экономику страны. Этим вкладом могут быть инвестирование в недвижимость, покупка акций инвестиционных фондов или государственных облигаций. Спустя 3 года приобретенную недвижимость можно продать, при этом гражданство остается у инвестора.

Также одним из ключевых факторов, делающих Турцию привлекательной страной для россиян, является её географическая близость к России, что делает ее привлекательной для людей склонных к частым поездкам за границу. Климат и природа также оказывают влияние на выбор страны для покупки недвижимости. Россияне чаще предпочитают теплые страны с мягкой погодой и большим количеством солнечных дней. В то время как большая часть России находится в холодных северных широтах, что делает ее холодной страной, Турция же предлагает более комфортный и благоприятный климат для жизни и отдыха.

Нельзя не отметить, что Турция является развитой страной. Турция является страной с высоким уровнем урбанизации, современные жилые комплексы, спортивные площадки и охраной, а также развитая городская инфраструктура делают жизнь в Турции благоприятной. Турция также продолжает активно инвестировать в развитие транспортной, энергетической и социальной инфраструктуры.

Турция считается туристической страной, поэтому недвижимость может доставлять высокие доходы со сдачи её в аренду. Например, по статистике ЕМИСС можно наблюдать устойчивый рост числа российских граждан, выезжающих в Турцию с целью туризма.

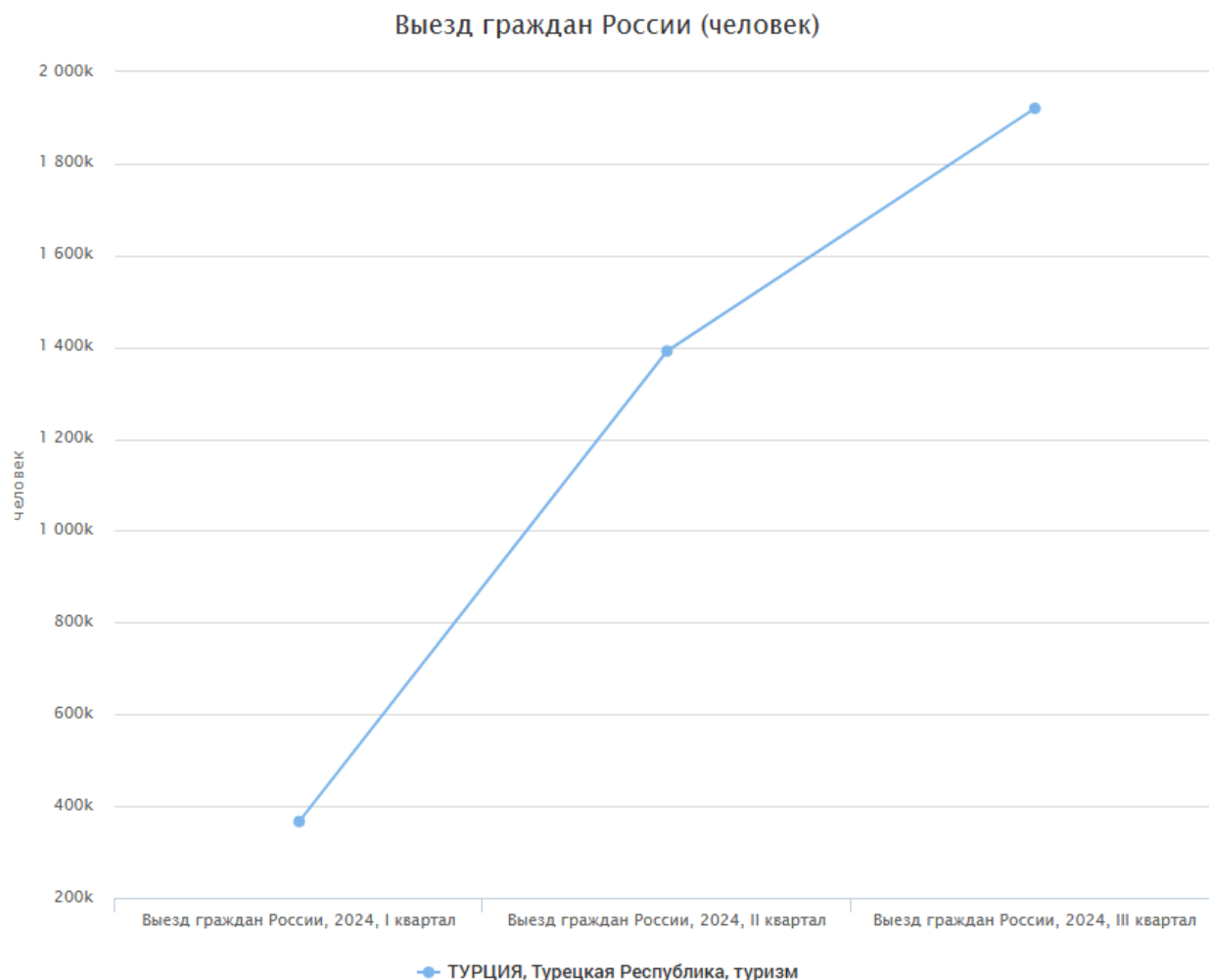


Рисунок 2 – Статистика выезда граждан России в Турцию с целью туризма за первые 3 квартала 2024 года

Однако сферы недвижимости и туризма непостоянны и подвержены изменениям под воздействием политических, экономических или социальных процессов. Необходимость в своевременном реагировании на эти изменения и принятии обоснованных решений порождает задачи регулярного отслеживания и актуализации данных. Однако ручной мониторинг новостных порталов и других источников требует значительных усилий, куда проще было бы это делать с помощью автоматически пополняемых, с какой-либо периодичностью, баз данных.

