



Сибирский государственный
университет науки и технологий
имени академика М.Ф. Решетнева

Тема ВКР: «Разработка программного обеспечения для анализа рынка строительных материалов»

Автор: Семенов Евгений Александрович
студент группы БПЦ21-01

Руководитель ВКР: Сенашов С.И. д.ф.-м.н.,
профессор, профессор кафедры ИЭС

Актуальность



Ваши конкуренты скрывают свои секреты? Раскройте их и используйте в своих интересах.

Начать

Скрапинг не запрещён - отставание запрещено! Получайте конкурентные преимущества уже сегодня

Руководство

Тенденции на рынке строительных материалов:

- Быстрый доступ к актуальным данным.
- Много времени на ручной сбор информации.
- Рост онлайн-продаж стройматериалов.
- Динамичное изменение цен.

Стейкхолдеры:

- Компании (строительные, поставщики).
- Интернет-магазины и маркетплейсы.
- Частные клиенты и подрядчики.
- Аналитики рынка.



Проблемы

Основные проблемы:

1. Разрозненные данные.
2. Неэффективный ручной сбор.
3. Быстро меняющийся рынок.

Результат для бизнеса:

1. Переплаты
2. Простои
3. Потеря конкурентоспособности



Цели и задачи

Цель: разработать программное обеспечение для анализа рынка строительных материалов

Задачи:

- Проанализировать существующие подходы и инструменты для сбора данных.
- Определить ключевые требования к программному обеспечению.
- Разработать архитектуру программы.
- Реализовать основные алгоритмы веб-скрейпинга и обработки текстовой информации.
- Провести тестирование программы.
- Оценить эффективность программы и её практическую применимость.



Сравнительный анализ решений, представленных на рынке

Тип решения	Примеры	Преимущества	Недостатки	Проблемы
Каталоги стройматериалов	Стройплощадка , Пульс Цен	Готовые базы данных, Удобный интерфейс	Ограниченный набор поставщиков, Нет автоматического обновления данных	Не охватывают нишевых поставщиков, Требуют ручного поиска
Строительные ERP-системы	1C, Procore, PlanRadar	Интеграция всех бизнес-процессов, Точный учет материалов	Не собирают внешние данные, Сложность внедрения	Не решают проблему мониторинга рынка, Высокий порог входа
Универсальные парсеры	Scrapy, Octoparse, ParseHub	Сбор данных с любых сайтов, Гибкость настроек	Требуют программирования, Платные подписки	Сложность поддержки Нет встроенной аналитики



