



Кировское областное государственное профессиональное  
образовательное бюджетное учреждение  
«Слободской колледж педагогики и социальных отношений»

## **Разработка программного обеспечения для извлечения информации из больших объемов текстовых данных**

**Выполнила:**

Поглазова Валерия Владимировна

**Группа: 21П-1**

**Специальность:**

09.02.07 Информационные  
системы и программирование

**Руководитель:** Калинин Арсений  
Олегович

# РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ

## Цели и задачи

Цель дипломного проекта - заключается в разработка и реализации программного обеспечения, способного эффективно извлекать заданную информацию из больших объемов текстовых данных, преобразовывать ее в удобный для анализа и дальнейшей обработки структурированный формат, обеспечивая высокую скорость и точность извлечения.

### Задачи

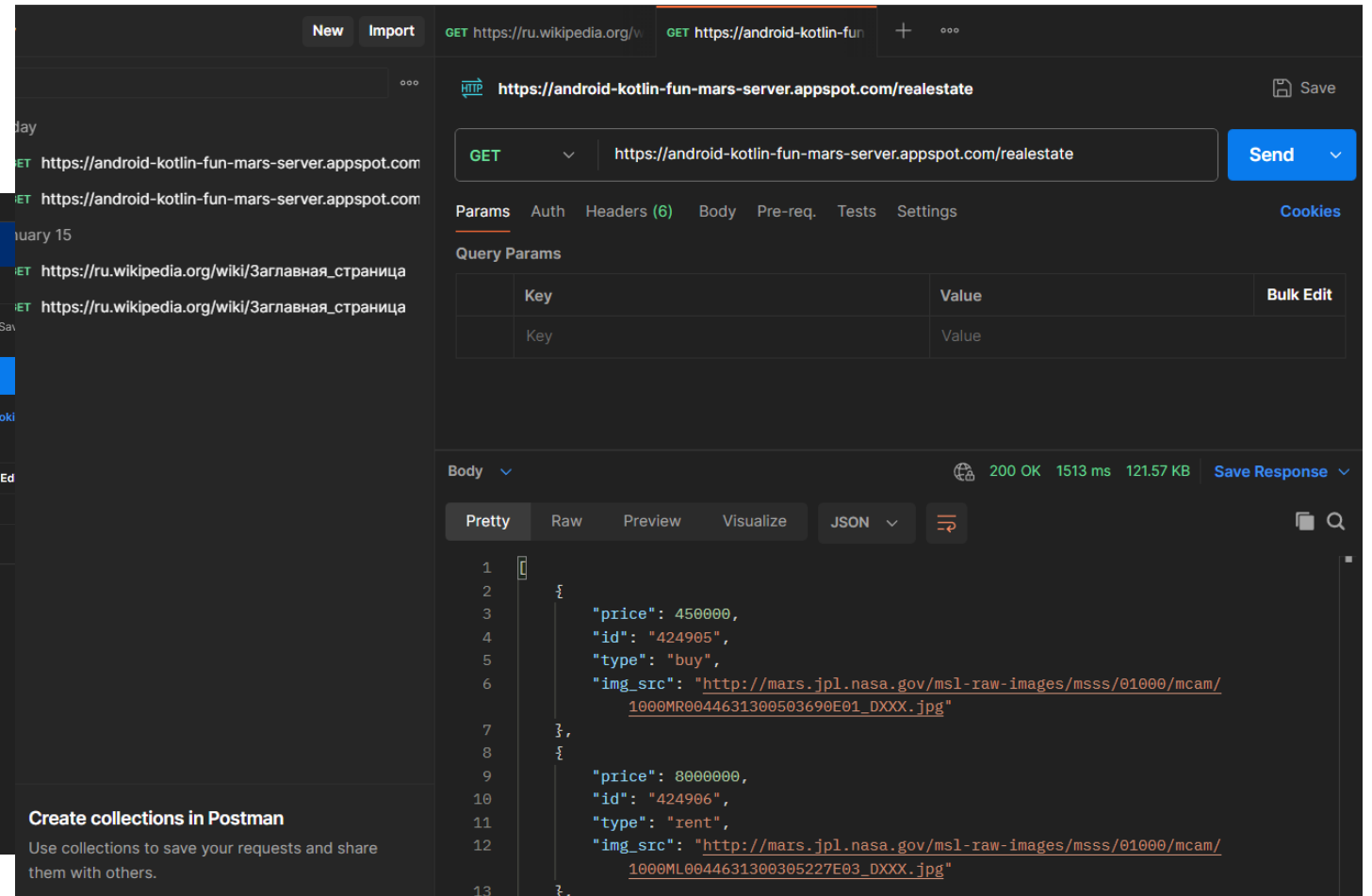
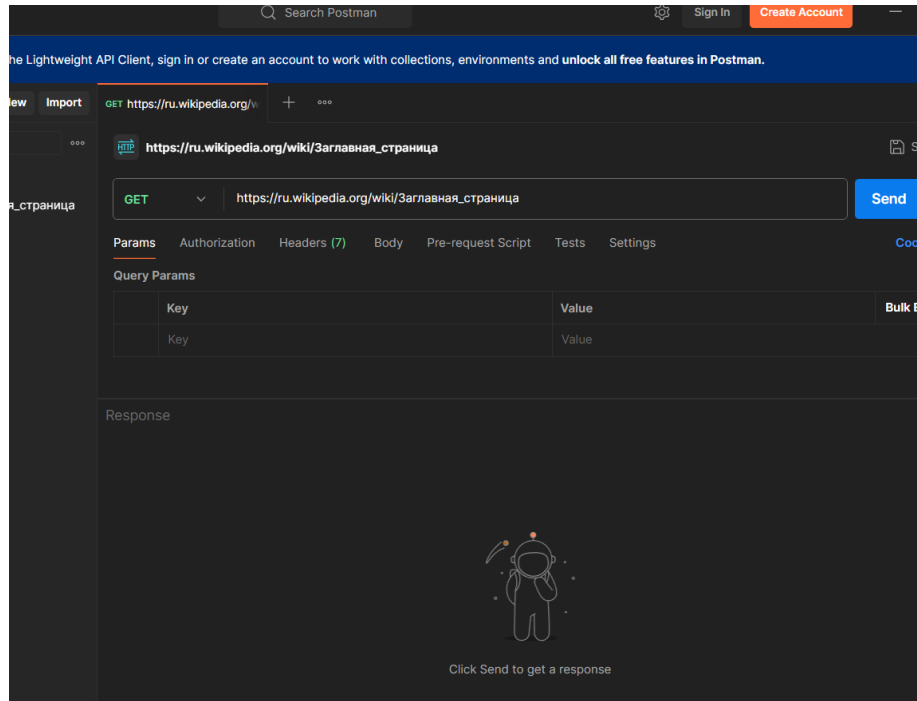
- Описать предметную область
- Разработать техническое задание на создание программного продукта
- Описать архитектуру программы
- Описать алгоритмы и функционирование программы
- Провести тестирование и опытную эксплуатацию
- Разработать руководство оператора

# РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ

## Программа аналог



# Postman



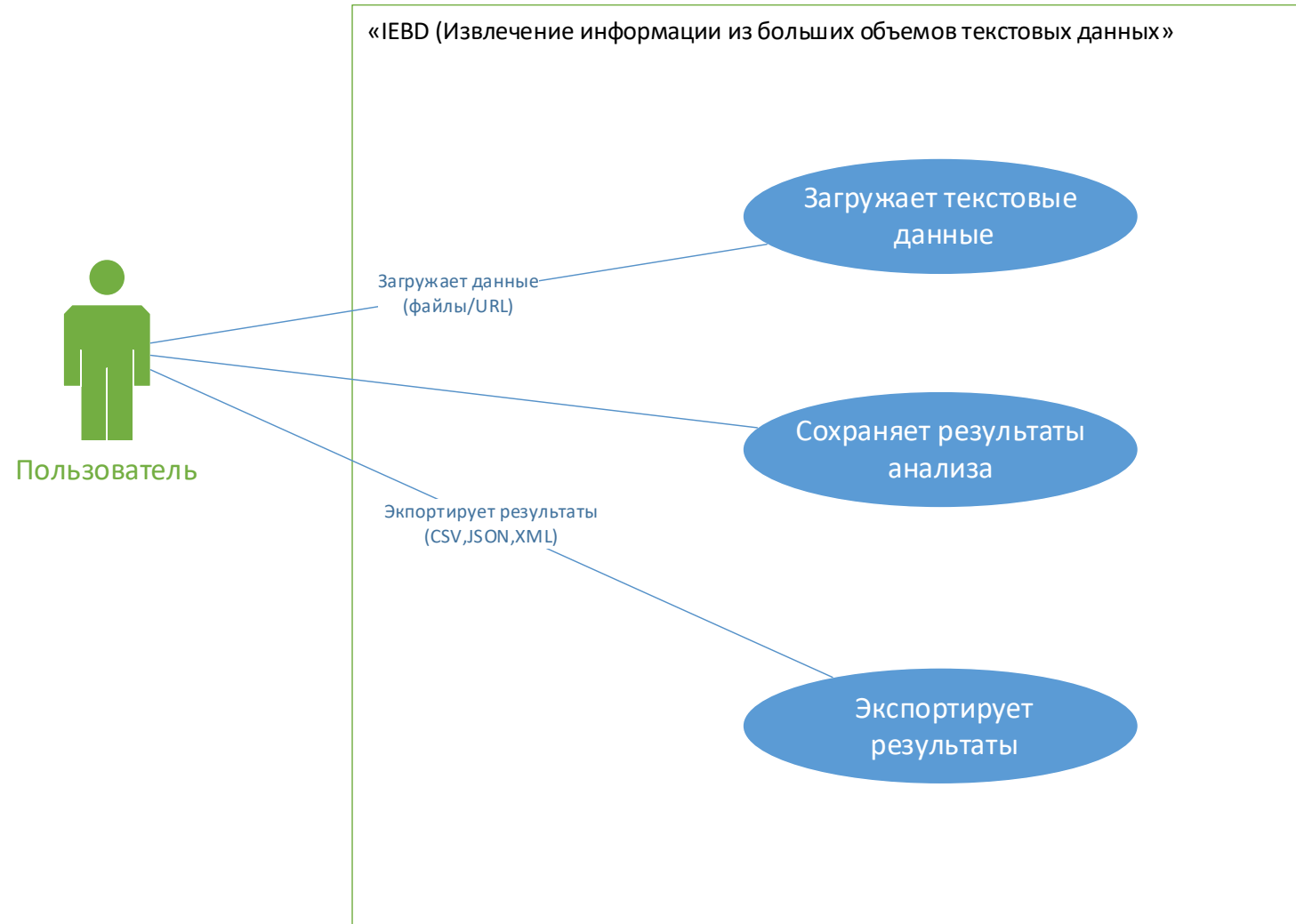
## Анализ предметной области

Извлечение информации из больших объемов текстовых данных (IEBD - Information Extraction from Big Data )— это процесс автоматического извлечения структурированной информации из неструктурированных или полуструктурированных текстовых данных.



