

User Manual

CSS4IMPACT

July 2023



Table of Contents

Introduction.....	3
Installed Software.....	3
Frontend & User Interface Guide	3
Overview.....	3
Functionality & Features.....	4
Filtering Language.....	4
Priority.....	4
Category.....	5
Actors and Organisation	6
Columns	6
Accessing File	7
Backend and Scraper	7
Project Dependencies.....	7
Overview.....	8
File Location.....	8
Code Structure.....	8
Editing and Running the Program.....	8
Opening Folders in VS Code.....	9
Opening a New Window or Folder from an Existing Project	9
Opening the Terminal in VS Code	10
To get the latest code from GitLab for any existing project:	10
Running the Database	11
Running the Scrapers.....	12
Running the Website	14
Closing the Database	14
Closing the Scrapers	14
Closing the Website.....	15

Introduction

This documentation provides an overview of our CSS4Impact project that incorporates database functionality, web scraping, natural language processing (NLP), and front-end development. It includes information about the software required for development, the installation of dependencies, and instructions on running the core parts.

Installed Software

To work on this project, ensure that the following software are installed on your system:

VSCode - Integrated development environment for editing and running Python scripts. Visit <https://code.visualstudio.com/> to download and install VSCode.

Docker Desktop - Containerization platform used to run the project's database. Before proceeding, make sure to activate Hyper-V on Windows and update the Windows Subsystem for Linux. Visit <https://www.docker.com/products/docker-desktop/> to download and install Docker Desktop.

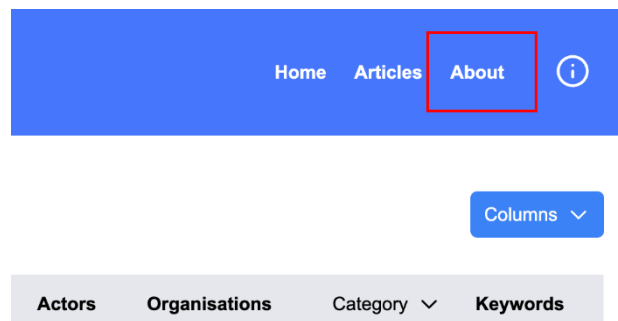
Sublime Text - Text editor used to open and edit text files. You can download and install Sublime Text from <https://www.sublimetext.com/>.

Frontend & User Interface Guide

Overview



Articles: All the articles, file sorting and filtering is available under this section.



About: This section contains overview of the team and the project.

Functionality & Features

Filtering Language

Previous 1/3 Next

	Date Publication	Language	Priority	Uri	Actors	Organisations	C
ქვემოთ დევნილები ფოცხოვანურში - JAMnews		<input checked="" type="checkbox"/> Georgian <input type="checkbox"/> Azerbaijani <input type="checkbox"/> Ukrainian	21		Actors (+)	Organisations (+)	Pt
			9		Actors (+)	Organisations (+)	Pt
აფი კავშირები - JAMnews	2017-05-09	Georgian	1.8		Actors (+)	Organisations (+)	Pt
საგან გაყოფილი სოფლიდან - JAMnews	2019-12-07	Georgian	1.4		Actors (+)	Organisations (+)	Pt
საოგანო. როგორ ცხოვრობენ საქართველოში	2022-01-19	Georgian	1.2		Actors (+)	Organisations (+)	Pt
	2016-04-03	Georgian	1		Actors (+)	Organisations (+)	Pt

You can filter between three languages i.e. Georgian, Azerbaijani, and Ukrainian by pressing on language button and selecting one of the languages.

Note: The language filter is also the '**mainFilter**' i.e. you will see articles only after selecting one language. Only one language can be selected at a time and changing language will reset the page.

Priority

Home Articles About ⓘ

Previous 1/3 Next Columns

	Date Publication	Language	Priority	Uri	Actors	Organisations	Category	Keywords
ვნილები ფოცხოვანურში - JAMnews	2017-11-28	G	<input type="text" value="12"/> Highest available priority: 25.2		Actors (+)	Organisations (+)	Pension	Keywords (+)
	2016-05-27	G	<input type="text" value="12"/> Reset Set		Actors (+)	Organisations (+)	Pension	Keywords (+)
არები - JAMnews	2017-05-09	Georgian	1.8		Actors (+)	Organisations (+)	Pension	Keywords (+)
ოფილი სოფლიდან - JAMnews	2019-12-07	Georgian	1.4		Actors (+)	Organisations (+)	Pension	Keywords (+)
როგორ ცხოვრობენ საქართველოში	2022-01-19	Georgian	1.2		Actors (+)	Organisations (+)	Pension	Keywords (+)

Priority can be set by either writing the numerical value or by using the **arrow keys** to increase or decrease the score.

