Instruction on how to use web scraper for lubimyczytac.pl.

The web scraper provides a functionality of checking data about top 150 books by popularity on the website and download detailed data about them to database.

Usage:

The project is sent with virtualbox image, that contains proper environment for running the program. To use it do following steps:

1. Login to osboxes.org user (password osboxes.org)
2. Go to project folder: cd WebScrapper
3. Source the virtual environment to have proper python packages available: source venv/bin/activate
4. Now you can run webscraper funnctions, run python3 ./webscraper_lubimyczytac.py -h for help.

Instruction for new fedora installation:

1. Install MySQL:

a. Update repository: sudo dnf update
b. Setup yum repository: vim /etc/yum.repos.d/mysql-community.repo
add following lines:

```
[mysql80-community]
name=MySQL 8.0 Community Server
baseurl=http://repo.mysql.com/yum/mysql-8.0-community/fc/$releasever/$basearch/
enabled=1
gpgcheck=0
```

c. Install MySQL server: sudo dnf install mysql-community-server

```
sudo systemctl enable mysqld.service
sudo systemctl start mysqld.service
```

d. Get temporary root password from log file: grep 'A temporary password is generated' /var/log/mysqld.log | tail -1
e. Run installation wizard and follow steps from there: sudo mysql_secure_installation
f. Connect to mysql with terminal: mysql -u root -p
g. Create database for websrcrapper project: CREATE DATABASE webscraper_lubimyczytac;

2. Install ChromeDriver

a. Install google chrome with sudo dnf install google-chrome-stable_current_*.rpm
b. Download new chromedriver version: wget https://chromedriver.storage.googleapis.com/78.0.3904.105/chromedriver_linux64.zip
c. Unzip: unzip chromedriver_linux64.zip
d. Copy chromedriver to proper folder: sudo cp chromedriver /usr/bin/chromedriver
e. Change owner of file to root: sudo chown root /usr/bin/chromedriver

f. Add run permissions for other users: sudo chmod 755 /usr/bin/chromedriver

3. Clone repository and install needed python packages:

a. Clone git repository with web scrapper: git clone https://github.com/Sauber59/Projekt-ETL.git
b. Go to project folder: cd WebScrapper
c. Create python3 virtual env: python3 -m venv venv
d. Activate virtual environment: source venv/bin/activate
f. Install required packages: python3 -m pip install -r requirements.txt
d. That is it, web scraper should be ready to run, you can display help on what you can do with =: python3 webscraper_lubimyczytac.py -h