**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: от структурированных таблиц и баз данных до неструктурированных текстов на сайтах производителей и маркетплейсах.

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 структурированных и неструктурированных 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 как структурированных данных (характеристики, цены), так и неструктурированных (описания) 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, имеющей как структурированную, так и неструктурированную природу.

# ОСНОВНАЯ ЧАСТЬ

# АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

# Анализ отрасли (Что там в сфере строительства и стройматериалов)

# Анализ имеющихся решений (Какие аналоги уже есть и почему они хуже нашего).

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

# ПРОЕКТИРОВАНИЕ СИСТЕМЫ

# Требования к системе(функциональные/нефункциональные).

# Функциональная модель системы (диаграмма прецедентов)

# Архитектура системы (клиент сервер)

# Моделирование данных (классы, данные)

Входные данные:

1. Главная страница — ссылка откуда стартуем

2. Ограничение на домен — домен сайта

3. Маршрут обхода — список css селекторов для обхода всех уровней вложенности

4. Селектор перехода на следующую страницу(пагинация)

5. Название css селектор\* — имя селектора

6. Цена css селектор\* — имя селектора

7. Ед.изм css селектор\* — имя селектора

8. Характеристики css селектор\* — имя селектора

Выходные данные:

1. Стройматериал:

1.1 Название\* — одна строка

1.2 Ссылка\* — одна строка на страницу самого товара

1.3 Цена\* — цифра

1.4 Ед.изм — текст

1.5 Характеристики — все что внутри целевого блока

# 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 json/html в текст превращали, где и как это сделать)

# Входные данные (параметры для веб-скрепинга)

# 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