Next

Language	Priority	Url	Actors
Georgian	12		Actors (+)
	Highest available priority: 25.2		
	Reset	Set	
Georgian	1.8		Actors (+)

The highest available priority is displayed under text input. Once you set the priority press **Set** to display the output (by default the output is displayed in ascending order) and press **Reset** to reset the priority.

Category

Next Columns

Language	Priority	Url	Actors	Organisations	Category	Keywords
Georgian	21		<input type="checkbox"/> Cash Benefits			Keywords (+)
Georgian	9		<input type="checkbox"/> Pension			Keywords (+)
Georgian	1.8		<input type="checkbox"/> IDP allowance			Keywords (+)
Georgian	1.4		<input type="checkbox"/> Unemployment benefits			Keywords (+)
Georgian	1.2		<input type="checkbox"/> Regulations for issuance of cash benefits			Keywords (+)
Georgian	1		<input type="checkbox"/> Social assistance			Keywords (+)
Georgian	1		<input type="checkbox"/> Provision of food			Keywords (+)
Georgian	1		<input type="checkbox"/> Provision of everyday goods			Keywords (+)
Georgian	1		<input type="checkbox"/> Medical services			Keywords (+)
Georgian	1		<input type="checkbox"/> Medical insurance			Keywords (+)
Georgian	1		<input type="checkbox"/> Financing education			Keywords (+)
Georgian	1		<input type="checkbox"/> Financing of higher education			Keywords (+)
Georgian	1		<input type="checkbox"/> Promotion of professional education			Keywords (+)
Georgian	1		<input type="checkbox"/> Provision of housing			Keywords (+)
Georgian	1		<input type="checkbox"/> Housing issue			Keywords (+)
Georgian	1		<input type="checkbox"/> Criteria for the provision of IDPs with living space			Keywords (+)
Georgian	1		<input type="checkbox"/> Legalization of residential areas			Keywords (+)
Georgian	1		<input type="checkbox"/> Compact accommodation facility			Keywords (+)
Georgian	1		<input type="checkbox"/> IDP status seeker			Keywords (+)
Georgian	1		<input type="checkbox"/> Places of resettlement of IDPs			Keywords (+)
Georgian	1		<input type="checkbox"/> Long-term settlement of IDPs			Keywords (+)
Georgian	1		<input type="checkbox"/> Employment Promotion			Keywords (+)
Georgian	1		<input type="checkbox"/> Childcare			Keywords (+)
Georgian	1		<input type="checkbox"/> Family of IDPs			Keywords (+)
Georgian	1		<input type="checkbox"/> Help with household responsibilities			Keywords (+)
Georgian	1		<input type="checkbox"/> Assistance in dealing with state bureaucracy			Keywords (+)
Georgian	1		<input type="checkbox"/> Counselling			Keywords (+)
Georgian	1		<input type="checkbox"/> IDP rights			Keywords (+)
Georgian	1		<input type="checkbox"/> Integration			Keywords (+)

Clicking on **Category** you can see all the categories. Selecting on a category, articles having selected category will be displayed.

Home Articles About




Previous 1/1 Next Columns

Title	Date Publication	Language	Priority	Url	Actors	Organisations	Category	Keywords
„იმ მინდორზე ის ბიჭებთან დევხეურის ვიამაშობდით“. რეპორტაჟი გუგუტიანთაგან - JAMnews	2019-08-23	Georgian	0		Actors (+)	Organisations (+)	Cash Benefits	Keywords (+)
სამსახური პოლიონეში - ახალგაქმნილი გზა ევროპაში საქართველოს მოქალაქეებისთვის - JAMnews	2019-03-29	Georgian	0		Actors (+)	Organisations (+)	Cash Benefits	Keywords (+)

As example above only articles with **Cash Benefits** were displayed when Cash Benefit was selected.

Actors and Organisation

Next

Priority ▾	Uri	Actors	Organisations	Cat
18.83		Actors (-) Омбудсмана Романом Лихачовим Роман Лихачов	Organisations (+)	Non orge
16.43		Actors (+)	Organisations (+)	Non orge
...		...	Organisations	Non

All the actors and organisations from the articles can be found in **Actors** and **Organisations** column. You can click on **Actors (-)** button to expand the row and view all the actors. Similarly, you can expand **Organisation (-)** to view all organisations present in article.

Note: The results for actors and organisation are automatically detected by NER algorithm that uses specific language models for all three languages. So, it may not give accurate results all the time.

