



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 разработке 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

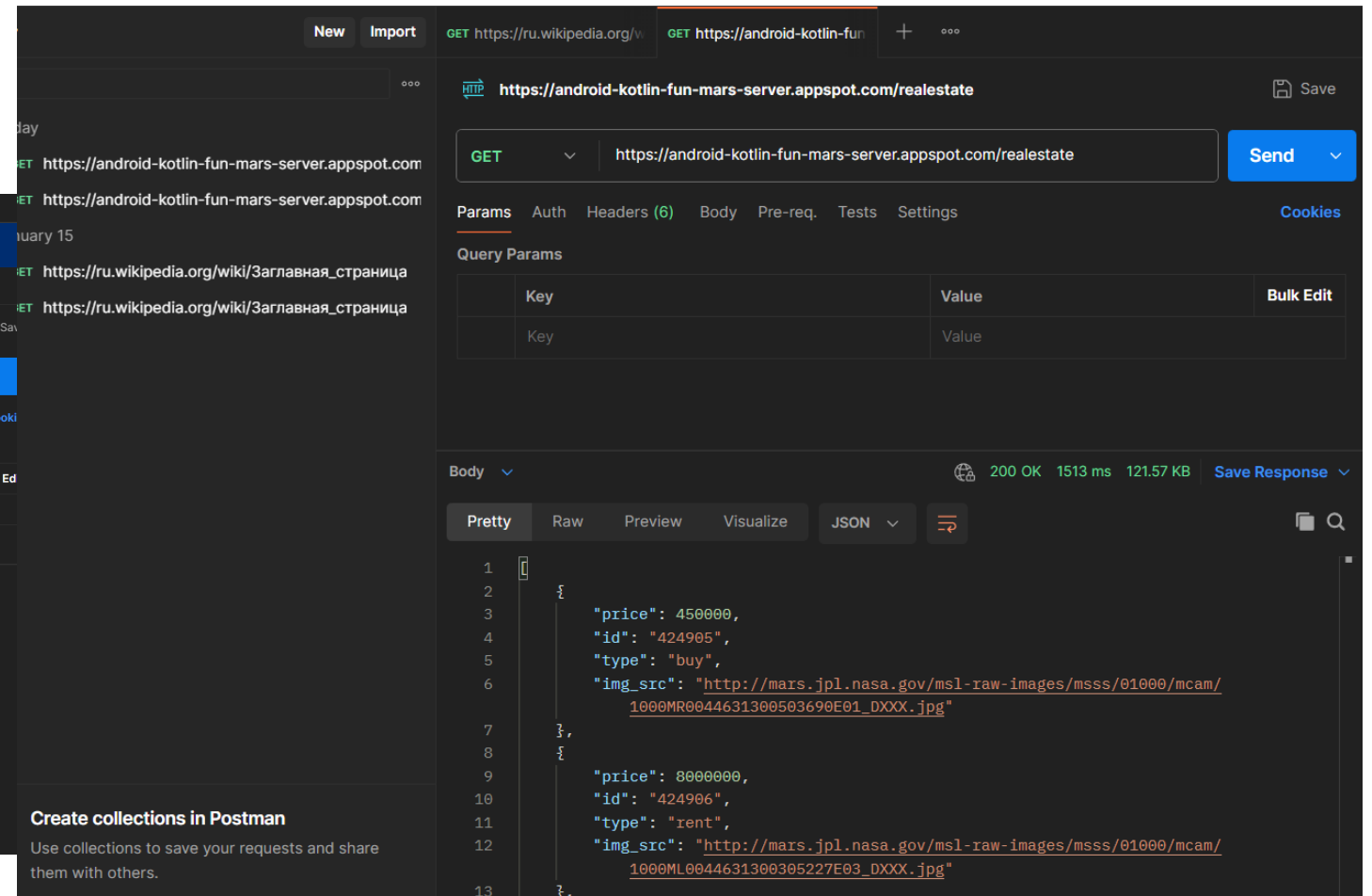
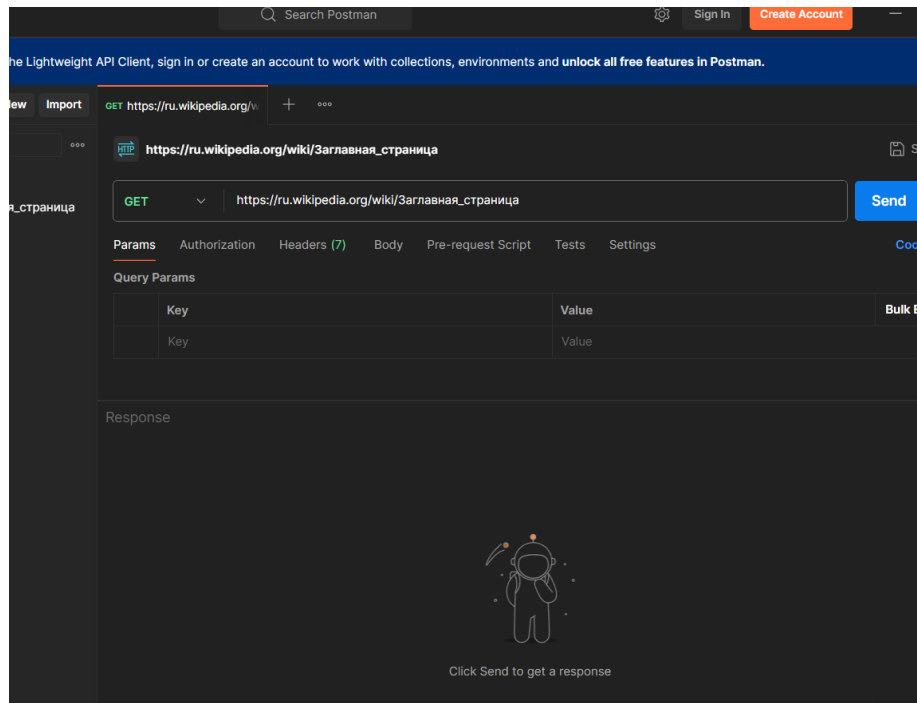
- Описать предметную область
- Разработать техническое задание на создание программного продукта
- Описать архитектуру программы
- Описать алгоритмы и функционирование программы
- Провести тестирование и опытную эксплуатацию
- Разработать руководство оператора

