Web Crawler

Project Setup

**1. Introduction**

Web crawler is a software that searches for links in the web that is potentially tracking your web activities. This document is provided for the requirement and design specification so that developer can get the clear idea of the system and also the design specification of the system. It contains also prototype UI which is very useful for the users to know about front-end view.

**2. Requirements**

1. Download web page content

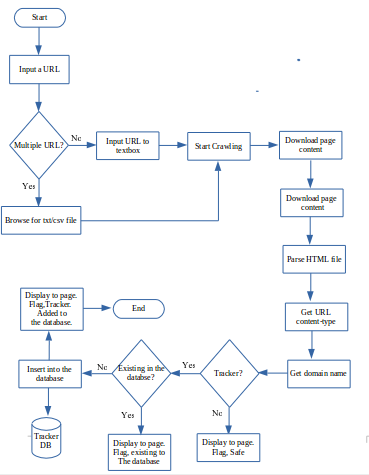
2. Parse html file and search for URL strings

3. If a third Party URL is found, check if its existing in the Tracker Database

4. Add third party URL to the database if its not existing

5. Data should be exported as txt/csv file

**3. Process Flow**



**4. Software development tools**

1. Text Editor

2. Php5

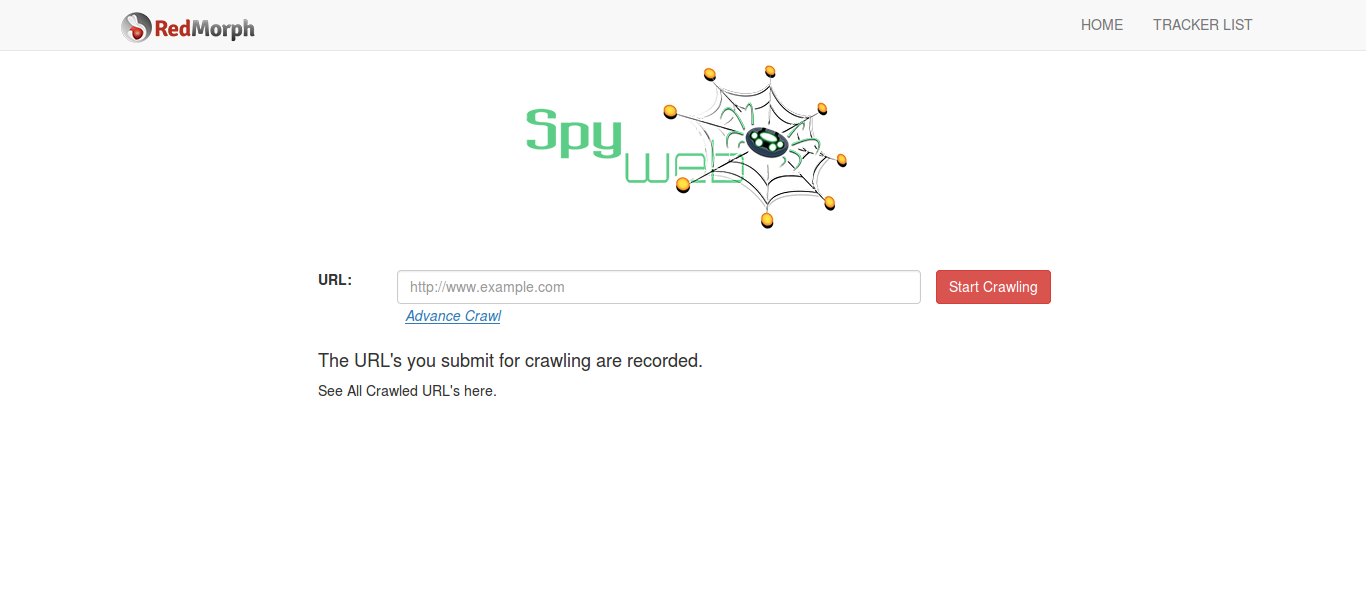
3. MySQL-server

4. phpMyAdmin

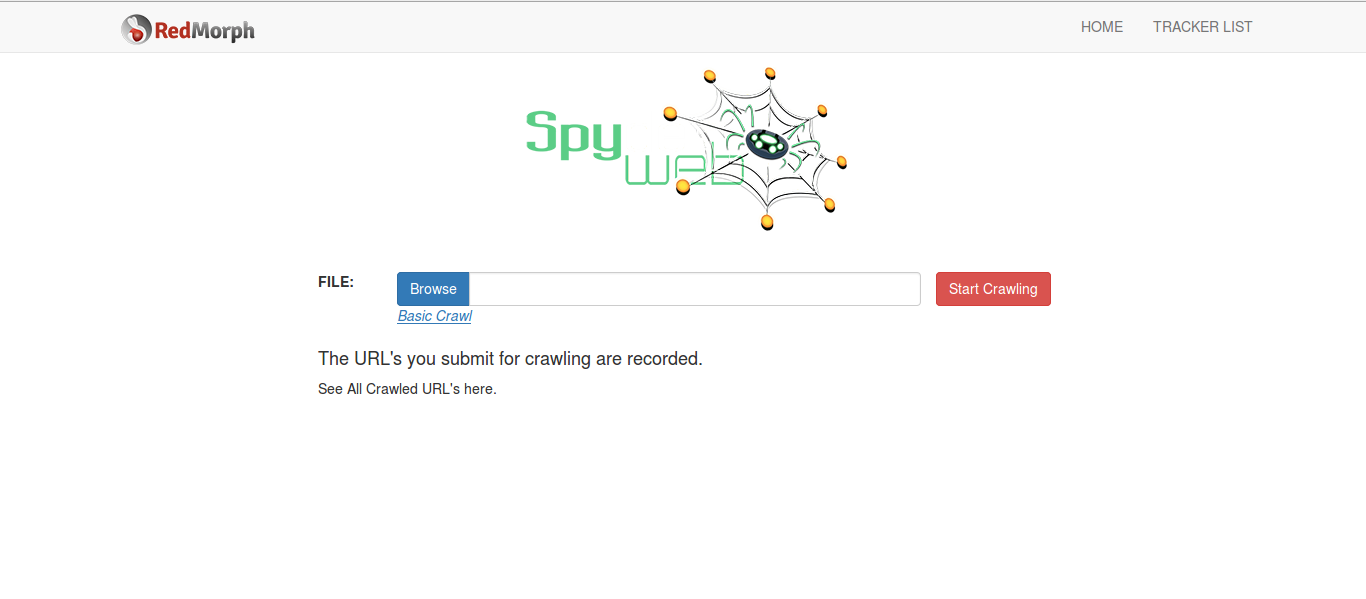
**5. Human Interface Design**

**5.1 Screen Images**

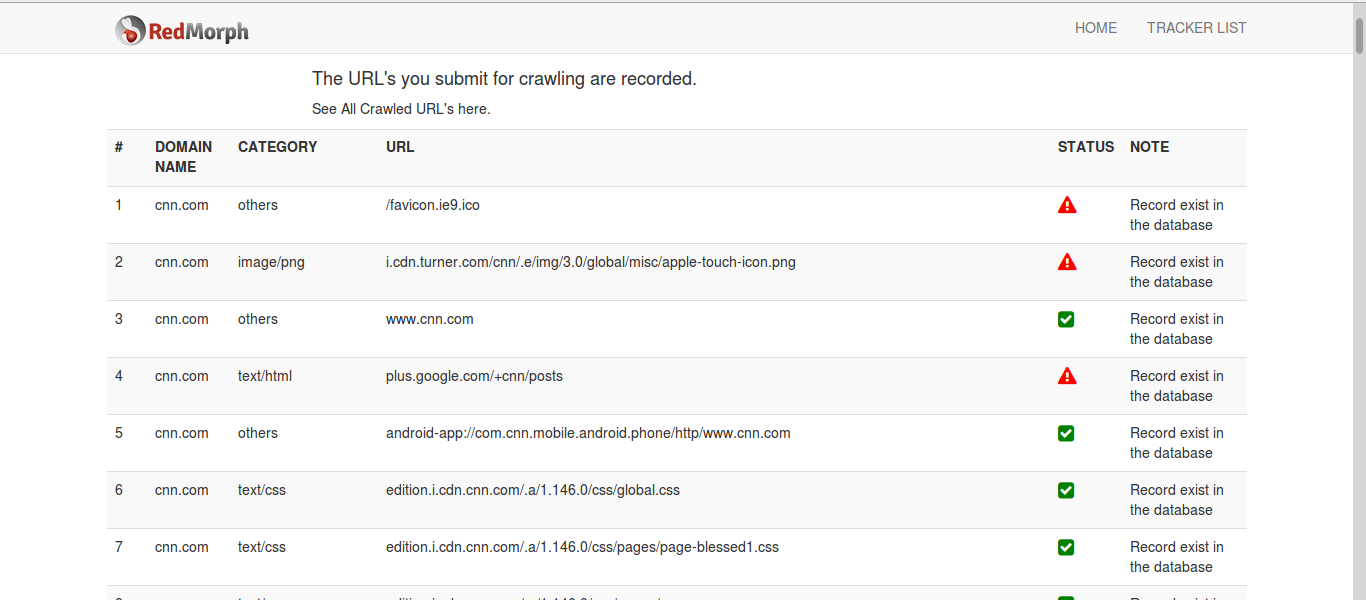
**A.** Basic Crawling Screen



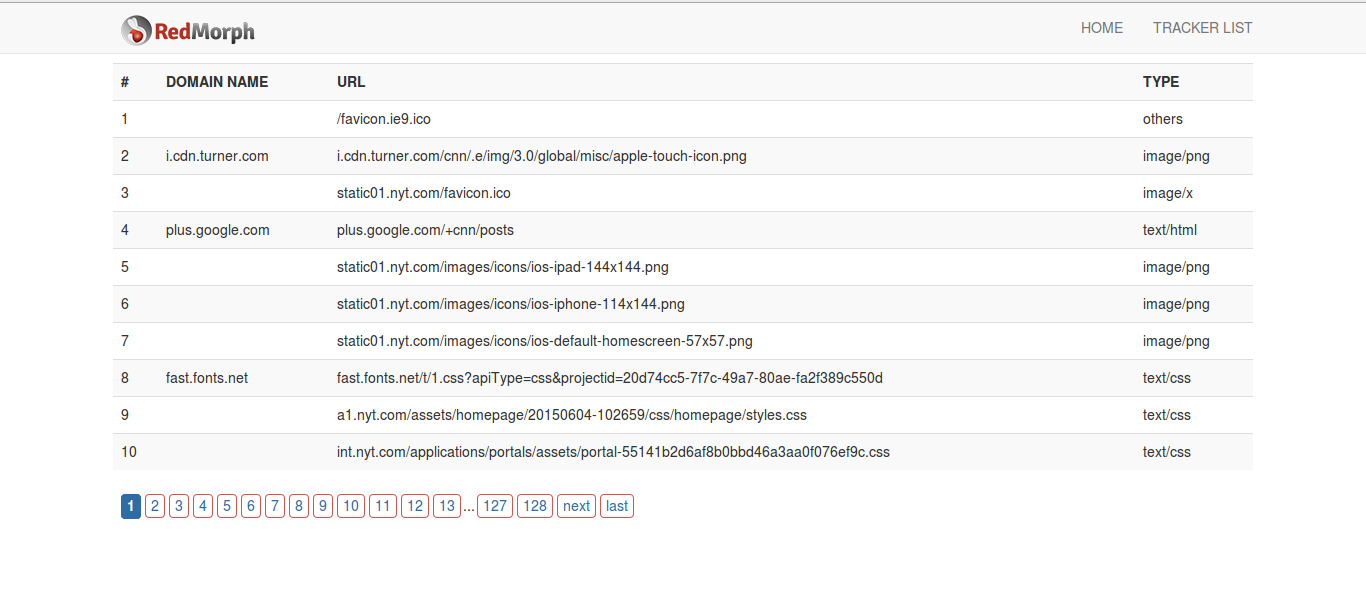
B. Advance Crawling Screen

****

C. Result's Page



D. Tracker List



**6. Local Setup: Step by Step**

1. Install all the development tools (reffer to number four(4) ).

2. Enable CURL Library, install php-curl.

3. Download/clone the repository file in gitlab under the project Web Tracker.

4. Copy the 'web-tracker' directory to '/var/www/html'.

5. Create a database in phpMyAdmin, name it 'tracker' then export the data table 'tracker\_list.sql' located inside database directory.

6. Open 'db\_connection.php' inside the web-tracker directory, change the value of server name, username, password, database name depending on your connection information.

7. Access the index.php uisng any browser installed in your machine. Type in <hostname>/web-tracker and your done!