

Forum:[Hack Forums](#)

Thread: [How can I download a whole forum?](#)

KiNGTyreZe - August 19, 2015, 7:10 am

Thread subject says it all. I'm user in a private forum and would like to leech the informations to have offline access.

Thanks in advance.

Viktory - August 19, 2015, 7:17 am

Have you considered typing the same into Google and clicking the first result?

KiNGTyreZe - August 19, 2015, 7:33 am

(Aug 19, 2015 07:17 AM)Viktory Wrote: [\[CITE\]](#) Have you considered typing the same into Google and clicking the first result?

Yeah, but it doesn't work for private forums since HTTrack for example has no access to them.

XMR - August 19, 2015, 8:34 am

(Aug 19, 2015 07:33 AM)KiNGTyreZe Wrote: [\[CITE\]](#) Yeah, but it doesn't work for private forums since HTTrack for example has no access to them.

If you don't have access to a forum, you can't download it. End of story.

KiNGTyreZe - August 19, 2015, 9:22 am

(Aug 19, 2015 08:34 AM)Arccotangent Wrote: [\[CITE\]](#) If you don't have access to a forum, you can't download it. End of story.

I think you misunderstood me. I have access to the forum.

However the tool isn't trying to download the website over my browser client.

In fact it's trying to download the private forum over a new instance.

My question is still: How can I download the whole forum?

I couldn't find something yet which could help me in my concern.

nondescriptoldie - August 19, 2015, 9:32 am

(Aug 19, 2015 09:22 AM)KiNGTyreZe Wrote: [\[CITE\]](#) I think you misunderstood me. I have access to the forum.

However the tool isn't trying to download the website over my browser client.

In fact it's trying to download the private forum over a new instance.

My question is still: How can I download the whole forum?

I couldn't find something yet which could help me in my concern.What do you mean the database? Accounts...

I don't get you here

Maxi - August 19, 2015, 9:33 am

Maybe it is a TOR website or a website with some kind of protection so HTTracker can't access it.

Nuclear3mbargo - August 19, 2015, 9:33 am

Python + Selenium is your answer

<http://stackoverflow.com/questions/8792591/crawling-a-web-site-using-python>

<http://stackoverflow.com/questions/7861775/python-selenium-accessing-html-source>

Do more research next time.

KiNGTyreZe - August 19, 2015, 10:58 am

(Aug 19, 2015 09:33 AM)Maxi Wrote: [\[CITE\]](#) Maybe it is a TOR website or a website with some kind of protection so HTTracker can't access it.

This! I just want to download the pages inside there.

(Aug 19, 2015 09:33 AM)Nuclear3mbargo Wrote: [\[CITE\]](#) Python + Selenium is your answer

Doesn't it require login first before I can download it? It seems like it's also creating a new instance.

Goodies - August 19, 2015, 12:00 pm

You can use an HTTP spider, but you can't really know if you have everything. Also, any server admin worth his weight in dog shit will prevent spiders like this.

KiNGTyreZe - August 19, 2015, 2:38 pm

(Aug 19, 2015 12:00 PM)Goodies Wrote: [\[CITE\]](#) You can use an HTTP spider, but you can't really know if you have everything. Also, any server admin worth his weight in dog shit will prevent spiders like this.

Does it work over my browser client instance? Remember a login is required to see the content.

Goodies - August 19, 2015, 3:00 pm

(Aug 19, 2015 02:38 PM)KiNGTyreZe Wrote: [\[CITE\]](#) Does it work over my browser client instance? Remember a login is required to see the content.

There are spiders that support HTTP logins and cookies, absolutely.

Nuclear3mbargo - August 20, 2015, 1:54 am

(Aug 19, 2015 10:58 AM)KiNGTyreZe Wrote: [\[CITE\]](#) This! I just want to download the pages inside there.

Doesn't it require login first before I can download it? It seems like it's also creating a new instance.

Code pip install selenium

Yes it creates new instance if you want to make it in the current instance in your browser your only choice is JS and i'm not sure if it's possible(to save page with javascript). However you can simulate every action with Selenium.

KiNGTyreZe - August 20, 2015, 7:50 am

Thanks everybody. I appreciate the help.

LogoiLab - August 20, 2015, 10:29 am

you need a web crawler, try dcrawl.