



WebCollector Introduction

Input a keyword into WebCollector and it will post a query against a chosen search engine such as Lycos, InfoSeek, etc. Once that search engine returns matches, Webcollector will page through each match, going to the specified web page, and collect all email addresses within that site.

This includes all email addresses on underlying pages (within the site). WebCollector sifts through the site, searching for every possible link on that site, then collects any email address that it finds.

WebCollector is very careful not to venture off the site. For example if you are looking for email addresses related to "automobiles" and WebCollector is searching "http://www.cars.com", you don't want to follow a link pointing to Netscape as this would have nothing to do with automobiles.

Procedures

Here are the steps you should follow to get started quickly:

- The first thing you want to do is select which Search Engine you want to use from the "Search Engines" menu.
- You should then enter at least one "Keyword" as a search parameter. When you want to use more than one keyword, just separate each word with a space. You can use lowercase or uppercase, it doesn't matter.
- Adjust the "Sensitivity" slider. The "looser" you set it, the more matches the search engine will return. For example: In the first position (to the left), you could get as many as 1.8 million hits to search, whereas if you move the slider all the way to the right (tightest), you might get 25-50 hits.
- Now click on the "Search" button.

The rest is automatic. Just sit back and watch as WebCollector does ALL the work.

Reference

[Sample Searches](#)

[How do I grab all the addresses from a domain or ISP?](#)

[What if I'm in the middle of a big search and my modem loses connection?](#)

[Where are the email addresses stored?](#)

[Some web servers take forever to respond - How can I change this?](#)

[I only want to save email addresses that are related to my keywords](#)



[Back to Contents](#)

Earthonline Corporation develops, markets and supports an integrated set of TCP/IP application and business development software for the Microsoft Windows operating platform.

The company is located at 13400 Northup Way, Suite 37, Bellevue, WA 98005

Phone: 206-865-9000

Fax: 206-865-9100

Internet: <http://www.earthonline.com>

Sample Searches

[Back to Contents](#)

See Also

[Collecting from ISPs](#)

You can enter up to 10 different keywords to search with. We call this a "Query". Forget about the rules and query structure that you might have learned while working with the search engines - it has all been built into WebCollector for two reasons: One, each engine needs different query structures and this makes it difficult to remember which structure works with which engine. Two, you can use the same parameters with every WebCollector search, with no modifications.

Each keyword that you enter into the "Keyword(s)" box must be separated with a space. To put it simply - WebCollector will ask the search engine that you choose to locate web pages that contain these keywords.

Here are some sample search parameters to give you an idea of what to type in. If for example, you wanted to retrieve email addresses that were related to "automobiles":

```
cars trucks autos  
automobiles  
chevy ford dodge  
mustang corvette camaro
```

Here is an example of a few WRONG ways to do the same search:

```
cars AND trucks AND autos  
chevy OR ford NOT chrysler
```

Collecting from Domains or ISP's

[Back to Contents](#)

It is possible to collect all the email addresses from a given ISP (Internet Service Provider) simply by pointing WebCollector to that ISP's "Main Page". You only have to find the main url off of which all the customers are hosted.

For example, let's say that you want to collect all the email addresses from all or most of the customers at `www.example.com` and they have 5000 customers. The first thing you must know is where the main page is hosted. Example.com hosts all their customer web pages under `http://www.example.com/users/`.

To collect from these web pages, click the "Search Domain" button and enter `http://www.example.com/users/` in the box provided. Note that the more information you give WebCollector the faster it will collect. For example, at `http://www.example.com/users/` there has to be an HTML file. This would be the index file that lists all the users. So by adding "index.html" to the end (`http://www.example.com/users/index.html`) you could significantly reduce the search time.

Once you press "OK", you will notice that the URL (`http://www.example.com/users/`) is entered in the "URL's To Search:" box at the top of WebCollector. Now just press the "Start" button.

You can enter as many domains as you wish by repeating the steps above or, you can create a batch file. Here's how:

Create a text file with each domain that you wish to search on a separate line and save this file as EXAMPLE.SRC (you can name it whatever you like, just make sure it has the extension of .SRC)

Now from WebCollector's "File" menu, select "Open Search" and choose the file EXAMPLE.SRC. You will notice that WebCollector loaded it into the listbox, now just press the "Start" button and sit back!

You can put as many ISP's as you like in each search (.SRC) file.

Open Search

[Back to Contents](#)

See Also:

[Collecting from ISP's](#)

If you are in the middle of a big search and you happen to lose your connection, don't worry. You can simply re-connect and start where you left off. Even if your computer crashes, or the power blinks off, you can restore the search very easily.

Let's take the worst case scenario and say the power blinks. We know that your computer would have rebooted, so follow these steps:

- 1 Get back into Windows and get back online with your Internet Service Provider.
- 2 Start WebCollector.
- 3 From the "File" menu, select "Open Search".
- 4 Choose the file named "backup.src". (After every web page, WebCollector backs up the search to the site it's on.) Press OK
- 5 Now, simply press the "Search" button.

Where are the addresses stored?

[Back to Contents](#)

All the email addresses that WebCollector gathers are stored in a file called EMAIL.TXT. This file is located in the WebCollector directory (usually C:\WC).

If the file exists when you start a session, all new addresses will be APPENDED to the end of the file, so don't worry about losing addresses.

You may notice that WebCollector, takes quite a while to go through thousands of web pages and gather hundreds of addresses. But just think about two things, how long would it take you to do this by hand, and these are highly qualified, TARGETED addresses.

Please be sure not to add any addresses that start with "webmaster@" or "hostmaster>". These are old timers on the net and there is a very good chance of getting flamed if you send them a commercial message.

If you are importing the addresses you collect with WebCollector into NetContact, you have nothing to worry about, NetContact will automatically remove webmasters, hostmasters and other possible irritable people from your list as you import it.

The Options Dialog

[Back to Contents](#)

Here are some helpful hints that should assist you in configuring WebCollector for optimum use! Choose the "Options" menu to set the your preferences on this dialog.

Only save email addresses from web pages that contain ALL of my keywords!

Checking this box will force WebCollector to make sure that when it finds an email address, the web site contained all keywords that you entered. If you have ever performed a query on one of those search engines, you know the value of this feature as search engines tend to return "hits" that have absolutely NOTHING to do with what you searched for.

Believe it or not, this will not slow WebCollector down. In fact, it speeds up the process as it doesn't have to scan for each link.

How long should WebCollector wait for a server to respond?

This slider will adjust the time WebCollector will wait for slow (or broken) web servers. If you have ever waited 5 minutes for Netscape to bring up a web page only to have an error come up and tell you that the "server is not responding" you should quickly grasp this functionality.

Experimenting around with this can speed up your entire search many times. Unfortunately, there is no "right" setting. Each user's setting is based upon several factors such as modem speed, line quality, ISP's computer speed, ISP's average load, etc

Glossary

≡
#
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

