Research Tools: Multiple Searches

What can you do with the Multiple Searches page?

- Get frequencies, plots or concordance lines for more than one word at a time;
- Get results with all other word forms;
- Get results using wildcard searches;
- Get results filtered in a number of different ways.

How does it work?

The results for the different words will be merged, and then the sample (e.g. 200 lines) will be selected using the fixed random value for each word (token) in the corpus.   After the results have been downloaded, this sample will then be sorted using the selected ranking order.

Further details

If you want to merge results of two or more words (e.g. *he* or *she*) you can use the Multiple Searches page on the Research Tools tab.   You can enter one or more words separated in space in up to 5 slots.   There is also a box where you can enter a list of words which must not occur in the results.   If you keep the Ordered List box ticked, this will mean that the words from top to bottom will appear from left to right in the results.

For example, to merge the results of *he said* with *she said*, you would enter *he she* in the top box and *said* in the second box.   Ticking the No gaps option will ensure no other words occur in between.   You can also select the "Include word families" option to extend the words you have entered to include different word forms of each word.

Known issue with *true* and *false*
At the present time, the word families are based on wordlists and/or an automatic stemming process.   Because of an issue when the lists were created, the words "true" and "false" are not currently linked to other word forms.   It is only these two words which have this problem, and it is something which will be slowly resolved as the corpora are re-processed.

Use the Frequencies button to see a table of different patterns and their raw frequencies along with normalized frequencies per million words.

Use the Plots button to see how words are distributed through texts.   You can alter the number of "buckets" and the width of the lines using settings on the Options tab.

You can also see the lines and cards by clicking the buttons to send results to the left or right.

Note: you can only send results to the right hand side of the display after you have already performed a search with results on the left.

If you enter words in one or more of the Co-text boxes, before retrieving results tPM will check whether the words in the "Node word(s)" box have the lowest frequency.
If the words in one of the Co-text boxes have a lower frequency, it is much faster to retrieve results if these are treated as the Node word(s).   If the message below appears, you can simply click "Optimize Node" to adjust the node for your search to make it more efficient.

If you want to keep your currently selected node word(s), you can simply click the Frequencies or Cards & Lines button again to skip node optimization.
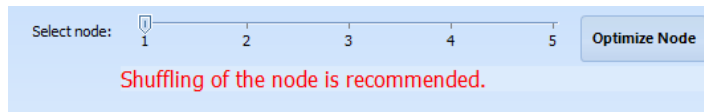
Multi-word searches are performed by first locating all of the instances of the node word(s) and then going through each to check the nearby Co-text words.   For example, in the BNC: Academic sub-corpus, *of* has a raw frequency of 701,144 while *value* has a raw frequency of 4,846.

Going through 4,846 instances of *value* to check for the word *of* will be much quicker than going through 701,144 instances of *of* to check for the word *value* in the co-text.

That's why allowing tPM to shuffle the node will mean results are obtained more quickly.

Filtering results

There are a number of ways in which you can filter the results.

Single category
Each of the texts in each corpus has been assigned to a major text category.   The way these are organized varies from corpus to corpus; you can find some additional information about each corpus in help file 003.   If you select this option, the results will only be included for the specified text category.