Возможности системы

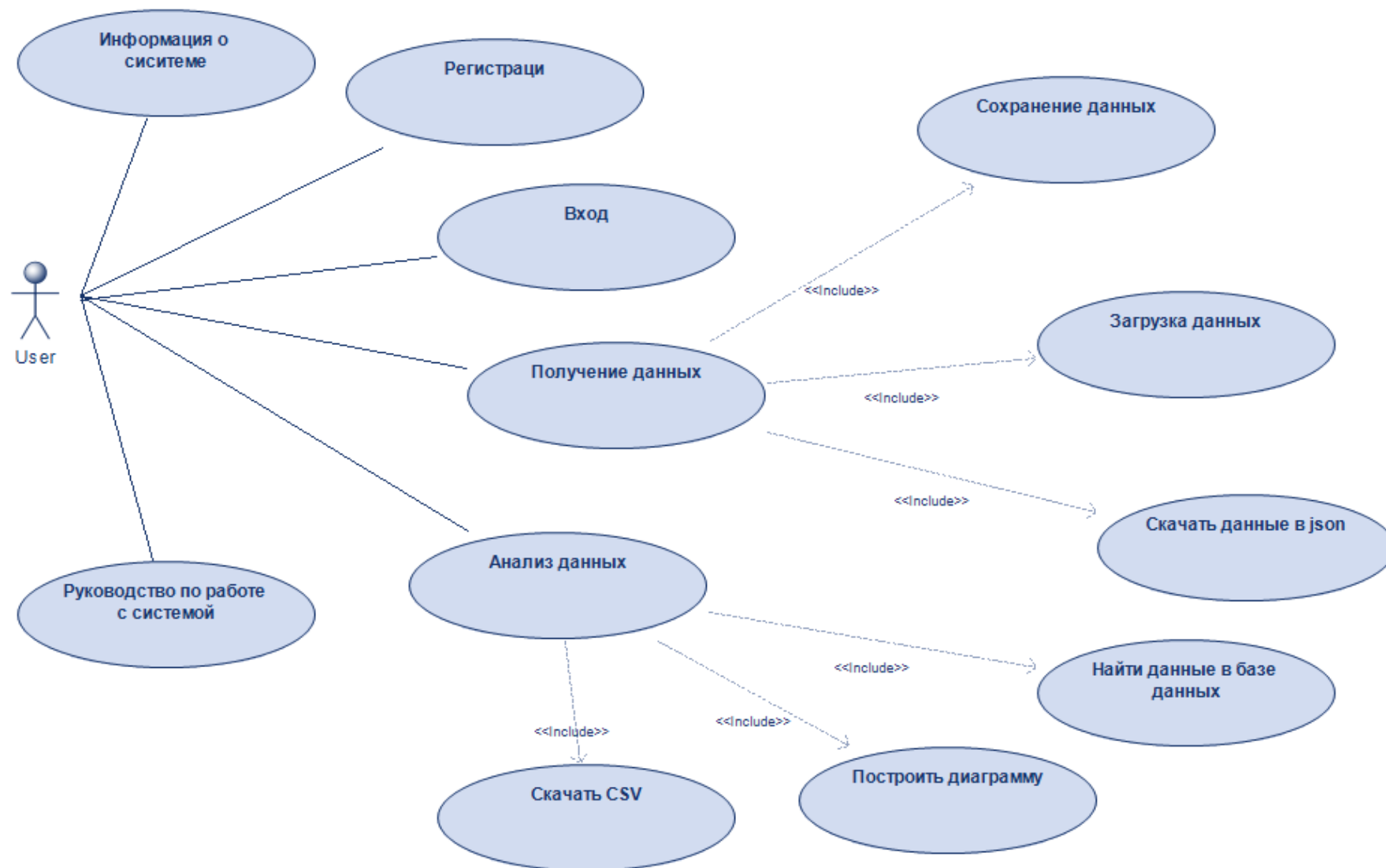


Диаграмма прецедентов

Модель данных

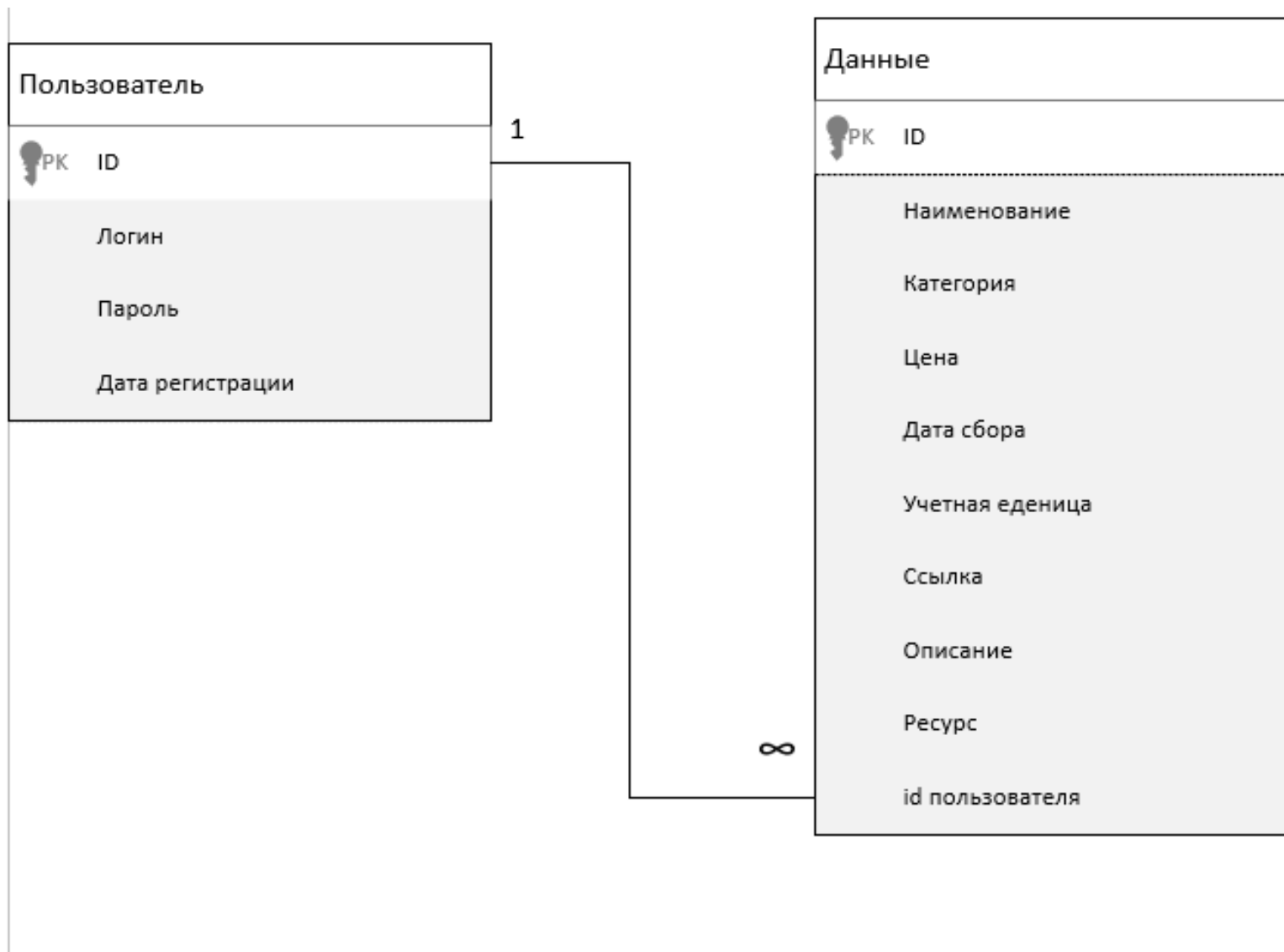


Схема хранения данных





Средства разработки системы

Фронтенд:

- React.js + Redux Toolkit
- Chart.js
- Sass + Animate.css
- Node.js



Бекенд:

- Python + Flask
- Scrappy + Playwright
- Pandas
- JWT



База данных:

- MySQL + SQLAlchemy
- Инфраструктура:
- Docker
- Linux/Ubuntu



Аппаратные требования:

- 4 ядра CPU, 8+ ГБ RAM, SSD 512ГБ
- Стабильный интернет (10+ Мбит/с)

Стартовая страница системы



testuser123

