

Forum:[Hack Forums](#)

Thread: [Crawly - a small scale webcrawler \(email, social media\)](#)

ClusterzInc - July 22, 2019, 1:39 pm

Hello, my friends,

I'm still learning VB.Net diligently. For practicing I dared to do a small fun project. I tried to build a small and rudimentary webcrawler. It can currently only read and save e-mail addresses.

It works like this:

1. you define some search terms (construction company in Dallas, painter in dallas)
2. you start the crawl process
3. extract the email addresses from the found pages.
4. (Coming soon) You start with the found email addresses an advertising campaign with automatic email sending (with attachments, html templates etc.)

Currently the version still contains many flaws:

1. threading works, but the thread remains active after the run.
2. it happens that links to images are recognized as email.
3. it takes an incredibly long time to crawl >150 pages, because .Net.WebClient waits forever for non-existent websites until it finally timed out.

As I said I'm still learning all the stuff. But I would be very grateful for tips (even critical ones).

I hope you can do something with it.

Link to Github: [hxxp://github .com/clusterzx/crawly](https://github.com/clusterzx/crawly)

(maybe a mod can edit this post and correct the link in case of insufficient permissions)

Have a nice day

Your Clusterz