1.4 Источники данных о турецкой недвижимости

Для того чтобы наполнить необходимую базу данных, нужно обращаться к различным источникам информации. Существует целый ряд источников информации, как специализированных, так и агрегаторы информации, например:

- Калатогии, брошюры и буклеты остаются традиционным инструментом продвижения на рынке недвижимости, где всегда есть ведущие компании застройщиков, или риэлторские агентства. Один из вариантов заявить о себе, тем самым заинтересовав покупателя, это брошюры, буклеты или каталоги, цели которых продвижение продукта или услуги, или компании в целом. Тогда можно собирать информацию о недвижимости с помощью этих самых брошюр, буклетов или каталогов, однако у них есть существенные недостатки, например, поддержание актуальности, неполные данные или трудности с анализом;
- Видеоролики в интернете. Видео хостинговые платформы такие как Youtube, Rutube или Дзен пользуются особой популярностью, так как на них агентства способны демонстрировать, а блогеры обозревать квартиры, дома и инфраструктуру. Однако использование видеоконтента для сбора информации и создания базы данных также имеет ряд неудобств, например, не структурированность информации, субъективность контента и также временные затраты на просмотр видео для поиска конкретных данных;
- Веб-сайты являются одним из наиболее удобных и эффективных источников информации для сбора данных в том числе о рынке недвижимости. В настоящее время большая часть компаний имеет свой сайт, на котором они размещают информацию о недвижимости. Веб-сайты доступны из любой точки мира, что позволяет собирать

информацию без необходимости физического присутствия. Данные на сайтах обновляются регулярно и без каких-либо трудностей, что позволяет поддерживать актуальность информации. И также структурированность данных на веб-сайте позволяет без трудностей собирать данные о недвижимости.

1.5 Веб-сайты с турецкой недвижимостью

С помощью веб-сервиса SimilarWeb, который является аналитическим инструментом для изучения веб-трафика, и который помогает определить лидеров в конкретной отрасли, были выбраны 3 наиболее популярных и подходящих веб-сайта с турецкой недвижимостью. SimilarWeb позволяет проанализировать посещаемость веб-сайтов, время, проведенное на сайте и географическое распределение аудитории, а также позволяет сравнить несколько веб-сайтов между собой. Отбор веб-сайтов производился по следующим критериям:

- общее количество посещений в месяц на момент обращения;
- глобальный рейтинг веб-сайта;
- региональный рейтинг веб-сайта для региона Турция;
- отраслевой рейтинг.

Исходя из этих критериев были выбраны следующие 3 веб-сайта: hepsiemlak.com, zingat.com, remax.com.tr.

Статистика по каждому представлена на рисунках 3, 4 и 5. Также на рисунке 6 показан график сравнения веб-сайтов по количеству посещений в месяц.

