**Wikipedia Graph**

This program receives a Wikipedia page, and spans a directional, weighted graph, using the links in the article. This program allows the user to enter a url, and then select two site. If there is a path, the program will display the shortest path taken, as well as the most similar site of each site visited.

The graph structure is an array list of the Node class, which contains the site url, name, contents, and links extracted. The node also consists of an array list of edges that are an array index, and a weight indicating the difference between the pages.

To traverse, I use Dijkstra’s shorted path algorithm to find the path between two sites. To find the shorted path, I use a priority queue, which allows for a sorted set, and also allows updating of what is already in the queue.

To find the similarity between two sites, I used cosine word frequency to find the similarity. This is used to calculate the weight of the each edge, as well as finding the most similar site when displaying the most similar site of each one that was visited.

To get certain contents from each url, I used JSoup libraries to extract contents, as well as the links from each article. To filter and limit the links, I use a stream for link filtering, as well as limiting to a desired amount.