Website monitor

Sanna Elomaa, 2022-11-06

This program periodically goes through a list of websites and does the following:
- Makes a GET request to the site - Checks the status code and text content of the response - Checks the elapsed time of the request - Logs the results to a log file

Running the program

The program can be stated by running

```
python website_monitor.py
```

Once started, the program runs until terminated manually (by `KeyboardInterrupt`, `Ctrl+C` on Windows).

Requirements

Python version 3.9+

Required Python dependencies are listed in the `requirements.txt` file. These need to be installed before running the program. This can be done by running the following command:

```
pip install -r requirements.txt
```

(Using a virtual environment is highly recommended!)

Usage

```
usage: website_monitor.py [-h] [-s SLEEP] [-v]

Monitor a list of websites given in config.toml

options:
  -h, --help            show this help message and exit
  -s SLEEP, --sleep SLEEP
                        Time to sleep (in seconds) between checks
  -v, --verbose         Log progress also to stdout
```

List of websites to monitor

The websites and their expected contents are defined in the configuration file
`config.toml`. The content is a string that is expected to be found in the response
text.

```
# config.toml

[[websites]]
url = "https://www.f-secure.com/en"
content = "F-Secure makes every digital moment more secure, for everyone."

[[websites]]
url = "https://en.wikipedia.org/"
content = "This text is not found in the content, so the content check fails"
```

Status code check

For all the websites we expect the GET request to return status code 200. If the
status code is something else, this is logged as an ERROR.

Content check

For all the websites we expect the GET response text to contain the content
string defined in the configuration file. If the status code check passes but the
text does not contain this content string, this is logged as a WARNING.

Request exceptions

If the request results in an exception, this is logged as an ERROR.

Checking interval

The checking interval in seconds is defined in the configuration file `config.toml`,
and it can also be given as a command line parameter when running the program.
This specifies the time (in seconds) the program sleeps before starting the website
checks again.

```
# config.toml

sleep = 5
```

Logging

The program logs the following information for each website: - URL - Status code - Content check result - Time elapsed

The log file is defined in the configuration file `config.toml`.

```
# config.toml

log_file = "website_monitor.log"
```

By running the program with the `--verbose` option, everything is also logged to stdout.

Example log

```
2022-11-06 15:23:19,472 - INFO - https://www.f-secure.com/en -
Status code: 200 - Content check passed: True - Time elapsed:
0:00:00.047692 2022-11-06 15:23:19,621 - WARNING - https://en.wikipedia.org/
- Status code: 200 - Content check passed: False - Time elapsed:
0:00:00.033539 2022-11-06 15:23:20,268 - ERROR - https://dummy.restapiexample.com/
- Status code: 406 - Content check passed: False - Time elapsed:
0:00:00.643957 2022-11-06 15:23:20,271 - ERROR - https://www.example.com/
- Request failed: HTTPSConnectionPool(host='www.example.com',
port=443): Max retries exceeded with url: / (Caused by NewConnectionError('<urllib3.connecti
object at 0x000002071ED99CA0>: Failed to establish a new connection:
[Errno 11001] getaddrinfo failed')) 2022-11-06 15:23:30,331 -
INFO - https://www.f-secure.com/en - Status code: 200 - Content
check passed: True - Time elapsed: 0:00:00.036651 2022-11-06
15:23:30,478 - WARNING - https://en.wikipedia.org/ - Status code:
200 - Content check passed: False - Time elapsed: 0:00:00.035405
2022-11-06 15:23:31,058 - ERROR - https://dummy.restapiexample.com/
- Status code: 406 - Content check passed: False - Time elapsed:
0:00:00.577696 2022-11-06 15:23:31,061 - ERROR - https://www.example.com/
- Request failed: HTTPSConnectionPool(host='www.example.com',
port=443): Max retries exceeded with url: / (Caused by NewConnectionError('<urllib3.connecti
object at 0x000002071ED99AF0>: Failed to establish a new connection:
[Errno 11001] getaddrinfo failed'))
```