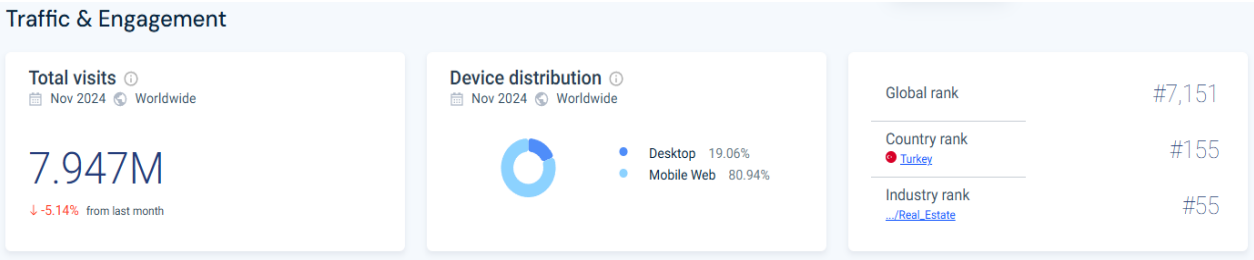


Рисунок 3 – Статистика для сайта hepsiemlak.com

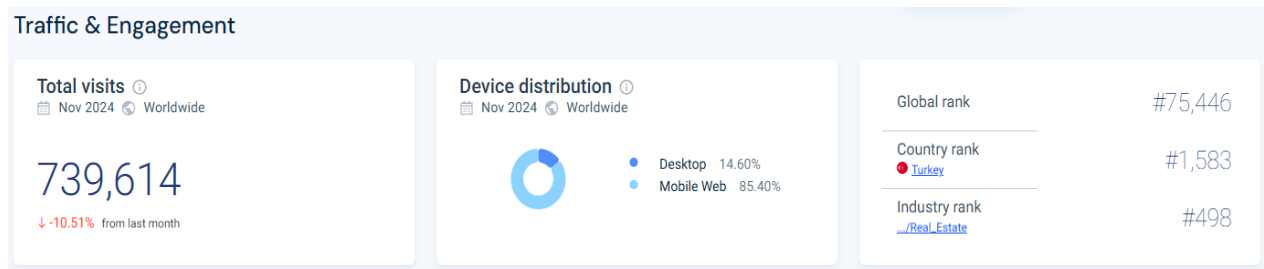


Рисунок 4 – Статистика для сайта zingat.com

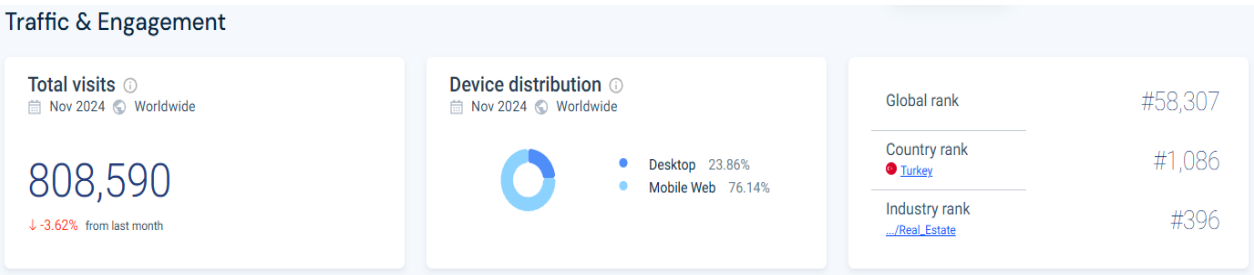


Рисунок 5 – Статистика для сайта remax.com.tr



Рисунок 6 – График сравнения посещаемости сайтов.

Исходя из данных представленных на рисунках, можно сделать вывод, что наиболее популярным и подходящим веб-сайтом для оценки и сбора информации о недвижимости, является hepsiemlak.com.

1.6 Способы формирования баз данных

Чтобы база данных была полезной и ее можно было анализировать или составлять заключения, её необходимо правильно наполнять информацией. Существуют следующие способы это сделать:

- Ручной сбор информации, с помощью которого производится поиск, а затем изучение материала. Этот процесс занимает очень много времени, так как необходимо обработать гигантские объемы информации, и не факт, что не будут допущены ошибки или упущены важные детали.
- Использование API, некоторые сайты и сервисы предоставляют API для получения интересующей информации с помощью специальных запросов в структурированном виде, например, JSON или XML.
- Веб-скрейпинг (Web-scraping) является методом автоматизированного сбора информации с веб-сайта путем извлечения HTML-страницы, требует реализации программного обеспечения для конкретной задачи.
- Использование готовых источников данных. Некоторые платформы и сервисы предоставляют данные в удобных форматах, например, CSV, Excel или JSON, которые можно загрузить напрямую в базу данных.
- Автоматизированный сбор информации (Парсинг) с HTML-страницы. Этот метод похож на веб-скрейпинг, однако он не

только собирает данные о HTML-странице, но и способен ее обработать и структурировать, что позволяет в дальнейшем работать с полученными данными.

Таблица №1 – Преимущества и недостатки различных способов наполнения базы данных

Метод сбора информации	Преимущества	Недостатки
Ручной сбор информации	Не требует специальных навыков, полный контроль процесса, предпочтителен для данных небольшого объема	Риск ошибок, временные затраты, низкая масштабируемость
Использование API	Структурированные данные, актуальность	Ограниченный доступ, лимиты запросов, уязвимость к изменениям
Веб-скрейпинг	Автоматизация сбора данных, получение данных с любого источника	Юридические ограничения, изменения на веб-сайтах, блокировки доступа, техническая сложность
Использование готовых	Экономия времени, высокое качество данных, удобный доступ, актуальность	Ограниченный контроль, ограниченный

источников данных		доступ, юридические ограничения, сложность интеграции
Парсинг	Доступность данных, возможность настроить под конкретные потребности, автоматизация, масштабируемость	Техническая сложность, сайты с защитой от парсинга, зависимость от структуры сайта

1.7 Автоматизированный сбор информации (Парсинг)

Парсинг является одним из наиболее популярных и универсальных способов сбора и обработки информации с любых текстовых данных, включая документы, файлы или веб-сайты. Он позволяет эффективно извлекать и структурировать данные. Парсинг также позволяет автоматизировать процесс получения и обработки данных, для дальнейшего анализа, поэтому парсинг применим для решения широкого спектра задач.

Процесс парсинга сводится к следующим этапам:

- получение исходного кода HTML-страницы веб-сайта, с помощью Get-запроса;
- предобработка веб-страницы, с помощью разных методов. Убрать все лишнее, отделить текст от разметки, найти данные на странице, удовлетворяющие цели парсинга;
- формирование необходимой информации страницы;
- передать данные на выход, для дальнейших манипуляций.

Для обработки данных и поиска нужной информации на веб-странице используется XPath. XPath представляет собой язык запросов, который особенно хорошо себя показывает при обработки HTML-страниц. Суть работы данного инструмента заключается в навигации по дереву элементов документа и выборе узлов, соответствующих определенным критериям. Узлом может быть атрибут, элемент, текст и т.д.

Также для обработки текста при парсинге может быть использован Regex, представляющий собой инструмент для обработки текста на основе шаблонов. Регулярные выражения могут быть использованы для поиска, удаления или замены текста на веб-странице.

Поэтому парсинг является наиболее удобным способом для сбора информации о зарубежной недвижимости, обеспечивая точность и полноту данных о объектах недвижимости. Процесс парсинга автоматизирован, что исключает необходимость вручную просматривать и сортировать большое количество информации, значительно экономя время и усилия, а также минимизируя риск ошибок.

1.8 Требования к системе автоматизированного сбора информации

Поскольку объем данных о недвижимости постоянно растет, производить поиск среди них является узким местом производительности и эффективности. Соответственно парсер должен уметь сохранять собранные данные в БД, чтобы была возможность с помощью языка SQL производить операции поиска, фильтрации и анализа.

Поскольку для каждого объекта недвижимости есть свои изображения, и количество их в разы больше чем самих объектов, есть необходимость также сохранять информацию о изображениях в БД, а сами изображения непосредственно на жесткий диск. Необходимо чтобы система могла

ассоциировать изображения на жестком диске с записями в БД, по средству языка SQL.

