**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc183573339)

[ОСНОВНАЯ ЧАСТЬ 5](#_Toc183573340)

[1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ 5](#_Toc183573341)

[1.1. Анализ отрасли (Что там в сфере строительства и стройматериалов) 5](#_Toc183573342)

[1.2. Анализ имеющихся решений (Какие аналоги уже есть и почему они хуже нашего). 5](#_Toc183573343)

[1.3. Обоснование необходимости разработки и юридической правомерности (почему надо это разработать, не закроют ли за такое). 5](#_Toc183573344)

[2. ПРОЕКТИРОВАНИЕ СИСТЕМЫ 5](#_Toc183573345)

[2.1. Требования к системе(функциональные/нефункциональные). 5](#_Toc183573346)

[2.2. Функциональная модель системы (idef0, диаграмма прецедентов) 5](#_Toc183573347)

[2.3. Архитектура системы (клиент сервер) 5](#_Toc183573348)

[2.4. Моделирование данных (классы, база данных) 5](#_Toc183573349)

[2.5. Моделирование логики работы (диаграмма состояний, компонентов, последовательности) 5](#_Toc183573350)

[3. РАЗРАБОТКА СИСТЕМЫ 5](#_Toc183573351)

[3.1. Этапы разработки (Последовательность выполнения действий для разработки системы). 5](#_Toc183573352)

[3.2. Средства разработки (стек технологий и аппаратное обеспечение). 5](#_Toc183573353)

[3.3. Описание системы 5](#_Toc183573354)

[3.3.1. Интерфейс пользователя (прототипы в Фигме) 5](#_Toc183573355)

[3.3.2. Дизайн (красивое/не красивое) 5](#_Toc183573356)

[3.3.3. Функциональные блоки (Список страниц отличных по наполнению и их описание) 5](#_Toc183573357)

[3.3.4. Серверная часть системы (потоки http запросов от клиента к серверу и обратно) 5](#_Toc183573358)

[3.3.5. Описание контроллеров (конченых точек и их описание) 5](#_Toc183573359)

[3.3.6. Описание моделей (как и где тащим данные из базы или загружаем их) 5](#_Toc183573360)

[3.3. Алгоритмы обработки данных (как мы json/html в текст превращали, где и как это сделать) 5](#_Toc183573361)

[3.3.1. Входные данные (описание получаемых данных html, json,xml) 5](#_Toc183573362)

[3.3.2. Обработка данных (Определяем, что за тип файла, определяем структуру, отсекаем ненужное, тащим нужное) 5](#_Toc183573363)

[3.3.3. Выходные данные (Описание того, что вытащили) 5](#_Toc183573364)

[3.4. Тестирование системы (зачем тестируем, что тестируем) 5](#_Toc183573365)

[3.4.1. Методика тестирования системы (интеграционное, функциональное, модульное) 5](#_Toc183573366)

[3.4.2. Результаты тестов (тестировали то то получили то то) 5](#_Toc183573367)

[ЗАКЛЮЧЕНИЕ 7](#_Toc183573368)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 8](#_Toc183573369)

# ВВЕДЕНИЕ

В современном мире строительная отрасль активно использует цифровые технологии для оптимизации процессов проектирования, закупок и выполнения строительных работ. Информация о строительных материалах, их свойствах, производителях и ценах является ключевым ресурсом для принятия обоснованных решений. Однако проблема заключается в том, что данные об этих материалах зачастую разрознены, хранятся в различных форматах и на множестве платформ.

Создание программных решений для автоматизации сбора и обработки таких данных становится актуальной задачей. Автоматизированный сбор данных позволяет не только сократить временные затраты, но и повысить точность анализа, обеспечивая доступ к актуальной информации о строительных материалах в режиме реального времени.

С развитием онлайн-коммерции строительные материалы становятся все более доступными через интернет-магазины и специализированные платформы. Однако поиск и систематизация данных о них требует значительных усилий, особенно если речь идет о сравнении цен, изучении характеристик и отзывов. Компании и частные клиенты, принимающие решения на основе неполной или устаревшей информации, рискуют столкнуться с перерасходом бюджета или получением некачественного продукта.

