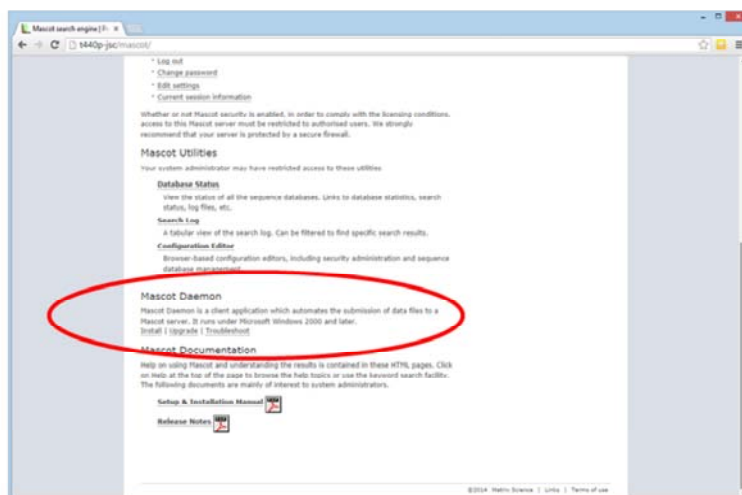


Automation with Mascot Daemon

MASCOT



Installing Mascot Daemon



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Mascot Daemon is our automation client. It's a part of the Mascot package that is only available when you have a Mascot server in-house.

It is only available for Windows because all of the mass spectrometry data systems are Windows packages, so this is always the platform where the data files originate.

Links to install and update Daemon can be found on your local Mascot home page, so that anyone with access to the server can easily get a copy of Daemon. You should try and make sure that you have the same version of Mascot Server and Mascot Daemon installed. An older version of Daemon will probably work, but it won't support features in later versions of Mascot Server.

Installing Mascot Daemon

Mascot Daemon: Preferences

Intranet | Data import filters | ODBC connection | Timer settings | Authentication | General

Mascot server URL, up to and including cgi directory,
(e.g. http://your_server/mascot/cgi/)

http://1440-jsc/mascot/cgi/

HTTP timeout (seconds) 30

HTTP access type Registry default

HTTP proxy server

HTTP protocol ☒ HTTP/1.0 ☐ HTTP/1.1

Save Cancel

MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



When Mascot Daemon is first launched, this dialog will be displayed and you have to enter the URL of your local Mascot server to proceed. Note that you cannot enter the public web site URL ... it simply won't work. If Mascot security is enabled, you will also need to enter a valid user name and password on the authentication tab.

Installing Mascot Daemon

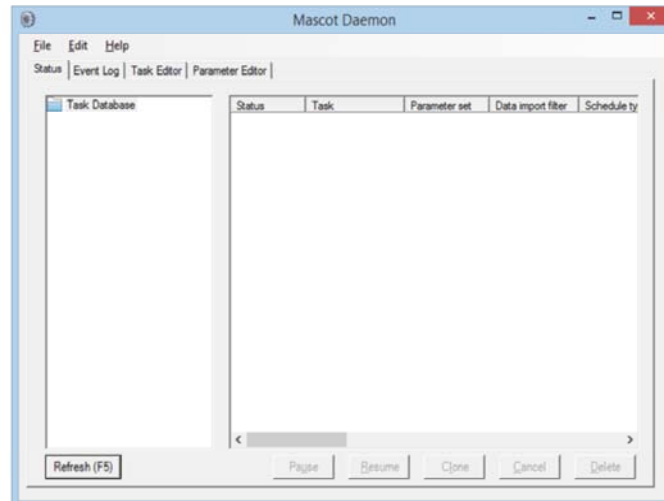
The screenshot shows the 'Mascot Daemon: Preferences' dialog box with the 'Authentication' tab selected. The dialog has a title bar with a close button. Below the title bar are tabs: 'Intranet', 'Data import filters', 'ODBC connection', 'Timer settings', 'Authentication', and 'General'. The 'Authentication' tab contains three sections: 'Web Server Authentication' with 'User name' and 'Password' fields; 'Proxy Server Authentication' with 'User name' and 'Password' fields; and 'Mascot Security' with 'User name' (containing 'nick') and 'Password' (containing '*****') fields. At the bottom right are 'Save' and 'Cancel' buttons.

MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Mascot security is not the same as web server security or having to tunnel through a firewall or authenticated proxy server. There are separate fields for each of these.