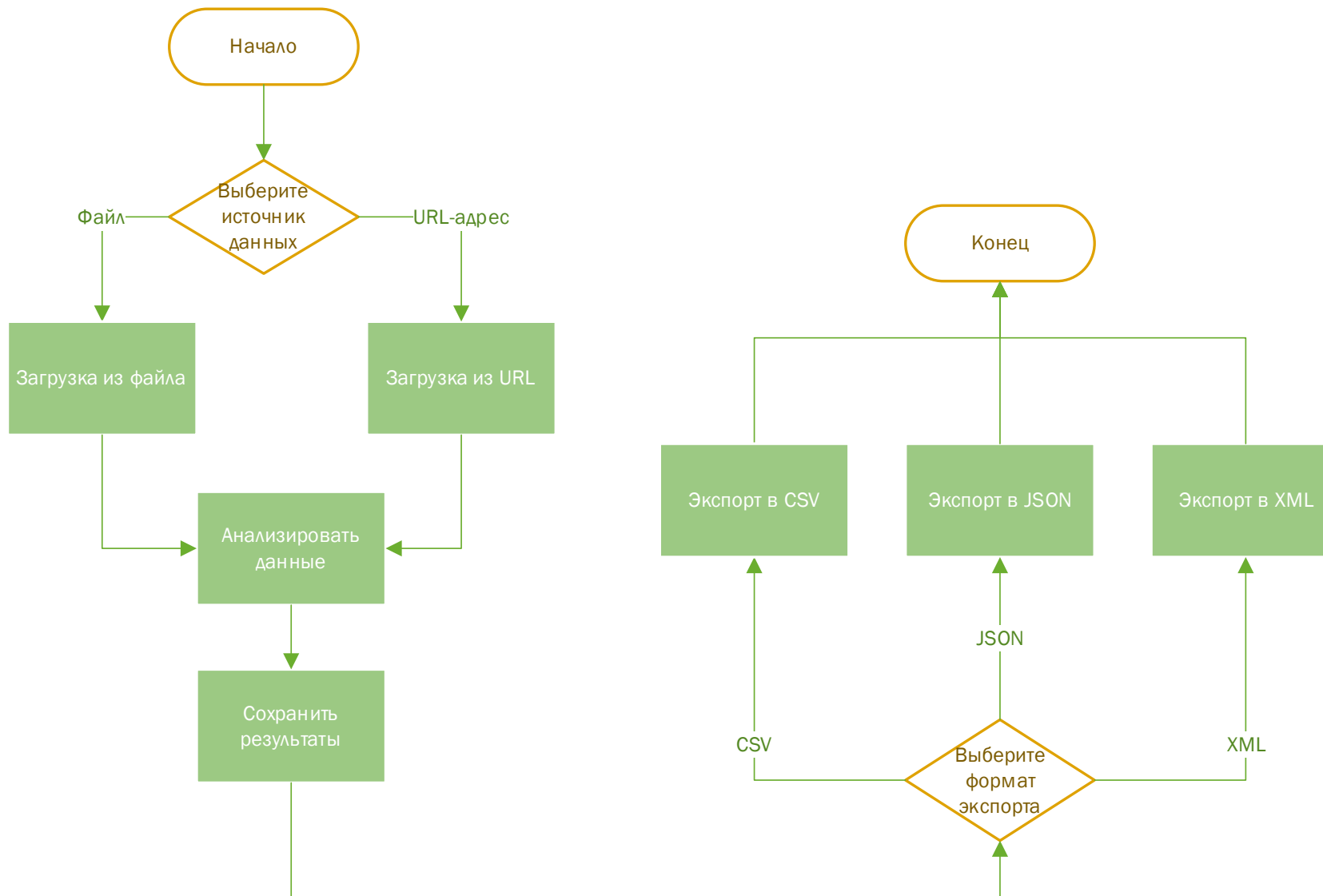
# РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ

## Диаграмма вариантов использования



РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ  
БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ

## Алгоритм работы программы

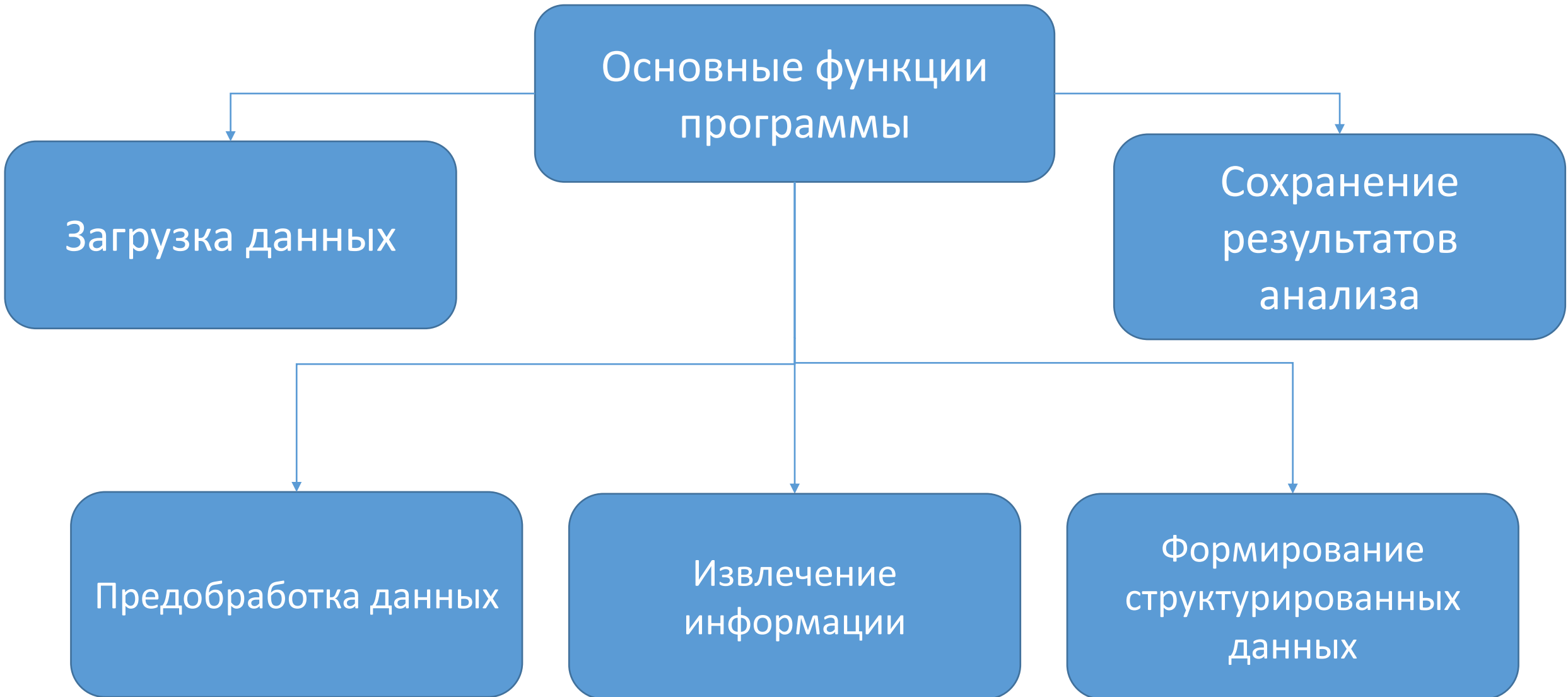


## Средства разработки

1. Среда разработки Visual Studio 2022
2. Языка программирования C#.
3. Windows Presentation Foundation (WPF)



РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ  
БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ





**РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ  
БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ**

Демонстрация программы

# РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ

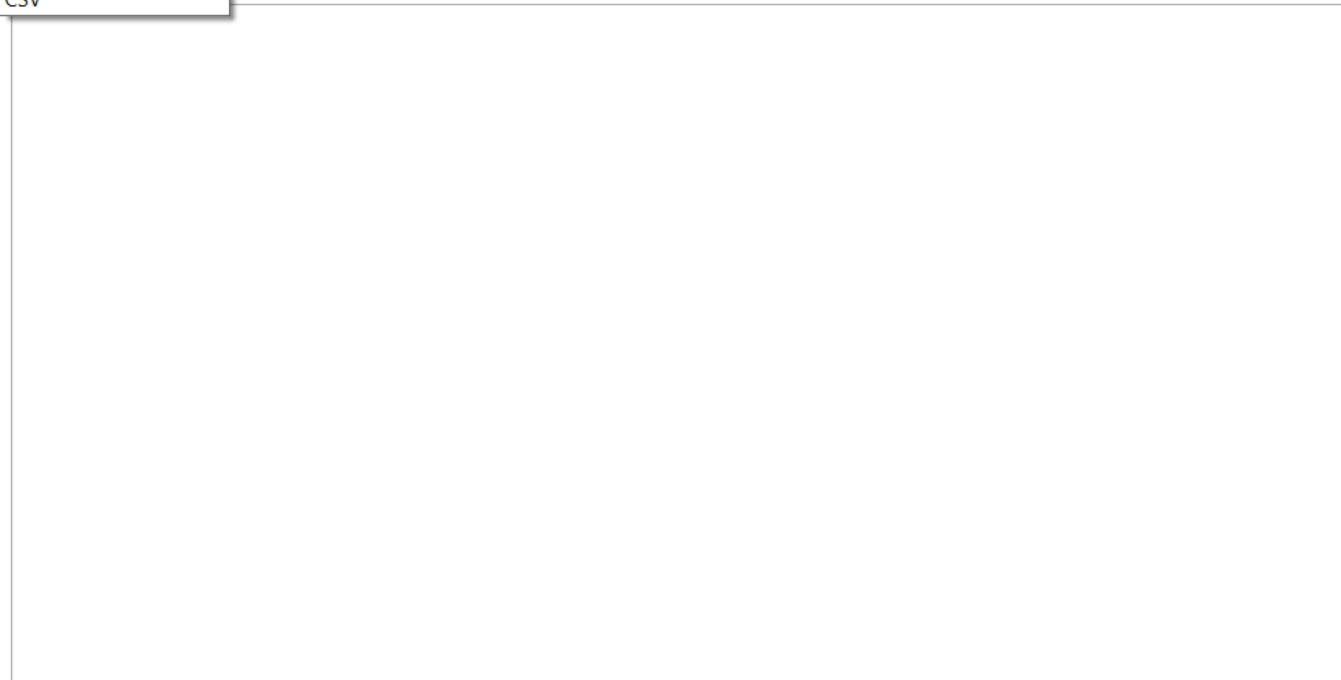
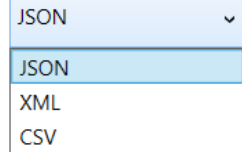
## Главное окно