На веб-странице объявлений присутствуют 2 вида информации: текстовая, содержащая описания и числовые данные, и так называемые теги, логические флаги наличия какой-либо характеристики объекта, например, парковки. Для эффективной обработки и хранения этих данных, парсер должен уметь различать эти 2 вида информации, и сохранять их в отдельные таблицы в БД.

Парсер должен быть настроен на двухэтапный процесс сбора данных. На веб-странице каталога объявлений, парсер должен уметь собирать информацию о ссылках объектов недвижимости, а затем по этим ссылкам собирать информацию непосредственно об объектах недвижимости.

Цены на недвижимость это - это информация с динамически изменяющимися показателями, которые подвержены колебаниям со временем из-за различных факторов. В связи с этим парсер должен иметь функциональность для настройки периодического запуска с целью обновления существующей информацией.

Анализом данных о недвижимости обычно занимаются специалисты, чьи компетенции лежат в областях маркетинга или аналитики, а не программирования. Поэтому для обеспечения эффективности аналитиков необходимо чтобы система предоставляла удобный графический интерфейс, с помощью которого можно запускать парсеры, а также отслеживать их состояние.

Поскольку количество парсеров в системе может быть велико, для обеспечения стабильной работы системы на странице графического интерфейса пользователей по парсерам должна быть реализована пагинация. Она позволит отображать заданное количество парсеров постранично, уменьшая объем одновременно загружаемой и показываемой информации.

Система должна предоставлять возможность логирования информации о работе парсера.

1.9 Обзор систем для автоматизированного сбора информации с веб-сайтов

1.9.1 Aviparser

Эта система предназначена для сбора информации об объявлениях товаров и услуг, а также объектах недвижимости на веб-сайте avito.ru. Данный парсер способен собирать ключевую информацию объявлений с любого раздела веб-сайта, включая дополнительные фильтры по поиску и ключевые слова.

Данный парсер умеет собирать информацию о телефоне объявления, заголовке, описании, цене, географическому расположению, категории, ссылке на объявление, дате добавления, кол-во просмотров, ссылки на изображения и др.

Парсер предоставляет графический интерфейс с настройками и логами по статусу и дате парсинга, внешний вид системы представлен на рисунках 7, 8 и 9.

