

Web Crawler Application

Created by:

Poolesville High School Team 1

Suriya Kandaswamy

Peiyi Lin

Umesh Padia

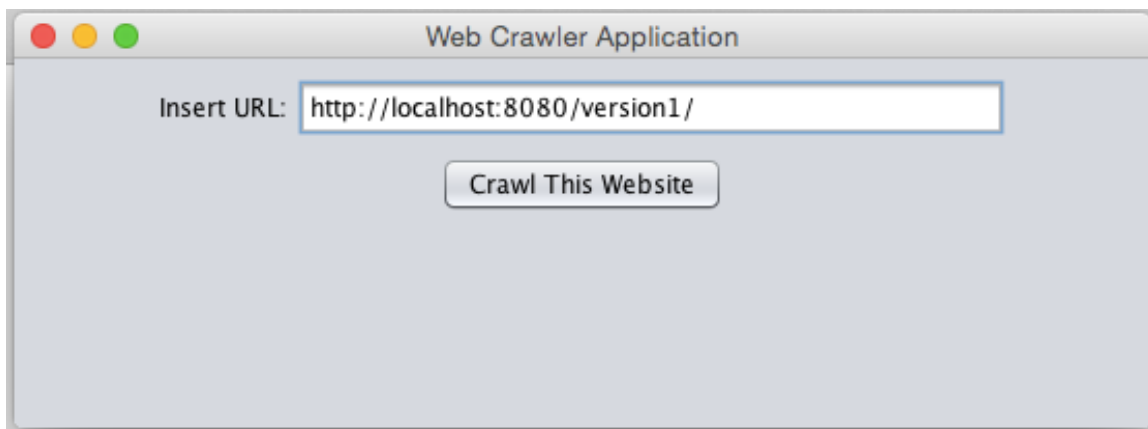
Seong Joon Yoo

Introduction

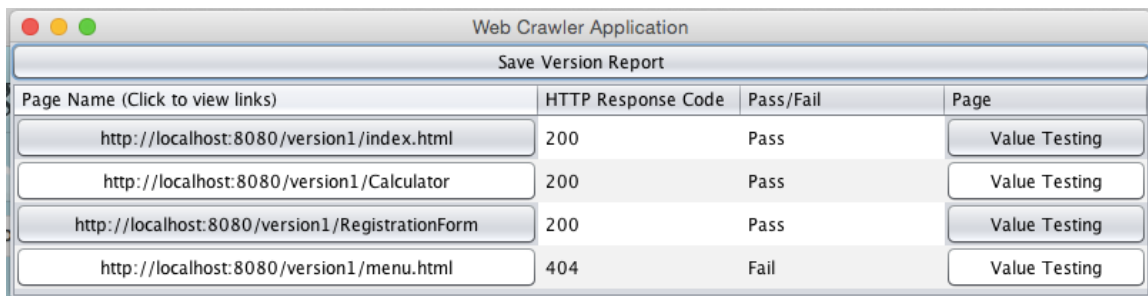
Our application is a site parser program that allows a user to give a valid site address and have the program parse through it for them, returning such information as internal links and input value testing options.

Application Usage

First enter a site address to parse and then clicks the button “Crawl This Website”:



You will then be led to another screen with a table of all the accessible pages in the site:

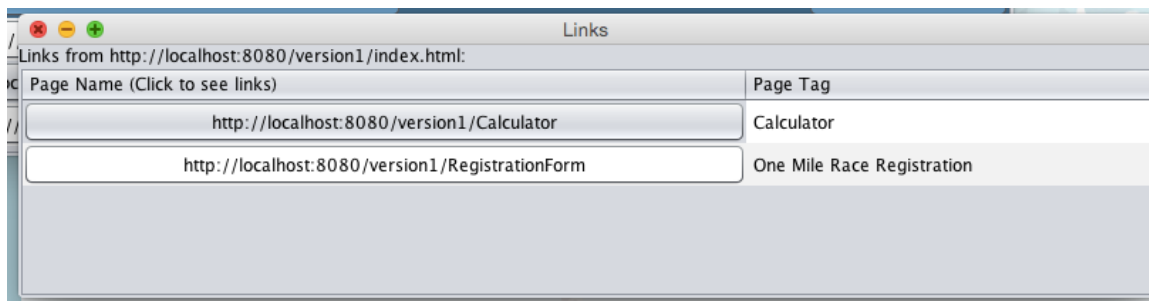
A screenshot of a web browser window titled "Web Crawler Application". The window displays a table with the title "Save Version Report". The table has four columns: "Page Name (Click to view links)", "HTTP Response Code", "Pass/Fail", and "Page". There are four rows of data. Each row has a button in the "Page" column labeled "Value Testing".

Page Name (Click to view links)	HTTP Response Code	Pass/Fail	Page
http://localhost:8080/version1/index.html	200	Pass	Value Testing
http://localhost:8080/version1/Calculator	200	Pass	Value Testing
http://localhost:8080/version1/RegistrationForm	200	Pass	Value Testing
http://localhost:8080/version1/menu.html	404	Fail	Value Testing

The first column has buttons that you can click on to be led to another window with all outgoing hyperlinks from that page. The second column

has the HTTP response code of the page. The third column provides information about whether the page has passed or failed. A page fail is defined by a 4** or 5** response code. The last column allows you to test various input values.

If you click on one of the buttons in the first column, it leads to a window with all linked pages:



You can then choose to follow the path of page links, or close the window and proceed to examine the other information in the previous window.

If you click on one of the buttons in the last column, it leads to a window for testing input values:

This window enables you to enter correct data for the program to automatically test.

A registration form with the following fields and controls:

- firstname
- lastname
- address
- city
- state
- zipcode
- email
- telephone
- miletime
- ☐ Click Register button.
- ☐ Click Clear Form button.
- Input Correct Values

The next window allows you to input the data to test against the correct data.

A web form with a light gray background. It contains the following elements:

- Input fields for: **firstname**, **lastname**, **address**, **city**, **state**, **zipcode**, **email**, **telephone**, and **miletime**.
- Two checkboxes with labels: ☐ Click Register button. and ☐ Click Clear Form button.
- A button labeled **Test Input Values** at the bottom.

If the state change of the site for the test data matches the test data the user enters as correct data, a value of true is stored for if the input is working.

When you are done evaluating the site, you can click on a button to produce a crawl report. This report is saved as a .csv file that can easily be opened in Microsoft Excel and be viewed as a spreadsheet. You can compare csv files from different versions of your site to see what changes have been made and what problems still exist or may have arisen in your site.