Ваши конкуренты скрывают свои секреты? Раскройте их и используйте в своих интересах.

[Начать](#)

Скрапинг не запрещён - отставание запрещено! Получайте конкурентные преимущества уже сегодня

[Руководство](#)

Парсер данных строительных материалов

Python-решение для автоматического сбора и анализа информации о строительных материалах с различных онлайн-источников

Основные возможности

Сбор актуальных данных

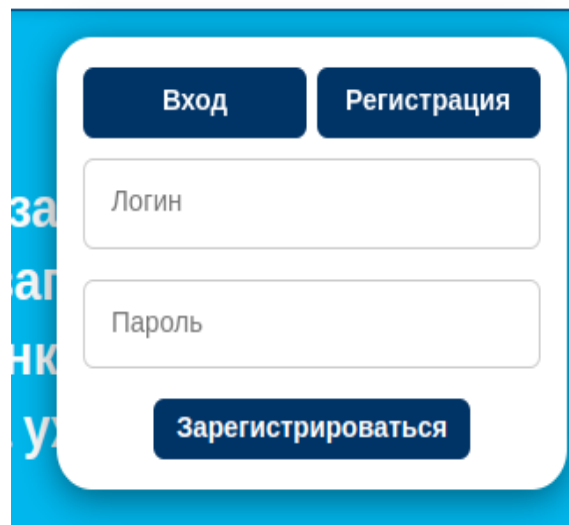
Автоматизированный сбор информации о ценах, наличии и характеристиках строительных материалов: кирпича, бетона, металлопроката, пиломатериалов и других

