

Laboratorium 2 – Python Basics. Working with functions. Standard input/output

Script languages

Aims

1. Familiarization with the basic methods of building abstractions in Python.
2. Familiarization with processing data from standard input.

Assignments

1. Data Preparation

Please download [the log file from the NASA Kennedy Space Center WWW Apache server](#) in Florida. Unpack the file. Each log line contains:

- the hostname or IP address of the client making the request,
- a timestamp in the format "DAY MON DD HH:MM:SS YYYY",
- the HTTP protocol method along with the path to the requested resource,
- the HTTP response code (200 - for successful retrieval of the resource, 302 - for temporary redirection, 404 - when not found),
- the number of bytes in the response.

2. Reading data from standard input

Write a program that allows processing of successive lines of data read from standard input.:

```
python lab_2.py < NASA
```

Let the program print to the standard output the data that was sent to the input, and upon reading EOF, let it terminate its operation.

3. Functions

Please extend the program from task 2 with the following functionalities in Python, taking into account the following remarks:

- Let each functionality be implemented as a function in a separate file.
- Ensure that each function in the program has only one responsibility.
- Ensure separation of responsibilities, distinguish functions processing data from top-level functions outputting text to standard output.
- In case several functionalities require the use of the same functions, place them in a separate module that will be reused.
- Protect yourself from potentially badly formatted input using the exception mechanism.

Reducing functions:

- a. A function that outputs to standard output the number of requests with the code:
 - a. 200,
 - b. 302
 - c. 404.
- b. A function that outputs the total amount of data sent to hosts in gigabytes to standard output.
- c. A function that outputs to standard output the path and size of the largest resource.
- d. A function that outputs the ratio of image downloads (resources ending in *.gif, *.jpg, *.jpeg, *.xbm) to other resources to standard output.

Filtering functions:

- e. A function that outputs to standard output only rows with a response code of 200.
- f. A function that outputs to standard output resources downloaded between 22:00 and 6:00.
- g. A function that outputs to standard output only resources downloaded on Fridays.
- h. A function that outputs to standard output only requests from Poland (i.e., hosts with a domain name ending in .pl).

4. __main__

Please familiarize yourself with the following Python documentation chapter:
https://docs.python.org/3/library/__main__.html

Modify the programs so that each script call from the terminal uses the `if __name__ == '__main__':` construct.

5. Piping

Show that the functions you developed allow you to send data between each other using pipes. For example, to list the number of requests with code 404 from Poland.

```
cat NASA | python lab_3_h.py | python lab_3_a_c.py
```

Additional materials

1. <https://www.digitalocean.com/community/tutorials/read-stdin-python>
2. Alex Martelli, Anna Martelli Ravenscroft, Steve Holden, Paul McGuire, Python in a Nutshell, 4th Edition, Published by O'Reilly Media, Inc., *Rozdział 3*
[URL: <https://learning.oreilly.com/library/view/python-in-a/9781098113544/>]
3. Ryan's tutorial, Piping & Redirection
[URL: <https://ryantutorials.net/linuxtutorial/piping.php>]