

# Managing Usage Statistics

- 1 [DSpace Log Converter](#)
- 2 [Filtering and Pruning Spiders](#)
- 3 [Routine SOLR Index Maintenance](#)

## DSpace Log Converter

With the release of DSpace 1.6, new statistics software component was added. DSpace's use of SOLR for statistics makes it possible to have a database of statistics. This in mind, there is the issue of the older log files and how a site can use them. The following command process is able to convert the existing log files and then import them for SOLR use. The user will need to perform this only once.

The Log Converter program converts log files from dspace.log into an intermediate format that can be inserted into SOLR.

Command used:	[dspace]/bin/dspace stats-log-converter
Java class:	org.dspace.statistics.util.ClassicDSpaceLogConverter
Arguments short and long forms):	Description
-i or -in	Input file
-o or -out	Output file
-m or -multiple	Adds a wildcard at the end of input and output, so it would mean dspace.log* would be converted. (For example, the following files would be included because of this argument: <i>dspace.log</i> , <i>dspace.log.1</i> , <i>dspace.log.2</i> , <i>dspace.log.3</i> , etc.)
-n or -newformat	If the log files have been created with DSpace 1.6
-v or -verbose	Display verbose output (helpful for debugging)
-h or -help	Help

The command loads the intermediate log files that have been created by the aforementioned script into SOLR.

Command used:	[dspace]/bin/dspace stats-log-importer
Java class:	org.dspace.statistics.util.StatisticsImporter
Arguments (short and long forms):	Description
-i or --	input file
-m or --	Adds a wildcard at the end of the input, so it would mean dspace.log* would be imported
-s or --	To skip the reverse DNS lookups that work out where a user is from. (The DNS lookup finds the information about the host from its IP address, such as geographical location, etc. This can be slow, and wouldn't work on a server not connected to the internet.)
-v or --	Display verbose output (helpful for debugging)
-l or --	For developers: allows you to import a log file from another system, so because the handles won't exist, it looks up random items in your local system to add hits to instead.
-h or --	Help

Although the DSpace Log Converter applies basic spider filtering (googlebot, yahoo slurp, msnbot), it is far from complete. Please refer to [Filtering and Pruning Spiders](#) for spider removal operations, after converting your old logs.

## Filtering and Pruning Spiders

Command used:	[dspace]/bin/dspace stats-util
Java class:	org.dspace.statistics.util.StatisticsClient
Arguments (short and long forms):	Description
-u or -update-spider-files	Update Spider IP Files from internet into /dspace/config/spiders. Downloads Spider files identified in dspace.cfg under property solr.spiderips.urls. See <a href="#">DSpace SOLR Statistics Configuration</a>
-f or -delete-spiders-by-flag	Delete Spiders in Solr By isBot Flag. Will prune out all records that have isBot:true
-i or -delete-spiders-by-ip	Delete Spiders in Solr By IP Address. Will prune out all records that have IP's that match spider IPs.
-m or -mark-spiders	Update isBog Flag in Solr. Marks any records currently stored in statistics that have IP addresses matched in spiders files
-h or -help	Calls up this brief help table at command line.

---

#### Notes:

The usage of these options is open for the user to choose, If they want to keep spider entires in their repository, they can just mark them using "-m" and they will be excluded from statistics queries when "solr.statistics.query.filter.isBot = true" in the dspace.cfg.

If they want to keep the spiders out of the solr repository, they can run just use the "-i" option and they will be removed immediately.

There are guards in place to control what can be defined as an IP range for a bot, in [dspace]/config/spiders, spider IP address ranges have to be at least 3 subnet sections in length 123.123.123 and IP Ranges can only be on the smallest subnet [123.123.123.0 - 123.123.123.255]. If not, loading that row will cause exceptions in the dspace logs and exclude that IP entry.

## Routine SOLR Index Maintenance

Command used:	[dspace]/bin/dspace stats-util
Java class:	org.dspace.statistics.util.StatisticsClient
Arguments (short and long forms):	Description
-o or -optimize	Run maintenance on the SOLR index. Recommended to run daily, to prevent your servlet container from running out of memory

#### Notes:

The usage of this this option is strongly recommended, you should run this script daily (from crontab or your system's scheduler), to prevent your servlet container from running out of memory.