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

## Программа аналог



# Postman



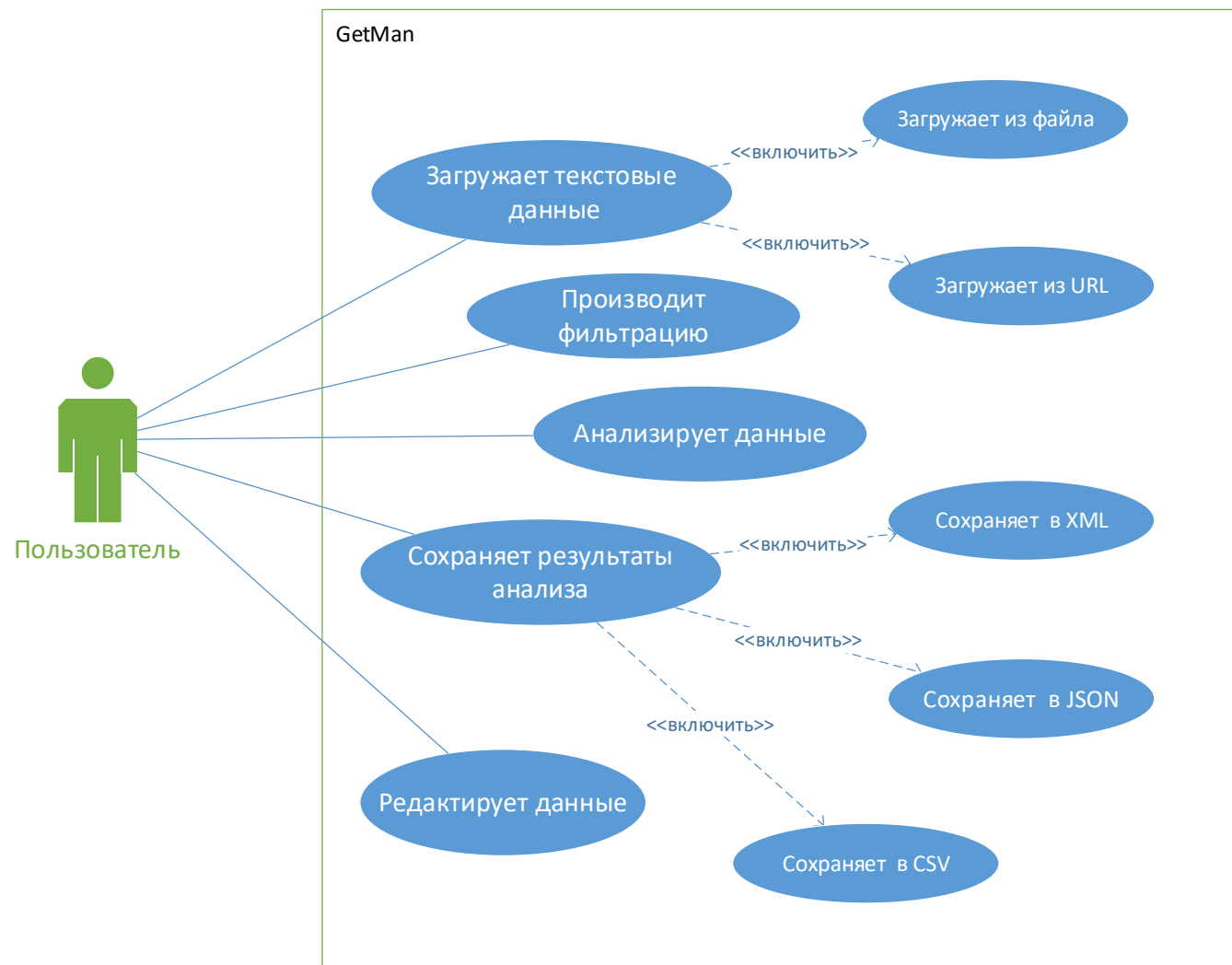
## Анализ предметной области

