

Forum:[Hack Forums](#)

Thread: [remove duplicate links url in a list of URL in listbox ?](#)

polas6 - May 18, 2014, 12:38 am

Hi,

I would like to identify duplicate in a list of links.

For example I would like to transform link such as

<http://domain.com/physical.aspx> into :

<http://domain.com/physical.aspx>

For example the 2 above are duplicates because they indicate the same website(valid URL) but the strings are different.

How to make sure there is no duplicate in a list of links like this :

<http://domain.com/physical.aspx>

<http://domain.com/physical.aspx>

<http://domain.com///physical.aspx>

I could have much more complex links like :

<http://domain.com/test///physical.aspx>

<http://domain.com/test/physical.aspx>

<http://domain.com/test/physical.aspx>

etc...

How to get rid of such duplicates from listbox in vb.net ?

Thanks

Dredd - May 18, 2014, 2:54 am

Use regex to remove any number of / grouped together apart from the http://

I am pretty certain regex can do this.

Cairns - May 18, 2014, 3:09 am

Code dim items as new list(of string)

dim cleanlist as new list(of string)

for each item as object in listbox.items

dim hostname as string = split(split(item.toString,"http://")(1),"/")(0)

dim path as string = split(split(item.toString,hostname)(1),"/")(0)

path = replace(path,"/",string.empty)

items.add(string.format("{0}/{1}",hostname,path))

next

cleanlist = items.distinct.tolist

This might replace the dodgy forward slashes and remove the duplicates. I have written it from the top of my head so it probably wont work, however I hope you know where to go from it.

If it does not work, I would be interested to code a function for you to do it via the IDE itself. just let me know :)

DBots - May 18, 2014, 3:36 am

You can use a HashSet(Of String) to avoid duplicate entries and if you don't care about a particular order. Take each item in the list, split it by "/", avoid any empty entries and join it back together:

```
Code Dim URLKeys As New HashSet(Of String)(URLList.Items.Cast(Of String).Select(Function(URL) String.Join("/", URL.Split({"/"c}, StringSplitOptions.RemoveEmptyEntries))))
```

Now, clear your list and add the hashset elements:

```
Code URLList.Items.Clear()  
URLList.Items.AddRange(URLKeys.ToArray)
```

Cairns - May 18, 2014, 3:44 am

(May 18, 2014 03:36 AM)DBots Wrote: [\[CITE\]](#) You can use a HashSet(Of String) to avoid duplicate entries and if you don't care about a particular order. Take each item in the list, split it by "/", avoid any empty entries and join it back together:

```
Code Dim URLKeys As New HashSet(Of String)(URLList.Items.Cast(Of String).Select(Function(URL) String.Join("/", URL.Split({"/"c}, StringSplitOptions.RemoveEmptyEntries))))
```

Now, clear your list and add the hashset elements:

```
Code URLList.Items.Clear()  
URLList.Items.AddRange(URLKeys.ToArray)
```

I have never used a hashset yet, seems very interesting. Hope the code works because it would be a much better solution than mine lol (from my head). Great job

polas6 - May 18, 2014, 4:56 am

thanks Cairns working just awesome ;)

Cairns - May 18, 2014, 5:01 am

You should also try out dbots method, his seems better. And no problem at all, I am surprised since I written it from the top of my head. Best of luck with your project buddy

DBots - May 18, 2014, 1:06 pm

Thanks, Cairns, appreciate the comment.

Polas just remember, there is more than one way of doing a single thing with .net. Keep it simple!