

Elite Concordance and Index Creator

User Manual

©2010 Elite Minds Inc, All Rights Reserved

What This Program Does

This program will analyze your PDF file and create a Concordance(a list of all words with the pages where they appear) which you can then convert to an Index(a list of relevant words with the pages where they appear). It is used to create an Index or Concordance for books and manuals.

Shareware Version

If you downloaded the shareware version, you are free to use it for commercial or non commercial purposes. You may redistribute the shareware version only. The shareware version is limited to processing the first 10 pages of a document.

Full Version

If you purchased the full version, you may NOT redistribute it, you may NOT share it with friends or family. The full version is commercial software and is copyrighted. See the license included with the commercial version for more details on how it may be used.

If you purchased the full version, uninstall or delete the Shareware version, then install or unzip the Full version in it's place.

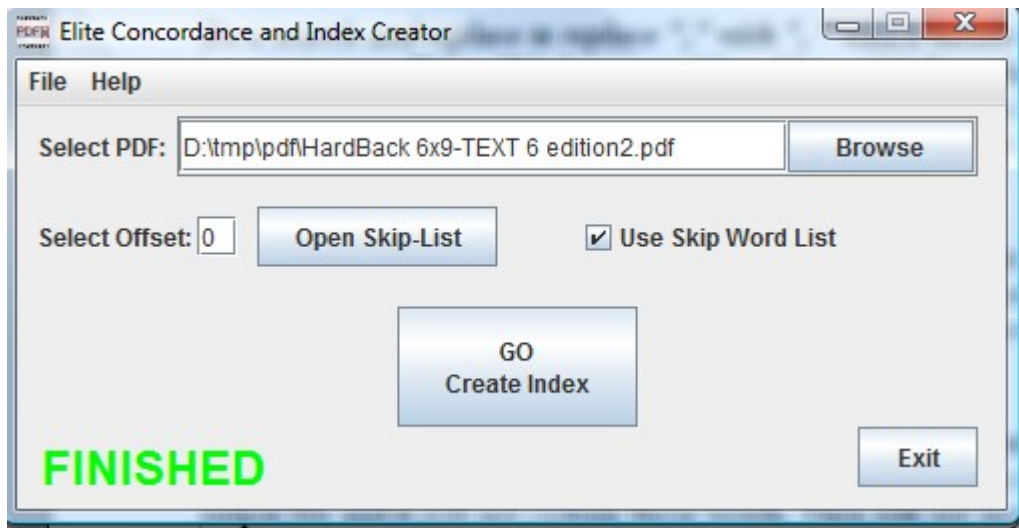
Compatibility

This program is written in Java. It will run on Windows Xp/Vista/7 or later, MacOSX, Linux and possibly other operating systems as long as you have Java installed correctly. If your computer does not have the latest Java version, download it for free from Java.com

Quick Start

1. Unzip the program to a folder on your computer.
2. Double click the EliteIndexer.jar file. It should run like any exe. If it does not run, see below.
3. Select Browse and locate the PDF file of your book.
4. Enter an offset value which is the first page where your document starts page numbers.
5. Click Go

The program will show PROCESSING while it is working and FINISHED when it finishes. The output file will be in the same folder as the PDF you selected.



NOTE: The program can take several minutes to run depending on how fast your computer is and how big the document being indexed is.

Details

The Elite Concordance and Index Creator software will analyze a pdf document and create a Concordance. A Concordance lists every word and the page number it is on. This differs from an Index because an Index only lists selected words. If you need an index for your book, you will need to edit the final output file and remove unwanted words. The program also uses a skip-list which contains common words that will not be included in the final output file.

Using the ECIC program is much easier than manually creating an index in Indesign or Word or any other program.

If you are the author of the document, it is assumed you can produce a PDF of your document which is required by this program. In order to create a PDF in Indesign, select File/Export. Microsoft Word does not usually support direct export to PDF, however OpenOffice does and OpenOffice is a clone of Word which is free. If you are using MS Word, download OpenOffice from www.openoffice.org and install. Then load your .doc or .docx file in Open Office and select File/Export As PDF to save a pdf of your book. You can then use the PDF to create an index. If you are using another program, see the documentation for that program to determine how to save your book as a PDF file.