Installing Mascot Daemon

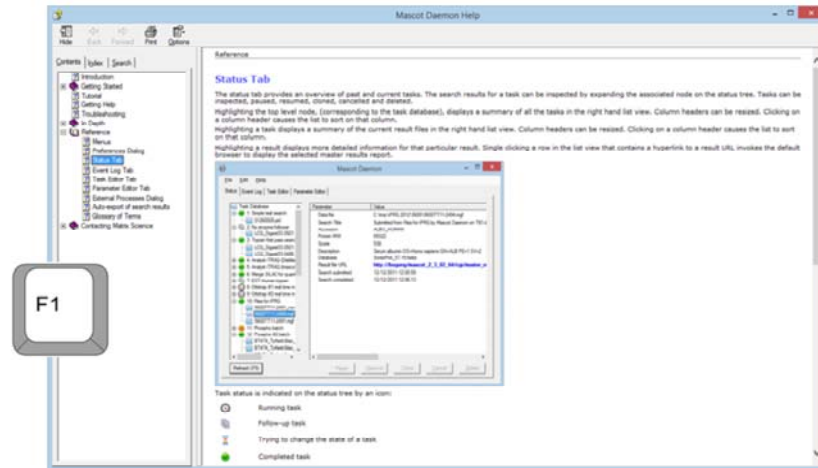


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Once Daemon has connected to the Mascot server, you will see a display like this. The user interface is a simple, tabbed dialog box

Installing Mascot Daemon

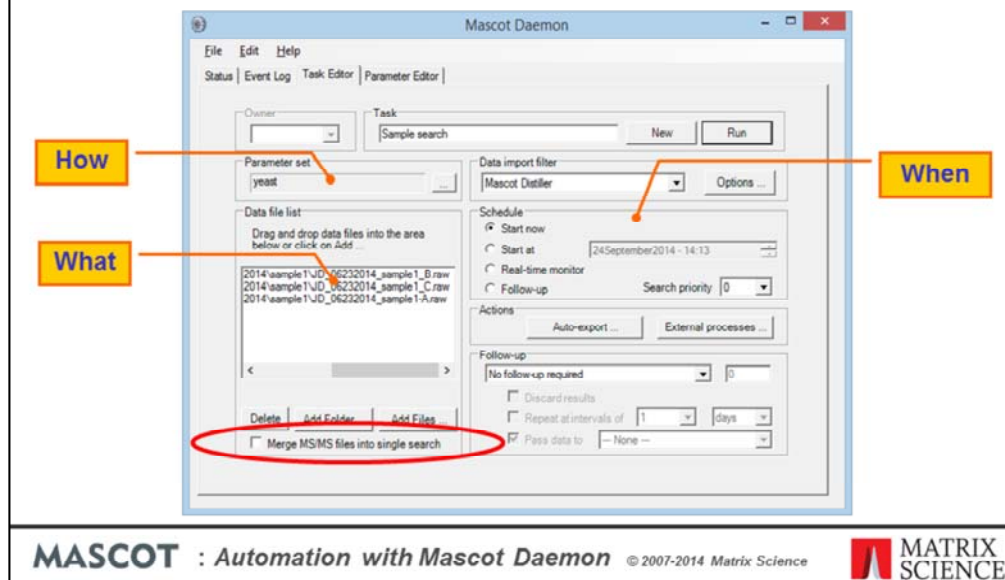


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Mascot Daemon has comprehensive, context sensitive help. Just press F1. There is a tutorial to get you up to speed and detailed reference material.

Task Editor tab



Lets look at the tabs in more detail.

The Task Editor tab is used to define each task. A task defines *what* data will be searched and *when* the search will take place. A Parameter set is selected that contains the search parameters. This determines *how* the data files will be searched.

Here we have a very simple batch task. A single data file has been selected, a parameter set has been chosen, and the task will run the search as soon as we press run.

The data files can be simple peak lists. In this case, we are using Mascot Distiller as a data import filter and the data file is a RAW file. More about data import filters later.

Note the checkbox at the bottom left. Merging a batch of files into a single search can be very useful. MudPIT fractions, for example

Search Parameter Editor

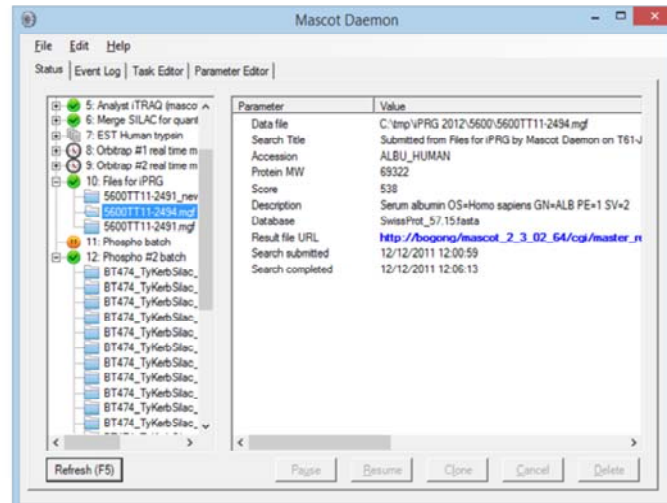
MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



The parameter editor allows sets of search parameters to be defined and saved to disk, so that they can be used over and over again by tasks. The search parameters define *how* the data will be searched. As you can see, the fields are very similar to those on the web browser search form.