Введите URL-адрес или вставьте текст



Отправить

Сохранить файл





```
← ↻ 🔒 api.open-meteo.com api.open-meteo.com/v1/forecast?latitude=52.52&longitude=13.41&hourly=temperature_2m
Автоформатировать ☐
{"latitude":52.52,"longitude":13.419998,"generationtime_ms":0.034928321838378906,"utc_offset_seconds":0,"timezone":"GMT","timezone_abbreviation":"GMT","elevation":38.0,"hourly_units":
{"time":"iso8601","temperature_2m":"°C"},"hourly":{"time":["2025-02-07T00:00","2025-02-07T01:00","2025-02-07T02:00","2025-02-07T03:00","2025-02-07T04:00","2025-02-07T05:00","2025-02-
07T06:00","2025-02-07T07:00","2025-02-07T08:00","2025-02-07T09:00","2025-02-07T10:00","2025-02-07T11:00","2025-02-07T12:00","2025-02-07T13:00","2025-02-07T14:00","2025-02-07T15:00","2025-02-07T16:00","2025-02-07T17:00","2025-02-
07T18:00","2025-02-07T19:00","2025-02-07T20:00","2025-02-07T21:00","2025-02-07T22:00","2025-02-07T23:00","2025-02-08T00:00","2025-02-08T01:00","2025-02-08T02:00","2025-02-08T03:00","2025-02-08T04:00","2025-02-
08T05:00","2025-02-08T06:00","2025-02-08T07:00","2025-02-08T08:00","2025-02-08T09:00","2025-02-08T10:00","2025-02-08T11:00","2025-02-08T12:00","2025-02-08T13:00","2025-02-08T14:00","2025-02-08T15:00","2025-02-
08T16:00","2025-02-08T17:00","2025-02-08T18:00","2025-02-08T19:00","2025-02-08T20:00","2025-02-08T21:00","2025-02-08T22:00","2025-02-08T23:00","2025-02-09T00:00","2025-02-09T01:00","2025-02-09T02:00","2025-02-
09T03:00","2025-02-09T04:00","2025-02-09T05:00","2025-02-09T06:00","2025-02-09T07:00","2025-02-09T08:00","2025-02-09T09:00","2025-02-09T10:00","2025-02-09T11:00","2025-02-09T12:00","2025-02-09T13:00","2025-02-09T14:00","2025-02-09T15:00","2025-02-09T16:00","2025-02-09T17:00","2025-02-09T18:00","2025-02-09T19:00","2025-02-09T20:00","2025-02-09T21:00","2025-02-09T22:00","2025-02-09T23:00","2025-02-10T00:00","2025-02-
10T01:00","2025-02-10T02:00","2025-02-10T03:00","2025-02-10T04:00","2025-02-10T05:00","2025-02-10T06:00","2025-02-10T07:00","2025-02-10T08:00","2025-02-10T09:00","2025-02-10T10:00","2025-02-10T11:00","2025-02-
10T12:00","2025-02-10T13:00","2025-02-10T14:00","2025-02-10T15:00","2025-02-10T16:00","2025-02-10T17:00","2025-02-10T18:00","2025-02-10T19:00","2025-02-10T20:00","2025-02-10T21:00","2025-02-10T22:00","2025-02-