Программное обеспечение, разработанное для автоматизированного сбора данных, предоставляет решение этой проблемы. Оно может осуществлять сбор информации с веб-ресурсов, таких как каталоги товаров, маркетплейсы и форумы, обрабатывать полученные данные и предоставлять пользователю структурированные результаты в удобном формате.

Кроме того, такие системы находят применение в мониторинге рынка, анализе конкурентов и оптимизации логистики. Для малых и средних предприятий внедрение подобных решений особенно важно, так как оно позволяет снизить затраты на ручную обработку информации и повышает конкурентоспособность.

Целью данной работы является разработка программного обеспечения для автоматизированного сбора данных о строительных материалах. Программа должна обеспечивать сбор данных (характеристики, цены) с различных веб-ресурсов и предоставлять их в удобной для анализа форме.

Для достижения данной цели необходимо решить следующие задачи:

1. Проанализировать существующие подходы и инструменты для сбора данных с веб-ресурсов, включая веб-скрейпинг и обработку данных.
2. Определить ключевые требования к программному обеспечению, включая целевые источники данных и формат их представления.
3. Разработать архитектуру программы, включающую модули сбора, обработки и хранения данных.
4. Реализовать основные алгоритмы веб-скрейпинга и обработки текстовой информации.
5. Провести тестирование программы на примере сбора данных с реальных веб-ресурсов.
6. Оценить эффективность программы и её практическую применимость.

Разработка программы для сбора данных о строительных материалах позволит автоматизировать трудоемкие процессы мониторинга и анализа информации. Компании смогут получать актуальные данные о товарах и ценах, анализировать характеристики материалов и изучать отзывы клиентов, что приведет к повышению качества принимаемых решений.

Кроме того, созданное программное решение может быть адаптировано для других отраслей, где требуется сбор и обработка данных из различных источников. Это делает проект универсальным инструментом для работы с информацией.

# ОСНОВНАЯ ЧАСТЬ

# АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

# Анализ отрасли

Строительная отрасль играет ключевую роль в экономике, формируя инфраструктуру, жильё, промышленные объекты и общественные сооружения. Это одна из самых динамично развивающихся сфер, которая тесно связана с рядом смежных отраслей, таких как производство строительных материалов, транспорт, логистика и архитектурное проектирование.

Особенности строительной отрасли:

1. Высокий уровень конкуренции среди компаний, предлагающих строительные материалы и услуги.

2. Зависимость от сезонности: пик активности приходится на тёплые месяцы года.

3. Постоянная потребность в оптимизации затрат на материалы, труд и логистику.

4. Широкий спектр используемых строительных материалов, который варьируется от стандартных (бетон, кирпич) до инновационных (композитные материалы, эко-решения).

Строительные материалы являются основой любой стройки, определяя качество и долговечность объектов. Рынок этих материалов характеризуется большим разнообразием и включает продукцию различных производителей: от крупных международных брендов до локальных компаний.

Основные категории строительных материалов:

1. Основные строительные материалы:

1.1. Бетон, цемент, песок, гравий.

1.2. Кирпич (керамический, силикатный).

1.3. Металлоконструкции.

2. Отделочные материалы:

2.1. Штукатурки, шпаклёвки, краски.

2.2. Плитка (керамическая, мраморная), ламинат, паркет.

2.3. Обои, декоративные панели.

3. Изоляционные материалы:

3.1. Теплоизоляция (минеральная вата, пенопласт).

3.2. Звукоизоляция.

3.3. Гидроизоляционные плёнки и мембраны.

4. Инженерные материалы:

4.1. Трубы и фитинги.

4.2. Электрические кабели, системы водоснабжения и вентиляции.

5. Инновационные и экологичные материалы:

5.1. Эко-материалы (солома, переработанный пластик, бамбук).

5.2. Самовосстанавливающийся бетон, «умные» стекла.

Факторы, определяющие выбор строительных материалов:

1. Качество и долговечность.

2. Стоимость. Это особенно важно для крупных проектов с жёстким бюджетом.

3. Эстетика и соответствие современным стандартам.

4. Экологичность и энергоэффективность.

Современный строительный бизнес ориентирован на внедрение новых технологий и материалов, что позволяет сокращать затраты и повышать производительность.

Основные тенденции:

1. Инновационные материалы.