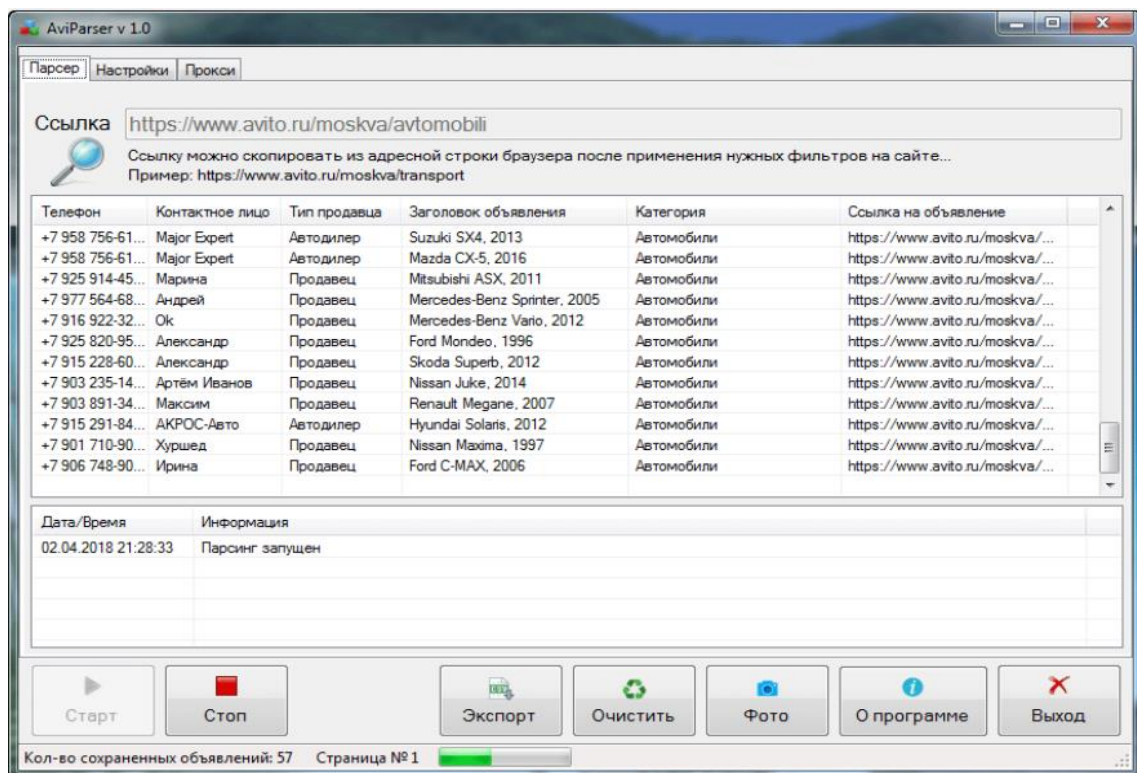


Рисунок 7 – Графический интерфейс системы AviParser

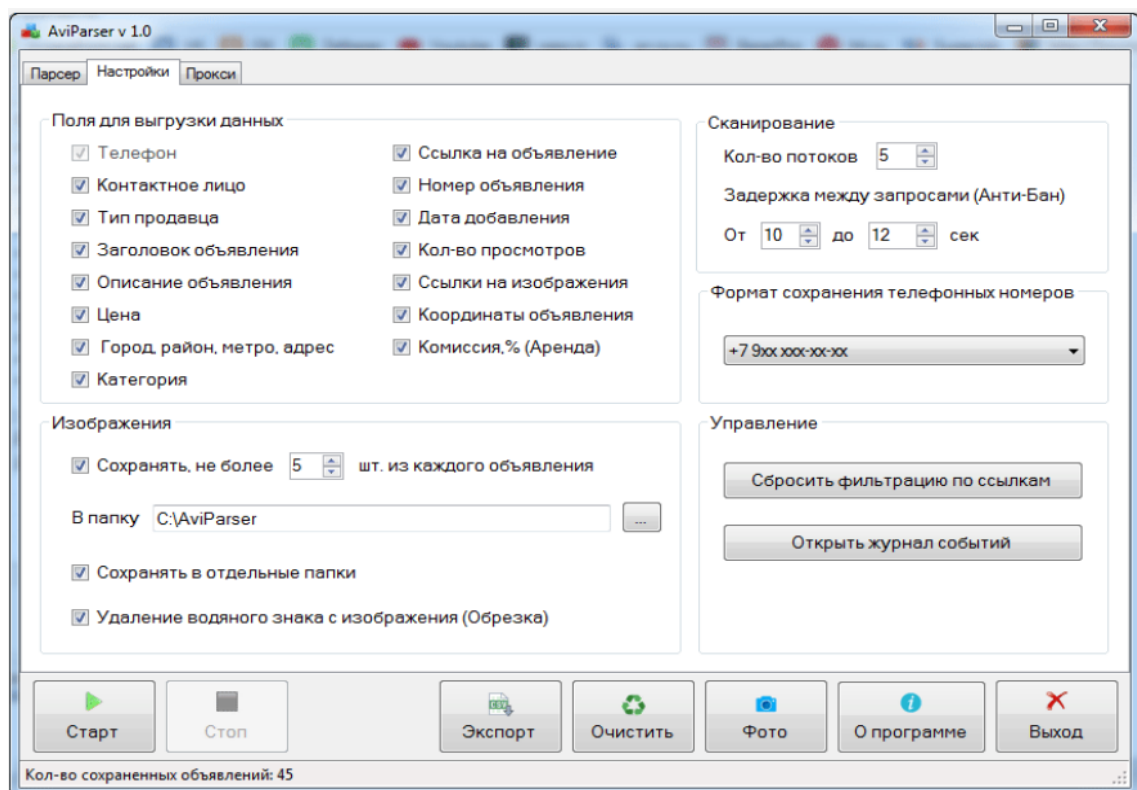


Рисунок 8 – Настройки парсера AviParser

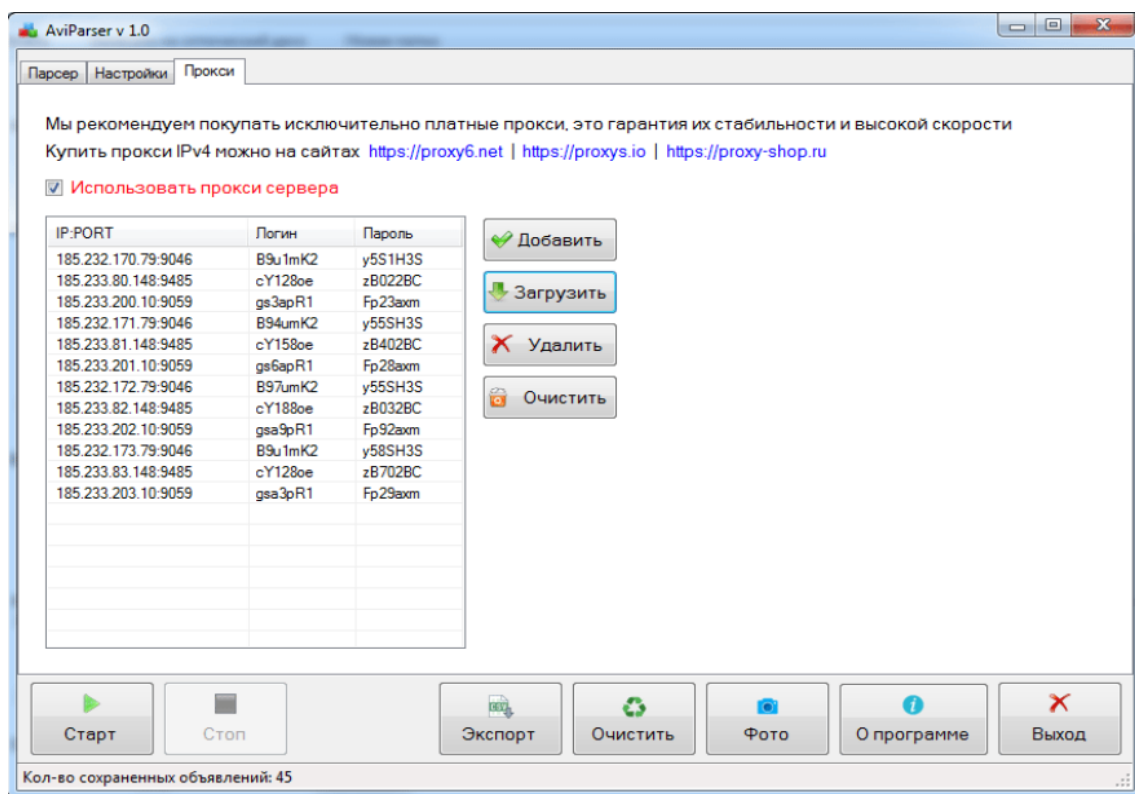


Рисунок 9 – Настройки прокси для системы AviParser

Как видно на рисунке 8, данная система предусмотрен экспорт собранной информации. Данные сохраняются в файл формата CSV. Что является не удобным для работы аналитиков.

