

Advanced Searching with QUOSA Information Manager

Full-Article Retrieval / Alerts / Advanced Full-Text Search / Cluster Analysis

Advanced Searching with QUOSA

This note describes the main advanced searching capabilities provided by QUOSA Information Manager. We do not describe the initial full text article retrieval from a PubMed or similar search; nor reviewing those articles, links with Citation managers, or folder management.

The subjects covered are:

- **Search alerts or automated updating of previous searches**
- **Secondary searches within existing search lists or folders**
- **Concepts4Clustering**
- **Batch queries of PubMed**
- **Terms4Clustering**
- **Using Terms4Clustering with Genes and Aliases search lists**

Search alerts

It is possible to set up Quosa automatically to update any search you have made with new additions to the search results returned by PubMed. At present search alerts works only when Quosa is running, although we hope to remove this constraint shortly. Search Alerts will work with any search gateway where you do not need to log on, including "General Web" with Google.

Note that PubMed ranks returns to a search according to the date of accession. This means that new entries to PubMed will be at top of a list of returns, so that if you had asked for the top 50 returns, a new search will differ from an old search by including the new PubMed entries. Quosa notes this difference and adds the new entries to your existing list. Optionally, QUOSA can be configured to send you an email alerting you to a new addition.

Search alerts works with the primary search that was originally made on PubMed. It is not available on secondary searchers that you perform later.

You can configure a search alert with the tool hammer icon on the Results pane. First select the search you want to be updated by clicking on the primary search name in My Searches and Folders. If you then click on the hammer icon, you will be able to click on Configure Search Alert. You can then set the time and frequency at which you want searches conducted. You could also set up an e-mail alert in case there is an update.

Once you have configured a Search Alert, the search list will be duplicated under My Alerts in MY Searches and Folders. It is otherwise unchanged. You can still select items from within the search results and extract them to EndNote or copy them to folders. Note that if you delete an item from a search alert list, then QUOSA will notice its absence at the next update and include the restoration of the item in the update.

Remember that QUOSA needs to be running at the designated update time in order for Search Alert to work.

Advanced Searching with QUOSA Information Manager

Full-Article Retrieval / Alerts / Advanced Full-Text Search / Cluster Analysis

Secondary searches within existing search lists or folders

Quosa offers powerful secondary search capabilities on either primary searches you have conducted on PubMed (or similar gateways) or on folders of full text articles which you have built up. Note that your primary searches will typically have been made on information only available in abstracts. Now that you have downloaded the full text article with the help of QUOSA, you have your first opportunity to search the full text all those articles.

QUOSA's features of analysis, automation, and speed enable you quickly to find what you need from a primary search. In doing so it also opens up the opportunity to retrieve large sets of full text articles from primary searches and to quickly and effectively use QUOSA search tools to identify the articles of interest to you.

Boolean queries

The simplest such tool is Search in Results. This conducts a secondary search with a Boolean query of your choice on any search list or folder available under My Searches and Folders. Help with formulating Boolean queries is provided from our website at <http://www.quosa.com/download/BooleanHelp.html>

The Search in Results tool is accessible from the magnifying glass icon on the Results pane toolbar. You need to have a search list or folder selected before using this, which will be the case if you have files listed in the Results pane. Clicking the magnifying glass icon will bring up a dialogue box for your Boolean search query on the full text> It also provides the option to add search terms to be applied to metadata, such as author or journal. When you click "Search" on the dialogue box, the secondary search is performed very rapidly. The dialog box stays open in case you wish to conduct a second search. Notice that a new search list has been created in My Searches and Folders, positioned as a sub search to the primary search. Notice that this new search list is the one currently selected. If you proceed to do a further search under the current set, that would be a search on the secondary list and not on the primary search. Reselect the primary search to search on it.

Note that in addition to being able to search on the currently selected list (the current set), you have the choice to search on all search lists, all folders, or both all folders and all search lists.

Having conducted a secondary search, you may wish to highlight the results with the search term used. To do this, click the highlight the icon. Having done this, you'll be able to use Document Summary as you would with the primary search to review the key text in each article in turn. Highlighting also has the effect of analyzing the concepts within a text and showing them in the Concepts pane of the document summary.

Regular Expression and WildCard Queries

These are Queries more oriented to sets of characters than sets of words. Help with formulating Regular Expression and WildCard Queries is provided from our website at <http://www.quosa.com/download/RegExpHelp.html>

The WildCard Queries are a subset of Regular Expression Queries for simple, character oriented WildCard Queries. Note that there are differences between these WildCard Queries and those available under the Boolean Search in Results.

