

Readme file for AntConc 2.4.0 (Windows)

Developed by Laurence Anthony, Ph.D.

Dept. of Information and Computer Engineering
Faculty of Engineering
Okayama University of Science
1-1 Ridai-cho Okayama 700-0001
Japan
May 29th, 2003
anthony 'at' ice.ous.ac.jp

AntConc started out as a relatively simple concordance program, but has been slowly progressing to become a rather useful text analysis tool. It is written in Perl 5.6 using ActiveState's excellent Komodo 1.2 development environment. The program can be launched by simply double clicking on the .exe executable file, which can be downloaded from the Laurence Anthony laboratory website at <http://antpc1.ice.ous.ac.jp/>. The program can run under any windows environment including Win 98/Me/2000/NT and probably will run with no problems on Win XP. An earlier version of the software for Linux is also available. (If anyone wants the latest version of AntConc ported to Linux please let me know). If a user finds any problem launching the program under a particular OS, please let me know using the email address below.

Currently, AntConc can generate concordance lines either from a single text or a set of texts stored in a single directory. A text or set of texts to be analysed can be chosen by selecting 'Open File' from the 'File' menu and then selecting the appropriate texts in the standard way. Similarly, a directory of texts can be chosen by selecting 'Open Dir' from the 'File' menu. A list of the chosen texts is displayed in the left frame of the main window, below the name of the chosen directory (if appropriate). The output display can be saved either to the clipboard or a .txt file by selecting 'Save Output' from the 'File' menu. The color used to highlight the search term in the concordance lines can be changed from the default color 'blue' by selecting the 'Highlight Color' option in the 'Options' menu. AntConc also has a sort feature, whereby the results can be realligned in alphabetical order according to either the target word, or words to the left or right of the target word. The words on which to rearrange the concordance lines are chosen via the button bar at the bottom of the main window. The highlight color for the Level 1 and Level 2 sort can be changed via the settings menu.

To produce a set of concordance lines of text, a user needs to perform 5 actions:

- 1) Select either a single text or set of texts stored in a single directory (see above)
- 2) Enter a search term on which to build concordance lines in the entry box on the left of the button bar.
- 3) Choose the number of text characters to be outputted on either side of

the search term, using the increase and decrease buttons on the right of the button bar (default value is 30 characters)

4) Click on the 'Start' button in the main window to start the concordance lines results generation (Note: The concordance generation can be halted at any time by clicking on the 'Stop' button in the main window)

5) Select a target word on which to rearrange the concordance lines. 0 is search word, 1L, 2L.. are words to the left of the target word, 1R, 2R .. are words to the right of the target word. Note that two levels of sort are possible, with the second level not-activated when the software is first launched.

6) Click on the 'Sort' button to start the sorting process.

New since AntConc2.3.0 is the ability to display a plot of the location of the concordance results for each file as a kind of barcode. The plot is produced automatically and can be viewed via the "Concordance Search Term Plot" page from the main window.

New concordance search criteria can be entered on the Plot page, and then pressing start will allow the user to see dynamically in what files and in what positions the concordance results are generated. The plot can also be enlarged or reduced in size using the 'Zoom' buttons in the button bar. Unfortunately, the Plot display cannot be saved at present.

Also new since AntConc2.3.0 is the ability to generate Wordlists for the files selected. This feature is accessed via the "Wordlist" page from the main window. By selecting the page, a wordlist will be generated automatically according to the criteria displayed in the button bar at the bottom of the page. These criteria can be changed to produced wordlists ordered by frequency or alphabetically, and with the frequencies displayed or omitted, or presented in inverted order. At any time during the wordlist generation, the user can stop the process.

A number of additions have been added since AntConc2.4.0. First, I have added the ability to see how far along the processing of files is. This is useful when processing a large number of texts as it tells the user what the time requirement will be. If the processing of files is stopped for any reason, the progress bar will indicate this, and by clicking on the reprocess buttons, a user can reset the system. Otherwise, the AntConc will only work with the files that have been processed.

