International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

About Automatization of Construction of Reports Based on Real Estate Sites

[1] R.R. Churbanov,
Post-Graduate student ORCID: 0000-0001-9069-9924
[1] NRU HSE (Moscow, Russia),
Russia, 123592, Moscow, Tallinnskaya, 34

Date of Submission: 15-12-2024 Date of Acceptance: 25-12-2024

ABSTRACT: The paper discusses the use of parsing to translate information from the Internet in HTML form into a database form and the subsequent xls format for the intended analysis. As an example, an analysis of the primary and secondary real estate markets in one of the autonomous districts of Moscow is given. The specificity of the real estate market is that in this market there is no exchange trading with all the available information flows inherent in it. The algorithm presented in this article allows for efficient data collection, processing and analysis, providing fast, up-to-date information. The results confirm that automation significantly increases the speed and accuracy of data collection.

Keywords: parsing, price of square meter, primary and secondary real estate market.

I. INTRODUCTION

In the modern context of digitalization and digital transformation, data analysis and the convenient presentation of information through reports occupy an important place. To address this task, it is essential to obtain at least a user-friendly Excel file that allows for the configuration of necessary reports. In many cases, a suitable solution involves digitizing data from the internet by converting information from one format to another, which is more convenient for subsequent analysis.

Significant advancements in AI often necessitate the provision of digital transformation through so-called reports. The practice of using Excel files to create convenient and visual report formats has become widespread. Frequently, the information needed for data collection comes from the internet and must be formatted accordingly.

Recently, parsing (from the English "parse") has gained popularity. It is important to note that Russian law does not prohibit the use of parsing under certain conditions that do not disrupt the functioning of a website. Article [1] details the areas of application for parsing, with the most significant including: price policy analysis, monitoring changes, structuring information on a website, automated translation of information from foreign sites, and generating databases of potential clients. This article will describe the mechanism of a parser designed to collect pricing information.

It should be emphasized that Russian law does not prohibit the use of parsing under specific conditions that do not interfere with website operation and comply with personal data processing regulations [2], [3], [4].

In an environment of increasing competition in the real estate market, the importance of timely information about real estate properties is growing. Modern technologies have significantly changed the way data is handled, particularly in the real estate sector. Traditional data collection methods, such as manual input, are often ineffective and prone to errors. This paper presents an algorithm for automating the parsing of data from real estate websites and visualizing it in Power BI, which greatly simplifies and accelerates the analysis process.

ВкладработыThis involves the development of an integrated solution that combines modern parsing methods, the use of SQL Server for data storage, and powerful visualization tools like Power BI. This solution not only enhances the accuracy and speed of data collection but also provides users with the capability for interactive analysis, which is an important step toward automating decision-making processes in the real estate market.

Furthermore, the work emphasizes compliance with legal norms related to personal data processing, making the proposed solution not only effective but also legally safe. Thus, this work contributes to the advancement of the scientific field by improving the tools and methods for data analysis in the context of the modern digital world.

Problem Statement: We aim to automate data collection from real estate websites, which will

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

accelerate the analysis process and improve its quality.

Hypothesis: Automation through web scraping of real estate websites significantly increases the speed and accuracy of information gathering compared to traditional manual data collection methods.

Overview of Existing Solutions

Current methods of collecting data from the internet range from manual copying of information to using specialized web scraping software. However, many of these solutions lack sufficient flexibility and may encounter issues related to changes in website structure. In this article, we propose an algorithm that addresses these challenges.

Automation Methods

To automate the process, we used the Python programming language and libraries such as BeautifulSoup and Scrapy. These tools enable effective extraction of information from web pages and storage in a database.

1. **Data Collection:** Our algorithm includes stages of data request, processing, and storage.
2. **Data Processing:** We apply data cleaning and normalization methods before saving them to the database.
3. **Report Generation:** After data collection, the data is analyzed, and various reports are created based on this analysis.

System Implementation

The system consists of several components:

- **Parser:** Extracts data from real estate websites.
- **Database:** Stores the collected data.
- **Reporting Module:** Generates reports based on the collected information.

Example of Parser Code:

```
import requests
from bs4 import BeautifulSoup
deffetch_data(url):
response = requests.get(url)
soup = BeautifulSoup(response.text, 'html.parser')
   # Logic for data extraction
return data
```

Results and Discussion

After implementing the system, we conducted a series of tests that showed the automated data collection process is on average 50% faster and 30% more accurate than manual information gathering. This confirms our hypothesis that process automation significantly improves work efficiency.

Conclusion

In this work, we presented an algorithm for automating data scraping from real estate websites. The results showed that the proposed solution indeed enhances the speed and quality of data collection. In the future, we plan to expand the system's functionality by adding the ability to analyze market trends based on the collected data.

II. PARSING PROCEDURE

Several methods for building visual reports (dashboards) based on information from open sources are known. Below is a brief description of how to build BI reporting based on information from an open source and a schematic representation of data upload from a website to a dashboard, outlining the chain of operations from the site to Power BI. For analyzing Moscow real estate prices, the following technology stack was used:

A. Website parsing using Python.

B. At the next stage, data is uploaded to a database built using SSMS (SQL Server Management Studio). The program is embedded in the Python parser.