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.



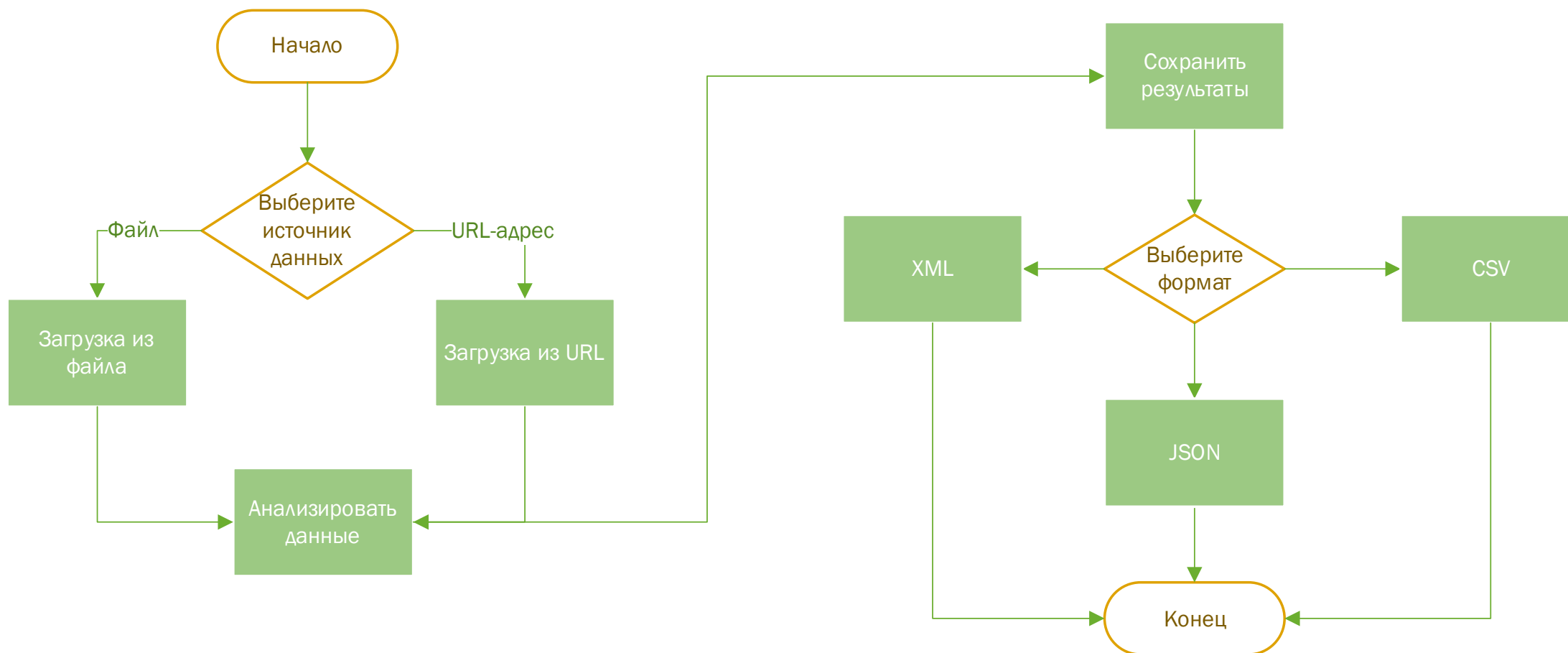
# РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ

## Диаграмма вариантов использования



РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ  
БОЛЬШИХ ОБЪЕМОВ ТЕКСТОВЫХ ДАННЫХ

## Алгоритм работы программы

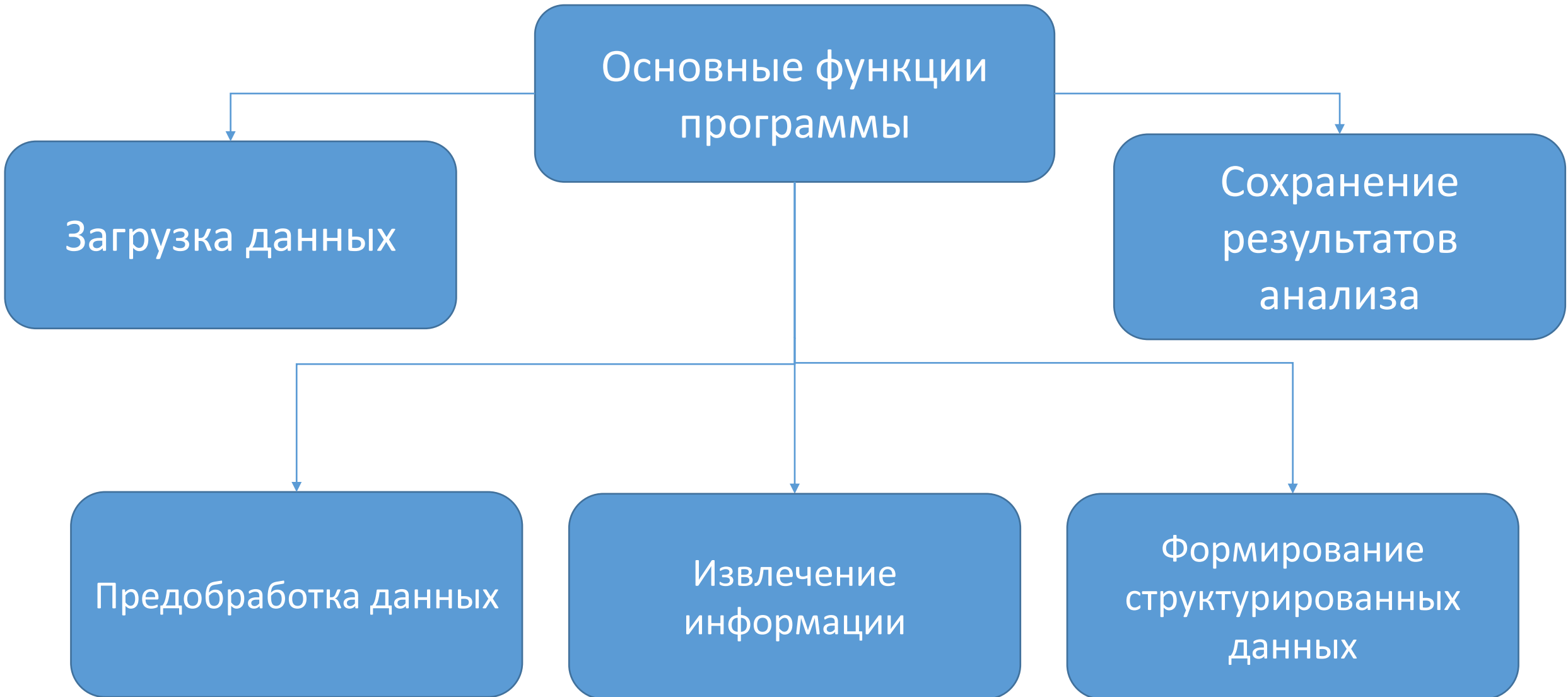


## Средства разработки

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

GetMan

Введите URL-адрес или вставьте текст

Отправить

JSON

Редактировать

Сохранить файл

Поиск

Сортировать по

(↑) (↓)

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

## Окно редактирования

Редактирование данных

<root>

<latitude>52.52</latitude>

<longitude>13.419998</longitude>

<generationtime\_ms>0.034928321838378906</generationtime\_ms>

<utc\_offset\_seconds>0</utc\_offset\_seconds>

<timezone>GMT</timezone>

<timezone\_abbreviation>GMT</timezone\_abbreviation>

<elevation>38</elevation>

<hourly\_units>

<time>iso8601</time>

<temperature\_2m>°C</temperature\_2m>

</hourly\_units>

<hourly>

← →

СохранитьОтменить

## Неструктурированные данные в CSV

**Автосохранение** **5mb.csv** **Поиск**

**Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Справка**

**Отмена** **Буфер обмена** **Шрифт** **Выравнивание** **Число**

**A1** **X** **✓** **fx** title,author,genre,year,publisher

	A	B	C	D	E	F	G	H	I	J	K	L
1	title,author,genre,year,publisher											
2	Voluptas veritatis et nam.,Golda Bailey,dignissimos,1992,https://examplefile.com											
3	Sapiente ea quae.,Hiram Cummerata,beatae,2023,https://examplefile.com											
