

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)`.

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

Zadanie 2.

Zaprogramuj system do monitorowania stron, 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

- przegląda strony i zapamiętuje liczbę wystąpień poszczególnych słów na poszczególnych stronach;
- zachowuje się podobnie jak pythonowy słownik, gdzie kluczem jest słowo, a wartością lista stron na których to słowo występuje (bądź lista pusta). Strony powinny być uszeregowane malejąco względem podanej liczby wystąpień. Możesz też zaproponować własną strategię rankowania stron.

Zakładamy, że indeksujemy tylko stronę wskazaną jako parametr odpowiedniej funkcji czy metody, oraz strony do których da się dojść po linkach a href w nie więcej niż z góry zadana liczba kroków.

Marcin Młotkowski