Стартовая страница



Управление доступом

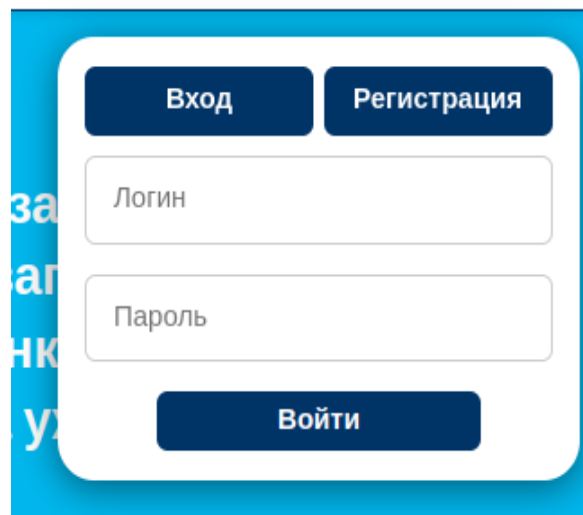
 Войти



UI for the registration block. It features a blue header with a user icon and the text 'Войти'. Below this is a white rounded rectangle containing two dark blue buttons: 'Вход' and 'Регистрация'. Under the buttons are two input fields labeled 'Логин' and 'Пароль'. At the bottom is a dark blue button labeled 'Зарегистрироваться'.

Блок регистрации

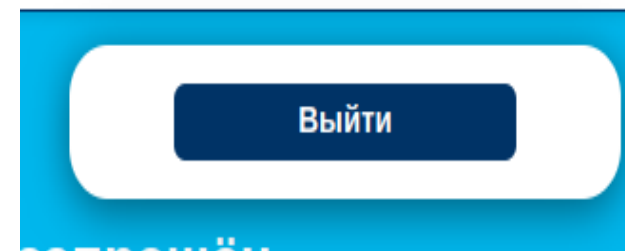
 Войти



UI for the login block. It features a blue header with a user icon and the text 'Войти'. Below this is a white rounded rectangle containing two dark blue buttons: 'Вход' and 'Регистрация'. Under the buttons are two input fields labeled 'Логин' and 'Пароль'. At the bottom is a dark blue button labeled 'Войти'.

Блок входа

 testuser123



UI for the logout block. It features a blue header with a user icon and the text 'testuser123'. Below this is a white rounded rectangle containing a single dark blue button labeled 'Выйти'.

Блок выхода из учетной записи



Руководство пользователя



testuser123

Скрапинг Анализ **Руководство**

Тutorial

Общая информация
Возможности платформы
Правила скрапинга
Выходные данные
Визуализация

Правила скрапинга

Правила скрапинга (с использованием Scrapy)

Обязательные параметры конфигурации

- **start_urls** — начальные ссылки для парсинга, например: ["https://domain.com"]
- **allowed_domains** — ограничение на обрабатываемые домены: ["domain.com"]
- **product_path** — относительный путь к страницам товаров: "/product/"
- **category_selector** — селектор категории товара (CSS)
- **name_selector** — селектор названия товара (CSS)
- **price_selector** — селектор цены товара (CSS)
- **unit_selector** — селектор единицы измерения (CSS)
- **block_selector** — блок, содержащий дополнительные параметры (XPath)
- **key_selector** — селектор ключа параметра внутри блока (XPath)
- **value_selector** — селектор значения параметра внутри блока (XPath)

Примеры CSS-селекторов

- `div` — все `<div>`
- `#price` — элемент с `id="price"`
- `.title` — элементы с классом `title`
- `div.product` — `<div>` с классом `product`
- `span[data-id="123"]` — `тег` с атрибутом `data-id="123"`
- `li:nth-child(2)` — второй элемент внутри списка
- `ul li` — вложенные теги (`li` в `ul`)
- `.container .price` — класс `price` вложен в `container`
- `h1, h2` — выбрать несколько тегов
- `.col, .row` — несколько классов
- `.product.featured` — элемент с двумя классами одновременно

Примеры XPath-селекторов

- `//div` — все `<div>` на странице
- `//*[@id="price"]` — элемент с `id="price"`
- `//*[contains(@class, "item")]` — элементы, содержащие класс `item`
- `//div[contains(@class, "card")]` — `<div>` с классом `card`



Сбор данных

testuser123

Скрапинг

Анализ

Руководство

Стартовая ссылка

https://baza.124bt.ru

Домен

baza.124bt.ru

Идентификатор продукта

/product/

Категория

р е м а

Наименование

[itemprop="name"]

Цена

.price.nowrap

Ед.из

.ruble

Блок характеристик

//table[@id="product-features"]

Ключ характеристик

./td[@class="name"]

Значение характеристик

./td[@class="value"]

Старт

Стоп

Сохранить

Скачать

Статус: Выполняется

03:54:48 [INFO] Страница не содержит данных: https://baza.124bt.ru/category/fanera/ (не содержит /product/)

Страница не содержит данных: https://baza.124bt.ru/category/fanera/ (не содержит /product/)

03:54:50 [INFO] Ignoring response <404 https://baza.124bt.ru/compare/>: HTTP status code is not handled

Ignoring response <404 https://baza.124bt.ru/compare/>: HTTP status code is not handled or not allowed

03:54:58 [INFO] Ключи: ['Производитель', 'Тип сайдинга', 'Толщина', 'Применение', 'Ударная вязкость, кгс/

Ключи: ['Производитель', 'Тип сайдинга', 'Толщина', 'Применение', 'Ударная вязкость, кгс/см2', 'Морозост

03:54:58 [INFO] Значения ['Тимлюйский шифер', 'Фиброцементный', '8 мм', 'Для наружной отделки', '3,9', 'F

Значения ['Тимлюйский шифер', 'Фиброцементный', '8 мм', 'Для наружной отделки', '3,9', 'F25', '30', '20x1

03:54:58 [INFO] Характеристики: {'Производитель': 'Тимлюйский шифер', 'Тип сайдинга': 'Фиброцементный',

Характеристики: {'Производитель': 'Тимлюйский шифер', 'Тип сайдинга': 'Фиброцементный', 'Толщина': '8 мм'

03:54:58 [INFO] Все вместе: {'category': 'Сайдинг',

'characteristics': {'Габариты (В x Ш x Г)': '8 x 220 x 3000 мм',

'Дугостойкость при токе 20 МА, с': '30',

'Морозостойкость': 'F25',

'Объемный вес, г/см3': '1,69',

'Площадь листа, м2': '0,66',

'Предел прочности при изгибе, МПА (кгс/см2)': '264',

'Применение': 'Для наружной отделки',

'Производитель': 'Тимлюйский шифер',

'Тип сайдинга': 'Фиброцементный',

'Толщина': '8 мм',

'Ударная вязкость, кгс/см2': '3,9',

'Штук в поддоне': '252',

'Электрическая прочность, кВ/м': '20x10^3'}},

'date_time': '2025-05-12 03:54:38',

'link': 'https://baza.124bt.ru/product/fibrocentenny-sayding-krasnyy-gost-18124-2012/',

'name': 'Фиброцементный сайдинг красный ГОСТ 18124-2012',

'price': '630',

'resource': 'https://baza.124bt.ru',

'unit': 'Р',

'user': 'testuser123'}

Все вместе: {'category': 'Сайдинг',

'characteristics': {'Габариты (В x Ш x Г)': '8 x 220 x 3000 мм',

'Дугостойкость при токе 20 МА, с': '30',

Страница «Сбор данных»



Поиск данных



testuser123

Скрапинг Анализ Руководство

Ресурс

Категория

Название

Начиная ☐

Заканчивая ☐

Найти

Диаграмма

Скачать CSV

Перемычка брусковая ЗПБ 39-8п ГОСТ 948-2016

Категория: Перемычки железобетонные

Ресурс: https://baza.124bt.ru

Цена: 3000.45 Р

Дата: 09.05.2025 12:00:00

Характеристики:

