**CS408 Project Backlog**

**Team 2 The Info Crawler**

**Hanxun Huang, Sun Kai, Yijie Wu, Longding Zhang, Di Yao.**

**1.Problem Statement**

We would like to create a tool that can help us grab information from websites. For example, we would like to collect all the names and abbreviations of student organizations from “Boilerlink.purdue.edu” or check the available seat for a course on “mypurdue.purdue.edu”. Also, we would like to make this tool available both on command line and GUI.

**2.Background Information**

We encountered a problem during the CS307 project last year. We wanted to store all the names and abbreviations of student organizations into the database; however, we were not able to achieve our goal due to the massive amount of data. Therefore, we wrote a simple web crawler to grab and store the information. Now, we would like to create a tool with the ability to collect and retrieve information from most of the websites.

**3.Environment**

We will be using Java and standard libraries. We may also need to set up a server to handle the email notifications.

**4.Requirements**

Backlog ID starting with “F” means functional requirement, “NF” means non-functional requirement.

**a) Functional Requirements**

|  |  |  |  |
| --- | --- | --- | --- |
| **Backlog ID** | **Functional Requirements** | **Estimate Time** | **Use Case** |
| **F-1** | As a user, I would like to provide a target URL for the program to use. | 5 | 3 |
| **F-2** | As a user, I would like to crawl the data by providing either keyword or HTML element. | 5 | 3 |
| **F-3** | As a user, I would like to choose between the repeat mode and periodic mode. | 8 | 3 |
| **F-4** | As a user, I would like to have all the information I need from the URL before search. | 5 | 3 |
| **F-5** | As a user, I would like to have the ability of using keywords to search and view the result. | 40 | 4 and 5 |
| **F-6** | As a user, I would like to choose a different method such as algorithm to do the crawling. | 20 | 3 |
| **F-7** | As a user, I would like to run different jobs at the same time. | 10 | 3 |
| **F-8** | As a user, I would like to receive notifications when the job is done. | 5 | 4 |
| **F-9** | As a user, I would like to access this tool using the command line. | 15 | 1 |
| **F-10** | As a user, I would like to access this tool through GUI. | 30 | 2 |
| **F-11** | As a user, I would like the ability to filter out the unwanted crawled data. | 30 | 4 |
| **F-12** | As a user, I would like to save the results into a file. | 1 | 5 |
| **Total** |  | 175 |  |

**b) Non-Functional Requirements**

|  |  |  |
| --- | --- | --- |
| **Backlog ID** | **Non-Functional Requirements** | **Estimate Time** |
| **NF-1** | As a user, I would like to see the help instruction | 5 |
| **NF-2** | As a user, I would like to finish the job in the least amount of time | 15 |
| **NF-3** | As a user, I would like to have accurate information with proper formation through using the tool | 10 |
| **Total** |  | 35 |

**c) Use cases**

**Case 1: Open command line and run the program**

|  |  |
| --- | --- |
| **User action** | **System Response** |
| 1.User open a command line |  |
| 2.User input command according to the instruction | 3.Program start to run |

**Case 2: Open GUI and run the program**

|  |  |
| --- | --- |
| **User action** | **System Response** |
| 1.User click the Icon of the program | 2.Program start to run |

**Case 3: Select program mode and set up**

|  |  |
| --- | --- |
| **User action** | **System Response** |
| 1. Run the program (Case 1 or Case 2) | 2.Program start to run |
|  | 3.Ask user to choice Repeat Mode or Periodic Mode |
| 4a. User choice Repeat Mode | 5a. Program ask for the URL and increment times |
| 4b. User choice Periodic Mode | 5b, Program ask for the URL and time interval |
| 6. User input keyword or select HTML element. |  |
| 7. User chooses method or algorithm. |  |
| 8. User chooses if he or she wants to receive notification when it’s done. |  |
| 8. User choose How many jobs want to run at the same time | 9. If it is more than one, repeat step 4-7 until the entire job is set. |
| 10. User start the job | 11. Program start to execute |

**Case 4: Using the filter**

|  |  |
| --- | --- |
| **User action** | **System Response** |
| 1a. User receive notification and open the program |  |
| 1b. User receive the result from the program | 2. Ask if the user need to use the filter |
| 3a. User do not need to use the filter |  |
| 3b. User need to use the filter | 4. Ask for the user to input remove keyword |
| 5. User input remove keyword | 6. Remove all the keyword and display the new result |
|  | 7. Ask if the user need to remove more keyword |
| 8a. User do not need to remove more keyword | 9a. Save and exit (Case 5) |
| 8b. User need to remove more keyword | 9b. Repeat Step 4-7 |

**Case 5: Save and exit**

|  |  |
| --- | --- |
| **User action** | **System Response** |
| 1. User confirm do not need to remove more word | 2.Ask if the user need to save the result |
| 3a. User need to save the result | 4a. Ask user to input file name |
| 5. User input the file name | 6. Save the result. |
| 3b. User does not need to save the result. |  |
| 7. User input command or click exit button in the GUI | 8. Exit |