4	Autem id consequatur voluptas.,Antonietta Goodwin,sit,2008,https://examplefile.com											
5	Voluptate dolore nemo.,Darrion Lind,est,2007,https://examplefile.com											
6	Sint asperiores ullam.,Dr. Westley Gottlieb Sr.,deleniti,2011,https://examplefile.com											
7	Repellendus ullam.,Asha Russel,dolores,2013,https://examplefile.com											
8	Eveniet rerum soluta.,Delia Denesik,omnis,1995,https://examplefile.com											
9	Eos repudiandae commodi.,Norval Hickles,aspernatur,2005,https://examplefile.com											
10	Perspiciatis dolores facilis quae.,Dr. Sofia Huels,esse,1983,https://examplefile.com											
11	Numquam corporis sapiente non error.,Janick Terry,quia,2005,https://examplefile.com											
12	Et nobis veritatis.,Nicolette Bashirian Sr.,est,2007,https://examplefile.com											
13	Maxime ut voluptatem.,Edgar Quigley DDS,incididunt,1998,https://examplefile.com											
14	Voluptas similique a omnis.,Odessa Schaefer,ipsa,1980,https://examplefile.com											
15	Nisi dolorum corrupti.,Juanita Douglas,dolor,1983,https://examplefile.com											
16	Ea quos fugiat sit.,Lilyan Braun Sr.,quam,1978,https://examplefile.com											
17	Hic beatae illum iusto et.,Naomi Johnston,culpa,2011,https://examplefile.com											
18	Aut maxime ad voluptas.,Branson Schinner,animi,2007,https://examplefile.com											
19	Ipsa excepturi dolores.,Marquise O'Conner,iste,1985,https://examplefile.com											