The search title includes some tokens, or variable, that will be replaced when the task runs. Here, for example, <taskname> will be replaced by the name of the task and <localhost> will be replaced by the name of the computer on which Daemon is running. This will make it easy to identify where the searches came from when looking at the Mascot search log on the server. There are many such tokens. You can find details in the help file.

Status tab



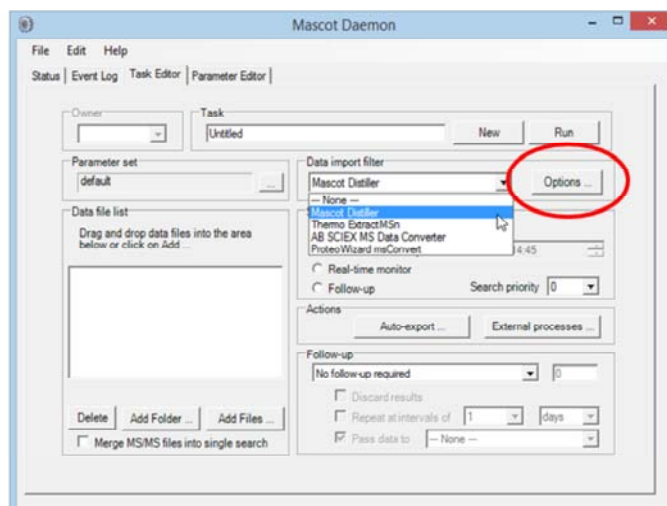
MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Each search result is added to an Explorer-like tree on the Status tab. There is a minimum of summary information, and a hyperlink to the full result report. Mascot Daemon isn't intended to be an interface to browse Mascot results, it is just a tool to automate search submission and keep track of results

Data files within a particular task are run serially, but multiple tasks can be set running at the same time. This means that several data files can be processed and search the same time.

Data Import Filters



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science

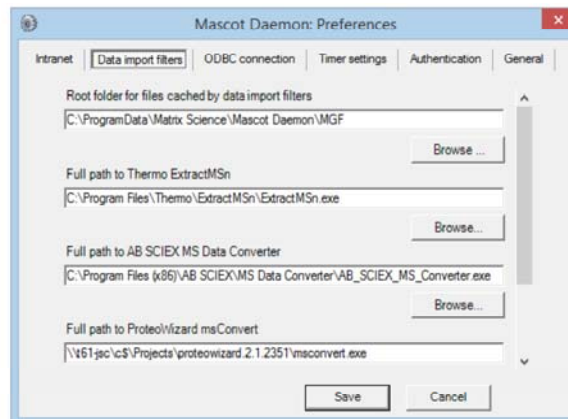


Going back to the task editor tab. For some file formats, we have data import filters. These allow Daemon to generate peak lists directly from the binary files created by the MS data system.

Next to the drop down list, you have a button to view the options for the selected data import filter

For Mascot Distiller, you need to have a licence for the Daemon Toolbox. If present on the Daemon PC, Mascot Distiller will be added to the list of import filters automatically.

Data Import Filters



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



All other data import filters must be configured on the Data import filters tab in Preferences. Install the package, browse to the executable, and save.

Data Import Filters

Thermo ExtractMSn data import options

First scan	
Last scan	
Min. precursor mass	600
Max. precursor mass	3500
Grouping tolerance	1.4
Intermediate scans	1
Precursor charge	A8
Min. scans / group	1
Min. peaks	10
Min. intensity	
Min. S/N	3
Min. major peaks	5

Reset OK Cancel

MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



A number of data import filters are pre-configured and their options are described in the Mascot Daemon help - just press F1

Thermo ExtractMSn is a utility that creates DTA peak lists from Thermo Xcalibur .RAW files. Mascot Daemon calls the utility using the parameters specified in this dialog, then merges the DTA files into a single MGF file. ExtractMSn can be downloaded from the Thermo web site, but Thermo is no longer developing it, so it cannot be used with their latest instruments, such as the Q-Exactive.

Data Import Filters

AB SCIEX MS Data Converter data import options

Raw data format: Wiff

Peak picking: ProteinPilot centroiding

Peak list format: MGF

Compression (mzML): zlib

Precision (mzML): 64-bit float single precision

Create index (mzML): ☒

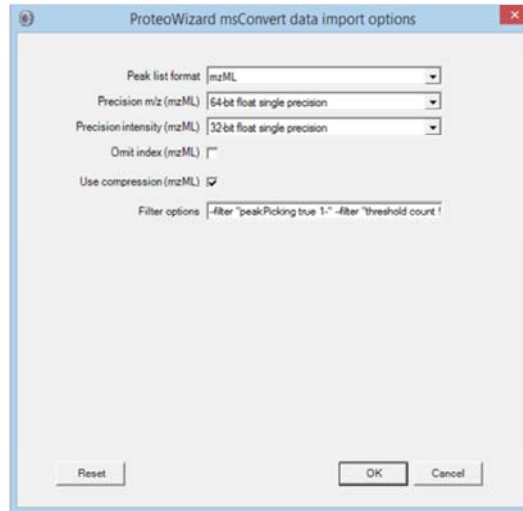
Reset OK Cancel

MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



The AB Sciex MS Data Converter supports both Wiff files and TOF/TOF data. Output can be an MGF peak list file or an mzML file containing a more complete representation of the data. For rapid peak picking, the converter exports the centroids determined by the instrument software during acquisition. For Wiff files, there is also an option to use the peak picking algorithm from ProteinPilot, which gives good results on TripleTOF data

