

Lab03. Lists and tuples.

Script Languages (INZ002025)

Wojciech Thomas

Summer 2020/2021

1 Learning goals

After this lab you should be able to:

1. Use lists and tuples to store and retrieve information.
2. Use `range()` to generate lists of numbers.
3. Perform operations on strings: searching, slicing, replacing.

Try to solve all exercises using approaches presented so far. Avoid using regular expressions or file reading functions. Treat it as an exercise for your brain.

2 Exercises

1. Use the same approach as in the previous lab: read content of the web server log file from standard input.
2. Place your main (start) code in a function `run()`.
3. Add the code that starts function `run()` only, if your Python file is run directly, eg.:
`py my_app.py < log.txt`.
4. Create function `read_log` that reads all lines of the input at once and returns list containing entries from the log. Inside function:

1. Split each line (entry) into separate attributes.
2. Convert strings to the proper data type e.g. resource name - string, HTML result code - int, and so on.

Assume that each line of input log contains always correct number of items or is empty (contains only whitespaces or nothing at all).

3. Store each entry (non-empty line) as a tuple in the list.
4. Log as a DEBUG message number of lines read. Remember to add meaningful content in the message.
5. Log as a DEBUG message number of entries in the list.

5. Create a function `successful_reads` that:
 - takes as a parameter list of all entries in the log,
 - logs as a INFO message the number of entries in the resulting list,
 - returns a new list containing only resources successfully read (HTTP result codes 2xx¹).
6. Create a function `failed_reads` that:
 - takes as a parameter list of all entries in the log,
 - creates two separate lists containing entries with HTTP codes 4xx and 5xx respectively,
 - merges both lists into one,
 - logs as an INFO message the number of entries with 4xx and (separately) with 5xx result codes,
 - returns merged list.
7. Create a function `html_entries` that returns a list containing entries of successfully retrieved pages with `.html` extension (e.g. `/news/index.html`, but not `/index.php` or `/feed/`).
8. Create a function `print_html_entries` that prints a list containing entries of successfully retrieved pages with `.html` extension (e.g. `/news/index.html`, but not `/index.php` or `/feed/`).
9. Explain: which of the above functions can be called multiple times, and which not? Why?

¹There are many result codes describing success in HTTP protocol, however the most popular is 200 OK