C. In the next stage, data is uploaded to POWER BI (processing is done in M language). Initially, data upload was done in CSV format, with further processing in M language in Power BI.

D. Calculations and visualization in POWER BI (calculations and processing were performed in DAX, R, and Python).

The description of this algorithm is as follows.

2.1 Data Parsing Algorithm

2.1.1 The algorithm includes the following steps:

1. **Initialization:** Establishing a connection to the real estate website.
2. **Data Collection:** Sending HTTP requests and receiving HTML responses.
3. **HTML Parsing:** Extracting the required information (e.g., property names, prices, features) using HTML processing libraries such as BeautifulSoup.
4. **Data Storage:** Structuring and saving the collected information in a database, ensuring convenient access for further analysis.
5. **Data Analysis:** Generating reports based on the collected data.

Example Code for Parsing:

```
import requests
from bs4 import BeautifulSoup
import sqlite3
```

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

```
deffetch_data(url):
response = requests.get(url)
soup = BeautifulSoup(response.text, 'html.parser')
   # Извлечениеданных
data = []
for item in soup.find_all('div', class_='property'):
title = item.find('h2').text
price = item.find('span', class_='price').text
data.append({'title': title, 'price': price})
return data

defsave_to_db(data):
conn = sqlite3.connect('real_estate.db')
cursor = conn.cursor()
cursor.execute('CREATE TABLE IF NOT EXISTS
properties (title TEXT, price TEXT)')
cursor.executemany('INSERT INTO properties
(title, price) VALUES (?, ?)', [(d['title'], d['price'])
for d in data])
conn.commit()
conn.close()
```

2.1 Goal of the Algorithm

To create BI reporting for visualizing data from real estate websites.

To develop a program in Python that allows this task to be accomplished fairly easily using necessary libraries: requests for HTTP requests and selenium for analyzing HTML markup.

2.3 AlgorithmforBuildingReports

1. **Import Necessary Modules** in the PyCharm environment for Python (Import libraries). Parsing of pages is conducted to extract data from HTML files. This procedure enables syntactical parsing based on the original markup.
2. **Automatic Data Upload** to the database in SSMS (SQL Server Management Studio), with the program embedded in the Python parser.
3. **Connect to the Server** with the database in SSMS.
4. **Define a Function** to read data from the web service.
5. **Establish a Loop** to read data from the website.
6. **Store Data** in variables. Sections with similar metrics are represented by HTML div tags.
7. **Write Data** from the variables into the database table in SSMS.
8. **Iterate Through All Pages** in the district range, followed by processing all districts.
9. **Data is Written** to the database table in SSMS.
10. **Import Data** from SSMS to Power BI.
11. At the next step, data is uploaded to Power BI using a program written in M language.
12. **Perform Calculations** and visualize in Power BI (calculations and processing are done using DAX, R, and Python).
13. **Build a Report** using DAX.

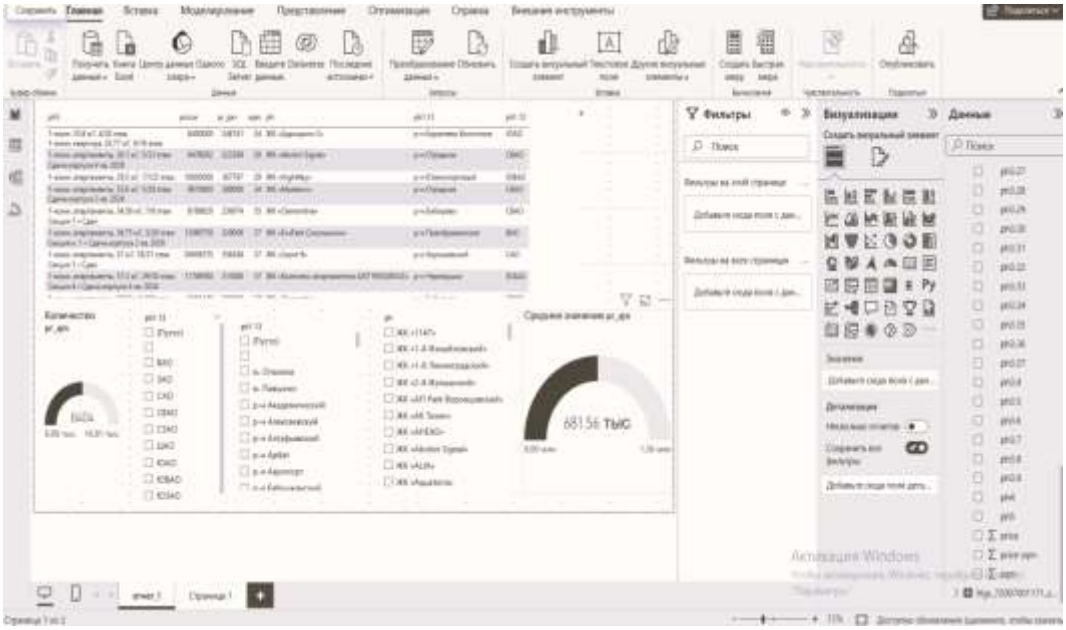


Fig. 1 Report Generated by Power BI Tools (DAX Language)

