

# PYTHON EXERCISE

## DATA PROCESSOR

### 03 – WEB SCRAPING FOR DATA STORAGE

As addition to the first exercises the user should also have the possibility to scrape information from a GitHub project sites, display the information and store them to the file system. The following requirements should be fulfilled:

- Extend either the Flask or PyQt5 application
- The user should have to possibility to insert a link of a GitHub project website.
- Check if the entered website is a GitHub project
- Scrape the following data and display it in the app:
  - General information (main\_app):
    - Name of the project
    - Number of repositories
    - Number of people
    - Top languages
  - For each repository (dataset):
    - Name
    - Short information
    - Programming languages and share of each
    - Stars
    - Watchers
    - Forks
    - Branches
    - Tags
    - Last updated
    - Topics

All of the requirements of the last exercises (store, edit, delete, etc.) should also work on the new version of the app.

Try to use the modules that were used in the previous lectures. If you have to use other modules describe in short words, why this is necessary.

A good documentation and useful test cases helps the lecturer to understand what you did and why and though helps you to get a better evaluation.

Files to be delivered:

- Source code
- Good documentation
- Test cases
- requirements.txt