Project Goals

Summary

We will be creating a program to graph the links between Wikipedia pages (inspired by the Wikipedia Game where each player has to get from one page to the other only using hyperlinks), and we plan on making this program be able to be queried by the user to find the most optimal path between pages. In the end, this should be a useful tool for people who want to get better at the wikipedia game. We will be using the 2013 Wikipedia link set, in the interest of space and memory.

Source of dataset: https://snap.stanford.edu/data/enwiki-2013.html

Format

File: enwiki-2013.txt
Size: 1.5 GB
Explanation: Contains the edge relationships from one Wikipedia link to another.
Shown: First six entries.

FromNodeId	ToNodeId1
2	0
0	2
4	3
7	5
5	6
7	8

File: enwiki-2013-names.csv
Size: 122 MB
Explanation: Contains the mapping from node_id to name.
Shown: First six entries.

node_id	name
0	Alexander Seton (d. 1332)
1	Alexander de Seton
2	Battle of Wester Kinghorn
3	Kharqan Rural District
4	Kharqan
5	Talageh-ye Sofla

Algorithms (in order of implementation)

Traversals:

1. BFS (Breadth First Search)

Graph Algorithms:

1. Dijkstra's Algorithm
2. Landmark Path (shortest path from a to b through c)