

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.

Login page

localhost/my-first-crawler-local/login.php

Sign In Form

Your Email

Your Password

Sign in

Don't have an account? [Sign up](#)

Registration page

localhost/my-first-crawler-local/register.php

Sign Up Form

Enter Name

Enter Your Email

Enter Password

Sign Up

Already have an account? [Sign in](#)

Landing page

localhost/my-first-crawler-local/crawl_index.php

userfive

Logout

Saved search Criteria

Mark as ReadAddDeleteImportStartExport CSV

Show10 entries

Search:

<input type="checkbox"/>	URL	Depth	Crawled time	Alerts	Schedule	Results	Properties
<input type="checkbox"/>	http://dryevanandi.pl/index.php/first-crawler/page1.html	1	2017-10-24 16:55:27	Anychange	7 4 ** 2	No change	Edit
<input type="checkbox"/>	https://www.barnstorne.com/	1	2017-10-24 16:55:27	Keyword	42 3 ** 5	No change	Edit
<input type="checkbox"/>	http://www.govliquidatic.com/	1	2017-10-24 16:55:27	Anychange	42 3 ** 5	New results	Edit
<input type="checkbox"/>	https://doa.az.gov/public-auctions	1	2017-10-24 16:55:27	Keyword	6 5 6 **	No new results found for specified keywords	Edit
<input type="checkbox"/>	http://www.usauctions.com/	1	2017-10-24 16:55:27	Anychange	4 3 ** 2	No change	Edit
<input type="checkbox"/>	http://aero.parts7.com/id=221/	1	2017-10-20 20:10:42	Anychange	4 2 ***	No change	Edit
<input type="checkbox"/>	http://www.aerolaska.com/for-sale.10003.html	1	2017-10-20 20:10:42	Anychange	4 4 ** 2	No change	Edit
<input type="checkbox"/>	https://threejs.org/learn/	1	2017-10-20 20:10:42	Anychange	5 2 ***	Error occurred (404)	Edit
<input type="checkbox"/>	http://www.in.gov/cgi-bin/doa/cgi-bin/bidad.pl	1	2017-10-20 20:10:42	Anychange	8 3 ** 3	No change	Edit
<input type="checkbox"/>	http://starmanauctions.com/	1	2017-10-20 20:10:42	Anychange	6 6 ** 3	No change	Edit

Showing 1 to 10 of 14 entries

Previous

12Next

Add new URL information

Mark as Read Add Delete Import Start

Show 10 entries

URL	Depth	Crawled time
http://dnyavanandi.pl/first-crawler/page1.html	1	2017-10-24 16:5
https://www.barnstorme	1	2017-10-24 16:5
http://www.govliquidatc	1	2017-10-24 16:5
https://doe.az.gov/public-auctions	1	2017-10-24 16:5
http://www.usauctions.r	1	2017-10-24 16:5
http://aero.parts/?id=221/	1	2017-10-20 20:1
http://www.aerolaska.c for-sale.10003.html	1	2017-10-20 20:1
https://threejs.org/exam	1	2017-10-20 20:1
http://www.in.gov/cgi-bin/doa/cgi-bin/bidad.pl	1	2017-10-20 20:1
http://starmansauctions.c	1	2017-10-20 20:10:42

Showing 1 to 10 of 14 entries

Previous 1 2 Next

userlive Logout

Edit Existing Information

Mark as Read Add Delete Import Start

Show 10 entries

URL	Depth	Crawled time
http://dnyavanandi.pl/first-crawler/page1.html	1	2017-10-24 16:5
https://www.barnstorme	1	2017-10-24 16:5
http://www.govliquidatc	1	2017-10-24 16:5
https://doe.az.gov/public-auctions	1	2017-10-24 16:5
http://www.usauctions.r	1	2017-10-24 16:5
http://aero.parts/?id=221/	1	2017-10-20 20:1
http://www.aerolaska.c for-sale.10003.html	1	2017-10-20 20:1
https://threejs.org/exam	1	2017-10-20 20:1
http://www.in.gov/cgi-bin/doa/cgi-bin/bidad.pl	1	2017-10-20 20:1
http://starmansauctions.c	1	2017-10-20 20:10:42

Showing 1 to 10 of 14 entries

Previous 1 2 Next

userlive Logout

Difference between read and unread messages

Saved search Criteria

Mark as ReadAddDeleteImportStartExport CSV

Show All 14 entries

Search:

