# Description of the Problem

## Outline of problem.

I intend to design a program to find paths between webpages. The program will include the following features: a help screen, text based ui to help user find their way, a database of webpages crawled that will be created using SQL, with a linked database of pages they have linked to and a pathfinding system.

The end users of my program will be people willing to find orphan links in webpages if a search system is not used, therefore will probably be tech literate.

My project meets the advanced higher computing requirements as it will have a UI suitable for tech literate users with validation for if the pages have valid urls by using a try catch with a get() procedure and checking the code sent from the sever is not 404 and that the domain exists. My project will interface with an SQL database, creating a database and writing and reading URL’s from it.

A piece of software that will be able to crawl webpages. The webpages crawled and the pages they link to should be stored in a database including the link to them using SQL. The database can then be read and a path can be found. And then the path is displayed to the user.

Scope the clearly defined outline of what the solution will deliver in terms of functionality

Boundaries: are the limitations of the project

Constraints the restrictions that apply to the development.

# UML

# Requirements

## End user requirements:

User must be able to use text-based UI to input a starting website, and an end website to find a path to.

User must be able to input the number of moves they wish it to be done in.

The user must be able to view the requested path, or receive an error message that there is no path

## Functional requirements:

The program must display a UI that can take in a starting page and end page with number of moves.

The program will be able to crawl a url and find all links on the url, follow them, and repeat the process until the maximum jumps is achieved.

The webpage’s url is to be stored then all the links leading off also need to be stored in a database using SQL.

These are to be stored in a database using sql queries.

The program will then call the separate path finding algorithm.

SQL queries then need to be written into a 2d array.

(possible: sort 2d array so that easier to read.)

Then a node map object is created using the data in the 2d array.

Then using a pathfinding algorithm, find A path (not shortest) from first link to second link in the maximum number of jumps given.

# Project plan

## Identified tasks: