

# Wyszukiwanie ścieżek w Wikipedii

## I Wprowadzenie

Działanie programu polega na znajdowaniu ścieżek pomiędzy dwoma pozycjami (stronami) w polskiej Wikipedii. Problem można sprowadzić do problemu grafowego, gdzie węzłami są strony Wikipedii, zaś krawędziami odnośniki (linki) na danej stronie do innych stron Wikipedii. Ścieżka między dwoma stronami istnieje, gdy na tej pierwszej stronie istnieje klikalny link do tej drugiej strony. Jest to więc rodzaj grafu skierowanego, gdyż nie wiadomo, czy istnieje link powrotny. Problem rozwiązuje zmodyfikowanym algorytmem BFS, w którym graf nie istnieje w chwili wywołania algorytmu, tylko jest tworzony dynamicznie, na bieżąco.

Algorytm Broad-First-Search jest jednym z najprostszych algorytmów przeszukiwania grafów. Przechodzenie grafu rozpoczyna się od zadanego wierzchołka początkowego  $s$  i polega na odwiedzeniu wszystkich osiągalnych z niego wierzchołków w pierwszej kolejności odwiedzając wszystkich sąsiadów  $s$ . Wynikiem działania algorytmu jest drzewo przeszukiwania wszerz o korzeniu w  $s$ , zawierające wszystkie wierzchołki osiągalne z  $s$ . Do każdego z tych wierzchołków prowadzi dokładnie jedna ścieżka z  $s$ , która jest jednocześnie najkrótszą ścieżką w grafie wejściowym.

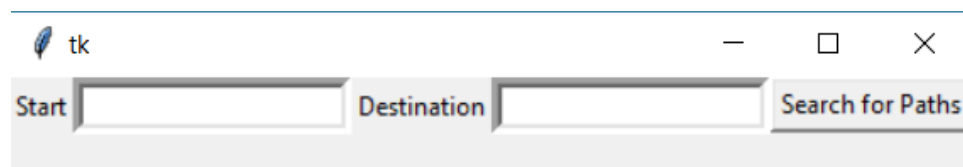
## II Opis interfejsu

Aby uruchomić program, należy ściągnąć i zainstalować bibliotekę Requests

<http://docs.python-requests.org/en/master/user/install/#install>

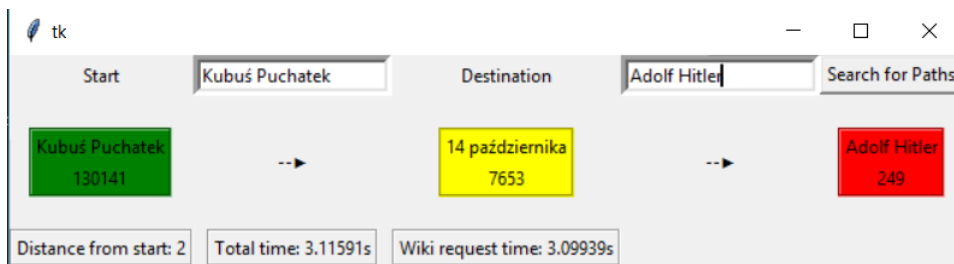
Następnie można uruchomić program z poziomu konsoli poleceniem "python wiki\_path\_finder.py"

Powinno pojawić się okienko aplikacji:



W polu tekstowym Start podajemy tytuł strony z polskiej Wikipedii, która będzie początkiem ścieżki. Następnie w polu tekstowym Destination podajemy tytuł strony, która będzie końcem ścieżki. Aby zacząć wyszukiwanie klikamy przycisk Search for Paths.

Gdy program znajdzie taką ścieżkę okienko aplikacji powinno wyglądać następująco:



Na zielono zaznaczona jest strona początkowa, na żółto strony pośrednie, a na czerwono strona końcowa. Każdy panel strony posiada jej tytuł w pierwszym rzędzie, oraz w drugim identyfikator strony. Identyfikator jest linkiem, który otwiera w domyślnej przeglądarce internetowej stronę na Wikipedii. Dzięki temu można samodzielnie sprawdzić poprawność rozwiązania.

Na dole okienka aplikacji znajdują się parametry wyszukiwania.

### III Uwagi

Bardziej szczegółowe informacje dotyczące implementacji znajdują się w plikach html generowanych za pomocą programu pydoc.

W skrócie działanie programu polega na budowaniu zapytań do API Wikipedii za pomocą biblioteki Requests. Każde zapytanie zwraca plik formatu json, który posiada tytuł strony oraz tablicę linków znajdujących się na tej stronie. Każda strona zostaje opakowana w obiekt typu Node. Tablica linków zostaje zgromadzona w liście sąsiedztwa węzła.

W zmodyfikowanym algorytmie BFS przeszukujemy tablicę sąsiedztwa danego węzła. Jeśli nie ma w niej poszukiwanej strony końcowej to wrzucamy zawartość listy na stos i zaczynamy ponowne przeszukiwanie każdego elementu ze stosu.

Do GUI aplikacji wykorzystano bibliotekę tkinter.

### IV Podsumowanie

Program działa poprawnie i potrafi odnajdywać ścieżki pomiędzy stronami, jednak jest w dużym stopniu ograniczany przez API Wikipedii. Na podstawie analizy parametrów wyszukiwania można dojść do wniosku, że najbardziej czasochłonne w odnajdywaniu ścieżek jest wysyłanie i odbieranie zapytań z API Wikipedii (ponad 90% całkowitego czasu). Również gdy program będzie kontynuował wyszukiwania przez dłuższy czas, API Wikipedii może w ramach bezpieczeństwa zerwać połączenie.

### V Literatura

-tutorial do wysyłania requestów:

<http://docs.python-requests.org/en/master/user/quickstart/#make-a-request>

-tworzenie zapytań w API Wikipedii

<https://www.mediawiki.org/wiki/API:Query>

-przykładowy plik json:

[https://pl.wikipedia.org/w/api.php?action=query&prop=links&meta=&titles=Harry%20Potter%20\(posta%C4%87\)&pllimit=500](https://pl.wikipedia.org/w/api.php?action=query&prop=links&meta=&titles=Harry%20Potter%20(posta%C4%87)&pllimit=500)

-tutorial biblioteki tkinter:

[https://www.tutorialspoint.com/python/python\\_gui\\_programming.htm](https://www.tutorialspoint.com/python/python_gui_programming.htm)

-algorytm BFS

[https://pl.wikipedia.org/wiki/Przeszukiwanie\\_wszerz](https://pl.wikipedia.org/wiki/Przeszukiwanie_wszerz)