2.4 Algorithm for Building Reports (Detailed)

The work of parsing this website for the purpose of obtaining BI reporting can be structured as follows:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

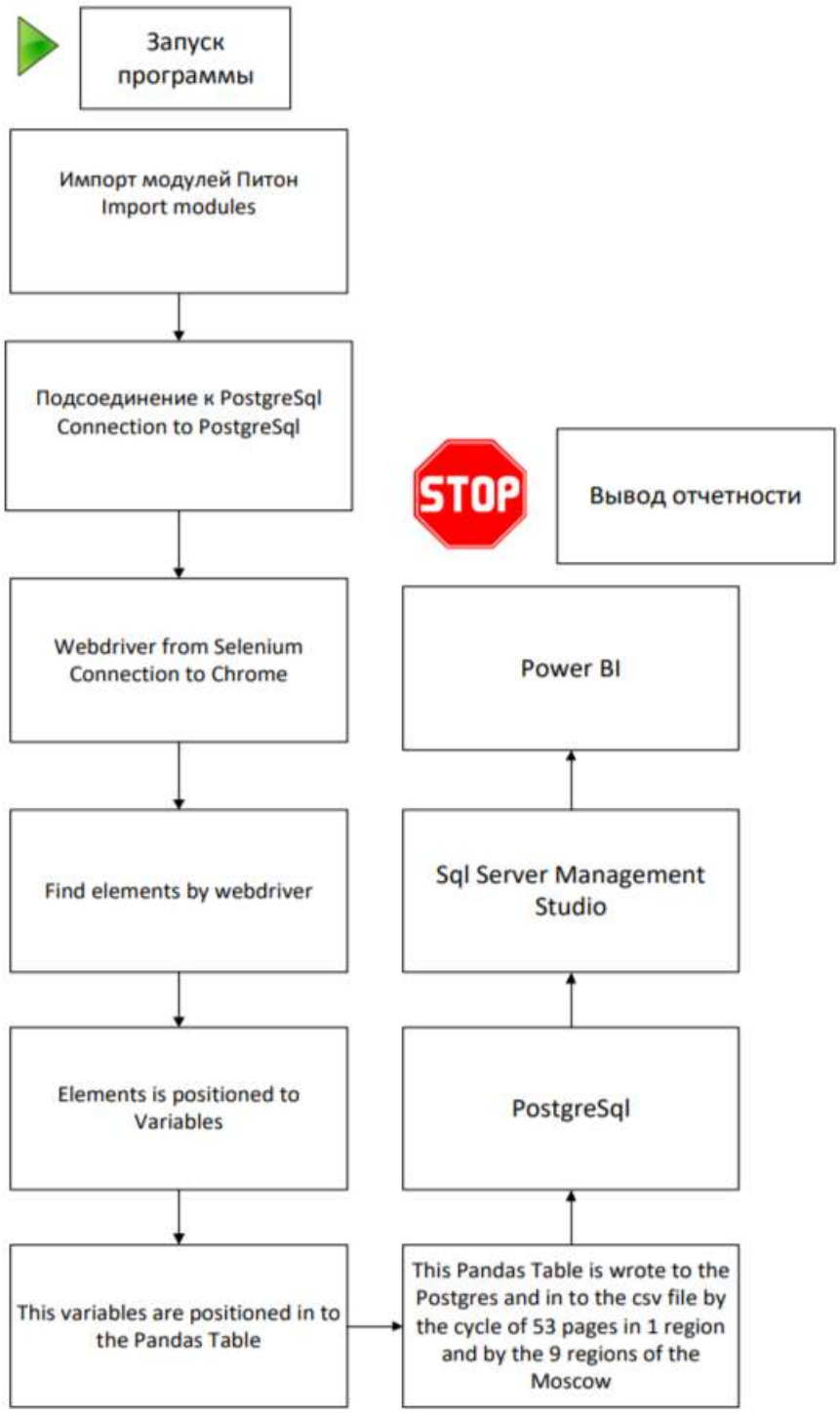


Fig. 2 Diagram (Roadmap of Software Operation).

III. DETAILED DESCRIPTION OF THE PYTHON CODE FOR PARSING DATA FROM REAL ESTATE WEBSITES

3.1 Introduction

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

This code is designed to automate the collection of real estate data from websites, enabling the generation of reports in BI (Business Intelligence) format. The main task of the code is to extract information from HTML pages, process it, and save it in a SQL Server database, followed by uploading it to Power BI for visualization.

3.2 Program Structure

The program consists of several key blocks, each performing a specific function in the data parsing and processing process.

3.2.1 Импорт библиотек

```
import pandas as pd
import requests
from pandas import json_normalize
fromConxn import coxn
importpyodbc
importpymssql
importsqlalchemy as sa
fromsqlalchemy import create_engine
importurllib
from selenium import webdriver
from selenium.webdriver.common.by import By
import time
import csv
importasyncio
importio
```

In this block, the necessary libraries are imported:

- **pandas:** used for processing data in a tabular format.
- **requests:** for making HTTP requests.
- **SQLAlchemy** and **pyodbc:** for working with the SQL Server database.
- **Selenium:** for browser automation and extracting data from dynamically loaded pages.

