

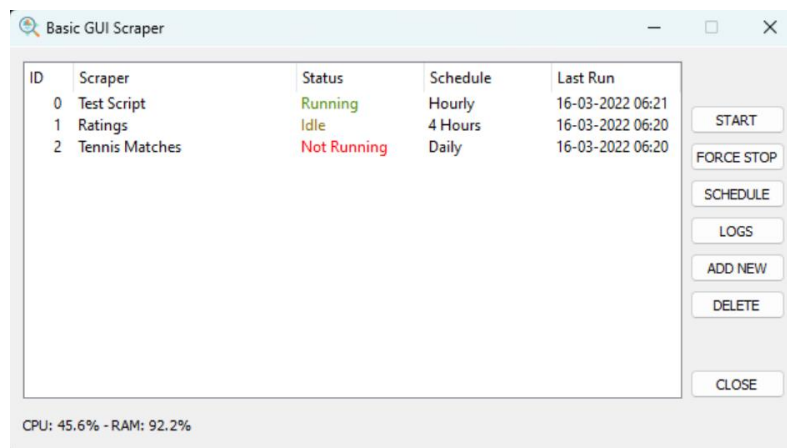
# Scrapers GUI

Documentation



## Introduction

This program monitors and manages web-scrappers. With the advantage of the easy-to-use GUI, you can list your scrapers in the program and schedule run them, monitor when they last ran and more!



## Usage

### Installation

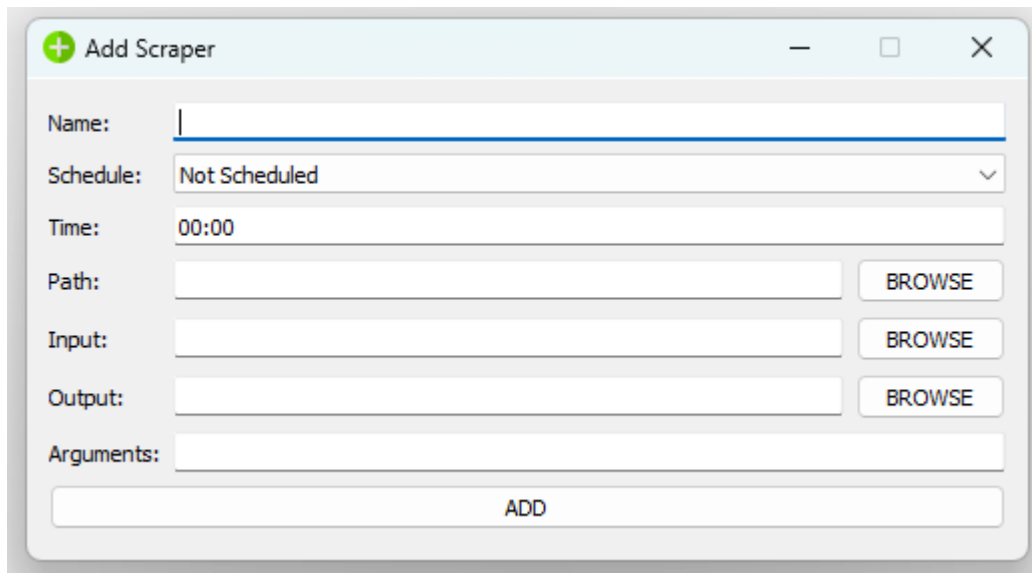
1. Copy the .exe file to a new folder and run the file.
  - When the program first starts, it creates a `settings.json` file where program settings and all scrapers are stored.
  - Moving the .exe file to another folder or deleting the settings file leads to losing the scrapers listed on the program and previous settings.

Or:

2. Execute the python code from the terminal using the command:  
`python scraper_gui.py`

Make sure to install the required libraries before executing the code, run the following command to install them: `pip install -r requirements.txt`

## Adding New Scraper



### Fields:

- **Name:** Name used in the program to call the scraper, can be any name.
- **Schedule:** How often the scraper runs (Hourly, every 4 hours or daily).
- **Time:** Works for “daily” scheduled scrapers only. Sets a time at which the scraper runs every day.
- **Path:** Where the scraper python code is located.
- **Input:** Location of input file.
- **Output:** Location of output file.
- **Arguments:** Any additional data passed to the scraper.

Finally, the command that the scraper calls is generated as follows:

```
python <Path> <input> <output> <arguments>
```

## Removing Scraper

Select the scraper from the list and press DELETE button.

## Pre-Requisites:

Input path, Output path, and arguments are passed to the scraper code and are read using the following code:

```
import sys
input_file = sys.argv[1]
output_file = sys.argv[2]
arguments = sys.argv[3]
```

An example of the previous code:

```
import sys
input_file = sys.argv[1]
output_file = sys.argv[2]

with open(input_file, 'w+') as file:
    file.write('This line is written to the input file')
```

## Program Settings:

Settings file is JSON, it is easy to modify when necessary (but remember to take a copy because wrong JSON format results in erasing all the data).

```
{
  "scrapers": [
    {
      "name": "Test Script",
      "path": "D:/Tim_projects/Scraper_GUI/output/scrapers/script.py",
      "cmd": "python D:/Tim_projects/Scraper_GUI/output/scrapers/script.py",
      "schedule": "Hourly",
      "time": "00:00",
      "lastRun": "16-03-2022 06:21",
      "status": 0
    }
  ]
}
```