Columns

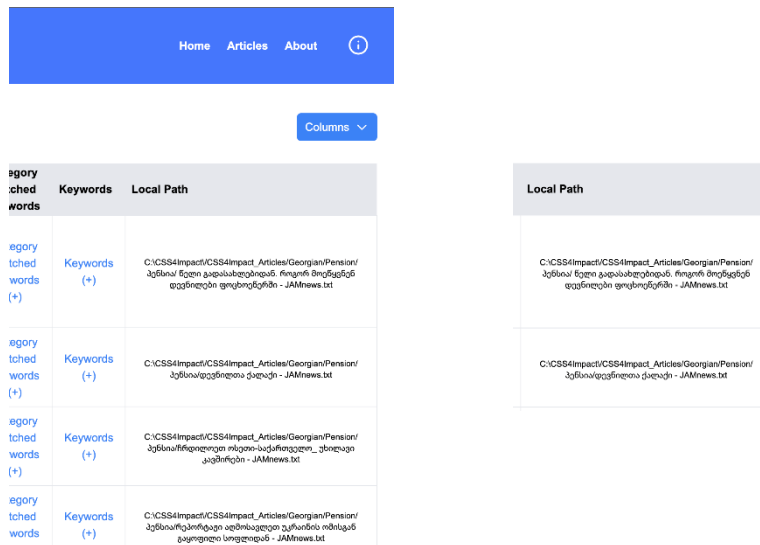
Home Articles About ⓘ

Columns ▾

Actors	Organisations	Category ▾	<input checked="" type="checkbox"/> Title
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Date Publication
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Language
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Priority
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Uri
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Actors
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Organisations
Actors (+)	Organisations (+)	Pension	<input type="checkbox"/> IDP Matched Keywords
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Category
Actors (+)	Organisations (+)	Pension	<input type="checkbox"/> Matched Keywords
Actors (+)	Organisations (+)	Pension	<input checked="" type="checkbox"/> Keywords
Actors (+)	Organisations (+)	Pension	<input type="checkbox"/> Local Path
Actors (+)	Organisations (+)	Pension	Keywords (+)
Actors (+)	Organisations (+)	Pension	Keywords (+)
Actors (+)	Organisations (+)	Pension	Keywords (+)
Actors (+)	Organisations (+)	Pension	Keywords (+)

On the upper right corner, you will find **Columns** that are being displayed. You can select or unselect any column you need.

Accessing File



The screenshot shows a web application with a blue header containing links for 'Home', 'Articles', 'About', and an information icon. Below the header is a 'Columns' dropdown menu. The main content area displays a table with three columns: 'Category', 'Keywords', and 'Local Path'. The 'Category' column contains the text 'Category', 'Keywords', and 'Local Path'. The 'Keywords' column contains the text 'Keywords' and a plus sign. The 'Local Path' column contains the text 'Local Path' and a plus sign. The table is partially obscured by a second table on the right, which also has a 'Local Path' column.

Category	Keywords	Local Path
Category	Keywords (+)	C:\CSS4Impact\CSS4Impact_Articles\Georgian\Pension/ პენსია/ წელი გადასახლებიდან. როგორ მოუწევნებ ფერწილში ფერწილში - JAMnews.txt
Category	Keywords (+)	C:\CSS4Impact\CSS4Impact_Articles\Georgian\Pension/ პენსია/ ფერწილში გადასახლებიდან - JAMnews.txt
Category	Keywords (+)	C:\CSS4Impact\CSS4Impact_Articles\Georgian\Pension/ პენსია/ როგორ მოუწევნებ ფერწილში ფერწილში - JAMnews.txt
Category	Keywords (+)	C:\CSS4Impact\CSS4Impact_Articles\Georgian\Pension/ პენსია/ როგორ მოუწევნებ ფერწილში ფერწილში - JAMnews.txt

You can access the file (article) from your local directory. The path to it is provided in **Local Path** column, which you can enable it from **Columns**.

Clicking on path will open the article in browser (local) where you can read, but the better option would be to copy the path and open from your local directory.

Backend and Scraper

Project Dependencies

The project requires the installation of the following main libraries:

Python 3.11.4 - Required for NLP and CS.

Node v18.16.1 - Required for website/frontend development.

Git - Required for version control. The project is maintained in GitLab.

Additionally, the following packages are needed for the scrapers and NLP functionality:

pip - Package manager for Python. It is installed automatically with Python.

bs4 (BeautifulSoup 4.12.2) - Used to extract HTML from webpages.

requests (2.31.0) - Used to retrieve the URL of a website.

selenium (4.10.0) - Used to extract data from websites.

trafilatura (1.6.1) - Used to extract article main text and headings.

spacy (3.5.4) - Required for Natural Language Processing.

uk-core-news-lg (3.5.0) - Spacy Language Model for Ukrainian.

xx-ent-wiki-sm (3.5.0) - Spacy Language Model for Azerbaijani and Georgian.

Overview

File Location

We have saved all of our files on the C drive in the CSS4Impact folder.