< > **5mb** +

Готово

# Неструктурированные данные в JSON

```
5mb.json
Файл  Изменить  Просмотр

[ { "title": "Voluptas veritatis et nam.", "author": "Golda Bailey", "genre": "dignissimos", "year": "1992", "publisher":
"https://examplefile.com" }, { "title": "Sapiente ea quae.", "author": "Hiram Cummerata", "genre": "beatae", "year": "2023", "publisher":
"https://examplefile.com" }, { "title": "Autem id consequatur voluptas.", "author": "Antonietta Goodwin", "genre": "sit", "year": "2008",
"publisher": "https://examplefile.com" }, { "title": "Voluptate dolore nemo.", "author": "Darrion Lind", "genre": "est", "year": "2007",
"publisher": "https://examplefile.com" }, { "title": "Sint asperiores ullam.", "author": "Dr. Westley Gottlieb Sr.", "genre": "deleniti", "year":
"2011", "publisher": "https://examplefile.com" }, { "title": "Repellendus ullam.", "author": "Asha Russel", "genre": "dolores", "year": "2013",
"publisher": "https://examplefile.com" }, { "title": "Eveniet rerum soluta.", "author": "Delia Denesik", "genre": "omnis", "year": "1995",
"publisher": "https://examplefile.com" }, { "title": "Eos repudiandae commodi.", "author": "Norval Hickie", "genre": "aspernatur", "year": "2005",
"publisher": "https://examplefile.com" }, { "title": "Perspiciatis dolores facilis quae.", "author": "Dr. Sofia Huels", "genre": "esse", "year":
"1983", "publisher": "https://examplefile.com" }, { "title": "Numquam corporis sapiente non error.", "author": "Janick Terry", "genre": "quia",
"year": "2005", "publisher": "https://examplefile.com" }, { "title": "Et nobis veritatis.", "author": "Nicolette Bashirian Sr.", "genre": "est",
"year": "2007", "publisher": "https://examplefile.com" }, { "title": "Maxime ut voluptatem.", "author": "Edgar Quigley DDS", "genre": "incididunt",
"year": "1998", "publisher": "https://examplefile.com" }, { "title": "Voluptas similique a omnis.", "author": "Odessa Schaefer", "genre": "ipsa",
"year": "1980", "publisher": "https://examplefile.com" }, { "title": "Nisi dolorum corrupti.", "author": "Juanita Douglas", "genre": "dolor",
"year": "1983", "publisher": "https://examplefile.com" }, { "title": "Ea quos fugiat sit.", "author": "Lilyan Braun Sr.", "genre": "quam",
"year": "1978", "publisher": "https://examplefile.com" }, { "title": "Hic beatae illum iusto et.", "author": "Naomi Johnston", "genre": "culpa",
"year": "2011", "publisher": "https://examplefile.com" }, { "title": "Aut maxime ad voluptas.", "author": "Branson Schinner", "genre": "animi",
"year": "2007", "publisher": "https://examplefile.com" }, { "title": "Ipsa excepturi dolores.", "author": "Marquise O'Conner", "genre": "iste",
"year": "1985", "publisher": "https://examplefile.com" }, { "title": "Non dolore aut labore.", "author": "Marie Emmerich", "genre": "voluptas",
"year": "1991", "publisher": "https://examplefile.com" }, { "title": "Dolores et est corporis.", "author": "Prof. Tess Wisozk DVM", "genre":
"facilis", "year": "2019", "publisher": "https://examplefile.com" }, { "title": "Mollitia illum adipisci et officia.", "author": "Eleanora
Gorczy", "genre": "aliquam", "year": "1997", "publisher": "https://examplefile.com" }, { "title": "Laborum deserunt.", "author": "Mrs.
Assunta Dietrich Jr.", "genre": "tempore", "year": "1992", "publisher": "https://examplefile.com" }, { "title": "Provident sunt harum.",
"author": "Prof. Clyde Bernhard PhD", "genre": "repudiandae", "year": "1975", "publisher": "https://examplefile.com" }, { "title": "Architecto
est.", "author": "Ken Wolff V", "genre": "sed", "year": "2000", "publisher": "https://examplefile.com" }, { "title": "Libero quas aut ipsam.",
"author": "Freddy Corkery", "genre": "quasi", "year": "1982", "publisher": "https://examplefile.com" }, { "title": "Illum alias est odio.",
"author": "Dayton Konopelski", "genre": "ea", "year": "1986", "publisher": "https://examplefile.com" }, { "title": "Cupiditate alias praesentium
fuga.", "author": "Kaylie Yost DDS", "genre": "et", "year": "1977", "publisher": "https://examplefile.com" }, { "title": "Aliquid et
voluptas.", "author": "Prof. Hipolito Bernhard", "genre": "occaecati", "year": "2004", "publisher": "https://examplefile.com" }, { "title":
"Esse assumenda occaecati omnis.", "author": "Dr. Vernie Sauer", "genre": "vel", "year": "1992", "publisher": "https://examplefile.com" },
{ "title": "Ut et.", "author": "Mrs. Clara Hoppe", "genre": "et", "year": "1984", "publisher": "https://examplefile.com" }, { "title": "Est
est et atque.", "author": "Tatyana Yost", "genre": "quo", "year": "1972", "publisher": "https://examplefile.com" }, { "title": "Odio optio et
officia.", "author": "Mr. Terrill Jacobi Jr.", "genre": "dignissimos", "year": "1989", "publisher": "https://examplefile.com" }, { "title": "Ut
dolorum nulla consequatur.", "author": "Aisha Leannon", "genre": "nam", "year": "1970", "publisher": "https://examplefile.com" }, { "title":
"Fugiat repellat unde.", "author": "Tatyana Brekke I", "genre": "qui", "year": "2001", "publisher": "https://examplefile.com" }, { "title":
"Sed ipsam consequatur.", "author": "Krystal Murphy Jr.", "genre": "qui", "year": "2009", "publisher": "https://examplefile.com" }, { "title":

Строка 1, столбец 1 | 4 506 035 символов | 100% | Windows (CRLF) | UTF-8
```

