Introduction:

Due to our middle/high schools' pesky internet blockers, we've all had the experience of playing a game called "Wikipedia Race", where we start on an arbitrary Wikipedia page and try to get to a target Wikipedia page using only Wikipedia hyperlinks. Our program is an automated Wikipedia Racer. Given two URLs, this program tries to find a path of Wikipedia links between them.

Interacting with the Program:

Getting Started:

First, find the Main.java file and run it. Then, proceed to the next step.

Welcome:

```
************************************
**********Wikipedia  Racer**********
************************************
Welcome to the Wikipedia Racer!
You may enter a start link and an end link
All input links *MUST* be Wikipedia links.
HIT ENTER TO START!
```

This is the welcome page to our program. To run the program, simply run the "Main.java" file included in the folder. Note that all links input into the program ***MUST*** be valid links of real Wikipedia articles. Here is an example of a valid link that is not a valid Wikipedia article here.

Input:

```
Please enter a start link :
```

The program will first ask you for a start link. You can type whatever you want, but the program may not be too happy about it...

```
Please enter a start link :
www.google.com
Oops. That's not on Wikipedia. Try again!
Please enter a start link :
```

If you enter a link that is not a part of https://en.wikipedia.org then it is rejected by the program! If you enter a correct link however, you can then enter the end link.

```
Please enter a start link :
https://en.wikipedia.org/wiki/Economic_history_of_Argentina
Please enter an end link:
```

Wow! Now you can enter an end link. Again, don't try to cheat and leave wikipedia...

```
Please enter a start link :
https://en.wikipedia.org/wiki/Economic_history_of_Argentina
Please enter an end link:
falafel
Oops. That's not on Wikipedia. Try again!
Please enter an end link :
```

Paths:

Once you've entered two wikipedia links, you'll get some path results! In this example, we've done a path of length 0 (the same link as the start and end node).

```
Please enter a start link :
https://en.wikipedia.org/wiki/91st_Academy_Awards
Please enter an end link:
https://en.wikipedia.org/wiki/91st_Academy_Awards
**********
**********Path Results**********
**********
[https://en.wikipedia.org/wiki/91st_Academy_Awards]
```

Then, the program will print out a list of links that a user could take to get from one page to another. Due to the constraints of web scraping speeds, we have limited the results to the first 10 links on each page.

```
Please enter a start link :
https://en.wikipedia.org/wiki/91st_Academy_Awards
Please enter an end link:
https://en.wikipedia.org/wiki/Hollywood_Boulevard

************************************
**********Path Results**************
************************************
[https://en.wikipedia.org/wiki/91st_Academy_Awards,
 https://en.wikipedia.org/wiki/Dolby_Theatre, https://en.wikipedia.org/wiki/Hollywood_Boulevard]
```

If a path hasn't been found after 1000 pages have been traversed, you'll get a little message telling you the path was too long.

```
Please enter a start link :
https://en.wikipedia.org/wiki/California
Please enter an end link:
https://en.wikipedia.org/wiki/British_people

************************************
**********Path Results**************
************************************
Whoops. That path is a lil too long.
```

Quitting:

Once you're satisfied with your path, you can quit the program by typing the keyword "quit"

```
Please enter a start link :
https://en.wikipedia.org/wiki/British_people
Please enter an end link:
https://en.wikipedia.org/wiki/British_people

************************************
**********Path Results**************
************************************
[https://en.wikipedia.org/wiki/British_people]
quit
Goodbye!
```

Miscellaneous Notes:

The number of outgoing links in an average Wikipedia article is high (some pages have >1000 outgoing links). Therefore, to prevent taking too much time or running out of memory, the program defaults to only considering 1000 pages total. In addition, for each page, the program only considers the first 10 outgoing links on the page. This limits the actual runtime of the program to around 1 minute per query.

IMPORTANT: As a consequence of these constraints, depending in the input, the output path of the program may neither be the shortest path or a path at all. If the program does not find a path with the allotted 1000 considered pages, then it will return null. Note