

All In Scraper - Quick Start Guide

Step 1 – Register

Simply put in your license code that was provided to you via email and click the Register button.

Your license code will be provided to you via email. Your license code will be your PayPal unique transaction id. This can be found in any of your emails from PayPal about your transaction or simply by viewing the details of the transaction when logged into your PayPal account.

You are allowed one installation per license. So after you activate your license you will not be able to activate it again on another computer. You can un-register a computer if you still have access to it by pressing the un-register button in the software. If for some reason you can't un-register you need to simply email joe@allinscraper.com with your license code from your PayPal email address and we will take care of it for you.

All In Scraper (AIS) will thank you for your purchase and notify you that the program must shutdown to complete the registration.

If you do not register you will be running in Demo Mode and you will be limited to importing 5 proxies and keywords each.

Step 2a – Use Proxies?

If you don't want to use proxies uncheck the Use Proxies checkbox and move on to step 3.

However we do recommend the use of proxies to prevent you from becoming blocked or banned from the search engines. From my experience you can't get away with many allintitle searches without getting a captcha ban and eventually a ban without captcha option.

So if you are using Proxies just make sure that the Use Proxies checkbox is checked.

Step 2b – Enter Proxies

Then you can import or copy and paste your proxy list into the program. We support text files with one proxy:port entry per line. You do this using either the Import button or the Copy/Paste button.

Step2c – Proxy Testing

Now that your proxies are listed in the data grid you should have more buttons enabled to work with.

- **Clear Button:**
Clears the data grid
- **Simple Test Button**
The simple test will check each proxy for access to the selected search engine in the settings section. If a response is made you will then be provided the response time and the status will be updated to ok.
- **Advanced Test Button**
This test will actually check each proxy for access to the selected search engine using an allintitle request. To use this test your keyword list must first be imported. The test will randomly use your keywords in the requests.

We offer the advanced test because we found that sometimes when doing “All In” requests the proxy will be blocked but when just visiting the home page like in the simple test there is no block. So this is a better test for “All In” runs. It returns the response times and it will tell you if the search engine is asking for captchas from that proxy or not.

Step2d – Remove Failed Proxies

After running your desired proxy test you will have a list that contains ok and failed proxies. You can press the Remove Failed button and all the proxies marked Failed will be removed.

You can also use the Remove Selected button to remove individually selected proxies.

Step2e – Proxy Settings

- **Requests**
This number represents the number of search engine request will be made against each proxy before moving on to the next one. We use proxies in the grid from top to bottom. So if it is set at 25 there will be 25 requests made to each proxy in the list. If the end is reached and there are still scrape requests to be made it will start back at the top again.
- **Timeout**
This is not used in scraping. It is only used in proxy testing. It represents the time in milliseconds you want to wait for the proxy to respond before you give up and mark it failed.
1000 milliseconds = 1 second
- **Import**
Use this to import proxies from a text file. Format is ipaddress:port and needs only one per line.
- **Copy/Paste**
Use this option to copy proxies from the web or somewhere on your computer and paste them into our program. The format is the same, ipaddress:port and one per line. Then click the import button to populate the list.
- **Export**
This allows you to export your proxy list from our program to either a text file or csv file. Both

formats export ipaddress:port and one per line.

- **Clear**
Pressing this button clears the proxy list.
- **Remove Failed**
This button will remove all proxies from the list that have a status of failed.
- **Remove Selected**
This button will remove the selected proxies from the list. Use of CTRL and ALT keys for multiple selections is allowed.

Step2f – Proxy Wrap-up

At this point you have a good list of proxies and you are ready to move on to the next step.

Often times at this point I will sort the proxy data grid by response time. You can do this by clicking on the response column header. I do this because the proxies are used in order from top to bottom. So I like to have my fastest proxies at the top.

