

# Python Advanced Course

## List 5.

In the tasks below, take care of error handling so that any problems with the network connection or temporary unavailability of the server do not cause the program to terminate unexpectedly.

For class, you must complete one of these tasks. Each task is worth 5 points.

### Task 1.

Program a function `crawl(start_page, distance, action)` that reads the page at `start_page`, calls the `action` function with the page's content as an argument, and then executes that function for other pages that are linked from that page.

Since there can be many links, we limit the depth of the search with the `distance` parameter. Also, make sure not to process the same page twice.

Implement the solution as an iterator that returns tuples of the form `(url, function_result_action)`

```
for url, result in crawl("http://www.ii.uni.wroc.pl", 2, lambda text : 'Python' in text):  
  
    print(f"{url}: {result}")
```

Demonstrate using the function to search pages for sentences containing the word Python.

### Task 2.

Program a web monitoring program that checks to see if any page has changed its content. We assume that our program can monitor more than one page; we also assume that the checking is done periodically (e.g., every minute). If a change has occurred, the program prints the address of the page and what has changed.

In this task, we assume that the page layout changes rarely, only individual elements of this layout change (e.g. a single element in the DOM tree), so if the program detects a change, let it only return what has changed and the path to that place.

The addresses of the pages to be monitored can be specified in the source code.

### Task 3.

Write your own system that:

- for a given list of web page addresses it will create an index of words along with the number of occurrences. The list of addresses can be given in the source code;
- will implement functions operating on the resulting index:
  - search for the most popular word;
  - search which pages contain a given word.

This quest will have a sequel.