

Kurs rozszerzony języka Python

Lista 5.

W poniższych zadaniach zadbaj o obsługę błędów, aby ewentualne problemy z połączeniem sieciowym czy chwilowa niedostępność serwera nie spowodowały niespodziewanego zakończenia programu.

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 przetwarzanie dwa razy tej samej strony.

Zaimplementuj rozwiązanie w formie iteratora, który zwraca wyniki zawierające `url` oraz `wynik funkcji action` (np. jako krotkę czy obiekt z odpowiednimi polami). Przykładowe zastosowanie takiej funkcji może wyglądać tak:

```
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). Jeśli nastąpiła zmiana, program wypisuje adres strony oraz co się zmieniło.

Zakładamy w tym zadaniu, że układ strony zmienia się rzadko, jedynie zmieniają się pojedyncze elementy tego układu (np. pojedynczy element w drzewie DOM), dlatego jeśli program wykryje zmianę to niech tylko zwróci to, co się zmieniło i ścieżkę do tego miejsca.

Lista adresów stron do monitorowania może być zadana w kodzie źródłowym.

Zadanie 3.

Napisz własny system który:

- dla zadanej listy adresów stron internetowych stworzy indeks słów wraz z liczbą wystąpień. Lista adresów może być podana w kodzie źródłowym;
- będzie implementował funkcje operujące na powstającym indeksie:
 - wyszukiwanie najpopularniejszego słowa;
 - wyszukiwanie, na jakich stronach występuje zadane słowo.

To zadanie będzie miało kontynuację.