This PC > Local Disk (C:)

Name	Date modified	Type	Size
SWinREAgent	01.07.2023 06:04	File folder	
CSS4Impact	30.06.2023 12:15	File folder	
Intel	06.07.2023 19:45	File folder	
PerfLogs	07.12.2019 10:14	File folder	
Program Files	04.07.2023 16:39	File folder	
Program Files (x86)	13.07.2023 14:19	File folder	
ProgramData	29.06.2023 19:15	File folder	
Python311	21.06.2023 09:32	File folder	
Users	29.06.2023 23:33	File folder	
Windows	06.07.2023 19:45	File folder	

Code Structure

Our codes are separated into the following three folders:

This PC > Local Disk (C:) > CSS4Impact >

Name	Date modified	Type	Size
Chromedriver	29.06.2023 23:33	File folder	
CSS4Impact_Articles	04.07.2023 15:16	File folder	
Database	29.06.2023 20:46	File folder	
Scrappers	29.06.2023 22:13	File folder	
Website	29.06.2023 22:53	File folder	

Database: Contains files related to the database functionality.

Scrappers: Contains files for web scraping.

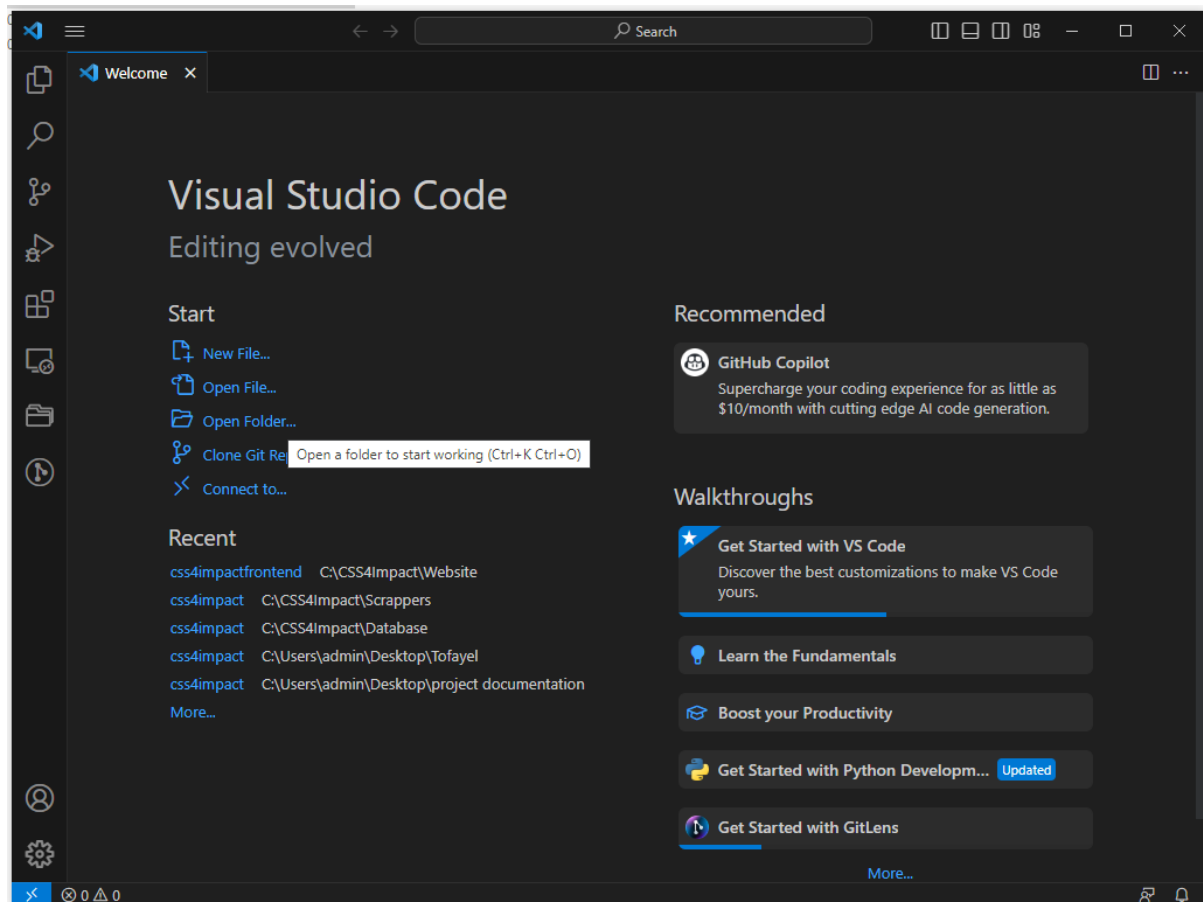
Website: Contains files for the website/frontend development.

Editing and Running the Program

To edit or run our program, we primarily utilize Visual Studio Code (VS Code). You will need to open the specific folders in VS Code to access the relevant files.

