

Kurs języka Python

Lista 5.

Zadanie 1. Zaprogramuj moduł służący do przeglądania stron WWW i wykonujący dla każdej strony pewną akcję. Akcja powinna być funkcją, przekazywaną jako parametr do modułu, a której argumentem jest strona html. Przyjmij, że przeglądanie rozpoczyna się od strony, której adres jest podany jako argument wywołania. Następnie przeszukiwane są te strony, do których linki znajdują się na bieżącej stronie. Zadbaj o to, aby przeszukiwanie się nie zapętlało. Przyjmij, że głębokość przeszukiwań jest ograniczona i zadawana każdorazowo przy inicjowaniu przeszukiwania. Również akcja, która ma być wykonywana dla każdej strony winna być parametrem wywołania odpowiedniej funkcji modułu.

Zaimplementuj rozwiązanie w formie iteratora, który zwraca wyniki działania funkcji będącej parametrem.

Zaprezentuj wykorzystanie modułu do wyszukiwania zdań zawierających słowo *Python*.

Zadanie 2. Napisz pakiet użytecznych narzędzi webowych, które pomogą pielęgnować i ulepszać prowadzony serwis:

1. moduł przeglądający strony WWW (pliki *.html) podanym katalogu i podkatalogach i sprawdzający, czy odnośniki do innych stron czy obrazków są aktywne;
2. moduł przeglądający strony serwisu (również przeglądając pliki) i dla każdej strony (pliku) wypisujący, w których plikach są odnośniki do niej.

Zaimplementuj te moduły jako iteratory zwracające wynik odpowiednich akcji.

Zadanie 3. Napisz własną miniwyszukiwarę internetową, która

- 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.

Wystarczy, że indeksowane będą tylko wybrane strony, np. znajdujące się we wskazanych podkatalogach albo tylko strony do których da się dojść po odwołaniach w nie więcej niż k krokach od zadanej strony początkowej.

Na zajęcia należy wykonać jedno z tych zadań. Każde zadanie jest warte 4 punkty.

Marcin Młotkowski