You can use the skip-list.txt file to define what words you do not want included in the index. This file already includes common words like "a", "an", "the", "about", etc. You can edit this file or you can uncheck the option to use the skip file and all words in your document will be included. You can also create your own skip file and use the File/Select Skip File option to select a different file for each document you want to create a Concordance or Index for.

It is assumed that the user of the program is the author of the PDF file. The ECIC program is intended for authors who wish to create a concordance for books and manuals.

Offset

The Offset value is the first page where your page numbers start. For example, let's say you have a

100 page book but the first 10 pages are copyright pages and an unnumbered introduction and on the 10th page is where you see the first page number of 1 at the bottom. You would want to enter 10 for the offset. This will start calculating page 10 as page number 1. Assuming you have "Chapter 1" on the page numbered 1, in your index it will say "chapter: 1,..." indicating the first occurrence of the word CHAPTER is on page 1. If you left the offset to zero, it would say the page number was 10. If you select the wrong offset value, your page numbers will be incorrect relative to the numbers printed in your book. Check your page numbers after generating an index file. If they are off, adjust your offset value and re-run ECIC.

NOTE: The Open-Skip-List button and Open-Last-Index-File options may not work on some computers. If you click these options and nothing happens, you will need to manually open your skip-list.txt file by double clicking on it and manually open the index file by double clicking on it.

FAQ: Why does your program only analyze PDF documents? Umm, because it would not make any sense to analyze other types of documents. If you are publishing a book or manual then you will have to create a final format PDF where your text falls on specific pages. It would not make sense to analyze a .doc or .txt file to generate an index because those page numbers can change when the book is type-set(or layed out) for printing or in the case of a txt file, there are no page numbers. Almost all publishers now require a PDF as the final document so a PDF makes sense to use for creation of the index.

Output

The output file is created in the same folder as the pdf file with the same file name except ending in txt. Make sure you do not have another file of the same name ending in .txt in the folder with your pdf.

Txt is the best output format because you can easily copy and past or import it into your editing program.

Workflow

The normal workflow would be to write your book, export your book to a PDF, then create an index using the ECIC program, edit the text file created to remove unnecessary words and consolidate listings, then import that index into your book authoring software(Indesign, Word, etc).

If you are unable to run the .jar file then your computer may not be configured correctly or it may be missing Java. Java is a free program and should have been included on your computer when you bought it. If it was not, then your computer was not configured correctly. You can still download Java for free from Java.com

If you have problems, download Java, install it, and then try running the ECIC program again.

If you have any problems or see any error messages, try updating to the latest Java version. It may resolve those problems.

You may want to create a special version of your book with no page numbers and if you have any