1.1. Использование композитных материалов с высокой прочностью и низким весом.

1.2. Внедрение 3D-печати для создания строительных конструкций.

1.3. Применение «зелёных» материалов, способствующих снижению углеродного следа.

2. Цифровизация отрасли.

2.1. BIM-технологии (Building Information Modeling) для проектирования и управления строительными объектами.

2.2. Использование программного обеспечения для оценки смет и управления поставками строительных материалов.

3. Автоматизация процессов.

3.1. Использование роботов и дронов для выполнения строительных задач и мониторинга.

3.2. Внедрение умных систем на объектах строительства (контроль за качеством материалов, автоматическое управление климатом).

4. Устойчивое строительство.

4.1. Переход на возобновляемые источники энергии в строительных процессах.

4.2. Разработка зданий с нулевым потреблением энергии (zero-energy buildings).

Проблемы и вызовы строительной отрасли:

1. Колебания цен на строительные материалы. Рынок материалов подвержен влиянию внешних факторов, таких как экономическая нестабильность, инфляция, логистические проблемы.

2. Качество материалов. Риск покупки несертифицированной или контрафактной продукции, что может негативно сказаться на долговечности строительства.

3. Дефицит инновационных материалов. В некоторых регионах существует ограниченный доступ к современным материалам, что замедляет внедрение инноваций.

4. Управление информацией. Из-за обилия поставщиков и материалов застройщики часто сталкиваются с трудностями в анализе и сравнении предложений.

5. Экологические стандарты. Ужесточение экологических требований вынуждает компании искать более экологичные материалы и технологии.

Важность автоматизации сбора данных:

Автоматизация процессов, связанных с выбором и покупкой строительных материалов, становится ключевым фактором успеха. Компании и индивидуальные застройщики нуждаются в инструменте, который:

1. Позволяет быстро находить и анализировать предложения поставщиков.

2. Обеспечивает доступ к актуальной информации о ценах, характеристиках и доступности материалов.

3. Снижает риск человеческой ошибки при работе с большими объёмами данных.

Создание программы для автоматизированного сбора данных даст следующие преимущества:

1. Снижение издержек. Экономия времени и ресурсов за счёт автоматизации анализа рынка строительных материалов.

2. Повышение конкурентоспособности. Быстрая реакция на изменения цен и условий поставки.

3. Улучшение качества строительства. Использование только проверенных и подходящих материалов.

Таким образом, строительный бизнес нуждается в эффективных цифровых решениях для обработки данных, что делает тему дипломной работы актуальной и практически значимой.

# Анализ имеющихся решений.

В современном строительном бизнесе широко используются цифровые инструменты для поиска, анализа и закупки строительных материалов. Эти решения позволяют автоматизировать рутинные процессы и принимать более обоснованные решения, основываясь на актуальных данных. На рынке представлено несколько типов программного обеспечения, которые могут быть разделены на три основные категории:

1. Каталоги строительных материалов (онлайн-платформы).

2. Программы для управления строительными проектами (строительный ERP).

3. Универсальные парсеры данных и платформы сбора информации.

Каталоги строительных материалов. Это специализированные онлайн-платформы, предоставляющие информацию о продукции, ценах и наличии. Основная цель таких решений – упростить процесс поиска материалов для строителей и проектировщиков.

Примеры решений:

1. "Стройплощадка.ру", "Пульс Цен", "Маркет строительных материалов".

Основной функционал: поиск материалов по категориям, сравнение цен от разных поставщиков, контакты для заказа напрямую у производителей.

Преимущества: простота использования, доступность информации о продуктах, возможность сравнения поставщиков.

Ограничения: нет автоматического сбора информации с сайтов поставщиков, часто требуется ручной поиск и выбор.

2. "AliExpress для стройматериалов" (международные площадки).

Преимущества: широкий выбор продукции, включая инновационные материалы.

Ограничения: сложность работы с большими объёмами данных, отсутствие локальных поставщиков.

ERP-системы интегрируют управление строительными процессами, включая закупки материалов, расчёты смет и контроль затрат.

Примеры решений:

1. 1С: Управление строительной организацией.

Возможности: управление проектами, автоматизация смет, интеграция с базами данных, отслеживание поставок строительных материалов.