Opening Folders in VS Code

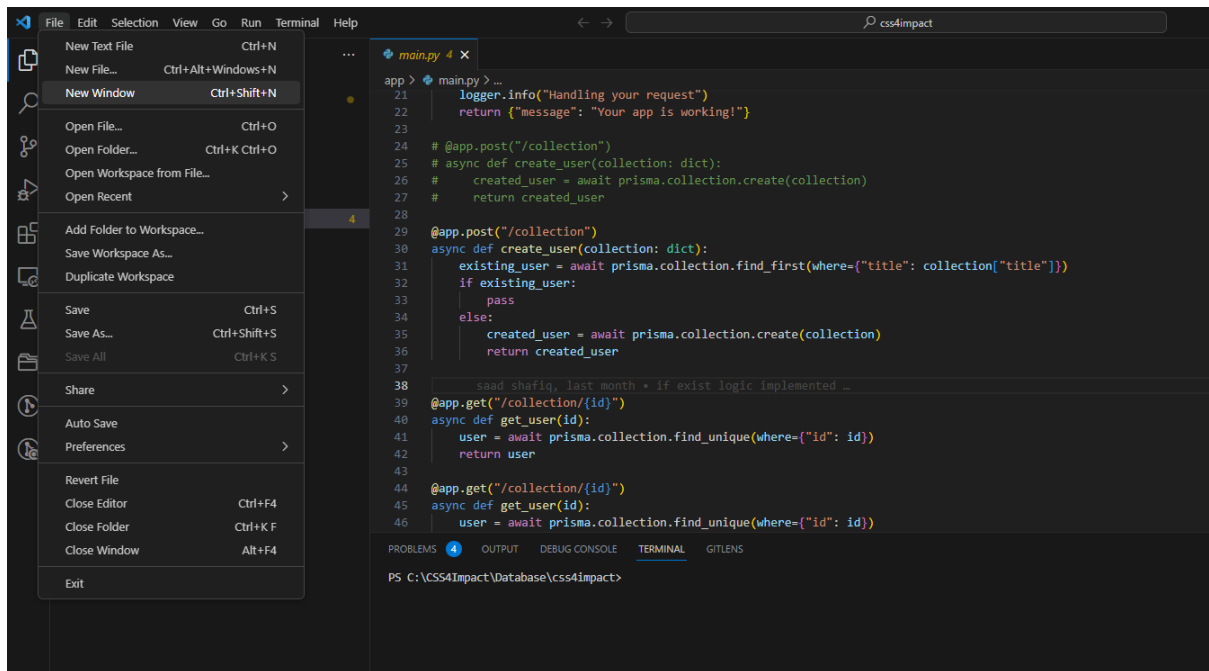
1. Open VS Code.
2. Select "File" from the menu and choose "Open Folder."
3. Navigate to the directory of the desired folder (Database, Scrappers, or Website).
4. Select the folder and click "Open."



Opening a New Window or Folder from an Existing Project

If you need to open a new window or folder from an existing project:

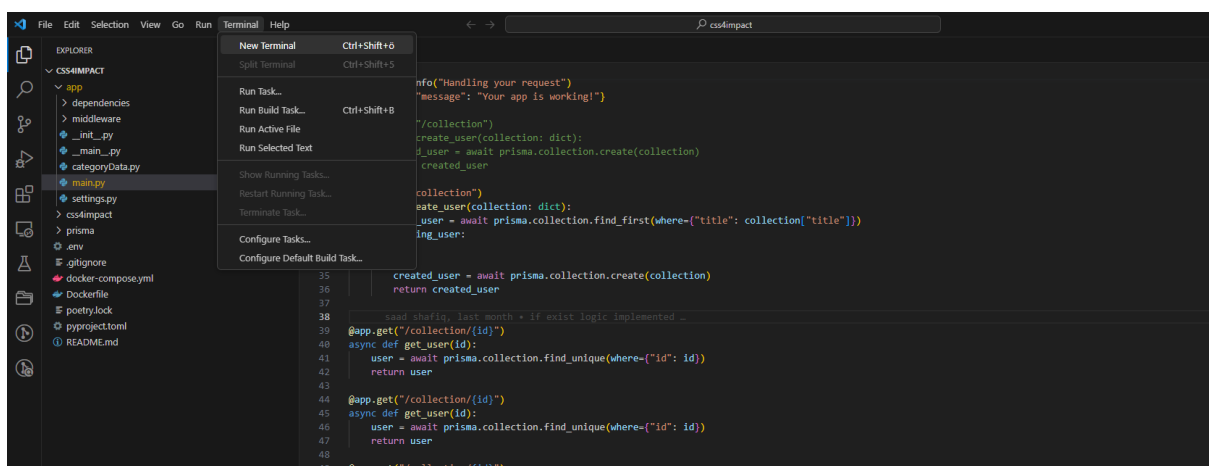
1. Open VS Code.
2. Select "File" from the menu and choose "**New Window**" or "**New Folder from Existing.**"
3. Navigate to the directory of the project and select it.