Data Import Filters



The screenshot shows a dialog box titled "ProteoWizard msConvert data import options". It contains several settings for data import:

- Peak list format: mzML
- Precision m/z (mzML): 64-bit float single precision
- Precision intensity (mzML): 32-bit float single precision
- Omit index (mzML): ☐
- Use compression (mzML): ☒
- Filter options: -filter "peakPicking true 1." -filter "threshold count 1"

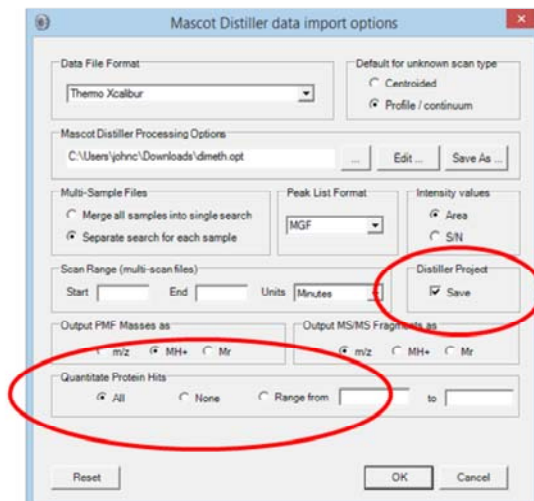
At the bottom of the dialog are three buttons: "Reset", "OK", and "Cancel".

MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



ProteoWizard msconvert is a command line tool for converting between various file formats. It has a large number of options. Note that the input file format is not specified; msconvert auto-detects the format.

Data Import Filters



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



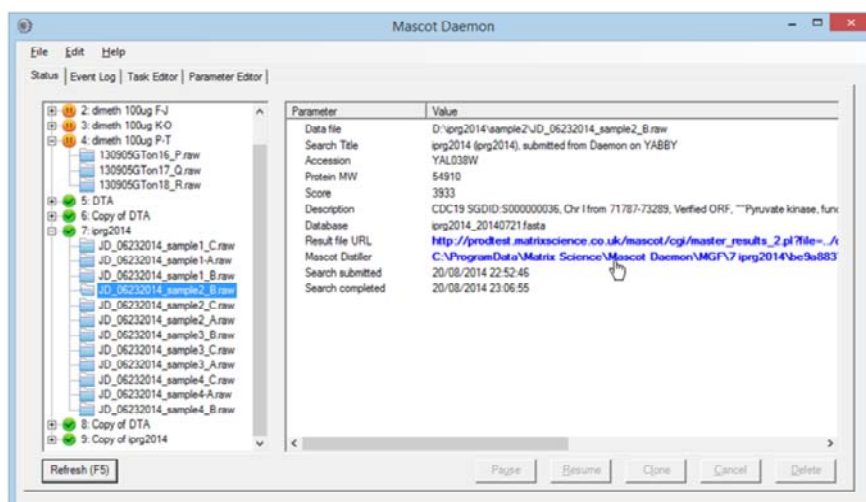
The most powerful data import filter is Mascot Distiller, because it can import all of the major raw file formats. The dialog is relatively simple because the peak picking options are contained in the selected processing options file.

The idea is that you adjust the processing options by working interactively with a typical data file in Mascot Distiller Workstation. You then save these options to a file and select it here.

You can ask Daemon to save a project file containing the Peak lists and Mascot search results.

If you have the search and quantitation toolboxes, you can also batch automate quantitation. We'll be discussing quantitation in detail in another presentation.

Data Import Filters

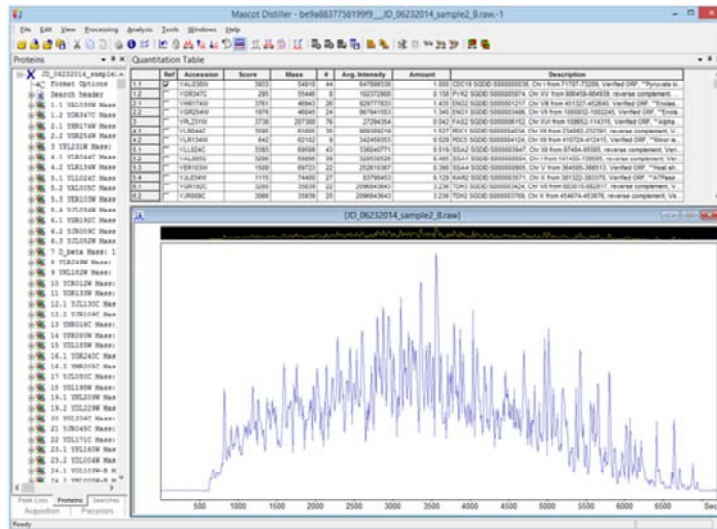


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



If Distiller is used for peak picking, once the search has completed, you get a hyperlink to open the project in Mascot Distiller. Click on the link and ...

Data Import Filters



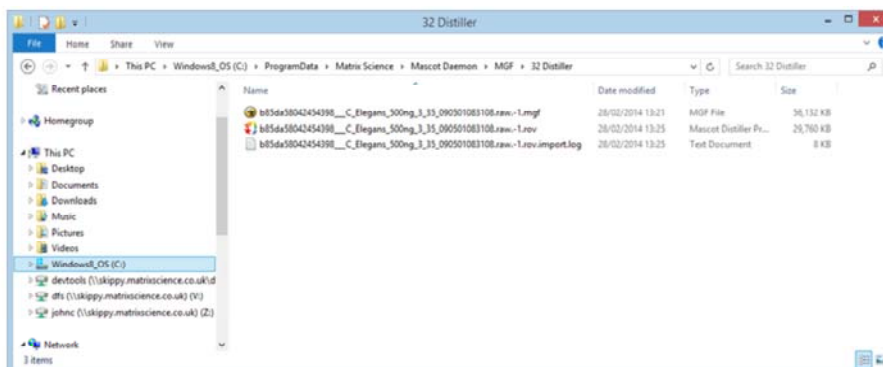
MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Shazam

Data Import Filters

- Peak lists and Distiller projects created by Daemon are saved to disk:
%allusersprofile%\Matrix Science\Mascot
Daemon\MGF\<task name>



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



- Peak lists created by Daemon are always saved to disk. This is the default path, which can be changed in preferences.
- When you delete a task, any peak lists are also deleted. If you have used Mascot Distiller as in import filter, then the distiller project file is also deleted. The raw data files are never deleted.
- You may also find Distiller project files, log files and result exports in the same folder

Types of task

Batch task

- A batch of data files to be searched immediately or at a defined time

Real-time monitor task

- New files on a defined path are searched as they are created

Follow-up task

- Accepts data files from another task. For example, to repeat a search against a different database

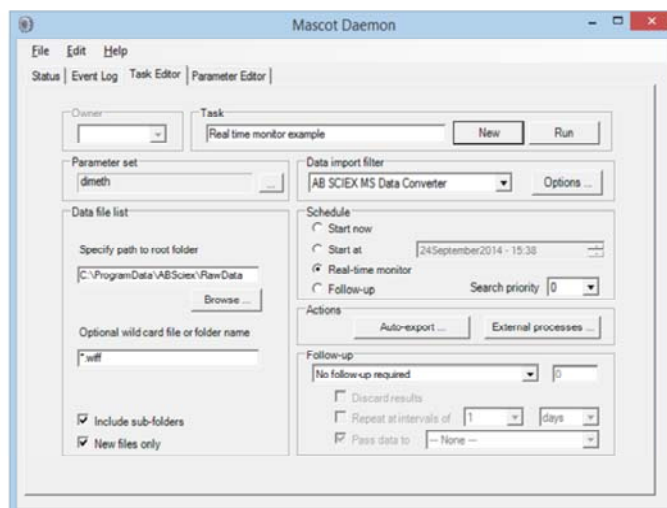
Lets talk in more detail about tasks. Mascot Daemon supports three kinds of task:

A batch task allows you to search one or more specified files immediately or at a defined date and time

A real-time monitor task is the key to real automation. You define a task using a file path for the data file which includes wild cards. Any file that matches this path is picked up and searched.

A follow-up task is a task where the data files are passed to it by another task. This is very powerful because it allows searches to be chained together to implement complex workflows. For example, a batch of data files might be screened against a contaminants database containing entries for keratins, BSA, trypsin, etc. Those data files which fail to find a match can then be automatically searched against a non-redundant protein database. Spectra which are still unmatched can then be searched against an EST database, etc., etc.

Real-time Monitor Tasks



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



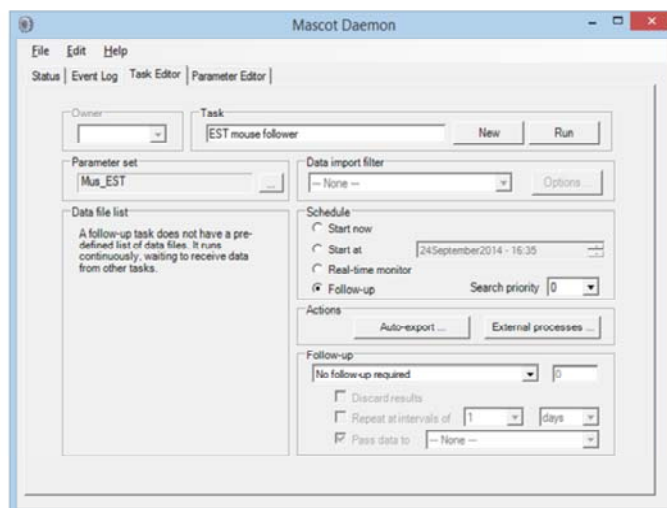
When you select real-time monitor in the schedule frame, the task editor tab looks like this. Here, we are looking in the specified directory and all its subdirectories, for any file called with the extension wiff.

Note that the final part of the path says “file or folder name”. This is because some raw files are actually folders, not files.

If Include sub-folders is checked, then Daemon will search down through all sub-directories of the root folder

If new files only is checked, then any files that exist when the task is started will be ignored. A common problem when using real-time monitor like a batch task is to have this checked and then wonder why nothing is happening

Follow-up Tasks



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science

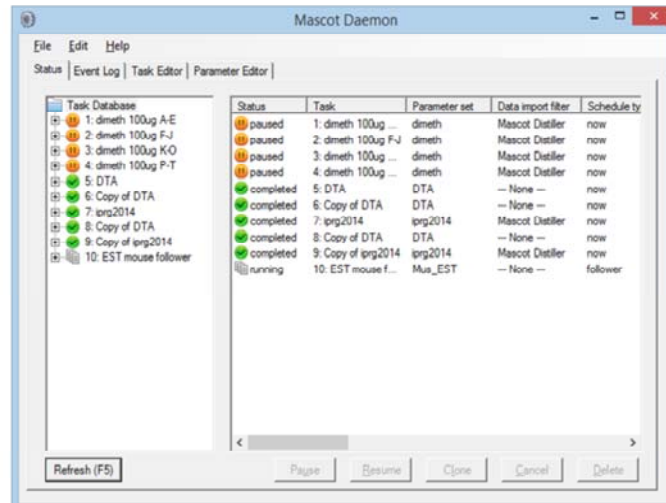


Follow-up tasks are a little more difficult to explain. Any task can pass data to a follow-up task to be searched again in a different way. Only follow-up tasks can receive data from other tasks in this way.

Lets illustrate with an example. Imagine we want to search our MS/MS data files against the mouse proteome as a first pass search. We then want to search all spectra that fail to get a significant match against a database of mouse EST sequences.

We start at the end of the chain, and define the EST search as a follow-up task. We just choose Follow-up in the schedule frame and a parameter set, created earlier in the parameter editor, that contains the required search parameters.

Follow-up Tasks

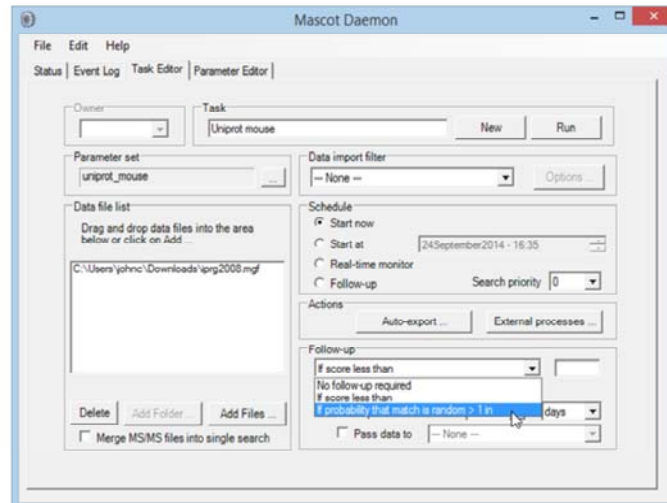


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



We press run, and we see our Follow-up task running. Although the task is running, nothing will happen until we feed some data into it

Follow-up Tasks



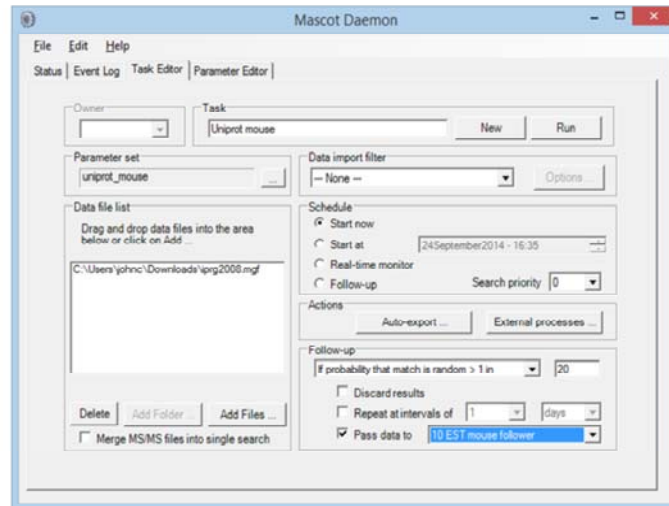
MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



We go back to the task editor and create a simple batch task. There is one data file and the search parameters are a different set, in which the databases are the Uniprot complete mouse proteomes plus second database of contaminant sequences.

Before pressing Run, we have to link this task to the follow-up task. We need to specify a condition for passing the data to the follow-up task. This can be score based or probability based.