Advanced Searching with QUOSA Information Manager

Full-Article Retrieval / Alerts / Advanced Full-Text Search / Cluster Analysis

These two Queries are accessed under Regular Expression Search via the hammer icon on the Results pane toolbar. If the results you want searched are not already in the Results pane, you need to select the relevant search list or folder. The dialog box that comes up when you click Regular Expression Search offers you a choice between a WildCard query and a wider scope Regular Expression query.

Like search in results, Regular Expression Queries create a sub folder with their results appended to the primary search list or folder. Note that in this case relevant text in the results will be automatically highlighted and reproduced in the document summary for each article. If you choose to re-highlight the secondary research with other text, you should bear in mind that the highlighter operates like a Boolean query.

Concepts4Clustering

Concepts4Clustering is a method for automatically extracting from a body of article concepts which occur in more than one article. It can be applied to any search list or folder within My Searches and Folders. It is useful for both prompting the user with concepts that could be relevant to his or her inquiry and generating a list of concepts for further searches of the types described in later sections.

To extract concepts from a list, first select the list and then click the Concepts4Clustering tab. This tab is grouped with the Document Summary, My Searches and Folders, and Terms4Clustering tabs on the Organizer. The Organizer will then switch to the Concepts4Clustering view. If no concepts are found, this will be indicated. Otherwise a list of concepts is presented, the most numerous at the top. For each concept listed, the number of articles with significant instances of that concept is given. Each concept is expandable to show the number of articles with significant instances of variants of the main concept.

When the user clicks on a concept the articles in the Results pane are reordered so that those including the concept are brought to the top of a list. A new field called "In Cluster" is shown in the Results pane, with a cluster icon against all articles including the relevant concept. This reordering occurs whenever a new concept is selected in the organizer pane.

If you right click on a concept in the Organizer pane, you are presented with a number of choices, including to reorder the concepts alphabetically rather than by cluster size and to save concepts to a file. Saving concepts to a file allows one to keep a record of generated concepts. It also allows one to use a set of concepts relevant to one body of articles for searches on the other lists of articles such as with the batch queries and Terms4Clustering described below.

Advanced Searching with QUOSA Information Manager

Full-Article Retrieval / Alerts / Advanced Full-Text Search / Cluster Analysis

Import Batch Query

This tool and all the QUOSA tools described later are only available in Professional versions of the QUOSA Information Manager.

Import Batch Query allows the user to automate a series of primary PubMed searches along with a further series of secondary searches on the full text articles retrieved in each of those primary searches. QUOSA can automate the whole batch of searches, saving time and making possible searches that would otherwise be impracticable.

The tool is accessed from the Import Batch Query choice on the Tools menu at the top of the QUOSA window. This offers the choice of:

- A batch query on the Author field in PubMed from a list of authors provided by the user, and
- A batch query on the full abstract in PubMed for a list of search terms.

There is also a guide for this tool on our WebSite at:
<http://www.quosa.com/download/ImportQueryHelp.htm>

The author and normal batch queries work in the same way. You can see from the Help note above an example query in Excel as follows:

	A	B	C	D
1	{pthrp}	expr	[R]S.h	[W]diff*
2	{heat stroke}	"effect of"	[W]bi*r results	[R]g.*th
3	{asthma}	pl*te	[B]"Annexin A8"	[R]a[0-9]+ b[0-9]+ c[0-9]+

If you use this list, QUOSA will retrieve a number of files under each search term in column A. Note that this search term is in parentheses - formatted as a PubMed search expression. You can copy and paste PubMed search expressions for this purpose.

After each such primary search, QUOSA will perform a number of secondary searches on the results of the primary search, according to the contents of columns B and above in the same row as the primary search expression.

In the example below, you see a row for which the primary search expression is {asthma} and then a number of secondary expressions in quotes (to signify that a hit with both terms is required) and a number which is a measure of the desired proximity of those terms.

{asthma}	"alias big" ~4	"candidate gene responsible" ~4	"detectable nr4a3 fusion" ~4
----------	----------------	---------------------------------	------------------------------

Search lists like these can be generated from saving to file concepts generated by Concepts4Clustering, as described in the earlier section. You can take any list like that and copy it into an Excel spreadsheet to get an automated batch query. In fact, the transformations needed to do this in Excel can themselves be automated in an Excel macro.

CSV files can also be used for this purpose.

Advanced Searching with QUOSA Information Manager

Full-Article Retrieval / Alerts / Advanced Full-Text Search / Cluster Analysis

The process of performing an Import Batch query consist of first setting the size of each of the primary searches (e.g. 100 articles) and then selecting the Excel or CSV file containing the primary and secondary search terms.