Opening the Terminal in VS Code

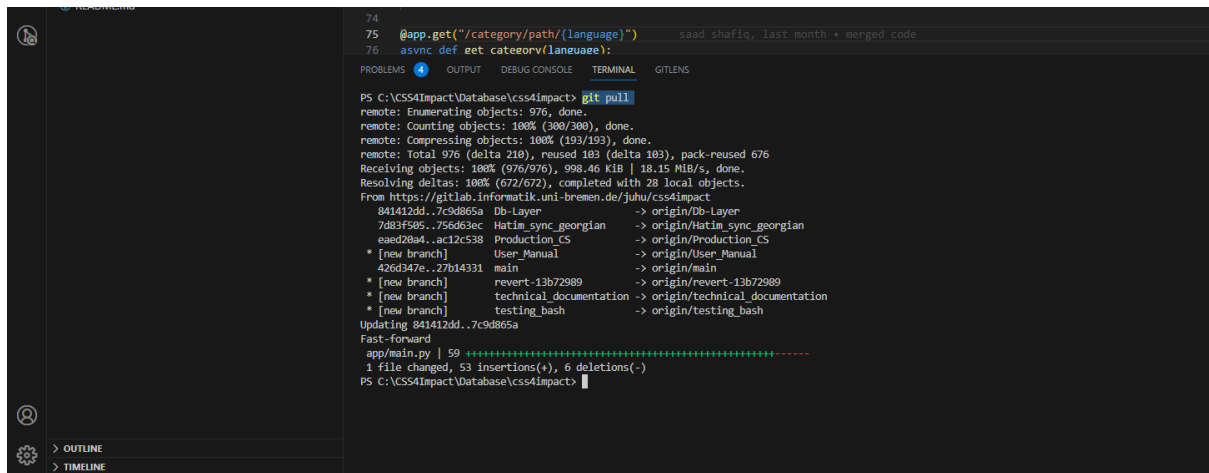
To open the terminal for any existing project in VS Code:

1. Open VS Code.
2. From the menu and choose "Terminal"



To get the latest code from GitLab for any existing project:

1. From the menu and choose "Terminal"
2. Execute the following command: **git pull**. It will pull all the latest code from GitLab. If there are any changes on Database, Scrappers or Website. Use this command to get the changes.



```
74
75 @app.get("/category/path/{language}") saad shafiq, last month • merged code
76 async def eet cateorv(laneuae):

PROBLEMS 4 OUTPUT DEBUG CONSOLE TERMINAL GITLENS

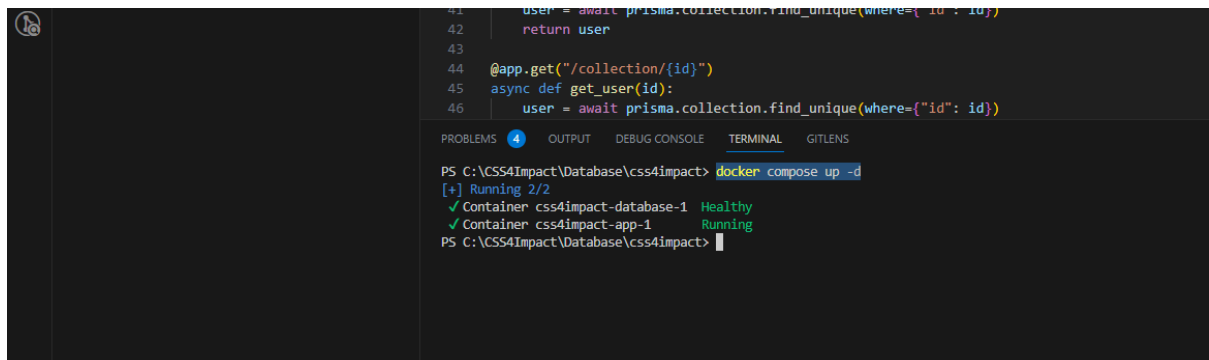
PS C:\CSS4Impact\Database\css4impact> git pull
remote: Enumerating objects: 976, done.
remote: Counting objects: 100% (300/300), done.
remote: Compressing objects: 100% (193/193), done.
remote: Total 976 (delta 210), reused 183 (delta 183), pack-reused 676
Receiving objects: 100% (976/976), 998.46 KiB | 18.15 MiB/s, done.
Resolving deltas: 100% (672/672), completed with 28 local objects.
From https://gitlab.informatik.uni-bremen.de/juhu/css4impact
  841412dd..7c9d865a Db-Layer      -> origin/Db-Layer
  7d83f585..756d63ec Hatim_sync_georgian -> origin/Hatim_sync_georgian
  eaed20a4..ac12c538 Production_CS   -> origin/Production_CS
  * [new branch]      User_Manual     -> origin/User_Manual
  426d347e..27b14331 main           -> origin/main
  * [new branch]      revert-12b72989 -> origin/revert-12b72989
  * [new branch]      technical_documentation -> origin/technical_documentation
  * [new branch]      testing_bash    -> origin/testing_bash
Updating 841412dd..7c9d865a
Fast-forward
 app/main.py | 59 ++++++
 1 file changed, 53 insertions(+), 6 deletions(-)
PS C:\CSS4Impact\Database\css4impact>
```

