Wyszukiwarka składa się zasadniczo z 3 części.

Klasa *DocumentParser* odpowiada za ściągnięcie strony, usunięcie z niej w całości tagów odpowiedzialnych za css oraz skrypty, wyciągnięcie zawartości pozostałych tagów i przetworzenie jej w listę słów. Następnie z tej listy odfiltrowywane są słowa nie mające znaczenia dla wyszukiwania (wczytywane z plików *stopwords_eng* i *stopwords_pl*).

Klasa *PageIndex* odpowiada za przechowywanie linków wraz z najczęściej występującymi w nich słowami (i ich liczebnościami). Tworzy słownik, którego kluczami są adresy stron a wartościami listy krotek (liczebność, słowo). Klasa automatycznie zapisuje swój stan do pliku pickle (domyślnie *index.pickle*).

Klasa *SearchEngine* to już właściwa klasa wyszukiwarki internetowej. Korzysta z obiektu *PageIndex*, w nim szuka słów. Jeśli znajdzie mniej niż 3 wyniki pyta użytkownika czy kontynuować wyszukiwanie dla podobnych słów (tzn. krótszych z każdą iteracją o 1 znak aż do 2/3 długości słowa w celu znalezienia tego samego słowa w innej formie).

Wyszukiwarka najlepiej działa na linkach do artykułów, wpisów na blogach itd. dlatego napisałam jeszcze generator linków do angielskiej wikipedii. Zapisuje on podaną jaką pierwszy argument liczbę wylosowanych (unikalnych) adresów do pliku podanego jaki drugi argument. Plik *pages* zawiera już przygotowane w ten sposób adresy, klasa *PageIndex* domyślnie czyta z niego adresy.

W celu włączenia wyszukiwarki należy uruchomić skrypt *SearchEngine.py*
Wersja Pythona: 2.7.12
Pisałam pod Linuxem, ale raczej powinno działać wszędzie.