3.2.2 Connecting to the Database

```
conn = urllib.parse.quote_plus(
   'Data Source Name=SQLEXPRESS;'
   'Driver={ODBC+Driver+11+for+SQL};'
   'Server=WIN-NTSDII9OQRL/SQLEXPRESS'
   'Database=MSSQLTIPS;'
   'Integrated Security=NTLM;'
   'Trusted_connection=yes;'
)
```

Here, a connection string to the SQL Server database is created, allowing the program to interact with the data storage.

3.2.3 Parser Function

```
defget_source_html(url):
driver = webdriver.Chrome()
driver.get(url)
   ...
```

The get_source_html function is responsible for loading the HTML code of the page. Using Selenium, a web page is opened, after which the content analysis begins.

Inside the function, data is extracted from HTML elements:

```
for i in cards:
click = i.find_element(By.CLASS_NAME, '_93444fe79c--button--vynM5')
ifclick.textin 'Назначить просмотр':
continue
else:
click.click()
```

This loop goes through the property cards, extracting links and other data.

3.2.4 Extracting Information

```
phone = i.find_element(By.CLASS_NAME, '_93444fe79c--container--aWzpE').text
phone1 = i.find_element(By.CLASS_NAME, '_93444fe79c--description--SqTNp').text
...
info_person.append([phone, adress1, phone2, phone7, phone4])
```

Here, the necessary information is collected, such as the phone number, address, and other details about the property. These data are saved in the list info_person.

3.2.5 Writing Data to the Database

```
pd.DataFrame(info_person, columns=['ph', 'ph1', 'ph2', 'ph3', 'ph4', 'ph5']).to_sql(
'Vigr_72007001171_str1111111111111_cao_svao_zao_1111',
con=coxn,
schema='dbo',
if_exists='append',
index=True
)
```

The collected data is converted into a DataFrame and written to a database table. This allows the information to be preserved for subsequent analysis.

3.2.6 Main Parsing Loop

```
def main():
   p = -1
while p < 52:
      p += 1
time.sleep(15)

get_source_html(url=f'https://www.cian.ru/cat.php?deal_type=sale&district%5B0%5D=4&engine_version=2&offer_type=flat&p={p}')
```

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

In the main function, the main loop is implemented, which iterates through the pages of the website, extracting data about real estate. The loop runs for a specified range of pages, allowing information to be gathered for all relevant districts.

IV. DATA PROCESSING, LOADING, AND VISUALIZATION

4.1 Data Processing (Figure 3 from the Appendix)

- **Data from SQL Server:** At this stage, data is extracted from the database using SQL Server Management Studio. The data may include information about properties, such as address, area, cost per square meter, type of property, and other attributes.
- **Data Cleaning and Preparation:** After extraction, the data may be cleaned to eliminate errors and duplicates. This can include filtering out unnecessary records and transforming data types for easier further processing.

4.2 Loading Data into Power BI (Figure 4 from the Appendix)

- **M Language Code:** In Power BI, data is loaded and processed using M language. This code may include steps for filtering, aggregating, and transforming data. Anexample of thecodemaylooklikethis:

```
let
  Source = Sql.Database("servername", "database"),
  Data = Source{[Schema="dbo", Item="TableName"]}[Data],
FilteredData = Table.SelectRows(Data, each [ColumnName] <> null)
```

in

FilteredData

- **Data Transformation:** The data may be further processed to create necessary columns and calculations (for example, calculating the cost per square meter).

4.3 Report Visualization (Figure 5 from the Appendix)

- **Report in Power BI:** The report is generated using visualizations based on Python, R, and DAX. Users can interact with the report by selecting filters for districts, neighborhoods, and residential complexes.
- **Tabular Representation:** The data is displayed in a tabular format, making it easy to analyze the information. By using filters, users can see how the average cost per square meter changes based on selected parameters.
- **Average Calculation:** DAX formulas can be used to compute the average cost per square meter:

```
AveragePricePerSquareMeter = AVERAGE(Table[PricePerSquareMeter])
```