Также парсер может сохранять изображения объявлений, но так как он сохраняет данные в файл CSV, и сохраняет ссылки на изображения, поиск и анализ самих изображений может стать очень трудоемкой и неудобной работой.

Для запуска парсера необходимо предоставить конкретную ссылку для парсинга страницы.

Сама система является платной с закрытым кодом.

В системе предусмотрен парсинг только 1 сайта – avito.ru. Нет парсера турецкой недвижимости.

1.9.2 Octoparse

Эта популярная универсальная система для автоматизированного сбора информации с различных веб-сайтов, без необходимости программирования. Она ориентирована на пользователей с различным уровнем технических знаний и предлагает интуитивно понятный интерфейс для сбора информации.

Данная система предлагает ряд шаблонов для парсинга веб-сайтов. Эти шаблоны могут быть настроены по ряду критериев. Например, шаблон для парсинга веб-сайта с недвижимостью Zillow.com представлен на рисунке 10.

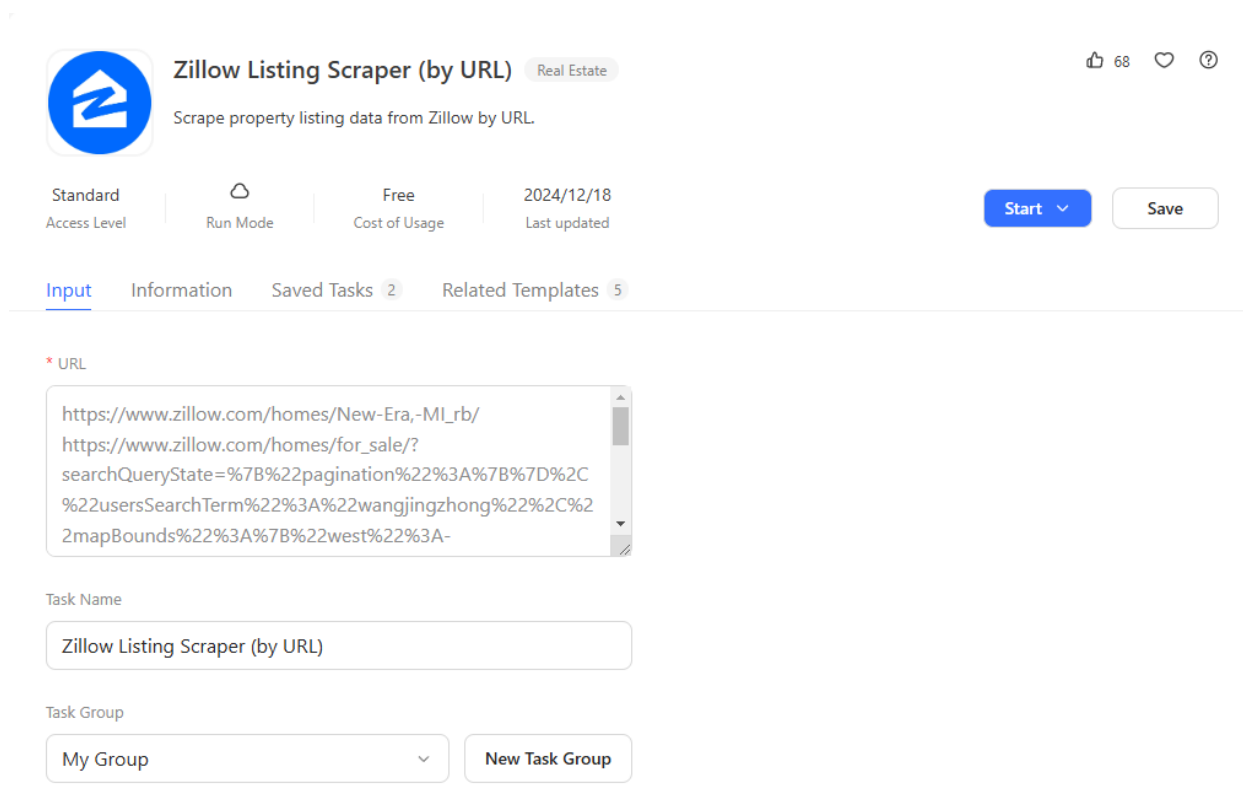


Рисунок 10 – Интерфейс шаблона для парсинга системы Octoparse

Также система предоставляет возможность сохранить результат парсинга в различном формате, а также экспортировать данные в БД. Пример изображен на рисунке 11.