Advanced filter
Clicking on the button next to this option will take you to another screen where you can add additional filters from "Priming filters" or "Label filters".
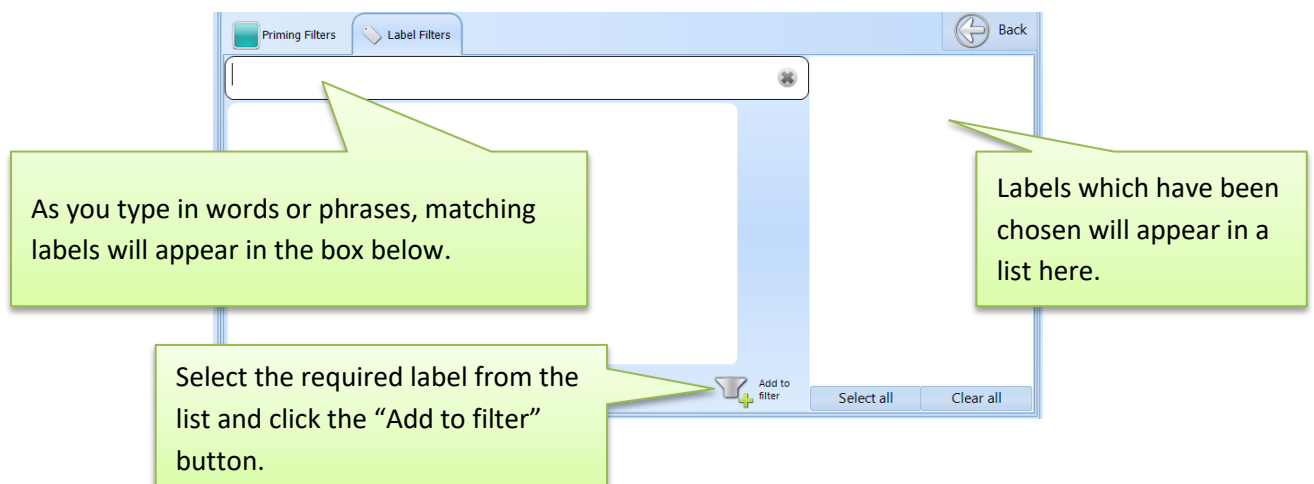
Priming Filters
By default, results will include instances in any contextual environment.   Using the Priming Filters tab, you can deselect specific features and thereby limit a search to any of the specific contextual environments which are displayed for normal searches on the Graphs Tab.

If you untick some features, results will only include concordance lines matching the ticked items.
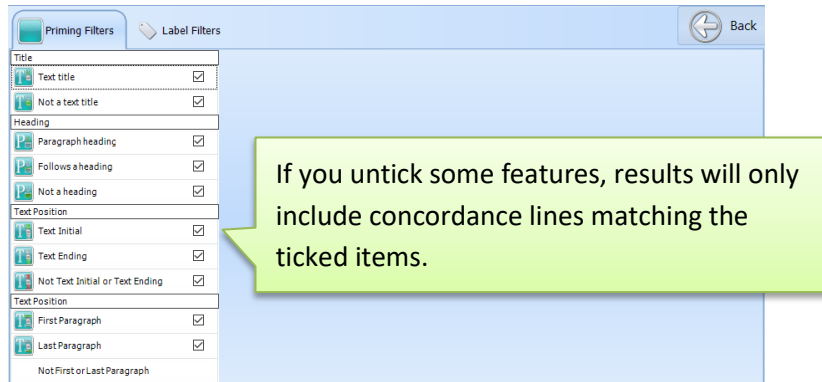
Label Filters

Labels include the major text categories, labels related to the text or producer (author or speaker), as well as section headings and neighbourhood labels.   As you start to type in a word or string of letters, matching labels will be displayed, with the total number of words (tokens) which match the label.   Clicking the Add to filter button will add the selected label to a list of filters which can be selected or deselected.

As you type in words or phrases, matching labels will appear in the box below.

Labels which have been chosen will appear in a list here.

Select the required label from the list and click the "Add to filter" button.

When you have finished setting the requirements for the advanced filter, you can return to the Multiple Searches page by clicking the **Back button**.

There are some important differences between normal searches and searches which are performed using the Multiple Searches page.   The concordance line results will not include any priming summary data, so the Frequencies, Collocations, Labels and Associates tabs will be empty.   You will be able to see some data on the Graphs tab, but these will only be for the selection of lines which have been downloaded.   No icons will appear at the bottom of the screen and no data appears in the table display on the Graphs tab.

Support

The Prime Machine is still undergoing development.

For further information see http://help.theprimemachine.com

Last Updated: Friday, October 26, 2018