10T23:00","2025-02-11T00:00","2025-02-11T01:00","2025-02-11T02:00","2025-02-11T03:00","2025-02-11T04:00","2025-02-11T05:00","2025-02-11T06:00","2025-02-11T07:00","2025-02-11T08:00","2025-02-11T09:00","2025-02-11T10:00","2025-02-11T11:00","2025-02-11T12:00","2025-02-11T13:00","2025-02-11T14:00","2025-02-11T15:00","2025-02-11T16:00","2025-02-11T17:00","2025-02-11T18:00","2025-02-11T19:00","2025-02-11T20:00","2025-02-11T21:00","2025-02-11T22:00","2025-02-11T23:00","2025-02-12T00:00","2025-02-12T01:00","2025-02-12T02:00","2025-02-12T03:00","2025-02-12T04:00","2025-02-12T05:00","2025-02-12T06:00","2025-02-12T07:00","2025-02-12T08:00","2025-02-12T09:00","2025-02-12T10:00","2025-02-12T11:00","2025-02-12T12:00","2025-02-12T13:00","2025-02-12T14:00","2025-02-12T15:00","2025-02-12T16:00","2025-02-12T17:00","2025-02-12T18:00","2025-02-12T19:00","2025-02-12T20:00","2025-02-12T21:00","2025-02-12T22:00","2025-02-12T23:00","2025-02-13T00:00","2025-02-13T01:00","2025-02-13T02:00","2025-02-13T03:00","2025-02-13T04:00","2025-02-13T05:00","2025-02-13T06:00","2025-02-13T07:00","2025-02-13T08:00","2025-02-13T09:00","2025-02-13T10:00","2025-02-13T11:00","2025-02-13T12:00","2025-02-13T13:00","2025-02-13T14:00","2025-02-13T15:00","2025-02-13T16:00","2025-02-13T17:00","2025-02-13T18:00","2025-02-13T19:00","2025-02-13T20:00","2025-02-13T21:00","2025-02-13T22:00","2025-02-13T23:00"],"temperature_2m":
[3.6,3.4,2.9,2.4,2.0,2.1,2.6,2.2,1.9,1.7,2.0,2.9,3.6,3.8,3.8,4.0,3.4,3.1,2.5,2.3,1.9,1.4,0.6,0.1,-0.3,-0.5,-0.7,-0.9,-1.0,-1.1,-1.1,-1.2,-1.1,-0.4,0.8,2.1,3.5,4.5,4.9,4.7,3.9,3.2,2.5,1.8,1.3,0.9,0.7,0.5,0.2,0.0,-0.
2,-0.3,-0.3,-0.4,-0.5,-0.7,-0.7,-0.2,0.3,1.2,2.0,2.7,3.2,3.2,2.6,1.7,1.1,0.7,0.5,0.3,0.1,0.0,0.1,0.1,0.2,0.2,-0.0,-0.2,-0.4,-0.7,-0.9,-0.8,-0.3,0.4,1.1,1.4,1.5,1.4,1.1,0.5,0.0,-0.4,-0.7,-1.1,-1.4,-1.8,-2.1,-2.4,-2.
7,-3.0,-3.2,-3.4,-3.5,-3.3,-3.1,-2.5,-1.5,-0.3,0.6,1.1,1.3,1.3,0.9,0.3,-0.3,-0.7,-1.2,-1.7,-2.4,-3.1,-3.7,-3.9,-3.9,-3.9,-3.9,-3.9,-4.6,-4.5,-4.1,-3.5,-2.6,-1.9,-1.4,-1.0,-0.8,-1.1,-1.6,-2.1,-2.4,-2.7,-3.0,-3.
3,-3.5,-3.8,-4.0,-4.2,-4.4,-4.7,-4.9,-4.9,-4.7,-4.3,-3.7,-2.6,-1.4,-0.3,0.3,0.6,0.8,0.7,0.5,0.3,-0.0,-0.3,-0.5,-0.5,-0.5]}}
```