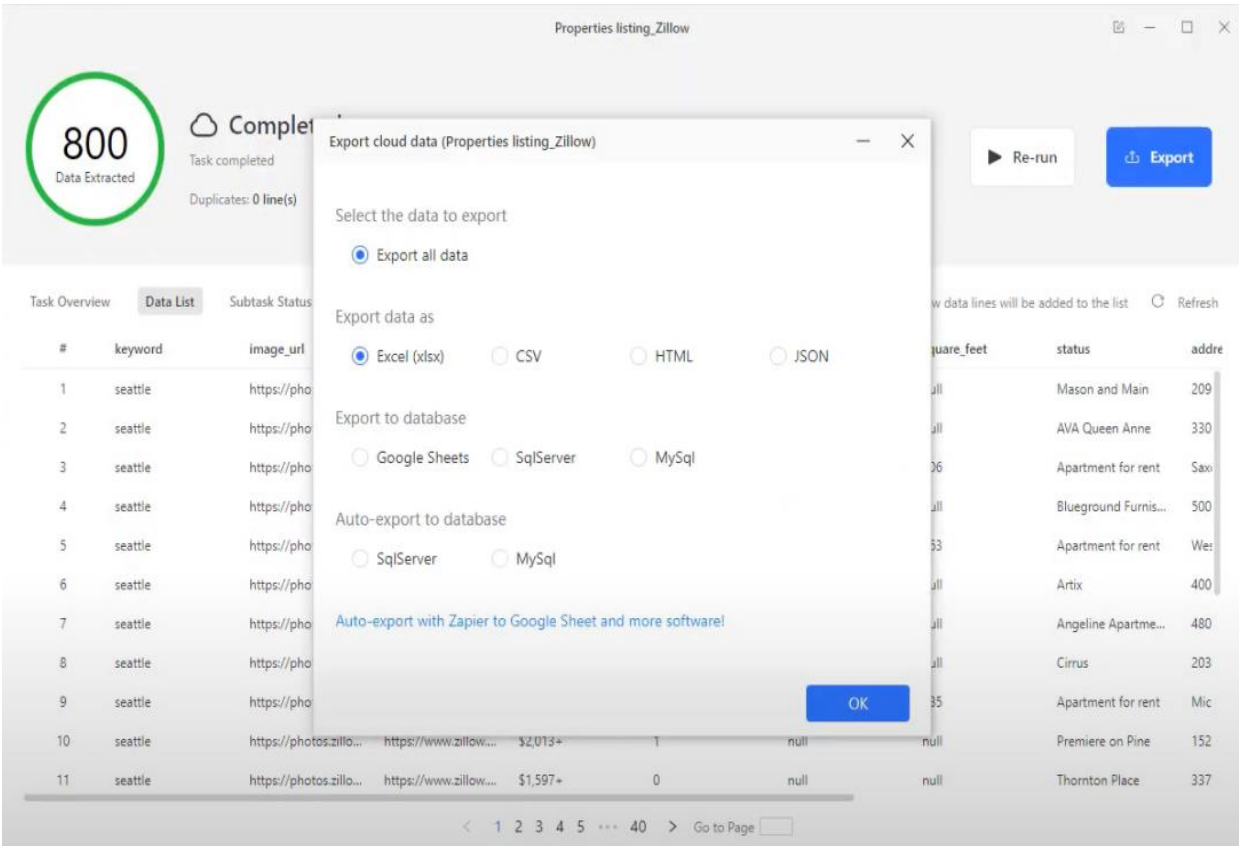


Рисунок 11 – Выгрузка собранной информации в системе Octorparse

Система Octorparse предоставляет возможность просмотра задач для парсинга, в которой описано количество собранных данных, запланированное выполнение, потраченное время, статус, и сами данные, находящиеся в облаке. Пример представлен на рисунке 12.

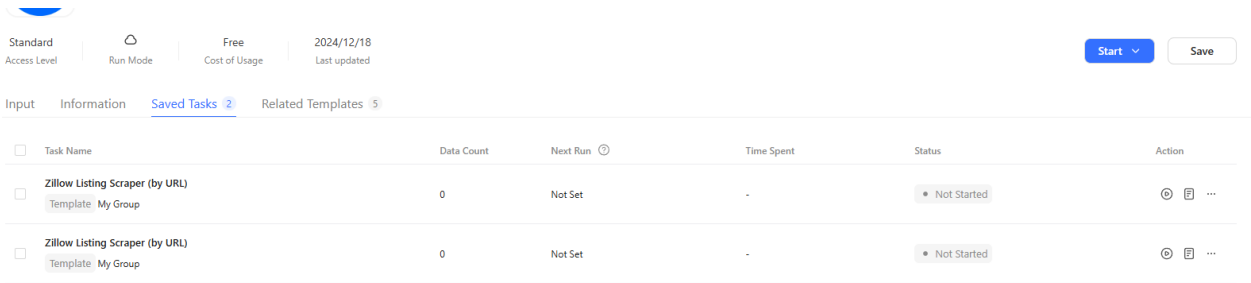


Рисунок 12 – Графический интерфейс пользователя для просмотра информации о парсинге система Octorparse

Данный система не сохраняет изображения на жесткий диск, а сохраняют их только как ссылку в таблице. Также данная система имеет ограниченную гибкость в управлении, и у нее отсутствует тонкая настройка под специфические требования.

Для запуска системе нужно передать список ссылок для парсинга, что является не удобным, потому что присутствует ручной ввод.

Данная система предоставляет функционал для парсинга с многих сайтов, однако на ней нет парсинга веб-сайта с турецкой недвижимостью.

Система является платной с закрытым исходным кодом.

1.9.3 Зарубежная недвижимость

Существует система «Зарубежная недвижимость», которая разработана на языке программирования Ruby с использованием библиотеки Nokogiri, которая позволяет легко работать с HTML и XML за счет CSS-селекторов и XPath, также поддерживающая параллельную работу. Система «Зарубежная недвижимость» предназначена для сбора информации о недвижимости из разных стран мира.

Система умеет собирать сначала все ссылки объявлений необходимого сегмента, а затем собирать информацию о конкретных объектах недвижимости по этим ссылкам.

Система собирает большие объемы данных о недвижимости и сохраняет их в БД, чтобы потом иметь к ним доступ посредством SQL.

«Зарубежная недвижимость» также умеет сохранять изображения объектов недвижимости на жестком диске, и умеет ассоциировать файлы изображений с записями в БД, т.е. можно получить изображение в результате SQL запроса.

Система собирает основные данные о недвижимости такие, как заголовок, местоположение, цена, площадь, этаж и описание, а также собирает данные о тегах и сохраняет их в отдельную таблицу в БД.

Система состоит из нескольких парсеров, которые собирают информацию с определенных веб-сайтов стран. Необходимо отслеживание работы каждого парсера, для этого в системе предусмотрена админ-панель для просмотра статуса парсеров и логи по самому парсеру. Пример админ-панели представлен на рисунках 11 и 12.

[Главная](#) / [Парсеры](#)

Парсеры

№	Агентство	Дата начала	Дата завершения	Статус	Логи	
4	real_estate_us	18.12.24 05:12	18.12.24 12:23	успешно	Логи	Начать парсинг
5	real_estate_germany	16.12.24 10:00	16.12.24 12:26	неуспешно	Логи	Начать парсинг

Рисунок 11 – Страница с отображением парсеров системы «Зарубежная недвижимость»

[Главная](#) / [Парсеры](#) / [Логи \(2\)](#)

Логи

№	Дата начала	Дата завершения	Статус	Количество добавленных объектов	Количество обновленных объектов	Количество удаленных объектов	Файл с логами
2	18.12.24 05:12	18.12.24 12:26	успешно	99	15	30	Открыть файл
1	18.12.24 12:23	18.12.24 12:23	неуспешно	0	0	0	Открыть файл

Рисунок 12- Страница для просмотра логов конкретного парсера системы «Зарубежная недвижимость»

На странице парсеров и логов отсутствует пагинация – постраничный вывод информации, из-за этого система может быть перегружена большим количеством информации.

В системе «Зарубежная недвижимость» отсутствует парсер веб-сайта hepsiemlak.com

1.10 Требования к логированию информации о процессе парсинга

Логирование – это процесс сбора информации о работе приложения. Эта информация записывается в реальном времени по мере работы приложения в хронологическом порядке, которая позволяет отслеживать события и состояния системы. Информация в логах дает представление о произошедшем процессе работы системы, что спровоцировало то или иное событие, какие ошибки возникли или результаты запросов.

Выделяется несколько уровней логов:

- unknown – неизвестные сообщения, которые логируются всегда;
- fatal – необработанные исключения, которые приводят программу к аварийному завершению;
- error – обработанное исключение;
- info – полезная информация о работе системы;
- debug – низкоуровневая информация для разработчиков.

Логи могут храниться в различных форматах в зависимости от требований системы и объема данных. Основными способами хранения логов выступают файлы, хранение логов в БД либо может использоваться централизованный сбор логов, куда будут транслироваться логи с различных систем.

1.11 Выбор и обоснование методов

В таблице №2 представлены результаты сравнительного анализа систем автоматизированного сбора информации.

Таблица №2 – Сравнительная таблица аналогов

	AviParse	Ostoparse	«Зарубежная недвижимость»
GUI с информацией по парсерам	+	+	+
Пагинация на странице с информацией про парсеры	-	+	-
Сохранение результатов парсинга в БД	-	+	+
Гибкость настройки парсинга	-	-	+
Доступность исходного кода	-	-	+
Сохранение изображений на жесткий диск	+	-	+
Получение изображений в результате SQL запроса	-	-	+
Логирование работы парсера	-	-	+
Парсинг турецкого веб-сайта hepsiemlak.com	-	-	-
Собирание ссылок и парсинг непосредственно объектов недвижимости с ссылок	-	-	+

На основании проведенного анализа, система «Зарубежная недвижимость» является наиболее подходящей и удовлетворяющей требованиям к системе автоматизированного сбора информации. Поэтому для реализации поставленных задач будет использоваться язык программирования Ruby и инструмент Nokogiri.

2 Предлагаемый процесс распространения информации о вузе среди абитуриентов

2.1 Решение, используемое для устранения критических мест существующего процесса

Выводы

3 Реализация виртуального тура

3.1 Требования к функциональным характеристикам

Выводы

4 Тестирование виртуального тура

4.1 Mind Map карта областей тестирования

Выводы

Заключение

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Морозова, Е. С. Технология создания виртуальных интерактивных туров / Е. С. Морозова, В. В. Лавров // Теплотехника и информатика в образовании, науке и производстве : сборник докладов I Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых (TIM2012) с Международным участием / УрФУ [и др.] ; под ред. Н. А. Спирина.– Екатеринбург, 2012.– С. 245-247.
2. Что такое виртуальный тур? [Электронный ресурс]. —Режим доступа : <https://3dturov.net> (дата обращения 28.04.2022).
3. PostgreSQL. Разработка баз данных : учебник / М. Ф. Ванина, А. Г. Ерохин, Н. В. Тутова [и др.]. — Москва : Русайнс, 2024. — 227 с. — ISBN 978-5-466-06974-7. — URL: <https://book.ru/book/954200>.
4. Базы данных : учебное пособие / Ратушняк Г.Я., Золкин А.Л., Никитин А.Л. / <https://book.ru/book/951769>.
5. Design and implementation of knowledge base for target recognition in remote sensing images / S. Chen, C. Cai, E. Li, M. Ding // 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery – Yantai, China, 2010. – P. 2604-2608.
6. Антипина, В. А. Разработка базы данных «Учебная база данных» для организации, оказывающей образовательные услуги / В. А. Антипина // Современные научные исследования и инновации. – 2023. – № 11(151).
7. Мищенко, Я. В. Базы данных: Современные тенденции в области баз данных и их значимость для различных сфер деятельности / Я. В. Мищенко // Вестник науки. – 2024. – Т. 4, № 7(76). – С. 252-256.
- 8.

Приложение А

Справка о результатах проверки выпускной квалификационной работы
на наличие заимствований

Приложение Б
Техническое задание

Приложение В
Руководство системного программиста