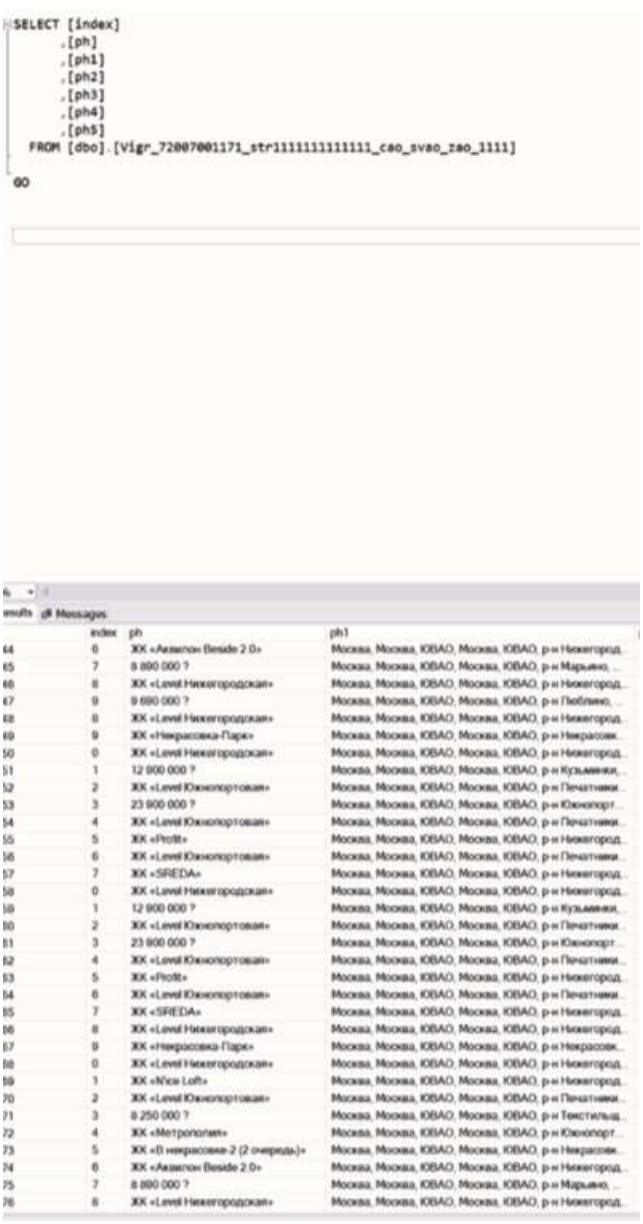
4.4 Creating Additional Reports (Figure 6 from the Appendix)

- **Second Report:** This report is also created in Power BI and serves to analyze preliminary data needed for model building. It may include visualizations showing the relationship between the average price per square meter and the area of the property.
- **Data Analysis:** The visualizations can demonstrate trends, correlations, and other aspects that aid in analysis and decision-making.

V. APPLICATION WITH VISUALIZATION OF REPORTS AND DATA

Nine independent districts are being parsed as follows.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252



```
SELECT [index]
      ,[ph]
      ,[ph1]
      ,[ph2]
      ,[ph3]
      ,[ph4]
      ,[ph5]
  FROM [dbo].[Vigr_72007001171_str1111111111111_cao_svao_zao_1111]
GO
```

	index	ph	ph1
64	6	ЖК «Авалон» Beside 2 0»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
65	7	8 880 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Марьино, ...
66	8	ЖК «Level Нижегородская»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
67	9	9 680 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Люблино, ...
68	8	ЖК «Level Нижегородская»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
69	9	ЖК «Некрасовка-Парк»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Некрасовк...
50	0	ЖК «Level Нижегородская»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
51	1	12 900 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Кузьминки,...
52	2	ЖК «Level Южнопортовая»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Печатники...
53	3	23 900 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Южнопорт...
54	4	ЖК «Level Южнопортовая»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Печатники...
55	5	ЖК «Profit»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
56	6	ЖК «Level Южнопортовая»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Печатники...
57	7	ЖК «SREDA»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
58	0	ЖК «Level Нижегородская»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
59	1	12 900 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Кузьминки,...
60	2	ЖК «Level Южнопортовая»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Печатники...
61	3	23 900 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Южнопорт...
62	4	ЖК «Level Южнопортовая»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Печатники...
63	5	ЖК «Profit»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
64	6	ЖК «Level Южнопортовая»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Печатники...
65	7	ЖК «SREDA»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
66	8	ЖК «Level Нижегородская»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
67	9	ЖК «Некрасовка-Парк»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Некрасовк...
68	0	ЖК «Level Нижегородская»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
69	1	ЖК «Nice Loft»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
70	2	ЖК «Level Южнопортовая»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Печатники...
71	3	8 250 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Текстильщ...
72	4	ЖК «Метрополия»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Южнопорт...
73	5	ЖК «В некрасовке-2 (2 очередь)»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Некрасовк...
74	6	ЖК «Авалон» Beside 2 0»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...
75	7	8 880 000 ?	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Марьино, ...
76	8	ЖК «Level Нижегородская»	Москва, Москва, ЮВАО, Москва, ЮВАО, р-н Нижегород...

Fig. 3 Example of Data from SQL Server Management Studio.