Running the Database

Follow the steps below to run the project's database:

1. Navigate to the CSS4Impact folder and then the Database folder.
2. Open the Database folder using VSCode.
3. Open the Terminal from VSCode.
4. Execute the following command: **docker compose up -d**.

This command will run the database, and you can access it from any browser using the link <http://localhost:8000/docs>.



```
41 user = await prisma.collection.find_unique(where={"id": id})
42 return user
43
44 @app.get("/collection/{id}")
45 async def get_user(id):
46     user = await prisma.collection.find_unique(where={"id": id})

PROBLEMS 4 OUTPUT DEBUG CONSOLE TERMINAL GITLENS

PS C:\CSS4Impact\Database\css4impact> docker compose up -d
[+] Running 2/2
  ✓ Container css4impact-database-1 Healthy
  ✓ Container css4impact-app-1 Running
PS C:\CSS4Impact\Database\css4impact>
```

Once the database is running, you can minimize VSCode.

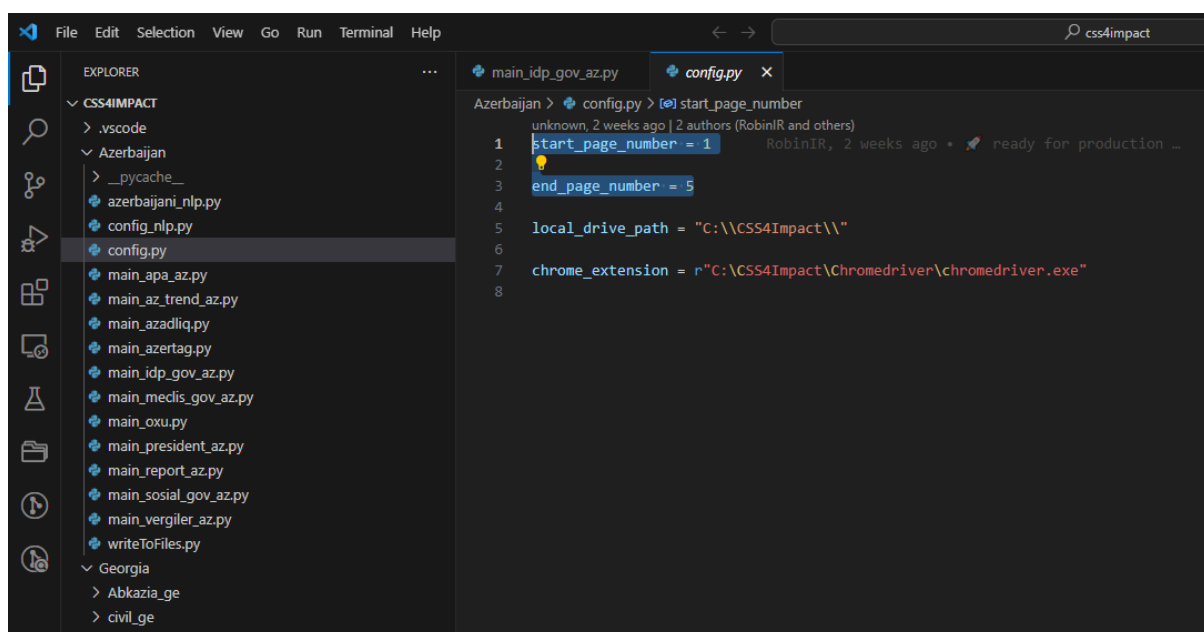
Note: Make sure that Docker Desktop is running and properly configured before executing the command.

Running the Scrapers

To run the scrapers for the project, follow these steps:

1. Navigate to the CSS4Impact folder and then the Scrappers folder.
2. Open the Scrappers folder using VSCode.
3. Go to the specific language folder that you want to run the scraper for.
4. Open the config.py file and update the start and end page numbers.

Note: The existing start and end page numbers indicate that articles from those pages have already been crawled and saved in the **CSS4Impact_Articles** folder. For example, if **start_page_number = 1** and **end_page_number = 5**, it means articles from the 1st to the 5th page have already been crawled. For the next crawling, change **start_page_number = 6** and **end_page_number = 10** or more. Then save the file using **Ctrl+S**.



