

Forum:[Hack Forums](#)

Thread: [My first Scraper\(Proxy Scraper\)](#)

i-manuel - August 17, 2016, 12:53 pm

Hi ?,

i coded my first scraper, its a proxy scraper, nothing special but here is the source:

Image:

[img]http://213.202.228.97/Images/ProxyScraper.PNG[/img]

Source: [Github](#)

SIM-2626129 - August 18, 2016, 6:20 am

Looks good man. Did you add there proxy checker as well, or this only scrapes them?

Ninja - August 18, 2016, 8:19 am

Why did you repost? You already posted this thread in the VB.net section.

Code System.Windows.Forms.MessageBox.Show(ex.ToString());just import the namespace fam.
Same goes for the System.Net namespace.

Code count++;

Window.Text = string.Format("Proxys: {0}", count);int++ is a function that returns the integer value and then increments it.

++int is a function that increases the integer value and then returns it.

So you could just do:

Code Window.Text = string.Format("Proxys: {0}", ++count);

Code ListViewItem.SubItems.Add(GetTdlItem(proxy, 2));You still keep splitting the string over and over again. Like I said, split it once.

Code var trnew = new string[] { "<tr>" };Really?

Also, you capitalize random variables. I suggest you read up on the OOP naming conventions. You use var (which is not necessarily a bad thing but can lead to problems in for example group projects or the use of wrappers).

Also, like I said, you can just do WebRequest.CreateHttp("url") instead of (HttpWebRequest)WebRequest.Create("url").

Also, you have a class called Proxy but it is not a model, rather a controller.

And just a tip, but look into HTML parsing for this purpose.

And finally, clean the solution before uploading to GitHub. Dont include the exes in your git.

This is just at first glance. There is much more wrong.

Now, I am not trying to be a dick, just pointing out the errors so you can fix them and learn from them. Granted, most of these errors are just things like naming conventions or casts that are redundant, but some errors should actually be fixed, like the disposing of objects.

i-manuel - August 18, 2016, 12:09 pm

(Aug 18, 2016 08:19 AM)Ninja Wrote: [\[CITE\]](#) Why did you repost? You already posted this thread in the VB.net section.

Code System.Windows.Forms.MessageBox.Show(ex.ToString());just import the namespace fam.
Same goes for the System.Net namespace.

Code count++;

Window.Text = string.Format("Proxys: {0}", count);int++ is a function that returns the integer value and then increments it.

++int is a function that increases the integer value and then returns it.

So you could just do:

Code Window.Text = string.Format("Proxys: {0}", ++count);

Code ListViewItem.SubItems.Add(GetTdItem(proxy, 2));You still keep splitting the string over and over again. Like I said, split it once.

Code var trnew = new string[] { "<tr>" };Really?

Also, you capitalize random variables. I suggest you read up on the OOP naming conventions. You use var (which is not necessarily a bad thing but can lead to problems in for example group projects or the use of wrappers).

Also, like I said, you can just do WebRequest.CreateHttp("url") instead of (HttpWebRequest)WebRequest.Create("url").

Also, you have a class called Proxy but it is not a model, rather a controller.

And just a tip, but look into HTML parsing for this purpose.

And finally, clean the solution before uploading to GitHub. Dont include the exes in your git.

This is just at first glance. There is much more wrong.

Now, I am not trying to be a dick, just pointing out the errors so you can fix them and learn from them. Granted, most of these errors are just things like naming conventions or casts that are redundant, but some errors should actually be fixed, like the disposing of objects.

i updated it already, btw look at the thread in the vb section..., but thanks for helping :)

(Aug 18, 2016 08:19 AM)Ninja Wrote: [\[CITE\]](#) Why did you repost? You already posted this thread in the VB.net section.

Code System.Windows.Forms.MessageBox.Show(ex.ToString());just import the namespace fam.
Same goes for the System.Net namespace.

Code count++;

Window.Text = string.Format("Proxys: {0}", count);int++ is a function that returns the integer value and then increments it.

++int is a function that increases the integer value and then returns it.

So you could just do:

Code Window.Text = string.Format("Proxys: {0}", ++count);

Code ListViewItem.SubItems.Add(GetTdItem(proxy, 2));You still keep splitting the string over and over again. Like I said, split it once.

Code var trnew = new string[] { "<tr>" };Really?

Also, you capitalize random variables. I suggest you read up on the OOP naming conventions. You use var (which is not necessarily a bad thing but can lead to problems in for example group

projects or the use of wrappers).

Also, like I said, you can just do `WebRequest.CreateHttp("url")` instead of `(HttpWebRequest)WebRequest.Create("url")`.

Also, you have a class called `Proxy` but it is not a model, rather a controller.

And just a tip, but look into HTML parsing for this purpose.

And finally, clean the solution before uploading to GitHub. Dont include the exes in your git.

This is just at first glance. There is much more wrong.

Now, I am not trying to be a dick, just pointing out the errors so you can fix them and learn from them. Granted, most of these errors are just things like naming conventions or casts that are redundant, but some errors should actually be fixed, like the disposing of objects.

you can't say "there is much more wrong" because the scraper is working, but "it's not perfectly coded" because i'm new ... nah 6-7 months into coding, so each coder was a beginner :P

P4R4DiSi4C - August 18, 2016, 5:51 pm

Congrats ! I do think that he tried to help you and point out where you could improve. It wasn't his intention to get you angry, you should just accept the suggestions, try to fix these issues, clean the code and continue learning.

That's not because it's working that the project is finish. You can't just say "It's working so everything is fine". Congratulations tho and yes everyone was a beginner once.

Regards buddy

Ninja - August 18, 2016, 6:10 pm

(Aug 18, 2016 12:09 PM)i-manuel Wrote: [CITE](#) i updated it already, btw look at the thread in the vb section..., but thanks for helping :)

you can't say "there is much more wrong" because the scraper is working, but "it's not perfectly coded" because i'm new ... nah 6-7 months into coding, so each coder was a beginner :P

You didn't update shit Im looking at the code on the Github and the same mistakes are there. When scraping a lot of proxies and not disposing objects that can be disposed they will stay in memory. More requests = more stuff in memory while not much in memory is getting freed. So memory usage goes up, eventually causing problems and crashing the program. Its just good practice, a good habit, and helpful to do.