Follow-up Tasks

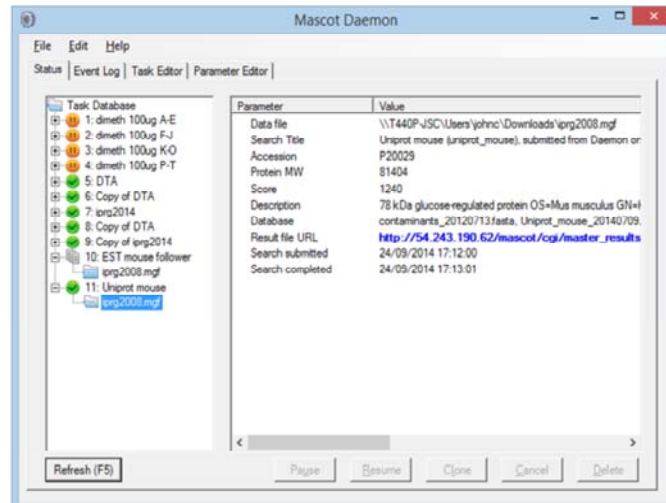


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



For this example, we will use the default significance threshold of 5%. If the score a peptide match in the first search is below a 5% significance threshold, then the spectrum will be passed to the follow-up task we created earlier. Note that you also have the option to repeat the search at intervals, maybe to see whether a better match can be found in some future update to the sequence database. The task is now configured, and we can press Run.

Follow-up Tasks



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Some time later, the status tab looks like this. The results for the first search appear under task 14. Those spectra that failed to obtain a significant match were passed to the follow-up task, task 13, where the result information appears once the search has completed

Follow-up Tasks

- Can chain follow-up tasks to create complex workflows
- To create a chain, start at the end and work back to the beginning
- Multiple batch or real-time monitor tasks can feed into a follow-up task
- Can also drag and drop result nodes onto follow-up task nodes

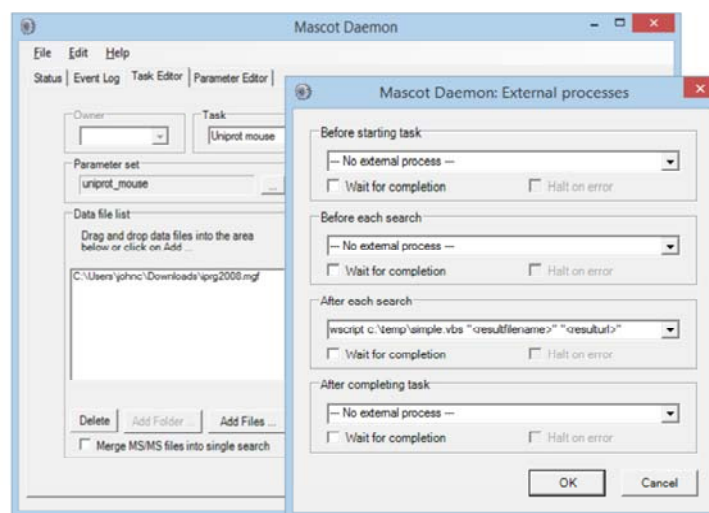
•One follow-up task can pass data to another follow-up task, creating a chain of tasks to implement a complex workflow. Thresholding applies to individual spectra. If there are 1000 spectra, and 100 get a good match, we take the 900 that didn't get a match and pass them to the follow-up task. You can think of the chain of tasks as a series of finer and finer sieves.

•Remember that, when you define the chain, you have to start at the end of the chain, with the last task, and work forwards to the first task in the chain

•Having created a follow-up task, or a chain of follow-up tasks, multiple batch or real-time monitor tasks can feed into it. It just sits there, waiting

•Note that any result node can be dragged and dropped onto a follow-up task node to repeat the search using a different parameter set

External Processes



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science

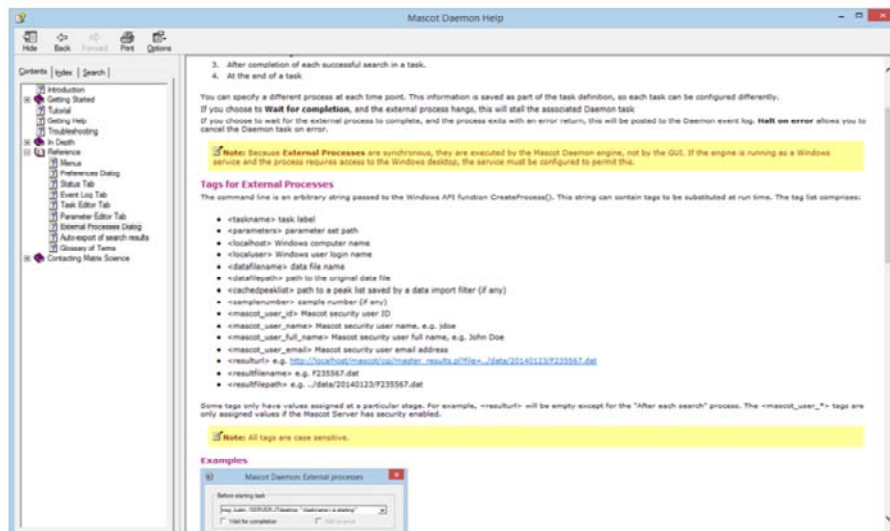


We'll briefly touch on a few other features of Daemon. The task editor tab includes a button labelled 'External processes'. A Mascot Daemon task can execute external processes at one or more of the following time points:

