

Kurs rozszerzony języka Python

Lista 6.

Na zajęcia należy wykonać jedno z tych zadań. Każde zadanie jest warte 5 punktów.

Zadanie 1.

Zaprogramuj funkcję `crawl(start_page, distance, action)` która odczytuje stronę pod adresem `start_page`, wywołuje funkcję `action` której argumentem jest treść strony, a następnie wykonuje tę funkcję dla innych stron których odnośniki są podane na tej stronie.

Ponieważ odnośników może być dużo, głębokość poszukiwań ograniczamy przez parametr `distance`. Zadbaj też, by nie przetwarzać dwa razy tej samej strony.

Zaimplementuj rozwiązanie w formie iteratora, który zwraca krotki postaci `(url, wynik_funkcji_action)`

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

Zaprezentuj wykorzystanie funkcji do wyszukiwania na stronach zdań zawierających słowo *Python*.

Zadanie 2.

Zaprogramuj program do monitorowania stron internetowych, który sprawdza, czy jakaś strona nie zmieniła zawartości. Zakładamy, że nasz program może monitorować więcej niż jedną stronę; zakładamy też że sprawdzanie odbywa się co jakiś czas (np. co 1 minutę).

Zakładamy w tym zadaniu, że układ strony zmienia się rzadko, jedynie zmieniają się pojedyncze elementy tego układu, dlatego jeśli program wykryje zmianę to niech tylko zwróci to, co się zmieniło.

Zadanie 3.

Napisz własny system do indeksowania stron internetowych, który tworzy indeks słów występujących na stronach. Lista stron, dla których tworzymy indeks jest zadana jako argument wywołania.

Zaprogramuj też funkcję, która dla podanego indeksu i słowa zwróci, na jakiej stronie dane słowo jest najpopularniejsze.

Marcin Młotkowski