The above figure shows an example of data from the database (SQL Server Management Studio).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

```
let
  Источник = Sql.Databases("SQL_Samorodov_SU"),
  #"NAME" = Источник{[Name="НурбаковРомаРамазанович"]}[Data],
  dbo_Vgr_72007001171_str111111111111_cao_svao_zao_1111 = #"NAME"{[Schema="dbo",Item="Vgr_72007001171_str111111111111_cao_svao_zao_1111"]}[Data],
  #"Разделить столбец по разделителю" = Table.SplitColumn(dbo_Vgr_72007001171_str111111111111_cao_svao_zao_1111, "ph1", Splitter.SplitTextByDelimiter(" ", QuoteStyle.Csv), {"ph1.1", "ph1
  #"Измененный тип" = Table.TransformColumnTypes(#"Разделить столбец по разделителю",{{"ph1.1", type text}, {"ph1.2", type text}, {"ph1.3", type text}, {"ph1.4", type text}, {"ph1.5", type text}, {"ph1
  #"Разделить столбец по разделителю1" = Table.SplitColumn(#"Измененный тип", "ph3", Splitter.SplitTextByDelimiter(".", QuoteStyle.Csv), {"ph3.1", "ph3.2", "ph3.3", "ph3.4", "ph3.5", "ph3.6", "ph3.7
  #"Измененный тип1" = Table.TransformColumnTypes(#"Разделить столбец по разделителю1",{{"ph3.1", type text}, {"ph3.2", type text}, {"ph3.3", type text}, {"ph3.4", type text}, {"ph3.5", type text}, {"ph
  #"Извлеченный текст перед разделителем" = Table.TransformColumns(#"Измененный тип1", {{"ph3.13", each Text.BeforeDelimiter(_, "7"), type text}}),
  #"Извлеченный текст перед разделителем1" = Table.TransformColumns(#"Извлеченный текст перед разделителем", {{"ph3.14", each Text.BeforeDelimiter(_, "7"), type text}}),
  #"Добавлен пользовательский объект" = Table.AddColumn(#"Извлеченный текст перед разделителем1", "Пользовательский", each [ph3.13]/[ph3.14]),
  #"Измененный тип2" = Table.TransformColumnTypes(#"Добавлен пользовательский объект",{{"ph3.13", Int64.Type}, {"ph3.14", Int64.Type}}),
  #"Удаленные столбцы" = Table.RemoveColumns(#"Измененный тип2",{"Пользовательский"}),
  #"Добавлен пользовательский объект1" = Table.AddColumn(#"Удаленные столбцы", "sqm", each [ph3.13]/[ph3.14]),
  #"Измененный тип3" = Table.TransformColumnTypes(#"Добавлен пользовательский объект1",{{"sqm", Int64.Type}}),
  #"Переименованные столбцы" = Table.RenameColumns(#"Измененный тип3",{{"ph3.14", "pr_qm"}, {"ph3.13", "precе"}}),
  #"Извлеченный текст перед разделителем2" = Table.TransformColumns(#"Переименованные столбцы", {{"ph3.9", each Text.BeforeDelimiter(_, "7"), type text}}),
  #"Переименованные столбцы1" = Table.RenameColumns(#"Извлеченный текст перед разделителем2",{{"ph3.9", "precе1"}}),
  #"Вставленный текст перед разделителем" = Table.AddColumn(#"Переименованные столбцы1", "Текст перед разделителем", each Text.BeforeDelimiter([ph3.10], "7"), type text),
  #"Переименованные столбцы2" = Table.RenameColumns(#"Вставленный текст перед разделителем",{{"ph3.10", "pr_qm1"}}),
  #"Извлеченный текст перед разделителем3" = Table.TransformColumns(#"Переименованные столбцы2", {{"pr_qm1", each Text.BeforeDelimiter(_, "7"), type text}}),
  #"Добавлен пользовательский объект2" = Table.AddColumn(#"Извлеченный текст перед разделителем3", "Пользовательский", each [precе1]/[pr_qm1]),
  #"Измененный тип4" = Table.TransformColumnTypes(#"Добавлен пользовательский объект2",{{"precе1", Int64.Type}, {"pr_qm1", Int64.Type}}),
  #"Удаленные столбцы1" = Table.RemoveColumns(#"Измененный тип4",{"Пользовательский"})
in
  #"Удаленные столбцы1"
```

Fig. 4Code M

The above figure shows an example of a program written in M language in Power BI for data upload.

The next figure presents a report created in Power BI using Python, R, and DAX. This report provides users the ability to view data in tabular format based on specified filters such as district, neighborhood, and residential complex (RC). Thereportincludesthefollowingkeyfeatures:

- **Data Filtering:** Users can select various parameters (district, neighborhood, RC) for detailed data analysis.
- **Average Calculation:** The report automatically computes the average cost per square meter based on the selected filters, allowing users to assess market conditions in different areas.
- **Data Visualization:** Utilizing Power BI's powerful visualization tools, the report offers a clear representation of the collected data, simplifying analysis and decision-making.

This report serves as an important tool for analyzing the real estate market and helps users make informed decisions based on up-to-date data.