- At the start of a task
- Before submitting each search in a task
- After completion of each successful search in a task.
- At the end of a task

A process is a one line command. This could execute a Windows program, or it could be a batch file or script to perform some more complex procedure. You can pass arguments to the commands by including tokens, such as <taskname>

External Processes

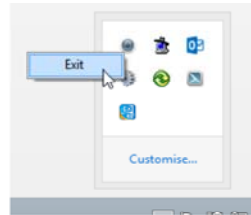


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



The help file contains a list of all the available tokens. It also gives examples of how to use External processes, such as a script to build an HTML page in real-time that contains hyperlinks to the search results.

Mascot Daemon Engine

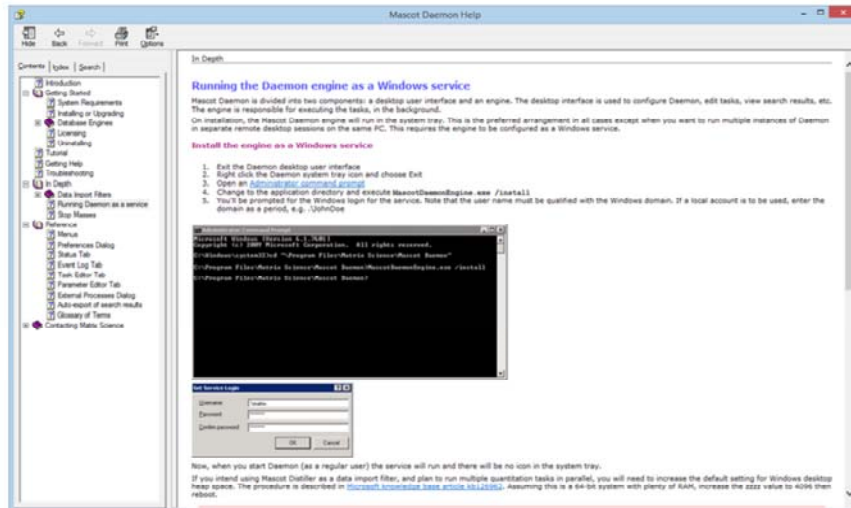


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Mascot Daemon is divided into two components: a Graphical User Interface (GUI) and an engine. We have been looking at the GUI, which runs on the desktop and is used to configure Daemon, edit tasks, view search results, etc. The engine is responsible for executing the tasks, in the background, whether the GUI is running or not. If something goes horribly wrong, you may need to stop and start the engine to clear the problem. Exit Daemon on the desktop then locate the Daemon icon in the system tray, right click it, and choose exit

Mascot Daemon Engine

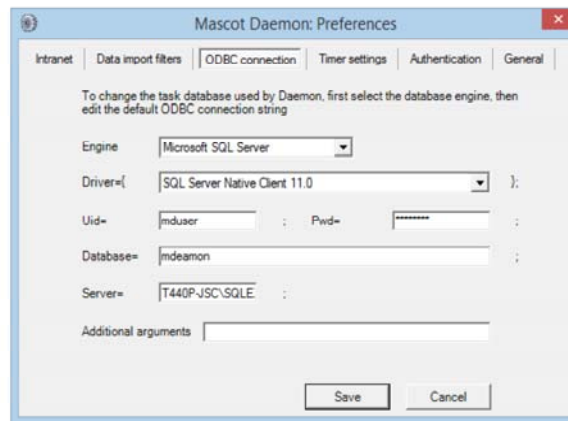


MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Running the engine in the system tray is the preferred arrangement in all cases except when you want to run multiple instances of Daemon in separate remote desktop sessions on the same PC. This requires the engine to be installed as a Windows service. You'll find full instructions in the help

The Mascot Daemon Database



MASCOT : Automation with Mascot Daemon © 2007-2014 Matrix Science



Under the hood, Mascot Daemon stores information in a relational database. The installation default is to use an embedded database engine called VistaDB. This is the recommended choice unless you wish to share a single task database between multiple instances of Damon running on different PCs, which requires a high performance database engine. You can choose from Microsoft SQL Server, MySQL, Oracle, and PostgreSQL. These are configured on the ODBC connection tab in Daemon preferences. When multiple instance of Mascot Daemon share the same task database, each task is "owned" by the instance that created it. You can only Pause, Resume, Cancel, or Delete your own tasks. However, any instance can Clone any task and can transfer data files to any follow-up task, either by dragging and dropping the result node or by specifying the task in follow-up criteria.