	URL	Depth	Crawled time	Alerts	Schedule	Results	Properties
<input type="checkbox"/>	http://dryavanandi.pl/interfirst-crawler/page1.html	1	2017-10-24 16:55:27	Anychange	7.4 *** 2	No change	Edit
<input type="checkbox"/>	https://www.barnstormer.com	1	2017-10-24 16:55:27	Keyword	42.3 *** 5	No change	Edit
<input type="checkbox"/>	https://www.govliquidator.com	1	2017-10-24 16:55:27	Anychange	42.3 *** 5	New results	Edit
<input type="checkbox"/>	https://dos.az.gov/public-auctions	1	2017-10-24 16:55:27	Keyword	6.5.6 ***	No new results found for specified keywords	Edit
<input type="checkbox"/>	http://www.usauctions.net	1	2017-10-24 16:55:27	Anychange	4.3 *** 2	No change	Edit
<input type="checkbox"/>	http://aero.parts?Id=221/	1	2017-10-20 20:10:42	Anychange	4.2 ***	No change	Edit
<input type="checkbox"/>	http://www.aerolaska.co.uk/for-sale-10003.html	1	2017-10-20 20:10:42	Anychange	4.4 *** 2	No change	Edit
<input type="checkbox"/>	https://threejs.org/example/	1	2017-10-20 20:10:42	Anychange	5.2 ***	Error occurred (404)	Edit
<input type="checkbox"/>	http://starmanauctions.co.uk	1	2017-10-20 20:10:42	Anychange	6.6 *** 3	No change	Edit
<input type="checkbox"/>	https://www.nyoor.com/auctions/	1	2017-10-20 20:10:42	Keyword	4.2 ***	No change	Edit
<input type="checkbox"/>	https://www.purplewave.com	1	2017-10-20 20:10:42	Keyword	6.7 *** 3	No change	Edit
<input type="checkbox"/>	https://www.rbauction.co.uk/utm_expId=331893-19_z7hw15GfRGf8_BzrF	1	2017-10-20 20:10:42	Anychange	4.3 ***	No change	Edit
<input type="checkbox"/>	http://www.in.gov/cgi-bin/idoa/cgi-bin/bidad.pl	1	2017-10-20 20:10:42	Anychange	8.3 *** 3	No change	Edit
<input type="checkbox"/>	http://www.publicsurplus.ca/tid=19	1	2017-10-20 20:10:42	Anychange	0.0 ***	No change	Edit

Showing 1 to 14 of 14 entries

Previous1Next


Email Notifications

System Services <daemon@fardapwksmac002.localdomain>
to me

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)

NEGOTIATED BID [Diesel Engine on Frame Trailer Pump 300-18-90495](#) 1 9/6/2017 10/4/2017 3:00:00 PM [Greg Moorman](#)
Added +
(Last updated: Oct/5/2017)

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

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

The following page has changed!
<http://dnyavanandi.pl/internal/my-first-crawler/page3.html>

change percentage 66%

deleted -

~~we do not manufacture any aeroplanes and engines~~
~~we repair turbines,aircrafts,aeroplanes~~
~~We are ready to ship your aeroplane parts~~
~~Auction starts today at 9AM. get ready~~
~~We are selling aircraft parts and aeroplane parts for low price~~
~~We repair turbines~~

Added +

We are selling aircrafts,bell for low price
auction starts today at 9AM






To take action on multiple items at a time

Saved search Criteria

Mark as ReadAddDeleteImportStartExport CSV

Show 5 entries

Search:

<input checked="" type="checkbox"/>	<input type="checkbox"/>	URL	Depth	Crawled time	Alerts	Schedule	Results	Properties
<input checked="" type="checkbox"/>		http://www.in.gov/cgi-bin/tdos/cgi-bin/bidad.pl	1	2017-10-20 20:10:42	Anychange	8 3 * * 3	No change	Edit
<input checked="" type="checkbox"/>		http://www.publicsurplus.com/catid=19	1	2017-10-20 20:10:42	Anychange	0 0 * * *	No change	Edit
<input checked="" type="checkbox"/>		http://dnyavanandi.pl/internal/my-first-crawler/page1.html	1	2017-10-24 16:55:27	Anychange	7 4 * * 2	No change	Edit
<input checked="" type="checkbox"/>		https://www.barnstormer.com	1	2017-10-24 16:55:27	Keyword	42 3 * * 5	No change	Edit
<input checked="" type="checkbox"/>		http://www.govliquidatic.com	1	2017-10-24 16:55:27	Anychange	42 3 * * 5	New results	Edit

Showing 1 to 5 of 14 entries






Previous123Next

Child urls of each URL

Mark as ReadAddDeleteImportStartExport CSV

Show5entries

Search:

