

Wikipedia

Naho Yokoyama

Overview

This is a program that analyzes characteristics of the maps (linkage of pages) in the Wikipedia web page. This program summarizes details of the longest titles, most linked pages, shortest path between two pages, shortest titles and how many are them. This document will focus on the functions for most linked pages, shortest path, and shortest titles algorithm.

Most linked pages

The implementation of the most linked pages is based on the random surfer model of the page rank algorithm. Here, we will create a rating called “page rank” which has an initial value of 1.0 for each web page. Then, we will repeat trials of giving 85% of its page rank to its neighbors evenly and 15% of its page rank to the whole (100% if no neighbor pages) until there is no change to the page rank.

Shortest path

This program is based on the BFS algorithm, which is a searching algorithm using a queue because this is an effective way to implement a shortest path function. However, to trace back the route, I am using another array to keep track of the route (where did I come from) so that after I find the shortest path, I can go back from the end to the beginning for a traceback.

Shortest titles

The shortest title algorithm is similar to the one for longest titles algorithms; instead of getting the ones with the longest length, we will get the shortest ones. Since I found out that there are many titles with the length=1, I decided to only show some examples and just print out the total number of pages.

Output result

The output of the program is shown below:

The longest titles are:

日本国とアメリカ合衆国との間の相互協力及び安全保障条約第六条に基づく施設及び区域

並びに日本国における合衆国軍隊の地位に関する協定の実施に伴う刑事特別法
一般社団法人及び一般財団法人に関する法律及び公益社団法人及び公益財団法人の認定等
に関する法律の施行に伴う関係法律の整備等に関する法律案

日本国とアメリカ合衆国との間の相互協力及び安全保障条約第 6 条に基づく施設及び区域
並びに日本国における合衆国軍隊の地位に関する協定

日本国とアメリカ合衆国との間の相互協力及び安全保障条約第六条に基づく施設及び区域
並びに日本国における合衆国軍隊の地位に関する協定

国際的な協力の下に規制薬物に係る不正行為を助長する行為等の防止を図るための麻薬及
び向精神薬取締法等の特例等に関する法律

民放 5 局史上最大のコラボレーション!地デジ夏祭り 2006 全部見せます!ナゴヤのテレビ"
過去""現在"そして"未来"

アイルランドの貧民の子供たちが両親及び国の負担となることを防ぎ、国家社会の有益な
る存在たらしめるための穏健なる提案

ドナウダンプフシファールトゼレクトリツィテーターテンハウプトベトリープスヴェルクバウ
ウンターベアムテンゲゼルシャフト

マルキ・ド・サドの演出のもとにシャラントン精神病院患者たちによって演じられたジャ
ン＝ポール・マラーの迫害と暗殺

くりいむしちゅーも観ながらいろいろゴチャゴチャ言ってますけども...笑いのタマゴ L サ
イズ（おひとり様何回でも）

偽造カード等及び盗難カード等を用いて行われる不正な機械式預貯金払戻し等からの預貯
金者の保護等に関する法律

ルイージ・アメデーオ・ジュゼッペ・マリーア・フェルディナンド・フランチェスコ・デ
ィ・サヴォイア＝アオスタ

中居正広のテレビ 50 年名番組だヨ!全員集合笑った泣いた感動したあのシーンをもう一度
夢の総決算スペシャル

タウマタファカタンギハンガコアウアウオタマテアトゥリプカカビキマウンガホロヌクボ
カイフェヌアキタナタフ

アウグステ・ヴィクトリア・フォン・シュレーズヴィヒ＝ホルシュタイン＝ゾンダーブル
ク＝アウグステンブルク

The most linked pages are:

ISBN 52641

['渋谷', 'マクドナルド', 'Twitter', 'パレートの法則']

The total path length was: 4

The most popular page is: 英語

Examples of shortest titles are:

暦

C
水
島
氷
海
川
河
陸
本

There are 2174 titles having the same length