ive\_humidity\_2m,dew\_point\_2m&models=icon\_seamless

C:\Users\vip66\Downloads\jsontoxml (1).xml

C:\Users\vip66\Downloads\open-meteo.csv

JSON



```
{
  "latitude": 52.52,
  "longitude": 13.419998,
  "generationtime_ms": 0.095248222351074219,
  "utc_offset_seconds": 0,
  "timezone": "GMT",
  "timezone_abbreviation": "GMT",
  "elevation": 38.0,
  "hourly_units": {
    "time": "iso8601",
    "temperature_2m": "°C",
    "relative_humidity_2m": "%",
    "dew_point_2m": "°C"
  },
  "hourly": {
    "time": [
      "2025-02-08T00:00",
```

JSON



```
{
  "root": {
    "latitude": "52.52",
    "longitude": "13.419998",
    "generationtime_ms": "0.034928321838378906",
    "utc_offset_seconds": "0",
    "timezone": "GMT",
    "timezone_abbreviation": "GMT",
    "elevation": "38",
    "hourly_units": {
      "time": "iso8601",
      "temperature_2m": "°C"
    },
    "hourly": {
      "time": [
        "2025-02-07T00:00",
        "2025-02-07T01:00",
```

JSON



```
[
  {
    "latitude": "52.52",
    "longitude": "13.419998",
    "elevation": "38.0",
    "utc_offset_seconds": "0",
    "timezone": "GMT",
    "timezone_abbreviation": "GMT"
  },
  {
    "latitude": "time",
    "longitude": "temperature_2m (°C)",
    "elevation": null,
    "utc_offset_seconds": null,
    "timezone": null,
    "timezone_abbreviation": null
  },
]
```

**РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ  
БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ**

**Спасибо за внимание!**