Matthew Ulmer

# Degrees of Wikipedia

## Algorithm

For this assignment we are given two text documents. The first contains Wikipedia link titles on separate rows and the second contains links to other Wikipedia articles, represented by the row the link title appears on. The goal is to find the shortest possible path between two articles and to do so I used Dijkstra's algorithm. But before using Dijkstra's algorithm, it is necessary to populate the graph. In order to do this I iterate through the links and set directed edges from each article to the article they point to. It is important that they are directed edges as it is not necessarily true that there will be a link to go back. Afterwards, Dijkstra's algorithm is performed and the path array is returned. In this path array, the value at a specific index represents the next index to jump to in the path. If a value at a particular index is null, that means that the link represented by that index value cannot reach the target link as they are in disjoint sets. Because of this, it is only necessary for me to check if the value in path at index source is equal to null once. If it is equal to null, that means that there is no path from the source to target. Otherwise, I just keep updating the value of source with the value at position path[source] until the target is reached.

## Format

- Start of first line is the source
- End of last line is the target
- In the middle of two link titles there is: (cur_source_index -> path[cur_source_index])

**1.**

Invincible_(Michael_Jackson_album) (651->4328) Venezuela

Venezuela (4328->4575) World_War_I

World_War_I (4575->2571) Peter_O'Toole

Peter_O'Toole (2571->1260) Laurence_Olivier_Awards

**2.**

Halle_Berry (320->501) Hugh_Jackman

Hugh_Jackman (501->935) John_Travolta

John_Travolta (935->1958) Moment_by_Moment

**3.**

Jennifer_Dunn (836->1019) July_29

July_29 (1019->799) January_7

January_7 (799->2183) Nicolas_Cage

Nicolas_Cage (2183->7) Forrest_Gump_(film)

**4.**

President_of_Argentina (2705->1782) May_25

May_25 (1782->2342) October_28

October_28 (2342->980) Julia_Roberts

Julia_Roberts (980->3928) The_Player

**5.**

Peter_Twinn (2578->1486) London

London (1486->1500) Los_Angeles,_California

Los_Angeles,_California (1500->1502) Los_Angeles_Daily_News

**6.**

The_Peninsula_Manila (3923->4262) United_States

United_States (4262->2320) October_17

October_17 (2320->2524) Paxton_Whitehead

**7.**

Laura_Branigan (1255->3203) Saturday_Night_Live

Saturday_Night_Live (3203->1887) Mike_Myers_(actor)

Mike_Myers_(actor) (1887->3116) Rupert_Everett

**8.**

Guy_Bolton (286->3310) September_6

September_6 (3310->2620) Pink_Floyd

Pink_Floyd (2620->4217) UK_Singles_Chart

**9.**

Thriller_(genre) (4037->3557) Stephen_King

Stephen_King (3557->995) July_16

July_16 (995->1032) June_10

June_10 (1032->2760) Prince_Hashim_bin_Al_Hussein

**10.**

Varsity_Blues_(film) (4322->2481) Paramount_Pictures

Paramount_Pictures (2481->4342) Viacom