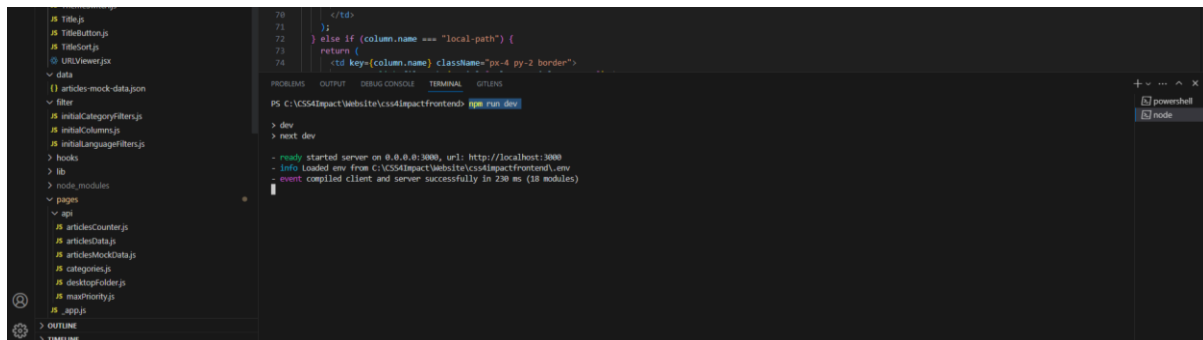
5. Only run the scripts with the name **main_***.py**, such as **main_azadliq.py**. The other Python scripts are helper functions and do not need to be run.
6. Click on the desired **main_***.py** script and run it using the play button in the top-right corner of VSCode.

Note: These main scripts may take some time to execute and may open the Chrome browser automatically. Do not close the browser, but you can minimize it. You can also view the logs from the VSCode terminal.

Running the Website

To run the website for the project, follow these steps:

1. Navigate to the CSS4Impact folder and then the Website folder.
2. Open the Website folder using VSCode and open the VSCode terminal.
3. Execute the following command: **npm run dev**.



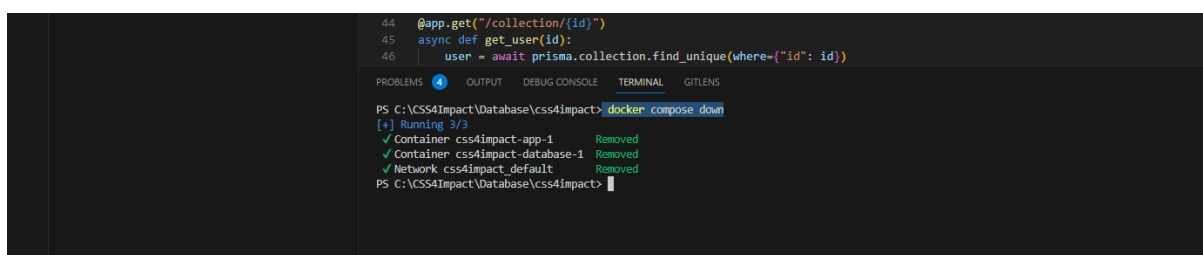
4. Access the website by visiting the link <http://localhost:3000/> in your browser.

Closing the Database

To close the database, follow these steps:

1. Open the VSCode window for the database.
2. Execute the following command: **docker compose down**.

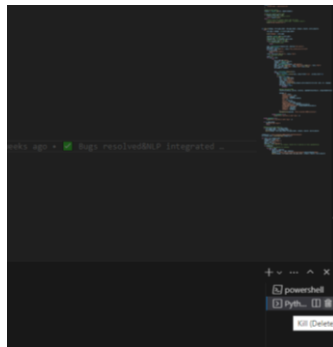
This will close the database. Don't worry, when you run it again, you can retrieve the data.



Closing the Scrapers

The scrapers will automatically close when the scraping process is finished. There will be no logs on the VSCode terminal. If you wish to terminate the process before it completes, you

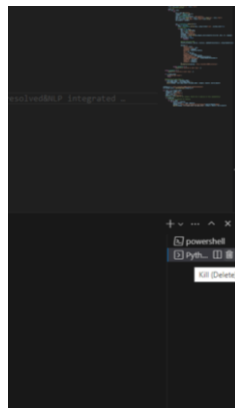
can go to the VSCode terminal delete the python terminal or press **Ctrl+C**. The files will be saved up until that point.



Closing the Website

To close the website, follow these steps:

1. Open the VSCode window for the website.
2. Delete the node terminal or Press **Ctrl+C** on the VSCode terminal.



```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  GITLENS

PS C:\CSS4Impact\Website\css4impactfrontend> npm run dev

> dev
> next dev

- ready started server on 0.0.0.0:3000, url: http://localhost:3000
- info Loaded env from C:\CSS4Impact\Website\css4impactfrontend\.env
- event compiled client and server successfully in 230 ms (18 modules)
Batchvorgang abbrechen (Y/N)?
Y
```