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.

#LancsBox v 3.0			- 🗆	X Top panel: Searching corpora
orpora Whe X	Search	KWIC Whe ts 15 ▼ Corpus LOB	elk GraphColl Words Te 3 ▼ Context 7 ▼ Display T	You can: Search, sort and filter.
Index File 1 A_Press_rep 2 A_Press_rep	n run the obvious ris		Right k. One of his tasks will be erican administration. And, since this is ele alth charges, says George Brown" PROBE T	 Use 'smart' searches.
5 A_Press_rej 6 A_Press_rej 7 A_Press_rej 8 A_Press_rej 8 A_Press_rej	London, Three of them— Can the door open for r in view of fresh p village was still held by th Cabinet are basing th	nada, Australia, and New Zea modifications to the new Con problems facing the new and le pro-communists. NEW IMA(heir speeches on a new Con	aland — will have strong delegations at an nstitution provided law and order is maintain i Independent countries of Artica, that the VGE OF LIBERAL PARTY LEADER FOR CAN nservative booklet called The Record Speak Inswick, and in the four federal by elections.	✓
4 A_Press_rej 5 A_Press_rej 6 A_Press_rej 7 A_Press_rej 8 A_Press_rej	London. Three of them— Can the door open for r in view of fresh p village was still held by th Cabinet are basing th Conservative provincial governm Tokens	nada, Australia, and New Zea modifications to the new Cord roblems facing the new and le pro-communists. NEW IMA heir speeches on a new Cord ments. Quebec and New Brur Frequency	aland — will have strong delegations at an stitution provided law and order is maintain independent countries of Adrica, that the use OF ULEBERA, PARTY LEADER FOR CAN inservative booklet called The Record Speak inswick, and in the four federal by-elections Relative frequency per 10k	~
4 A_Press_rej 5 A_Press_rej 6 A_Press_rej 7 A_Press_rej 8 A_Press_rej 8 A_Press_rej File File	London, Three of them— Can the door open for i in view of fresh p village was still held by h Cohsine are basing t Conservative provincial governm Tokens 88805	nada, Australia, and New Zea modifications to the new Con roblems facing the new and le pro-communists. NEW IMA- heir speeches on a new Con ments. Quebec and New Brun VV Frequency 181	aland — will have strong delegations at an nstitution provided law and order is maintain independent countries of Africa, that the NGE OF LUBERAL PARTY LEADER FOR CAN- servative bookter called The Record Speak inswick, and in the four federal by-elections. Relative frequency per 10k 20.381737	
4 A_Press_rej 5 A_Press_rej 6 A_Press_rej 7 A_Press_res 8 A_Press_rej 9 A Press_rej File A_Press_report.bt 3_Press_edit.bt	London. Three of them— Can the door open for in view of fresh in view of fresh willage was still held by th Cabled are basing th conservative provincial povernm Tokens 88805 54367	nada, Australia, and New Zea modifications to the new Con roblems facing the new and le pro-communists. NEW IMA heir speeches on a new Con ments. Quebec and New Brur Frequency 181 110	aland — will have strong delegations at an stitution provided law and order is maintain independent countries of Africa, that the vice OF LIEBERAL PARTY LEADER FOR CAN neareative booktet called The Record Speak inswick, and in the four federal by-elections Relative frequency per 10k 20.381737 20.232662	Bottom panel: Displaying distribution
4 A_Press_re; 5 A_Press_re; 6 A_Press_re; 7 A_Press_re; 8 A_Press_re; 9 A Press_ress_re; 9 A Press_report.bt 9 Press_edit.bt 2_Press_review.bt	London. Three of them— Can the door open for in view of fresh village was still held by th Cabinet are basing uConservative provincial dovernor 88805 54367 34289	ada, Australia, and New Zea modifications to the new Con roblems facing the new and le pro-communists. NEW IMA heir specthes on a new Con ments. Quebec and New. Brun VVV Frequency 181 110 55	aland — will have strong delegations at an estitution provided law and order is maintain 5 independent countries of Africa, that the NGE OF LUBERAL PARTY LEADER FOR CASA inswirck, and in the four federal by-elections. Relative frequency per 10k 20.331737 20.232862 16.040129	
4 APress_re; 5 APress_re; 6 APress_re; 7 APress_re; 8 APress_re; 8 APress_respontat 9 Press_edittd 2 Press_review.td 2 Press_review.td	London. Three of them— Can London. Three of them— Can In view of fresh p In view of fresh p Cabinet are basing th Conservative provincial overant Tokens 88805 54367 34289 34287	ada, Australia, and New Zea modifications the new Con roblems facing the new and is pro-communists. NEW Mak- heir speeches on a new Con ments. Quebec and New Brur V Frequency 181 110 55 54	aland — will have strong delegations at an nstitution provided law and order is maintain in dependent countries of Africa, that the USE OF LUBERAL PARTY LEADER FOR CAN- servative bookter called The Record Speak- inswick, and in the four federal by-elections Relative frequency per 10k 20.381737 20.232862 16.040129 15.763202	
