

Lab04. Files and dictionaries

Script Languages (INZ002025)

Wojciech Thomas

Summer 2020/2021

1 Learning goals

After this lab you should be able to:

1. Read and write text files.
2. Use dictionaries to store and retrieve data.

2 Exercises

In this exercise, you will use a real-life Apache web server log file to be analysed.

1. Download the real-life log file from the course web site and save it in the project folder.
2. Review the page <https://httpd.apache.org/docs/2.4/logs.html> to learn the structure of the Access Log.
3. Write `run()` function that will call the remaining functions in this assignment. Call this function only if the file is run directly.
4. Write function `read_log` that will read the content of the log file and store it in the dictionary returned as a result. Which one of the log entry fields can be used as a key? How do you store the remaining data of each entry?
5. Write `ip_requests_number` function. Create a data structure that will keep the number of requests for the given IP address. Return this structure as a result.
6. Write `ip_find` function to find IP addresses that issued:
 - the biggest number of requests, if function is called as `ip_find()` or `ip_find(most_active=True)`
 - the smallest number of requests, if called as `ip_find(most_active=False)`Remember that there could be many IP addresses with the same number of requests.
7. Write function `longest_request()` that will return the longest request string (e.g. `GET /index.html`) along with its IP address. If there are many such addresses, return the one arbitrarily chosen.
8. Write `non_existent` function that will return all request strings with HTTP result code `Page not found`. In the result, each request string should appear only once, even if there are many entries with the same request string.