Second, a new feature has been added that enables a user to view a file in it's original form, with search termes (if any) highlighted. The location of search terms in a particular file can also be navigated using buttons at the bottom of the window.

Finally, I have had many requests for the ability to generate not only wordlists, but also Keyword lists for a particular set of files. This Keyword generation feature has now been added to AntConc2.4.0. To generate keywords, the user must first select a reference corpus of texts which the system compares with the target files. Usually, this reference corpus will contain standard English, but any corpus can be used. The system then compares the frequencies of words in the target files with those in the reference corpus and identifies usually frequent (or infrequent words) using a choice of two statistical methods. The most effective is proably Log-Likelihood (the default option), but the popular Chi-Squared method can also be used. The choice of a reference corpus, statistical method, and keyword threshold cut-off point etc, can be changed via the Keyword Preference menu.

Many small bugs have also been identified and corrected in AntConc2.4.0. Please let me know if you find a bug in the program and I will correct it in an updated version.

AntConc can be used freely for individual use for non-profit research purposes, and freely distributed on the condition that this readme file is attached in an unaltered state. If the software is planned to be used in a group environment, please let me know how you plan to use the software, and I will then give you permission. The software comes on an 'as is' basis, and the author will accept no liability for any damage that may result from using the software. Although there is no support available for AntConc, the author would be happy to hear comments and/or suggestions from anyone using the software, and will try to fix any bugs etc. when time is available.

Known Issues

1. When the file or directory selection dialog boxes are accessed, if they are dragged across the top of the AntConc main window, they leave 'ghost' traces. These traces are removed when the dialog boxes are closed. This appears to be a problem with the perl modules I use to access the Windows API. I have yet to find a way to solve this problem. Any help or suggestions would be grateful.
2. When a large number of concordance lines are generated (or words or keywords), the scrollbar becomes sensitive to where on the bar the user clicks and drags to view lower down entries. Sometimes this results in a user not being able to view the last lines unless the cursor is repositioned on the scrollbar. There is an annoying bug in the scrollbar subroutine (not mine!) and I am waiting for someone to fix this.

Revision History

2.4.0 A major upgrade since 2.3.0

First, progress indicators were added to 'pages' of AntConc.

Second, a new file view feature was added to view target files in their original state.

Third, a keyword generation feature has been added using log-likelihood and chi-squared methods.

Finally several bugs were found, in particular, bugs centered around the wordlist generation feature. This feature of the software should work much quicker now.

Also, the user can interrupt the processing of files in any 'page' of the software.

2.3.0 A major upgrade since 2.2.3

First, the ability to view concordance search results as a barcode plot graph and a feature to produce wordlist according to different criteria were added. Numerous bugs centered around the way the software entered a 'Busy' mode were corrected.

The main core of the software was also updated resulting in a quicker, 'cleaner' processing of the data. Performance improvements should be noticed as a result.

2.2.3 Updated file and directory selection dialog boxes to run smoothly in a Windows environment. Also, changed the default colors for sort highlighting, and search window frame size. A number of small bugs were also corrected

2.2.2 Corrected critical fault with compiler that caused program to expire when evaluation version of ActiveState Perl Development Kit expired. Sorry folks!! I didn't realize this would happen!!

2.2.1 Corrected bug which prevented new concordance lines being generated if the search term was left the same and then new files were selected. Port to Linux also completed.

2.2 Designed new subroutines for selecting directories and files to solve rendering of dialog windows problems. This also enables an easier port to Linux.

2.1

Added a second level of sort.

Added ability to restrict searches to full-words only, case sensitive

Added ability to search using full Perl implemented regular expressions

Added ability to save results either to a file or the clipboard

2.0 Added new sort feature, for rearranging concordance lines.

Tidied up the interface.

Made the system more robust for novice users. (Now bad input will not cause the system to crash so easily).

1.1 Added binding to allow return key to launch concordance search.

Also, recompiled software so that no console is required.

1.0 First version

Laurence Anthony

anthony 'at' ice.ous.ac.jp