Преимущества: высокая надёжность, глубокая интеграция с бухгалтерскими и складскими процессами.

Недостатки: сложность настройки, не подходит для сбора данных с внешних источников (нужен ручной ввод или интеграция с другими программами).

2. PlanRadar, Procore.

Это системы управления строительными проектами, включая закупки материалов.

Ограничения: акцент на управлении проектами, а не на сборе данных с внешних источников.

Универсальные парсеры данных и платформы. Для автоматизации сбора информации многие компании используют универсальные инструменты веб-скрейпинга, которые можно адаптировать под сбор данных о строительных материалах.

Примеры решений:

1. BeautifulSoup, Scrapy (Python-библиотеки).

Используются для веб-скрейпинга HTML-страниц и извлечения неструктурированных данных.

Преимущества: гибкость и возможность адаптации, подходит для сбора данных с любых сайтов.

Ограничения: требуется высокая квалификация для настройки и поддержки, отсутствие встроенного анализа данных.

2. Octoparse, ParseHub. Платформы с графическим интерфейсом для веб-скрейпинга данных.

Преимущества: простота настройки для пользователей без навыков программирования. визуализация результатов.

Недостатки: ограниченные возможности обработки больших массивов данных. платная подписка.

Результаты сравнения указаны в таблице 1.

Таблица 1 – Результаты сравнения аналогов

|  |  |  |
| --- | --- | --- |
| Категория решений | Преимущества | Недостатки |
| Каталоги строительных материалов | Удобный интерфейс, информация о продуктах доступна сразу. | Ограничение данными только определённых поставщиков. |
| Строительный *ERP* | Интеграция всех процессов в компании, точность данных. | Сложность настройки, нет сбора данных с внешних сайтов. |
| Универсальные парсеры | Гибкость, возможность автоматического сбора больших объёмов данных. | Требуют навыков программирования или платных подписок. |

Категория решений Преимущества Недостатки

Каталоги строительных материалов Удобный интерфейс, информация о продуктах доступна сразу. Ограничение данными только определённых поставщиков.

Строительный ERP Интеграция всех процессов в компании, точность данных. Сложность настройки, нет сбора данных с внешних сайтов.

Универсальные парсеры Гибкость, возможность автоматического сбора больших объёмов данных. Требуют навыков программирования или платных подписок.

Проблемы существующих решений.

1. Ограниченность функционала каталогов. Каталоги предоставляют только уже готовую информацию, отсутствует гибкость в сборе данных из разных источников.

2. Сложность настройки универсальных парсеров. Самостоятельная настройка веб-скрейпинга требует знаний в программировании и создания сложных алгоритмов.

3. Отсутствие интеграции между системами. Например, ERP-системы не позволяют автоматически анализировать актуальные данные с сайтов поставщиков.

4. Неполная информация. В каталогах и ERP часто отсутствуют данные об уникальных или локальных поставщиках, что ограничивает возможности пользователей.

Потенциал разрабатываемого решения. На основании анализа можно выделить ключевые особенности, которые должны присутствовать в эффективной системе сбора данных о строительных материалах:

1. Автоматизация сбора информации. Интеграция данных из HTML-страниц (неструктурированные данные) и API поставщиков (структурированные данные).

2. Фильтрация и структурирование информации. Преобразование неструктурированных данных в удобный для анализа формат.

3. Удобство использования. Простота настройки и возможность быстрого анализа данных.

4. Гибкость. Возможность адаптации под потребности различных пользователей: строительных компаний, проектировщиков, частных застройщиков.

Предлагаемое решение будет эффективно сочетать возможности универсальных парсеров с пользовательской простотой, характерной для каталогов, и интеграционным потенциалом ERP. Это создаст новый уровень эффективности и конкурентоспособности для участников строительного рынка.

# Обоснование необходимости разработки и юридической правомерности.

В условиях современного строительного рынка, характеризующегося высокой волатильностью цен и постоянным обновлением ассортимента материалов, проблема оперативного получения достоверной информации приобретает особую актуальность. Традиционные методы сбора данных через ручной мониторинг сайтов поставщиков или работу с бумажными каталогами демонстрируют свою неэффективность, приводя к существенным временным и финансовым издержкам.