When QUOSA has finished the batch query, each primary search will be shown in My Searches and Folders, with each of the secondary searches shown as a sub-search to the relevant primary search. It is exactly as if a series of manual Searches in Results had been performed on each of the primary searches.

Terms4Clustering

This is sometimes informally referred to as a dictionary query. It allows the user to see how many articles in any search list or folder contain any instances of each of a list of search terms, i.e. to see the cluster size for each search term. The results are presented like Concepts4Clustering results: the relevant search terms (in this case) ranked in descending cluster size. (The resulting search terms may also be ranked alphabetically. Terms not found in any article are not displayed in the results.)

As with Concepts4Clustering, when the user clicks on a search term the articles in the Results pane are reordered so that those including the search term are brought to the top of the list. The In Cluster field is shown in the Results pane, with a cluster icon against all articles which include the relevant search term.

You can use any list stored on your computer as a txt, csv, or xls file. An example would be a text file created by saving to file a list of concepts generated by Concepts4Clustering, such as the following, saved as concepts.txt:

```
alias big  
candidate gene responsible  
detectable nr4a3 fusion genes  
differential gene  
differentiation potential  
gene mutation  
human neuroblastoma cells  
mutation alg3 gene causing  
net gene promoter polymorphic  
nr4a3 fusion genes  
polymorphism gene  
rara gene fusion products  
sert gene promoter polymorphism  
tp53 gene mutations
```

To apply this list to Terms4Clustering, go to "Tools" menu at the very top of the QUOSA window and select Tools/ Terms4Clustering/Configure Clustering. Then browse for concepts.txt (in this example) and click OK.

Terms4Clustering will continue to use this list until another is selected. So if one list is used repeatedly, there is no need to configure the list each time. There is also an option to edit the list.

Once the right list loaded, select the folder to search, make sure files from it are showing in the Results pane, and then click the Terms4Clustering Tab on the left (stacked with the Documents

Advanced Searching with QUOSA Information Manager

Full-Article Retrieval / Alerts / Advanced Full-Text Search / Cluster Analysis

Summary, Concepts4Clustering, and My Searches & Folders tabs). The Terms4Clustering view will open and display the name of the search list and the number of hits for each search term.

As mentioned earlier, the articles in the Results pane will be re-ordered when any search term in the Terms4Clustering pane is selected, with the Cluster icon showing against those articles including the selected search term. The user can highlight the relevant text with the highlighter tool and then find it via the Document Summary tab, which shows the passages with hits for all words in the search expression and also those with only a partial hit.

(Note that if one does not highlight the articles, they may be highlighted with a search term used earlier for that folder. Also, if for any reason you lose the Cluster-ordered view in the Results pane, you can retrieve it by clicking first on another tab (e.g. Document Summary) and then clicking back to Terms4Clustering and then the relevant search term. This won't affect the highlighting.)

Highlighting will highlight all articles in the search list. One advantage of this is that it reveals the incidence of hits which are not selected as being in the cluster. This includes articles with only partial hits and articles that have passages containing all search words but spaced further apart than the default proximity in Terms4Clustering.

Searching with Genes and aliases

QUOSA can perform a Terms4Clustering search with a search term list automatically generated for from a query on NCBI's Gene database. So – as a hypothetical example – one could get a gene term list from an “aging” query on the Gene database and quickly see the hits for those gene terms in a large group of articles on asthma.

To do this, go to Tools/ Terms4Clustering/Create Gene Dictionary. You get a dialog box with an explanation of these next steps. OK that and Entrez Gene appears in the browser pane. We are going to enter “aging” into the search term field. The return page lists the first 20 returns. In the toolbar at the top of the browser pane, we set the “Get” number to 29 and click on the Sigma button.

We then get a dialog box asking for a list file name and location and offering a choice between a dictionary with grouping and a flat list. The grouped dictionary will result in an hierarchical search and tiered search returns. A flat list will treat all search terms as equal and primary. Once this choice is made QUOSA starts to compile the list. This may take some time. If a dialog box warns of a NCBI Internal Error, repeat the process from the point of clicking on “Go” in Entrez Gene” to retrieve genes for your search term.

The result of this either a flat or hierarchic list of genes and aliases drawn from the search on Entrez Gene in a CSV file, which can be configured as the selected list for Terms4Clustering. When an hierarchic list is used in Terms4Clustering, the results are first shown at the top level of the hierarchy and are expandable to show the results for subsidiary terms.

7-Dec-06