	URL	Depth	Crawled time	Alerts	Schedule	Results	Properties
<input type="checkbox"/>	 http://dryavanandi.pl.internal/my-first-crawler/page1.html	1	2017-10-24 16:55:27	Anychange	7 4 * * 2	No change	<div>Edit</div>
<div>http://dryavanandi.pl.internal/my-first-crawler/page2.html</div> <div>http://dryavanandi.pl.internal/my-first-crawler/page4.html</div> <div>http://dryavanandi.pl.internal/my-first-crawler/page5.html</div> <div>http://dryavanandi.pl.internal/my-first-crawler/page3.html</div>							
<input type="checkbox"/>	 https://www.barnstorme	1	2017-10-24 16:55:27	Keyword	42 3 * * 5	No change	<div>Edit</div>
<input type="checkbox"/>	 http://www.govliquidatic	1	2017-10-24 16:55:27	Anychange	42 3 * * 5	New results	<div>Edit</div>
<input type="checkbox"/>	 https://doa.az.gov/public-auctions	1	2017-10-24 16:55:27	Keyword	6 5 6 * * *	No new results found for specified keywords	<div>Edit</div>
<input type="checkbox"/>	 http://www.usauctions.r	1	2017-10-24 16:55:27	Anychange	4 3 * * 2	No change	<div>Edit</div>

Showing 1 to 5 of 14 entries

Previous123Next

Import functionality

Import New Monitors Here

Home

Choose File

No file chosen

Import CSV

Results	Status	Depth	URL
---------	--------	-------	-----

No file chosen

Database

Server: localhost » Database: web_crawl » Table: childurls

Browse Structure SQL Search Insert Export Import Privileges Operations Tracking

Showing rows 0 - 24 (2377 total, Query took 0.0008 seconds.)

SELECT * FROM `childurls`

1 > >> | Number of rows: 25 Filter rows: Search this table Sort by key: None

+ Options

	ID	url_id	childurl
	9562	39	https://www.barnstormers.com/cat.php?mode=listing&...
	9563	39	https://www.barnstormers.com/eFLYER/eflyerARCHIVES...
	9564	39	http://www.barnstormers.com/cat.php?mode=listing&m...
	9565	39	https://www.barnstormers.com/fraud/fraud.html
	9566	39	https://www.barnstormers.com/guestbook.php
	9567	39	https://www.barnstormers.com/ad_manager/my_ads.php
	9568	39	https://www.barnstormers.com/cat.php?mode=search
	9569	39	https://www.barnstormers.com/HELP/HELP.php
	9570	39	https://www.barnstormers.com/acct.php
	9571	39	https://www.facebook.com/BarnstormersInc
	9572	39	https://twitter.com/BarnstormersInc
	9573	39	https://plus.google.com/{+PagId}
	9574	39	https://www.barnstormers.com/banner_manager/click....
	9575	39	https://www.barnstormers.com/banner_manager/click....
	9576	39	https://www.barnstormers.com/banner_manager/click....
	9577	39	https://www.barnstormers.com/banner_manager/click....
	9578	39	https://www.barnstormers.com/banner_manager/click....
	9579	39	https://www.barnstormers.com/banner_manager/click....

Server: localhost » Database: web_crawl » Table: CHILDURLS

Browse Structure SQL Search Insert Export Import Privileges Operations Tracking Triggers

Showing rows 0 - 19 (20 total, Query took 0.0010 seconds.)