* + - 1. Основная проблема заключается в том, что при ручном сборе информации:
      2. Существенно возрастает вероятность ошибок ввода данных;
      3. Теряется оперативность получения информации;
      4. Отсутствует возможность комплексного анализа большого массива данных;
      5. Затруднено сравнение предложений различных поставщиков.

Разрабатываемое программное решение призвано устранить эти недостатки путем автоматизации процесса сбора и обработки информации. Его внедрение позволит строительным компаниям существенно сократить временные затраты на анализ рынка, повысить точность принимаемых решений и, как следствие, оптимизировать бюджет строительных проектов. Особенно важным представляется возможность оперативного реагирования на изменения рыночной конъюнктуры, что в современных экономических условиях является критически важным фактором конкурентоспособности.

С юридической точки зрения разработка подобных систем требует особого внимания к вопросам правомерности сбора и использования данных. В данном проекте предусмотрены следующие меры для обеспечения соответствия законодательным требованиям:

Во-первых, система будет работать только с общедоступной информацией, не требующей авторизации или специальных разрешений для доступа. Это исключает нарушение положений о защите персональных данных и коммерческой тайне. При разработке алгоритмов сбора данных особое внимание уделяется соблюдению условий пользовательских соглашений веб-ресурсов.

Во-вторых, частота запросов к источникам информации будет ограничена таким образом, чтобы не создавать избыточной нагрузки на серверы поставщиков. Это соответствует принципам добросовестного использования интернет-ресурсов и исключает возможность квалификации таких действий как несанкционированный доступ.

В-третьих, в алгоритмы обработки данных закладываются механизмы фильтрации и исключения любой информации, которая может быть отнесена к персональным данным или коммерческой тайне. Это обеспечивает соответствие требованиям Федерального закона "О персональных данных" и аналогичных международных нормативных актов.

Важно отметить, что разрабатываемая система не предусматривает хранение или обработку данных, защищенных авторским правом. Все собираемые сведения относятся исключительно к техническим характеристикам товаров и их стоимости, что соответствует практике добросовестной конкуренции.

Таким образом, предлагаемое решение не только отвечает насущным потребностям строительной отрасли в автоматизации процессов сбора и анализа информации, но и полностью соответствует действующим правовым нормам, что делает возможным его практическое применение в реальных бизнес-процессах строительных компаний.

# ПРОЕКТИРОВАНИЕ СИСТЕМЫ

# Требования к системе.

**Функциональные требования.**

Функциональные требования определяют, что система должна **делать**, то есть её основные возможности и поведение с точки зрения пользователя и разработчика.

1. **Парсинг данных с внешних сайтов**  
   Система должна автоматически собирать информацию о товарах (название, цена, описание, единица измерения и т.д.) со сторонних веб-ресурсов, специализирующихся на продаже строительных материалов. Источники должны быть настраиваемыми, а сам процесс парсинга — управляемым.
2. **Сохранение данных в базу данных**  
   Все полученные данные после парсинга должны сохраняться в централизованную реляционную базу данных с возможностью последующей обработки и визуализации. Должна обеспечиваться структурированная схема хранения информации, поддерживающая связи между товарами, категориями и источниками.
3. **Категоризация товаров**  
   Система должна классифицировать строительные материалы по категориям (например, цемент, кирпич, утеплители, пиломатериалы и т.д.). Классификация может быть выполнена автоматически (по ключевым словам) либо вручную через интерфейс администратора.
4. **Просмотр данных через веб-интерфейс.**  
   Пользователи должны иметь доступ к информации о товарах через интуитивно понятный веб-интерфейс. Интерфейс должен предоставлять базовую информацию о товаре, цену, дату последнего обновления, а также ссылку на оригинальный источник.
5. **Поиск и фильтрация.**  
   Интерфейс должен поддерживать возможность поиска товаров по ключевым словам, фильтрации по категориям, диапазону цен, датам изменения и другим параметрам.
6. **Загрузка данных**  
   Пользователи должны иметь возможность выгружать данные в форматах CSV, Excel или JSON для дальнейшего анализа или использования в сторонних системах.
7. **Построение графика динамики цен**  
   Для каждого товара должна быть возможность построения диаграммы изменения цены во времени за заданный период. Это позволит анализировать рыночные тенденции и сравнивать цены от разных поставщиков.
8. **История изменений**  
   При каждом обновлении цены или других характеристик товара система должна сохранять предыдущее состояние для формирования полной истории изменений.