Step 3 – Settings

- **Site Combo Box**
Select the search engine that you want to scrape for results.
- **Timeout**
This is used for scraping and is the timeout on a request to the search engine. This does not come into play much if your proxies are all tested good. Sometimes a proxy will drop during a run and this is the timeout before you give up on that proxy and move to the next. Once again this is in milliseconds.
1000 milliseconds = 1 second
- **Interval**
This is an important setting for scraping. Basically it is the amount of time to wait before making the next request to the search engine. Making rapid requests will quickly get your IP or proxy temporarily banned or blocked. Although setting this too high will make the scraping process take forever. I like to run mine anywhere from 250 to 1000 milliseconds.
250-1000 milliseconds = ¼ of a second to 1 second

Now we have to choose what you want to scrape. The checkbox options are:

- **Broad Results**
The number of results returned by the selected search engine without quotes around the keyword.
- **Phrase Results**
The number of results returned by the selected search engine with quotes around the keyword.
- **Exact Results**
The number of results returned by the selected search engine with brackets around the keyword.
- **AIT – allintitle**
The number of results in the specified search engine that has the keywords in the title of the

page. No quotes. Broad match.

- **AITQ – allintitle**
The number of results in the specified search engine that has the keywords in the title of the page. Quotes used. Phrase match.
- **AIA – allinanchor**
The number of results in the specified search engine that contain all query terms in the anchor text on the links to the page. No quotes. Broad match.
- **AIAQ – allinanchor**
The number of results in the specified search engine that contain all query terms in the anchor text on the links to the page. Quotes used. Phrase match.
- **AIU – allinurl**
The number of results in the specified search engine that contain all query terms in the URL. No quotes. Broad match.
- **AIUQ – allinurl**
The number of results in the specified search engine that contain all query terms in the URL. Quotes used. Phrase match.
- **RC – broad**
This was a last minute option that was requested by a member of the warrior forum. RC stands for “Real Competition”. This will return the number you see on the last page of Google. It is worded like this: “In order to show you the most relevant results, we have omitted some entries very similar to the 769 already displayed.”
- **RCQ – phrase**
Same as RC but using quotes for a phrase match. This is the option that the warrior member actually requested and said he used to check competition.

Real Competition has become so popular with AIS that we have now added RC to all scrape options. Here is the break down and you can reference the above descriptions as to what each means.

RC	Real Competition – broad search results
RCQ	Real Competition – phrase search results
RCAIT	Real Competition – allintitle – broad
RCAITQ	Real Competition – allintitle – phrase
RCAIA	Real Competition – allinanchor – broad
RCAIAQ	Real Competition – allinanchor – phrase
RCAIU	Real Competition – allinurl – broad
RCAIUQ	Real Competition – allinurl – phrase

Step 4 – Decaptcher

If you are using proxies and are scraping “All In” data you are going to be requested for captchas. AIS will display popups that you can manually enter these captchas into to continue scraping. Unfortunately then you have to sit there and answer them while it is running. So to resolve this I partnered with Decaptcher. They employ people to answer those captchas for you. It is all done in the background of the program! Using Decaptcher allows you scrape on 100% autopilot.

If you want to use Decaptcher you need to setup an account with them. Then check the Use Decaptcher checkbox. They provide you with an IP address and port that you will need to enter into the appropriate fields. You then also need to supply your Decaptcher username and password.

Press the Test button to see if you got it all right. If so you will get a popup box with your Decaptcher balance displayed.

This service is charged at \$3/1000 captchas solved and is paid in advance through decaptcher.com.

- **Decaptcher IP**
This is the IP or DNS name of the Decaptcher server. Check decaptcher.com for your updated settings.
- **Port**
This is the port number that goes along with the Decaptcher IP. Check www.decaptcher.com for your updated settings.
- **Username**
This is your Decaptcher username and identifies the account to be used to solve captchas.
- **Password**
This is your password for the specified username.
- **Test**
Pressing the test button will test your Decaptcher credentials and connection settings. If everything is set up correctly you will see a popup message that contains your balance.

Step 5 – Keywords

You can import your keywords much like you did your proxies. You can import from a text file with one keyword listed per line or you can use the Copy/Paste button to paste them into the program.

We also support Google Adwords Keyword Tool (GAKT) CSV files. You must export in CSV and NOT CSV for Excel! The nice thing about using these files is that we will import Global Searches, Local Searched, Competition and AVG CPC fields. This provides you with even more information in one single spot. By default in GAKT the AVG CPC field is not displayed. If you want it to export to the CSV file you have to display that field in GAKT before exporting.

Only columns that you are scraping are displayed. If you want to see GAKT columns be sure to check Show GAKT Columns. If you want to display all columns regardless of what you are scraping just check Show All Columns.

- **Import**
Use this to import keywords from a text file or Google Adwords Keyword Tool (GAKT). If you are using a text file the format is just one keyword per line. If you want to use GAKT you must export the file from GAKT using the CSV option.

WARNING: Do not use the CSV for Excel option as this is not compatible with our program. It must say just CSV not CSV for Excel! We recommend importing your keywords from GAKT because you can import extra details along with your keywords such as competition, global searches, local searches and AVG CPC. If you do not get AVG CPC when importing you will want to make sure you are displaying it in GAKT when you are doing your export. If you are not displaying it in GAKT then it will not export either.

- **Copy/Paste**
Use this option to copy keywords from the web or somewhere on your computer and paste them into our program. The format is one keyword per line. Then click the import button to populate the list.
- **Export**
This allows you to export your keyword list from our program to either a text file or CSV file. Both formats export the keyword list just as it is shown on the screen with the fields separated with a comma.
- **Clear**
Pressing this button clears the keyword list.

Once you have your keywords imported into the grid we are ready to start scraping!

Step 6 – Scraping

- **Start Scraping Button**
This will start scraping your complete keyword list from top to bottom using the specified settings you choose in step 3. The button changes to Stop Scraping while the scraper is running. Pressing this will stop the scraping after the current keyword is processed. You can then press the Start Scraping Button again and it will pick up where it left off.
 - **Scrape Selected Button**
Same as Start Scraping Button except that it only scrapes the selected row for the settings you choose in step 3.
 - **Scrape All Selected Button**
Same Scrape Selected Button in that it only scrapes the selected row but it will scrape all available columns regardless of the settings. It will even scrape columns that are not visible in the data grid. So be sure to check Show All Columns if you want to view this data after running.
-

Step 6 – Export

You can now analyze your results on the screen or export them to a text file or excel via a CSV file. You can also export to an AIS file format that is associated with our program. This way you can open up your scrape runs at any time to analyze the data.

Here is the column layout of the exported files. They are all comma delimited.

Keyword, AIT, AITQ, AIA, AIAQ, AIU, AIUQ, Broad Results, Phrase Results, Competition, Global Searches, Local Searches, AVG CPC, RC, RCQ, Selected, Exact results, colcom, colnet, colorg, RCAIT, RCAITQ, RCAIA, RCAIAQ, RCAIU, RCAIUQ

Any new columns added in the future will be added onto the end to maintain backwards compatibility.

Filters

Filters are available on all scrape options. Applying a filter will permanently delete the keywords that do not meet your set requirements. So be sure to backup first to that you can always revert back to the original copy.

You set the filter to what you want to see. If you only want AITQ under 1,000 then you would set AITQ to less than 1000 and apply the filter. Simple as that. There is also a word filter that you can use to drop off keywords with less than or greater than a set number of words.

Advanced Options

- **Keywords: Export Column Names**
This determines if the keyword files you export contain the column header names or not.
 - **Proxy: Export Column Names**
This determines if the proxy files you export contain the column header names or not.
 - **Proxy: Always re-test proxies that have been tested**
Check to always re-test, un-check to skip.
 - **Proxy: Flag failed proxies during scraping**
Check to flag during scraping, uncheck to keep as a good proxy and try again on next proxy pass.
 - **Errors: Don't show scrape errors**
Check to omit scrape errors (good for unattended runs)
 - **Errors: Number of failed keyword attempts before stopping scraper**
Set this to the number of failures in a row it takes to completely stop the scraper. Default is 10.
-

Questions or Suggestions?

If you have any questions about All In Scraper please direct them to joe@allinscraper.com.

At this time all support requests must go through this email and we will respond as quickly as we can. We welcome all comments about our products and especially encourage you to suggest future enhancements. We are very open to listening to your needs and will try to implement as much as we can in future updates to this program.

We have put a lot of work into AIS. We are proud of the product. If you have ideas to make it better let me hear them. I am the owner, creator and programmer so if you want something done I can make it happen!

Thanks for supporting AIS!

Joe | www.allinscraper.com | joe@allinscraper.com