SELECT * FROM `CHILDURLS`

Show all | Number of rows: 25 Filter rows: Search this table Sort by key: None

+ Options

	ID	USER	URL	DEPTH	EMAIL	keyword1	keyword2	keyword3	keyword4	Status	logic1	logic2	logic3	Alerts	Cronvalue	Results	Mark_read	sendemail	Domain
	1	userone	http://www.barnstormers.com/cat.php?mode=listing&...	1	programmer2@partlogistics.com	turbine,aircraft,airplane,bell,audion	audion	turbine,engine	fargo	2017-10-24 16:55:27	&&		&&	Anychange	7.4***2	No change	0	yes	
	8	userfive	http://www.barnstormers.com/cat.php?mode=listing&...	1	programmer2@partlogistics.com	aircraft,airplane	audion	turbine,engine	fargo	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	No change	0	yes	
	25	usertwo	http://www.barnstormers.com/cat.php?mode=listing&...	1	programmer2@partlogistics.com	airplane,aircraft,audion,turbine,engine	UH-1,seal	dakota,turbo	audion and audion,turbine and bell parts	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	New results	0	yes	
	39	userfive	https://www.barnstormers.com/	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Keyword	4.2***5	No change	0	no	
	42	userfive	http://www.govtupdate.com/	1	programmer2@partlogistics.com	aircraft,airplane	turbine,engine	audion	fargo	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	No change	0	yes	
	45	userfive	https://bbs.a2.gov/public-auctions	1	programmer2@partlogistics.com	aircraft,audion	airplane	turbine	engine,UH-1	2017-10-24 16:55:27	&&		&&	Keyword	6.6***1	No new results found for specified keywords	0	no	
	46	Testor	www.partlogistics.com	2	programmer2@partlogistics.com	T30,engine,parts	Manual			2017-10-24 16:55:27	&&		&&	Keyword	4.2***5	No change	0	no	
	47	userfive	http://www.usauctions.net/Listing.aspx	1	programmer2@partlogistics.com	aircraft,airplane	turbine	engine	fargo	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	No change	0	no	
	48	userfive	http://www.parts7402.com/	1	programmer2@partlogistics.com	UH-1,aircraft	airplane	turbine	fargo	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	No change	0	no	
	50	userfive	http://www.aeromodels.com/aeromodel-parts-for-sale	1	programmer2@partlogistics.com	aircraft,airplane	turbine	engine	fargo	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	No change	0	no	
	51	userfive	https://www.orgsamplesweblog.com/convert_test...	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Anychange	5.2***5	Error occurred (404)	0	no	
	52	userfive	http://www.ging-birds.org/birds-bird-pl	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Anychange	5.2***5	No change	1	yes	
	53	userfive	http://www.auctions.com/	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Anychange	5.2***5	No change	0	no	
	54	userfive	https://www.ppcos.com/online-auctions/	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Keyword	4.2***5	No change	0	no	
	55	userfive	https://www.purpleware.com	1	programmer2@partlogistics.com	airplane, bell 306, m250	airplane	turbine	engine	2017-10-24 16:55:27	&&		&&	Keyword	6.7***5	No change	0	yes	
	56	userfive	https://www.fishbase.org/species/331865-19_37	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	No change	0	yes	
	57	userfive	http://www.retailforce.com/	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Anychange	4.2***5	No change	1	no	
	58	userfive	http://www.publiboutlook.com/brownbrowns/	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Anychange	5.2***5	No change	1	no	
	59	userfive	http://www.auction.robertaautos.com/auctionlist.aspx	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Keyword	12.11***4	No new results found for specified keywords	0	no	
	60	userfive	https://www.hurleyauctions.com/auctioning-auctione...	1	programmer2@partlogistics.com	airplane	aircraft	turbine	engine	2017-10-24 16:55:27	&&		&&	Anychange	8.5***5	No change	0	no	

☐ Check all With selected:

Server: localhost » Database: web_crawl » Table: users

Browse

Structure

SQL

Search

Insert

Export

Import

Privileges

Operations

Tracking

Triggers

Showing rows 0 - 5 (6 total, Query took 0.0009 seconds.)

SELECT * FROM `users`

Show all

Number of rows: 25

Filter rows: Search this table

Sort by key: None

+ Options

userld

userName

userEmail

userPass

<div><div></div><div>Edit</div><div>Copy</div><div>Delete</div></div>	1	userone	userone@gmail.com	1e9d59ad9be1cb302e155d55b61c95b3b3db897da2ed9643b1...
<div><div></div><div>Edit</div><div>Copy</div><div>Delete</div></div>	2	usertwo	usertwo@gmail.com	f64c95e5e1f4537428da7ba9fd1ee87bd263091c38c52d52ce...
<div><div></div><div>Edit</div><div>Copy</div><div>Delete</div></div>	3	userthree	userthree@gmail.com	cfb893cfa457ee4202174b5d0872cc27a202bcc78b1b1c49a7...
<div><div></div><div>Edit</div><div>Copy</div><div>Delete</div></div>	4	userfour	userfour@gmail.com	ea172dcf943ad14b3a583ba2883c9a2a5fa2c3f12a4ca183ca...
<div><div></div><div>Edit</div><div>Copy</div><div>Delete</div></div>	5	userfive	userfive@gmail.com	21a986cf5de3f572fe539d590e25bb681b736e64845b5f22e9...
<div><div></div><div>Edit</div><div>Copy</div><div>Delete</div></div>	6	Tester	testthisonce@mailinator.com	bfa4c507b6d2db4fca591fadd1e8b328672d47e2d4cd3f7e16...

Check all

With selected:

Edit

Copy

Delete

Export