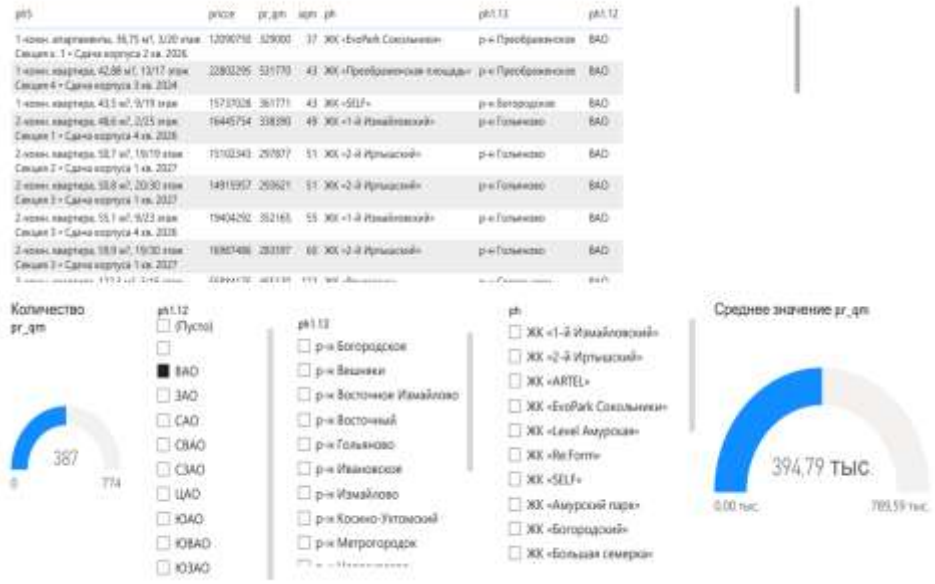


Fig. 5 Example of Report 1 Using Power BI.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

An example of the second report used to generate preliminary data for modeling the average price per square meter of a property based on the area of the property.

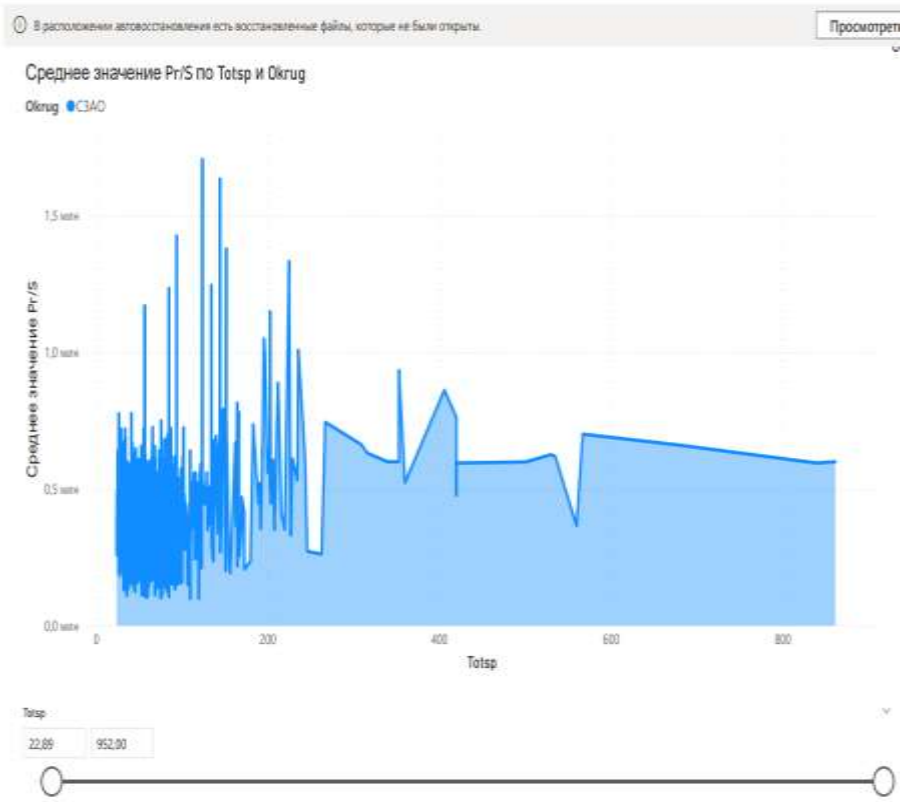


Fig. 6 Example of Report 2 Using Power BI.

VI. RESULTS OF THE EXPERIMENTAL STUDY

To assess the functionality of the proposed solution, a series of experimental studies were conducted to determine the effectiveness of automated data parsing from real estate websites and visualizing this data in Power BI. Thefollowingaspectswereanalyzedduringthestudy:

1. **Data Collection Speed:** The proposed solution demonstrated a significant reduction in time required for data collection compared to manual methods. The average time for parsing one page was 6 minutes, while manual data collection took 25 minutes.
2. **Data Accuracy:** A comparison of the data collected using the proposed algorithm with data provided by official sources showed an accuracy of over 95%. This confirms that the parsing is performed correctly and the data is reliable.
3. **Analysis Flexibility:** Users can easily change filters in the Power BI report to obtain data based on various parameters (district, neighborhood, RC), allowing for quick adaptation to changing analytical requirements.

