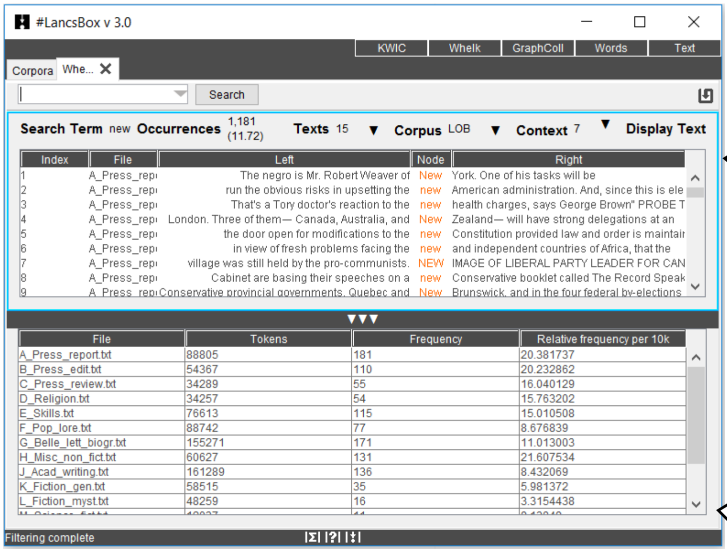
5 Whelk tool

The Whelk tool provides information about how the search term is distributed across corpus files.
It can be used, for example, to:

- Find absolute and relative frequencies of the search term in corpus files.
- Filter the results according to different criteria.
- Sort files according to absolute and relative frequencies of the search term.

5.1 Visual summary of Whelk tab



Search Term new Occurrences 1,181 (11.72) Texts 15 Corpus LOB Context 7

Index	File	Left	Node	Right
1	A_Press_rep	The negro is Mr. Robert Weaver of	New	York. One of his tasks will be
2	A_Press_rep	run the obvious risks in upsetting the	new	American administration. And, since this is ele
3	A_Press_rep	That's a Tory doctor's reaction to the	new	health charges, says George Brown" PROBE T
4	A_Press_rep	London. Three of them— Canada, Australia, and	New	Zealand— will have strong delegations at an
5	A_Press_rep	the door open for modifications to the	new	Constitution provided law and order is maintair
6	A_Press_rep	in view of fresh problems facing the	new	and independent countries of Africa, that the
7	A_Press_rep	village was still held by the pro-communists.	NEW	IMAGE OF LIBERAL PARTY LEADER FOR CAN
8	A_Press_rep	Cabinet are basing their speeches on a	new	Conservative booklet called The Record Speak
9	A_Press_rep	Conservative provincial governments. Quebec and	New	Brunswick, and in the four federal by-elections

File	Tokens	Frequency	Relative frequency per 10k
A_Press_report.txt	88805	181	20.381737
B_Press_edit.txt	54367	110	20.232862
C_Press_review.txt	34289	55	16.040129
D_Religion.txt	34257	54	15.763202
E_Skills.txt	76613	115	15.010508
F_Pop_lore.txt	88742	77	8.676839
G_Belle_lett_biogr.txt	155271	171	11.013003
H_Misc_non_fict.txt	60627	131	21.607534
J_Acad_writing.txt	161289	136	8.432069
K_Fiction_gen.txt	58515	35	5.981372
L_Fiction_myst.txt	48259	16	3.3154438

Filtering complete

Top panel: Searching corpora

You can:
- Search, sort and filter.
- Use simple and advanced searching functionality.
- Use 'smart' searches.

Bottom panel: Displaying distribution

You can:
- View the distribution of the search term in individual files.
- Sort, filter and copy/paste.

5.2 Top panel: KWIC

The top panel in Whelk has the same powerful search, sort and filter functionalities as the KWIC tool (see Section 4). It is directly connected to the bottom panel: any update in the top panel is immediately reflected in the bottom panel.

5.3 Bottom panel: Frequency distribution

The bottom panel in Whelk provides detailed information about the distribution of the search term.

1. 'File' column lists the name of the individual files in the corpus.
2. 'Tokens' column provides the information about the size of each file in running words (tokens).
3. 'Frequency' column provides absolute frequencies of the search term i.e. refers to how many instances of the search term there are in each file.
4. 'Relative frequency per 10k' provides relative frequency normalised to the basis of 10,000 tokens; this value is comparable across files and corpora.

▶ Did you know?

The Whelk tool (both the name and the functionality) is inspired by Kilgarriff’s (1997: 138ff) notion of the ‘whelks problem’. Imagine, says Kilgarriff, that you have a corpus which includes one text (a book) about whelks – small snail-like sea creatures (🐚). In this text, the word *whelks* will appear many times and hence will appear as a frequent word in the entire corpus, although its use is limited to one specific context. To overcome the problem and present more accurate information about word distribution, the Whelk tool shows the frequency distribution of search terms in individual corpus files.