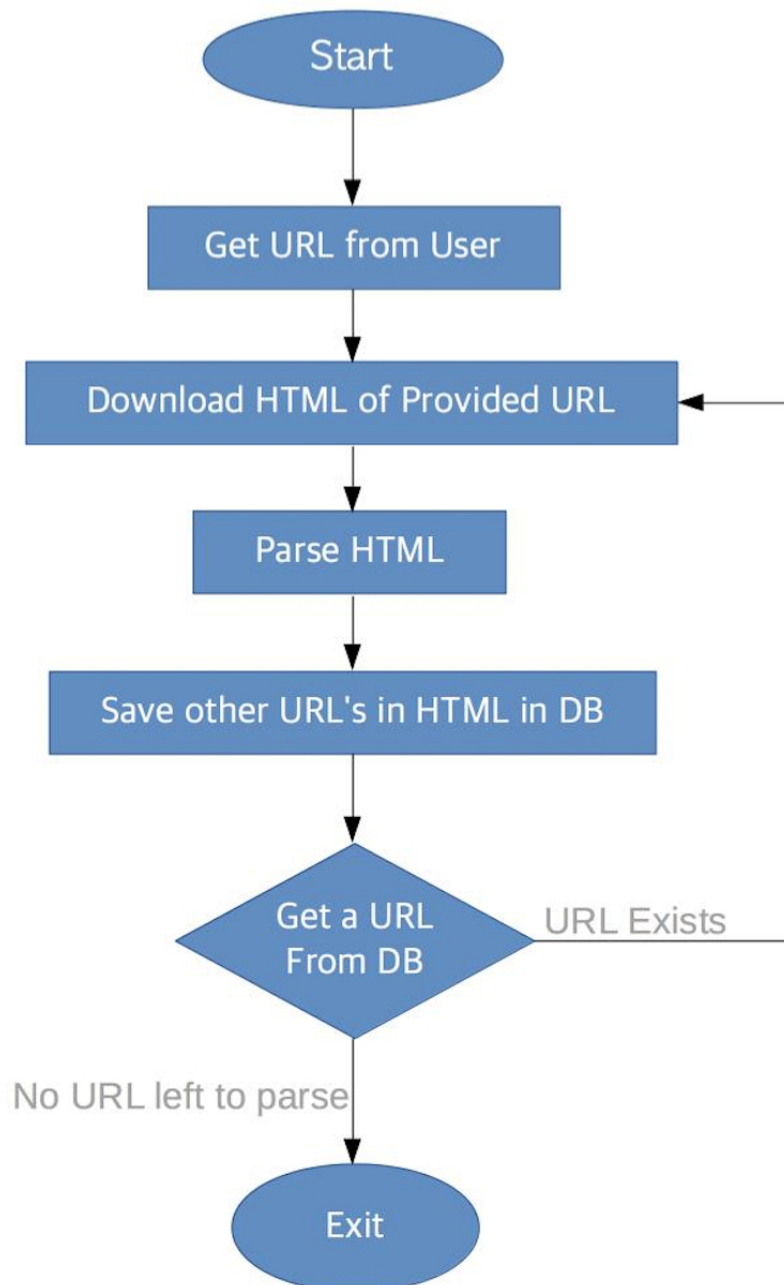


Assignment – Web spider and scrapper



Spider and scrapper

Create an app in any language where user can input a URL in a HTML form. The Application would parse the URL and extract each and every URL from it and URL's parsed will be stored and parsed in the same way recursively. Flow chart for the problem is given below:



Tasks

This app can be divided into following tasks which should be completed in sequence:

- Create a form with proper validation for URL input
- Fetch HTML content of the URL, Parse it to get more URL's and save each URL in a database table
- Fetch an unprocessed URL from table and apply the same process
- Create a **realtime dashboard** which will show following things
 - Total number of URL's parsed
 - Total number of Queued URL's (URL's stored but not yet parsed)
 - Data table showing list of latest URL's parsed
 - A graph showing number of URL's parsed per minute

Important points

- Tasks provided should be done in sequence, you can leave some tasks if time is running out
- Assignment should complete in provided time, late submission will not be considered
- Submitted app should work with all tasks completed, app can be submitted over Git or compressed file
- You can choose any design template of your choice
- Scoring will be based on efficiency of methods used, coding style, number of tasks completed

Points will be decided upon

- Completion of tasks
- Time taken to parse a full domain
- Coding style and standards
- Number of Bugs
- Tools used