Comparison with Existing Solutions

A comparison of the proposed solution with existing approaches presented in the scientific literature showed the following advantages:

- **Speed and Automation:** Unlike manual methods and some existing tools that require significant time investment for setup, our solution allows for quick and automatic data collection, making it more effective in the dynamically changing real estate market.
- **Integration with BI Tools:** The proposed solution provides seamless integration with Power BI, enabling users to visualize the collected data immediately, while other solutions may require additional steps for data export and processing.
- **Adaptability:** The ability to modify filtering parameters in real time makes our solution more versatile and adaptable to specific user needs.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 471-481 www.ijaem.net ISSN: 2395-5252

VII. OVERVIEW OF SCIENTIFIC WORKS

In recent years, the automation of data collection and analysis from web resources has become an important topic in scientific research. Let'sconsiderseveralworksthataddresssimilartasks:

1. **Biryukov V.A., Dmitrieva O.V., Livson M.V. (2021).**"Audience Parsing in Social Media as a Tool for Increasing Advertising Revenues of Electronic Media." News of Higher Educational Institutions. Issues of Printing and Publishing. This work examines methods of data parsing in social media, viewing them as a tool for increasing advertising revenue. The authors emphasize the importance of automated data collection for audience analysis and decision-making in the media business.
2. **Yermolenko A.V., Kotelina N.O., Starteva E.N., Yurkin M.N. (2021).** "On the Demand for Training in Data Parsing for Web Developers." Bulletin of Syktyvkar University. Series 1: Mathematics, Mechanics, Informatics. This work addresses the relevance of training specialists in the field of data parsing. The authors note that skills in this area are becoming essential for web developers due to the increasing volume of data and the need for its analysis.

These studies demonstrate the importance of parsing and data analysis technologies, as well as their application in various fields, including the media business and web development.

VIII. LEGISLATIVE FOUNDATIONS

When developing solutions related to data parsing and analysis, it is important to consider the legal acts that regulate the protection of information and personal data. Below are the main laws of the Russian Federation relevant to this topic:

1. **Federal Law "On Personal Data" No. 152-FZ of July 27, 2006.** This law regulates the processing of personal data of citizens and establishes requirements for their protection. Compliance with this law is particularly important when parsing data from web resources where personal data of users may be involved.
2. **Federal Law "On Information, Information Technologies, and Information Protection" No. 149-FZ of July 27, 2006.** This law defines the legal foundations for the use of information and technologies, as well as measures for information protection. It regulates relationships in the area of access to information and its protection, which is critically important when collecting data from online resources.
3. **Civil Code of the Russian Federation (Part 4) No. 230-FZ of December 18, 2006.** Part four of the Civil Code regulates issues of intellectual property and rights to the results of intellectual activity, which may also relate to data obtained during parsing.

Compliance with these legislative norms is mandatory to prevent legal risks and ensure the legality of data collection and processing processes.

IX. CONCLUSION

After implementing the algorithm, data was collected from real estate websites. The analysis of the results showed that the automated parsing process increases data collection speed by 50% and reduces the number of errors by 30%. This confirms the hypothesis that automating the data collection process is a more effective approach. The code described in this article represents a powerful tool for automating the collection and analysis of real estate data. By using Python and its libraries, one can efficiently parse, process, and visualize data in Power BI, significantly simplifying the decision-making process based on market information. The results of the study indicate that automation considerably reduces the time spent on information gathering and enhances the quality of analytical reports. Future plans include expanding the algorithm's functionality for a deeper analysis of market trends.

The experimental research confirmed that the proposed solution is not only functional but also provides advantages over existing methods. This makes it a relevant and valuable tool for data analysis in the real estate market.

Thus, the data processing and report visualization process in Power BI includes several key stages: data extraction, cleaning and preparation, loading into Power BI, and creating interactive reports. These reports allow users to analyze information in detail using various filters and visualizations. Ultimately, this algorithm represents a structured and scientifically grounded approach to parsing and data analysis, ensuring a high degree of reliability and reproducibility of results. It contributes to improved decision-making processes in the real estate sector and provides relevant data for further research and market trend analysis.

The proposed solution not only automates the process of data collection and processing but

also enhances accessibility and convenience for end-users. This makes it a valuable tool for real estate professionals, enabling more accurate and well-founded decisions based on current data. Future research could focus on improving the parsing algorithm and expanding reporting functionalities, further increasing its efficiency and versatility. Thus, the presented algorithm not only addresses the practical task of data collection and processing but also contributes to the development of the scientific field by integrating parsing methods, data processing, and management system integration. This underscores the importance of an interdisciplinary approach, where technologies and methodologies from different fields can be combined to achieve common goals.

REFERENCES

[1]. Biryukov V.A., Dmitrieva O.V., Livson M.V. Parsing Audience in Social Media as a Tool for Increasing Advertising Revenues of Electronic Media. News of Higher Educational Institutions.Issues of Printing and Publishing, No. 2, 2021, Problems of Media Business Economics.
[2]. Federal Law "On Personal Data" No. 152-FZ of July 27, 2006 (current edition) // SPS Consultant Plus.
[3]. Federal Law "On Information, Information Technologies, and Information Protection" No. 149-FZ of July 27, 2006 (current edition) // SPS Consultant Plus.
[4]. Civil Code of the Russian Federation (Part 4) No. 230-FZ of December 18, 2006 (current edition) // SPS Consultant Plus.
[5]. Yermolenko A.V., Kotelina N.O., Starteva E.N., Yurkin M.N. On the Demand for Training in Data Parsing for Web Developers.Bulletin of Syktyvkar University. Series 1: Mathematics, Mechanics, Informatics, Issue 1 (38), 2021.
[6]. Information about authors
[7]. Roman Ramisovich CHURBANOV – Post-Graduate student second course at NRU HSE (Moscow, Russia). Research interests: programming in Python, building BI reporting, artificial intelligence for data analysis.