Вес: 257 кг
Объем: 0.103 м³
Ширина: 120 мм
Назначение: Перемычки
Производитель: БТ
Марка прочности: М200
Габариты (В x Ш x Г): 220 x 120 x 3890 мм

Ссылка

Перемычка брусковая ЗПБ 39-8п ГОСТ 948-2016

Категория: Перемычки железобетонные

Ресурс: https://baza.124bt.ru

Цена: 2942.99 Р

Дата: 09.04.2025 12:00:00

Характеристики:

Вес: 257 кг
Объем: 0.103 м³
Ширина: 120 мм
Назначение: Перемычки
Производитель: БТ
Марка прочности: М200
Габариты (В x Ш x Г): 220 x 120 x 3890 мм

Ссылка

Перемычка брусковая ЗПБ 39-8п ГОСТ 948-2016

Категория: Перемычки железобетонные

Ресурс: https://baza.124bt.ru

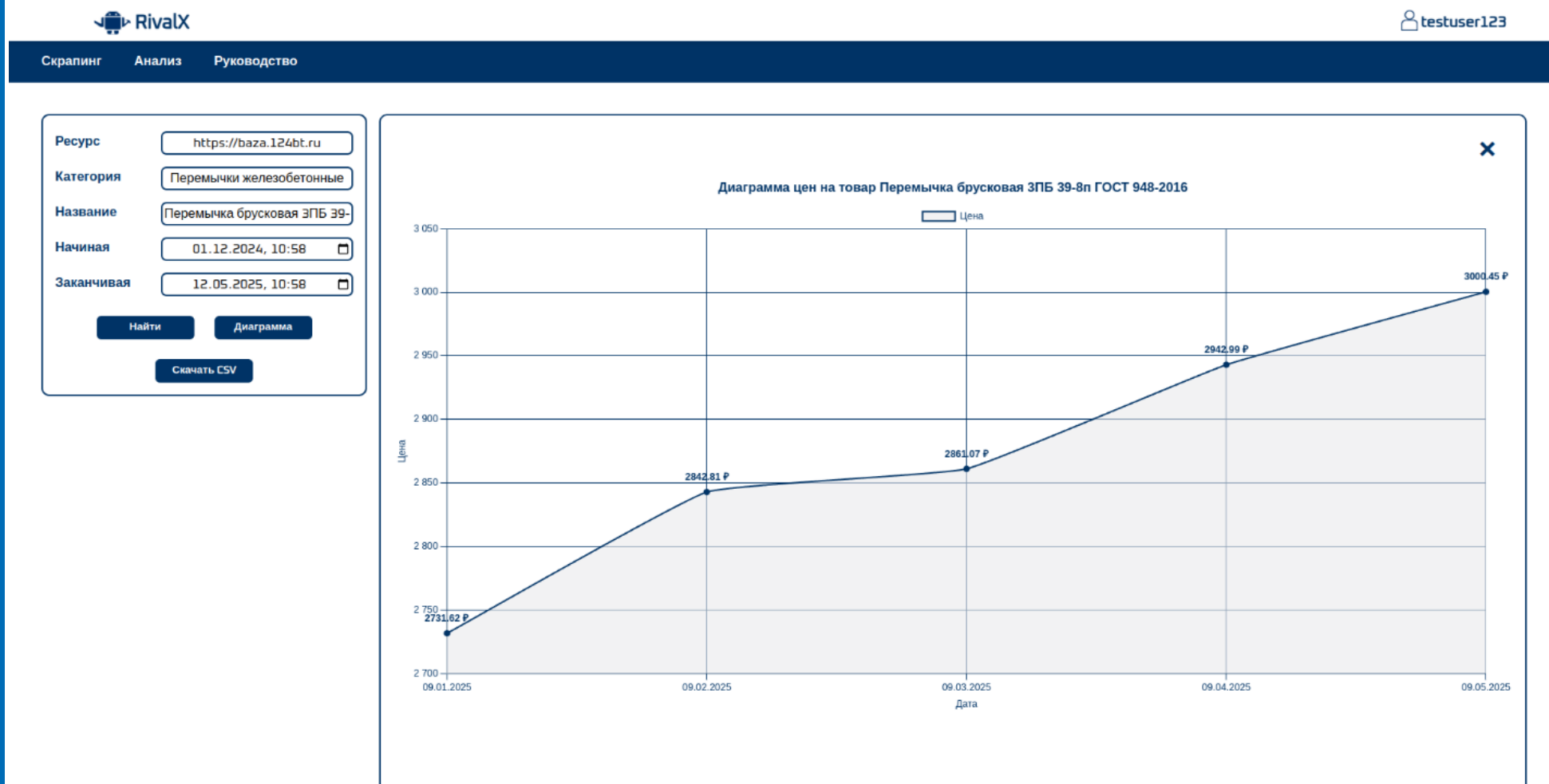
Цена: 2861.07 Р

Дата: 09.03.2025 12:00:00

Характеристики:

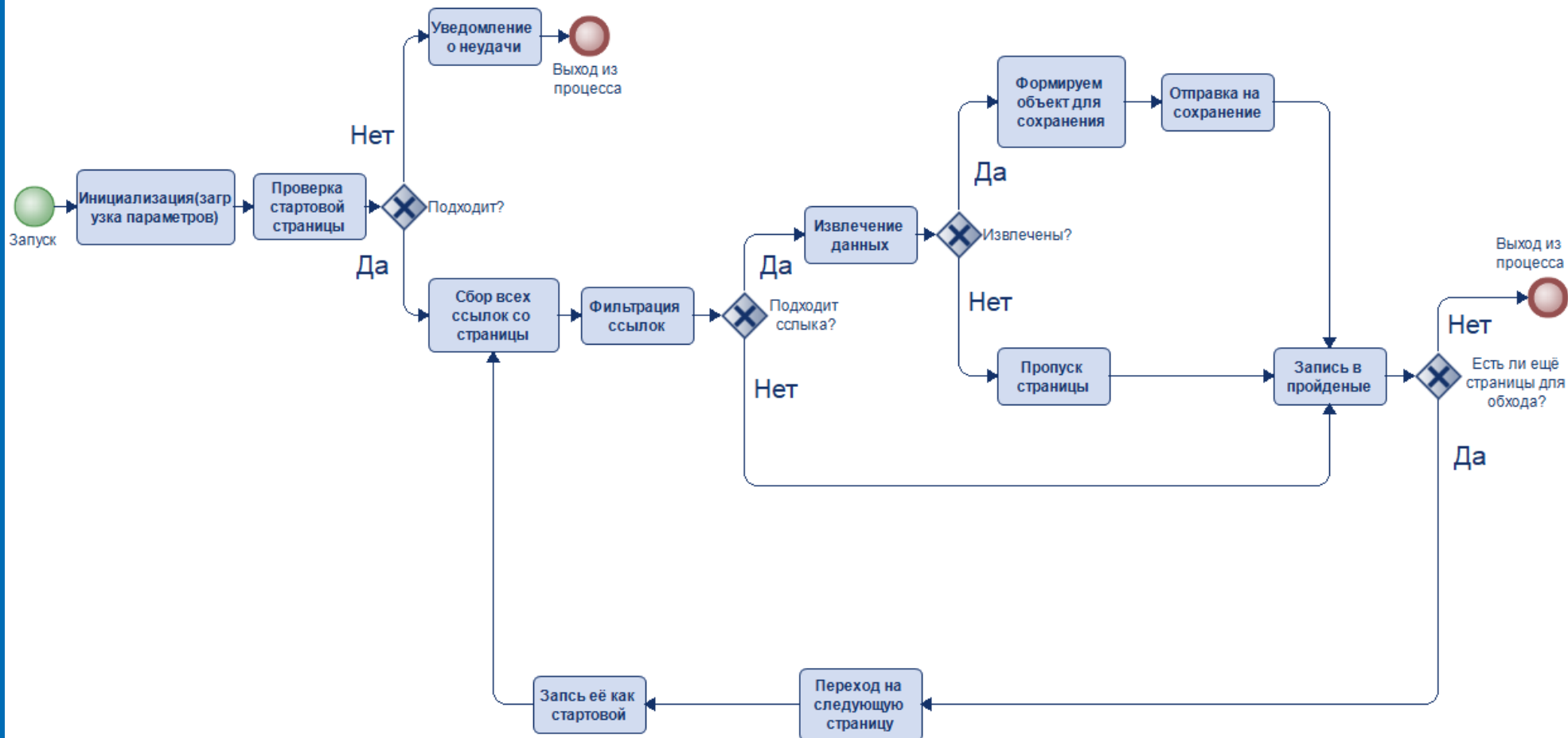
Страница «Поиск»