# Неструктурированные данные в XML

```
←  ↻  🔒  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]]}}
```

GetMan

Введите URL-адрес или вставьте текст  
C:\Users\vip66\OneDrive\Рабочий стол\файл\5mb.json

XML

✎ Редактировать

Сохранить файл

Поиск

title

(1)

(1)

<root>

<root>

<title>Voluptas veritatis et nam.</title>

<author>Golda Bailey</author>

<genre>dignissimos</genre>

<year>1992</year>

<publisher>https://examplefile.com</publisher>

</root>

<root>

<title>Sapiente ea quae.</title>

<author>Hiram Cummerata</author>

Введите URL-адрес или вставьте текст

C:\Users\vip66\OneDrive\Рабочий стол\dip\5mb.xml

Отправить

CSV

Редактировать

Сохранить файл

Поиск

title

(1)

(4)

title	author	genre	year	publisher
Voluptas veritatis et nam.	Golda Bailey	dignissimos	1992	https://examplefile.com
Sapiente ea quae.	Hiram Cummerata	beatae	2023	https://examplefile.com
Autem id consequatur voluptas.	Antonietta Goodwin	sit	2008	https://examplefile.com
Voluptate dolore nemo.	Darrion Lind	est	2007	https://examplefile.com
Sint asperiores ullam.	Dr. Westley Gottlieb Sr.	deleniti	2011	https://examplefile.com
Repellendus ullam.	Asha Russel	dolores	2013	https://examplefile.com
Eveniet rerum soluta.	Delia Denesik	omnis	1995	https://examplefile.com
Eos repudiandae commodi.	Norval Hickie	aspernatur	2005	https://examplefile.com
Perspiciatis dolores facilis quae.	Dr. Sofia Huels	esse	1983	https://examplefile.com
Numquam corporis sapiente non error.	Janick Terry	quia	2005	https://examplefile.com
Et nobis veritatis.	Nicolette Bashirian Sr.	est	2007	https://examplefile.com

Введите URL-адрес или вставьте текст

C:\Users\vip66\OneDrive\Рабочий стол\файл\5mb.csv

Отправить

JSON

Редактировать

Сохранить файл

Поиск

title

(1)

(4)

```
[
  {
    "title": "Voluptas veritatis et nam.",
    "author": "Golda Bailey",
    "genre": "dignissimos",
    "year": "1992",
    "publisher": "https://examplefile.com"
  },
  {
    "title": "Sapiente ea quae.",
    "author": "Hiram Cummerata",
```



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

**Спасибо за внимание!**