**Нефункциональные требования.**

Нефункциональные требования описывают **качество работы системы**, её надёжность, производительность, удобство использования и другие технические аспекты.

1. **Надёжность и сохранность данных**  
   Система должна обеспечивать устойчивое хранение и защиту информации. Все данные должны сохраняться в отказоустойчивой базе данных с регулярным резервным копированием. В случае сбоев не должно происходить потери данных.
2. **Регистрация и аутентификация пользователей**  
   Должна быть реализована система регистрации и входа с поддержкой безопасного хранения паролей (например, хеширование с солью).
3. **Многопоточность и параллельная обработка**  
   Система должна использовать многопоточность или асинхронную обработку при парсинге, чтобы одновременно обрабатывать данные с нескольких сайтов без снижения производительности.
4. **Отказоустойчивость**  
   В случае ошибок при подключении к источникам данных, сбоев в работе парсера или базы данных, система должна корректно логировать ошибки, сохранять текущее состояние и перезапускать процессы при необходимости.
5. **Производительность**  
   Интерфейс пользователя должен работать без заметных задержек, парсинг — происходить в разумные сроки (например, обработка 100 страниц за несколько минут), а построение графиков — выполняться мгновенно при наличии кэширования.
6. **Масштабируемость**  
   Архитектура системы должна предусматривать возможность расширения (например, добавление новых сайтов для парсинга, интеграция дополнительных аналитических модулей) без необходимости кардинальной переработки кода.
7. **Безопасность**  
   Система должна быть защищена от типовых веб-уязвимостей (SQL-инъекции, XSS, CSRF). Также следует ограничить количество попыток входа для предотвращения атак перебором.
8. **Интерфейс пользователя (UI/UX)**  
   Веб-интерфейс должен быть удобным, адаптивным под разные устройства (десктоп, планшет, смартфон), с поддержкой тёмной темы, если это предусмотрено.
9. **Кроссбраузерность**  
   Система должна корректно работать в популярных браузерах (Google Chrome, Mozilla Firefox, Microsoft Edge, Safari).
10. **Логирование и мониторинг**  
    Все ключевые действия системы (например, успешный парсинг, ошибки, входы в систему, скачивание файлов) должны логироваться. Это поможет в отладке, аудите и выявлении проблем.
11. **Локализация и язык интерфейса**  
    Интерфейс системы должен быть реализован минимум на русском языке с возможностью добавления других языков в будущем.
12. **Обновляемость**  
    Код и компоненты системы должны быть разработаны с учётом возможности обновления — добавление новых функций, исправление багов и обновление зависимостей не должно требовать полной переработки.

# Функциональная модель системы (диаграмма прецедентов)

# Архитектура системы (клиент сервер)

# Моделирование данных (классы, данные)

# Моделирование логики работы (диаграмма состояний, компонентов, последовательности)

# РАЗРАБОТКА СИСТЕМЫ

# Этапы разработки (Последовательность выполнения действий для разработки системы).

# Средства разработки (стек технологий и аппаратное обеспечение).

# Описание системы

# Интерфейс пользователя (прототипы в Фигме)

# Дизайн (красивое/не красивое)

# Функциональные блоки (Список страниц отличных по наполнению и их описание)

# Серверная часть системы (потоки http запросов от клиента к серверу и обратно)

# Описание контроллеров (конченых точек и их описание)

# Описание моделей (как и где тащим данные из базы или загружаем их)

# Алгоритмы обработки данных (как мы html в текст превращали, где и как это сделать)

# Входные данные (параметры для веб-скрапинга)

# Обработка данных (Определяем, что за тип файла, определяем структуру, отсекаем ненужное, тащим нужное)

# Выходные данные (Описание того, что вытащили)

# Тестирование системы (зачем тестируем, что тестируем)

# Методика тестирования системы (интеграционное, функциональное, модульное)

# Результаты тестов (тестировали то то получили то то)

# ЗАКЛЮЧЕНИЕ

# СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