МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «КубГУ»)

Факультет компьютерных технологий и прикладной математики

Кафедра информационных технологий

ОТЧЕТ О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ № 3

по дисциплине  
 «ТЕХНОЛОГИИ ПРОЕКТИРОВАНИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ»

Выполнил студент группы 34                                       В.В.Несветайлов

Отчет принял канд. техн. наук, доц. КИТ                                                                                       А.Н. Полетайкин

Краснодар

2025 г.

Тема: Техническое задание на создание программного продукта.

Цель: Освоение методики предварительного анализа разрабатываемой программы; освоение задач формулирования функциональных и нефункциональных требований к программной реализации отдельных задач и к программе в целом; выработка навыков разработки технического задания. Индивидуальная тема: Разработка парсера для сайта Wildberies

Задание

1. Установить назначение и общую цель создания программы.
2. Определить структуру программы и состав функциональных задач.
3. Разработать функциональные требования к программе: требования к входным и выходным данным; требования к программной реализации задач; специальные требования к математическому обеспечению программной реализации задач;
4. Разработать модель требований в нотации UML
5. Разработать требования к информационному обеспечению (к базе данных).
6. Разработать требования к инструментальному программному обеспечению (к системе управления базой данных (СУБД), к средству разработки программ (IDE), средствам автоматизированного проектирования ПО)
7. Установить нефункциональные требования к программе. Дать не менее 5 наиболее очевидных для данной системы требований

1. Назначение и общая цель

Программное средство автоматизирует процесс сбора и анализа данных с маркетплейса Wildberries, включая информацию о товарах, ценах, остатках и отзывах.

Вид автоматизированной деятельности:

* Парсинг (сбор) данных с веб-ресурсов.
* Автоматизация мониторинга цен и ассортимента.

Объект автоматизации:

* Компании, занимающиеся аналитикой e-commerce.
* Продавцы на Wildberries, нуждающиеся в конкурентном анализе.

Цели:

1. Снижение временных затрат на ручной сбор данных.
2. Повышение точности и актуальности информации о товарах.
3. Оптимизация процессов ценообразования и управления ассортиментом.

Ожидаемые улучшения показателей:

1. Скорость сбора данных – сокращение времени парсинга в 10–20 раз.
2. Точность данных – минимизация ошибок по сравнению с ручным вводом.
3. Автоматизация отчетности – генерация аналитических отчетов в реальном времени.
4. Масштабируемость – возможность обработки большого количества товарных позиций.

2. Структура программы и состав функциональных задач

Программа состоит из следующих модулей:

1. Модуль авторизации и обхода защиты – работа с прокси, ротация User-Agent, обработка капчи.
2. Модуль парсинга данных – сбор информации о товарах (название, цена, рейтинг, отзывы).
3. Модуль обработки данных – очистка от дубликатов, форматирование.
4. Модуль хранения данных – сохранение в базу данных (SQLite/PostgreSQL) или файлы (CSV/JSON).
5. Модуль визуализации и отчетности – построение графиков динамики цен, экспорт в Excel.

Функциональные задачи:

1. Загрузка и обработка входных данных (URL категорий/товаров Wildberries).
2. Парсинг HTML/API Wildberries с учетом защиты от блокировок.
3. Фильтрация и структурирование данных.
4. Сохранение данных в выбранном формате.
5. Визуализация результатов (графики, таблицы).
6. Уведомление пользователя об ошибках (напр., при блокировке IP).

Структура программы представлена на рисунке 1.

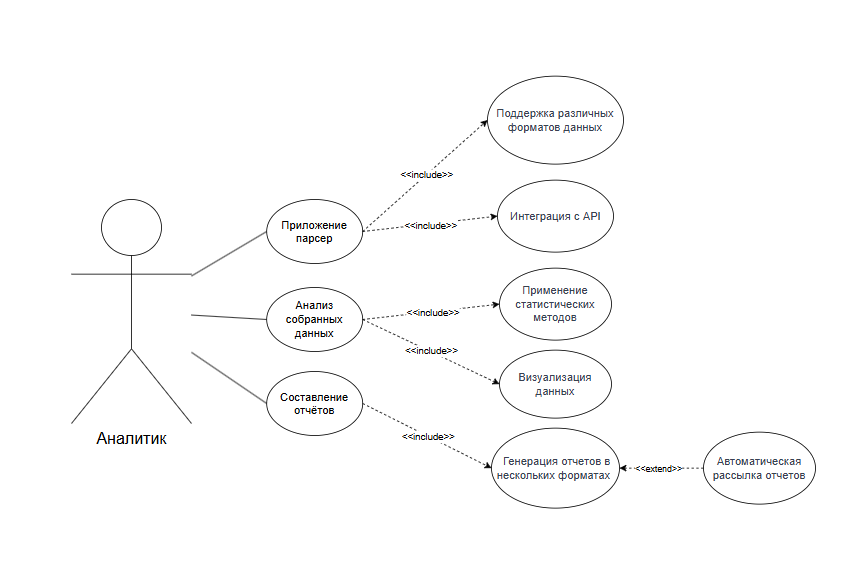


Рисунок 1 — Модель требований UML

Требование к инструментальному программному обеспечению

1. Средства разработки (IDE и редакторы кода)

* Основная IDE:
  + PyCharm Professional (для полноценной разработки с поддержкой БД и отладки)
  + VS Code (с расширениями Python, Docker, SQL Tools) – как легковесная альтернатива
* Обязательные плагины:
  + Поддержка Python 3.10+
  + Инструменты для работы с SQL (для СУБД)
  + Git-интеграция
  + Литера (для статического анализа кода)