Визуализация данных



Страница «Построение диаграммы»

Алгоритм получения данных



Алгоритм получения данных



Тестирование системы

1. API (автотесты)

- Регистрация/авторизация.
- Защищенные endpoints.
- Обработка ошибок.



2. Парсер (ручные тесты)

- Сценарии парсинга.
- Валидация данных.

3. Интеграция с БД

- Сохранение данных.

Пример вывода тестов.

Успешный результат:

```
{test_01_register (__main__APITestCase) ... OK  
test_02_login (__main__APITestCase) ... OK  
test_03_protected_valid_token  
(__main__APITestCase) ... OK  
test_04_protected_no_token (__main__APITestCase)  
... OK  
Ran 10 tests in 1.542s  
OK}
```

Ошибка:

```
{test_04_protected_no_token (__main__APITestCase)  
... FAIL  
AssertionError: 200 != 401}
```

1. **Автотесты** проверяют API и базовую логику.

Ручные тесты покрывают парсинг и обработку данных

2. **Ручные тесты** покрывают парсинг и обработку данных.



Экономическая эффективность разработки системы

1. Затраты на разработку

Статья расходов	Стоимость (руб.)
Фронтенд (React, Redux)	150 000 – 300 000
Бекенд (Python, Flask)	200 000 – 400 000
Парсинг (Scrapy, Playwright)	100 000 – 200 000
База данных (MySQL)	50 000 – 100 000
Тестирование и DevOps	100 000 – 200 000
Итого	600 000 – 1 200 000

2. Модель монетизации

Тариф	Цена (руб./мес.)	Кол-во клиентов	Годовой доход
Базовый	1 000	50	600 000
Профессиональный	5 000	20	1 200 000
Корпоративный	7 000	10	840 000
Итого	-	80	2 640 000

Срок окупаемости

- Первоначальные затраты: 1 200 000 руб.
- Чистая прибыль (после налогов и оплаты хостинга): ~1 848 000 руб./год
- Окупаемость: 8 месяцев.



Обоснование юридической правомерности

1. Законодательная база:

- *Федеральный закон **№149-ФЗ** "Об информации, информационных технологиях и о защите информации":*
- *Судебная практика (дело № А40-18827/17-110-180 ВК **предъявило иск к Дабл**):*

2. Наше ПО не нарушает закон, так как:

- Собирает только общедоступные данные
- Не требует обхода авторизации или CAPTCHA
- Соблюдает ограничения robots.txt

3. Что мы НЕ делаем (чтобы избежать нарушений)

- Не собираем персональные данные
- Не взламываем защиту сайтов
- Не нарушаем пользовательские соглашения

4. Техническая и юридическая безопасность

- Частотность запросов — без перегрузки серверов
- Открытые источники — только публичные страницы
- Структурирование данных — без искажения оригинальной информации



Результаты работы

Достигнутая цель - разработано программное обеспечение для автоматизированного анализа рынка строительных материалов.

Выполненные задачи:

- Анализ решений.
- Определение требований.
- Разработка архитектуры.
- Реализация системы.
- Тестирование.
- Оценка эффективности.



1. Интеллектуализация платформы

- Внедрение ИИ и машинного обучения

2. Максимальная автоматизация

- Самообучающийся парсер
- Упрощенный интерфейс

3. Глубокая интеграция

- Подключение к бизнес-системам
- Расширенные форматы выгрузки

4. Расширенный мониторинг

- Трекер изменений в реальном времени





СПАСИБО ЗА ВНИМАНИЕ



<https://github.com/Evgenij1234/Diplom-work>

© Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева, 2025