4Press_res 5Press_res 7Press_res 8Press_res 9A_Press_res 9A_Press_res 6A_Press_res 6A_Press_res 7A_Press_res 6A_Press_res 7A_P	London. Three of them— Can London. Three of them— Can In view of fresh p In view of fresh p Village was sith led by th Cabnetia the lab yth Cabnetia the lab yth Conservative provincial oovernr Tokens S88005 54367 34267 76613	ada, Australia, and New Zea modifications to the new Con- problems facing the new and e pro-communists. NEW IMA heir speeches on a new Con- ments. Quebec and New Brur VVV 181 181 110 55 54 115	aland — will have strong delegations at an estitution provided law and order is maintain 5 independent countries of Africa, that the NGE OF LUBERAL PARTY LEADER FOR CAN eservative bookfet called The Rever Of Speak inswirck, and in the four federal by-elections. Relative frequency per 10k 20 381737 20 232662 16 040129 15 763202 15 05068	Bottom panel: Displaying distribution
4 APress_rel 5 APress_rel 6 APress_rel 7 APress_rel 8 APress_rel 8 APress_rel 8 Press_editt CPress_relewitt D_Religion.tt E_Skills.tt Fop_lore.tt	London. Three of them— Can the door open for in view of fresh in view of fresh willage was still held by th Cabled are basing th Conservative provincial covern Tokens 88805 54367 34269 34267 766613 88742	ada, Australia, and New Zea modifications to the new Con roblems facing the new and p pro-communists. NEW IMAI heir speeches on a new Con ments. Quebec and New Brun Frequency 181 110 55 54 115 77	aland — will have strong delegations at an nstitution provided law and order is maintain independent countries of Africa, that the GE OF LUBERAL PARTY LEADER FOR CAN- servative booktet called The Revord Speak- inswick. and in the four federal by-elections. Relative frequency per 10k 20.381737 20.232862 16.040129 15.763202 15.010508 8.876839	
4 A_Press_rev 5 A_Press_rev 6 A_Press_rev 8 A_Press_rev 9 A_Press_rev File A_Press_rev 6 B_Press_ett 7 B_Press_ett 0 B_Press_ett 0 Revert 1 B_Press_ett 1 B_Press_ett 1 B_Press_review.td 1 B_Press_review.td	London, Three of them— Can donot pen for in view of forsh p village was still held by th Cabinet are basing th Conservative provincial dovemn Tokens 888005 54267 34289 34257 76613 88742 155271	ada, Australia, and New Zea modifications the new Con- zoblarms facing the new and e pro-communists. NEW IMA herrs speches on a new Con- ments. Juebec and New Brur VVV 181 181 155 55 54 115 177 171	aland — will have strong delegations at an independent countries of Africa, that the isde OF LIBERAL PARTY LEADER FOR CAN servative booklet called The Record Speak inswirck, and in the four federal by-elections Relative frequency per 10k 20.381737 20.222652 16.040129 15.753020 15.753020 15.753020 15.76339 11.013003	Bottom panel: Displaying distribution You can:
4 APress_rep 5 APress_rep 6 APress_rep 7 APress_rep 8 APress_res 8 APress_report 8 Press_redutt 9 APress_redutt 2 Press_retwewth D Religion bt 5 Kolls bt 5 Kolls bt 6 Gelie left, biogr.tbt 4 Misc_non_fict.tbt	London. Three of them— Can the door open for i in view of fresh in view of fresh in view of fresh in cabine are basing the cabine are basing the	ada, Australia, and New Zea modifications to the new Con roblems facing the new and perocommunists. NEW IMAI her speeches on a new Con ments. Quebec and New Brun Frequency 181 110 55 54 115 77 171 131	aland — will have strong delegations at an nstitution provided law and order is maintain independent countries of Africa, that the GG OF LUBERAL PARTY LEADER FOR CAN, servative bookist called The Record Speak inswick, and in the four federal by-elections. Relative frequency per 10k 20.381737 20.232862 16.040129 15.763202 15.763202 15.76339 11.013003 21.607534	Bottom panel: Displaying distribution
4 APress_rel 6 APress_rel 6 APress_rel 8 APress_ress_ 9 APress_ress_ 9 APress_ress_ 8 BPress_editd CPress_review.td 0 Religion.td 2 Skills.td 7 Jore tt 5 Belle_lett_biogr.tdt HMisc_non_fict.tdt 1,Acad_writing.td	London. Three of them— Can in the door open for in the doo	ada, Australia, and New Zea modifications the new Consolvemmunists. nebs/sectors nebs/sectors nebs/sectors NEW Makes nebs/sectors New Brur New Brur New Brur New Brur New Brur 110 55 54 115 77 171 131 136	aland — will have strong delegations at an stitution provided law and order is maintain in dependent countries of Africa, that the isore of LIBERAL PARTY LEADER FOR CAN newritive booklet called The Record Speak newritike booklet called The Record Speak newritike and in the four federal by-elections. Relative frequency per 10k 20 381737 20 222652 16 040129 15 763202 15 701500 8 67639 11 013003 21 607534 8 432069	 Bottom panel: Displaying distribution You can: View the distribution of the search term in
4Press_rel 5Press_rel 6Press_rel 8Press_rel 9Press_rel FID A_Press_rel FID A_Press_rel FID C_Press_revew.tt Revew.tt Revew.tt Revew.tt	London. Three of them— Can the door open for i in view of fresh in view of fresh in view of fresh in cabine are basing the cabine are basing the	ada, Australia, and New Zea modifications to the new Con roblems facing the new and perocommunists. NEW IMAI her speeches on a new Con ments. Quebec and New Brun Frequency 181 110 55 54 115 77 171 131	aland — will have strong delegations at an nstitution provided law and order is maintain independent countries of Africa, that the GG OF LUBERAL PARTY LEADER FOR CAN, servative bookist called The Record Speak inswick, and in the four federal by-elections. Relative frequency per 10k 20.381737 20.232862 16.040129 15.763202 15.763202 15.76339 11.013003 21.607534	Bottom panel: Displaying distribution You can:

5.1 Visual summary of Whelk tab

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.