2. Библиотеки и фреймворки

* + Парсинг:
    - requests, aiohttp (HTTP-запросы)
    - BeautifulSoup4, lxml (HTML-парсинг)
    - selenium (для JS-страниц)
  + Работа с данными:
    - pandas (анализ и обработка)
    - numpy (числовые операции)
  + Визуализация:
    - matplotlib, seaborn (графики)
    - reportlab (PDF-отчеты)

Требования к входным данным

В качестве входных данных будет приниматься текстовый файл в формате JSON или CSV, содержащий информацию о товарах с сайта Wildberries. Структура данных должна включать следующие поля:

1. Id товара: уникальный идентификатор товара на сайте Wildberries. Должен быть представлен в виде строки (string).
2. Название товара: название товара, представленное в виде строки (string).
3. Цена: стоимость товара, представленная в виде числа с плавающей запятой (float).
4. Ссылка на изображение: URL-адрес изображения товара, представленный в виде строки (string).
5. Категория: категория, к которой относится товар, представленная в виде строки (string).
6. Описание: текстовое описание товара, представленное в виде строки (string).

Требования к выходным данным

В качестве выходных данных будет предоставляться структурированный отчет в формате Excel (XLSX) или CSV, содержащий информацию о товарах, собранную с сайта Wildberries. Отчет должен соответствовать следующим требованиям:

1. Формат: файл должен быть сохранен в формате XLSX или CSV, который поддерживается большинством офисных приложений.
2. Содержимое: отчет должен включать в себя следующие столбцы:
   * Id товара
   * Название товара
   * Цена
   * Ссылка на изображение
   * Категория
   * Описание
3. Структура: данные должны быть организованы в виде таблицы, где каждая строка соответствует отдельному товару.
4. Форматирование: в случае формата XLSX, отчет должен содержать базовое форматирование (например, заголовки столбцов должны быть выделены).
5. Совместимость: выходной файл должен быть совместим с основными офисными приложениями, такими как Microsoft Excel и Google Sheets.

Функциональные требования для парсера сайта Wildberries

1. Требование к задаче «Загрузка и обработка данных»  
   Система должна поддерживать возможность загрузки данных с сайта Wildberries, включая информацию о товарах, такую как Id, название, цена, ссылка на изображение, категория и описание.
2. Требование к задаче «Парсинг данных»  
   Система должна автоматически извлекать и обрабатывать информацию о товарах с сайта Wildberries, обеспечивая корректное извлечение всех необходимых полей.
3. Требование к задаче «Визуализация данных»  
   Система должна иметь возможность визуализировать собранные данные о товарах в виде таблицы или графика для удобства анализа и оценки качества информации.
4. Требование к задаче «Редактирование данных»  
   Система должна предоставлять инструменты для ручного редактирования собранных данных, позволяя пользователю вносить изменения в информацию о товарах при необходимости.
5. Требование к задаче «Экспорт данных»  
   После завершения парсинга и обработки данных программа должна экспортировать собранную информацию в формате Excel (XLSX) или CSV для дальнейшего использования.
6. Требование к задаче «Генерация отчета»  
   При наличии ошибок или несоответствий в собранных данных система должна выдавать отчет об ошибках, в котором будут описаны нарушенные пункты, такие как отсутствующие поля или некорректные значения.

Требования к информационному обеспечению

Программная система должна содержать базу данных, в которой будет храниться информация о товарах с сайта Wildberries, включая их Id, название, цену, ссылку на изображение, категорию и описание. Должна быть реализована возможность добавления новых данных, изменения текущих, а также удаления элементов из базы данных. Данные должны включать информацию о каждом товаре, а также метаданные, такие как дата последнего обновления.

База данных должна обеспечивать:

1. Защиту данных от несанкционированного доступа: Доступ к базе данных должен быть ограничен и защищен с помощью аутентификации и авторизации пользователей.
2. Отсутствие дублирования информации: Система должна обеспечивать уникальность записей в базе данных, предотвращая дублирование информации о товарах.
3. Высокую скорость обработки запросов: База данных должна обеспечивать быструю обработку запросов на добавление, изменение и удаление данных, а также на извлечение информации о товарах.

Нефункциональные требования

1. Система должна обеспечивать проверку на соответствие актуальным данным с сайта Wildberries: Система должна регулярно проверять и обновлять информацию о товарах, чтобы гарантировать актуальность данных.
2. Система должна сохранять данные и продолжать работу в случае аварийного завершения: В случае сбоя или аварии система должна иметь механизмы резервного копирования и восстановления данных.
3. Время парсинга данных не должно превышать 10 секунд: Система должна эффективно обрабатывать запросы и извлекать данные с сайта Wildberries в установленные сроки.
4. Система должна быть доступна 24/7: Программное обеспечение должно обеспечивать круглосуточный доступ к функционалу парсинга и обработки данных.
5. Пользовательский интерфейс должен быть интуитивно понятным: Интерфейс системы должен быть простым и удобным для пользователей, позволяя легко выполнять операции по загрузке, редактированию и экспорту данных.

Вывод

В данном отчете представлено техническое задание на разработку парсера для сайта Wildberries, целью которого является автоматизация сбора и анализа данных о товарах. Программа включает модули для авторизации, парсинга, обработки и визуализации данных, а также обеспечивает надежное хранение информации в базе данных с защитой от несанкционированного доступа. Нефункциональные требования акцентируют внимание на быстродействии, доступности и удобстве пользовательского интерфейса. Ожидаемые улучшения в скорости и точности сбора данных делают проект актуальным для аналитики в сфере e-commerce.