Deepika project

Functionalities

- 1. Developed using php, mysql, AJAX, javascript, Jquery and MYSQL and Data tables
- 2. Completely web based
- 3. Ability to import and export URLS in bulk
- 4. User-friendly interface to add new URL information and edit existing information
- 5. Ability to crawl and delete multiple websites at a time
- 6. Ability to view the child pages of each URL
- 7. Ability to crawl the website based on depth and keywords
- 8. Notify the user via email if the user would like to receive the email
- 9. Crawler update results in the database and interface shows the results immediately
- 10. Results types
 - No change
 - New results:
 - If the Alert type="Anychange" and New information found on the website then it notify the user or update the database as New results found on the website
 - If the Alert type="Keyword" and New information has keyword in it then only it notify the user and update the database as New results found on the website
 - No new results found for specified keywords
 - Eventhough the website has updated with new information and do not have any specified keywords in it then it will not notify the user and it update the database as No new results found for specified keywords
 - Error occurred (404)
 - Error Occurred (Time Out Error)
- 11. If the Depth=1 then it crawls only the front page of the website.
- 12. If depth=2 then it crawls front page as well as all urls found in the front page
- 13. Ability to mark the message as read in bulk as well as individually
- 14. Email provides the clear difference between deleted and added content in each page
- 15. If the URL is crawled for the first time then child pages of the that URL will be added to the database and we can view them in the interface by clicking small green'+' button.

$\begin{array}{c} Login\ page \\ \rightarrow \ {\tt C} \ {\tt (0)\ localhost/my-first-cra} \end{array}$



Registration page

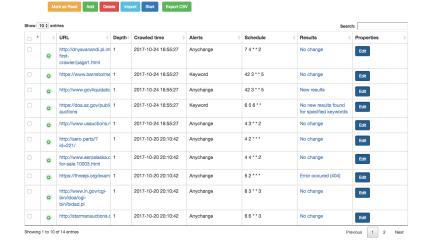


Landing page

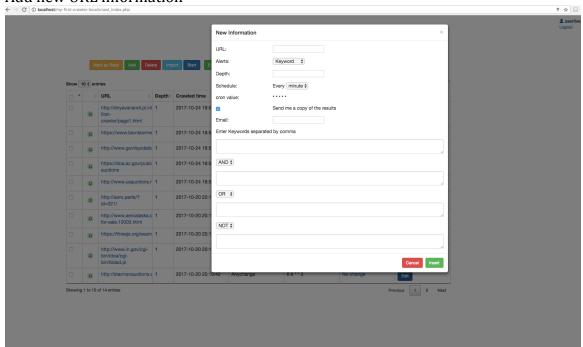




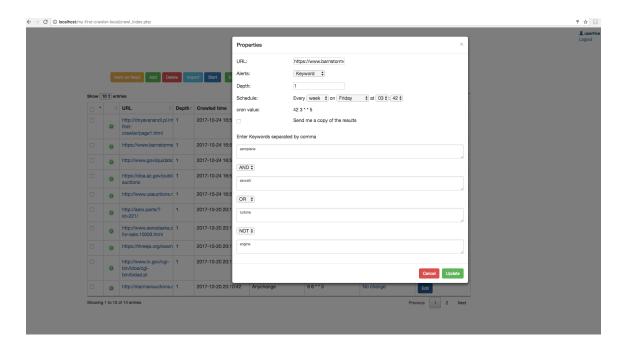
Saved search Criteria



Add new URL information

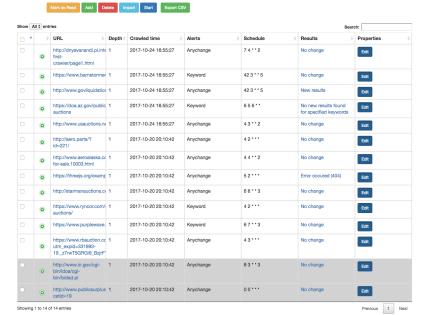


Edit Existing Information



Difference between read and unread messages





Email Notifications

← → ♂ ① localhost/my-first-crawler-local/crawl_index.php

System to me System Services <daemon@fardapwksmac002.localdomain>

The following page has changed! http://www.in.gov/cgi-bin/idoa/cgi-bin/bidad.pl

change percentage 380%

deleted -

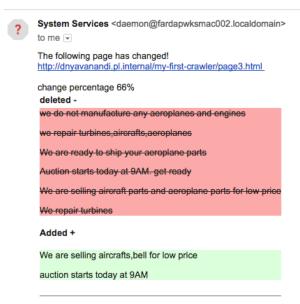
(Last updated: Oct/3/2017)

<u>Diesel Engine on Frame Trailer Pump 300-18-90495</u> 1 9/6/2017 3:00:00 PM Greg Moorman NEGOTIATED BID

Added +

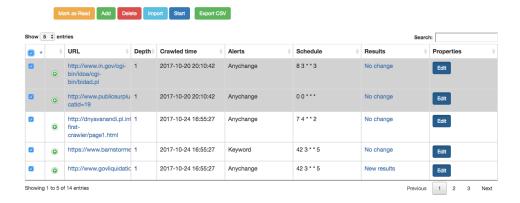
(Last updated: Oct/5/2017)

10/5/2017 11/15/2017 Jennifer Mayfield 3:00:00 PM NEGOTIATED BID Body Armor ASA-18-024

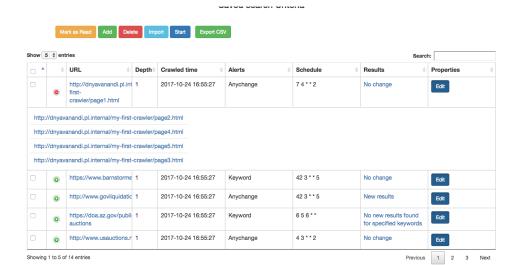


To take action on multiple items at a time

Saved search Criteria



Child urls of each URL



Import functionality



Database