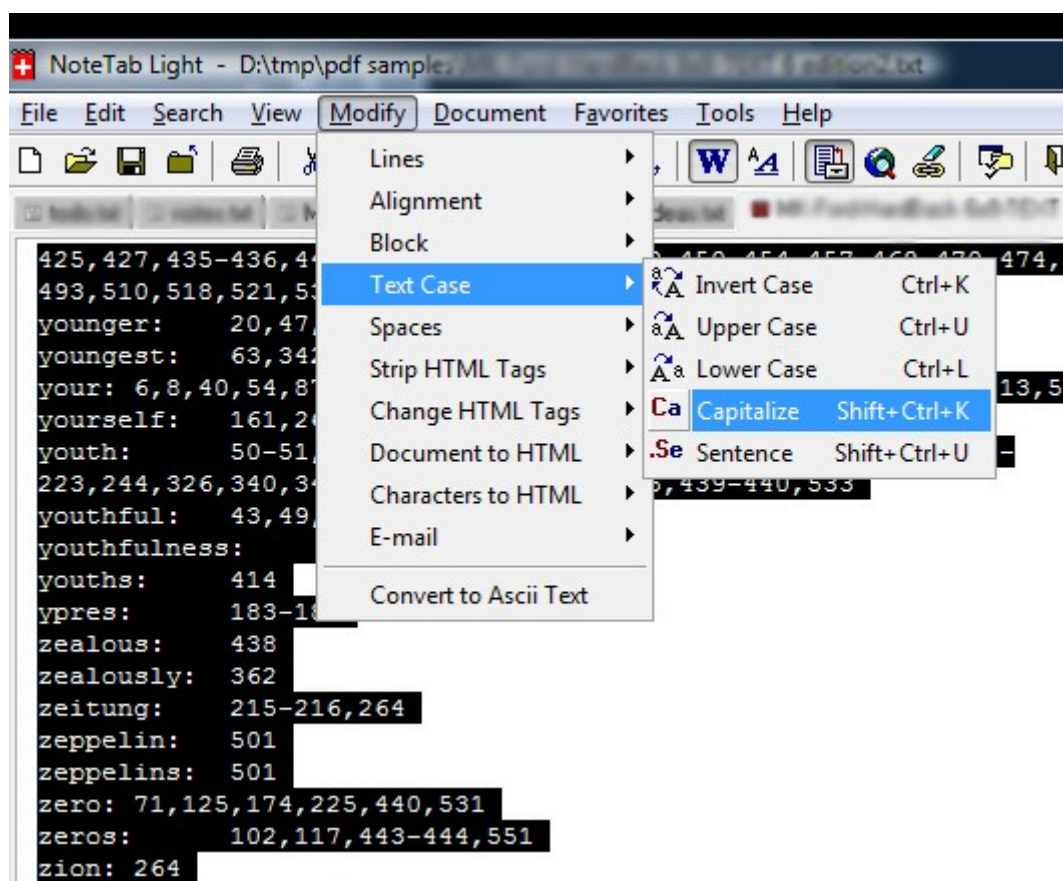
sections with lots of words such as an engineering reference which does not need to be indexed, remove those sections or make those pages blank. This will save you editing time for the final index file because it will list any page numbers and all information from all pages in your pdf document.

Editing

You will need to edit the output file. The output file is a Concordance which contains EVERY word in your book that was not listed in the skip-list.txt file.

For editing I recommend Note Tab Light which has a free version and is available at <http://www.notetab.com/> You can, of course, use any text program to edit the txt file.

The concordance converts all words to lower case. I recommend making all entries capital using the capitalize option in Note Tab Light(from the Note Tab Light menu select Modify/Text Case/Capitalize). By making all entries capital letters you do not have to worry about proper nouns being uncapitalized.



Use the Capitalize option after selecting all text(Ctrl-A) to capitalize all entries.

Do a search and replace to replace ", " with ", " which means replace commas with comma+space. This will help you avoid problems with lines that are too long to be displayed. The shorter form was meant to save space and it can if you have a short index. If you have a long index or references to many pages, it can cause problems. It is recommended to always do this replacement.

You may need to consolidate some entries. For example, if your book references John Smith, there will be index listings for "john" and for "smith". You would take both of these and combine them by creating a new entry for "John Smith". This is not necessary, however, and you can let the reader simply look up the person by his other last name.

When you go through your Concordance text output file, instead of just deleting the unwanted entries, you should add those words to the skip-list.txt file. You can then create a new concordance output file and it will not contain those words. Then edit out any excess numbers or remaining words to produce your final Index text file.

The Ctrl+Y key combination is very helpful because it allows you to delete a full line in Note Tab Light.

Editing Tip: look for common words that no one will look up. Remove those. This will make your index shorter which means fewer pages in your book. If you need to pad out your book, you can leave these in to create more pages.

Example:

```
abolishing:      233
abolition:       99,193
aborted:    135
about:      6-10,33,35,39-45,49,53,56,58,61,63-68,70,72-74,78-
79,81-84,87,89-91,95,102,104-106,112,114,116,120-
122,124,128,131,133,136,144,148,155,158-159,161,165-
169,171,177,183,185-186,188-189,193,195-197,199,204,206,209-
215,218,220,222-223,227-228,230,232-
233,235,239,241,246,250,253,256-257,261,263,268-
269,271,289,292,294,297,299,301-302,304-
312,314,316,319,324,328-329,332-333,336-
337,340,343,345,347,351-352,359,362,378,380,382-386,388,390-
392,394-398,400,403,407,409,411,413-416,420-
421,436,440,443,445-446,449-450,453,457,459,462-465,469-
472,474,477,480-481,483,492,495-498,500,503-507,513-
514,516,519,523-524,526,528-529,531,533,539-541,546-
548,555,557-558
above:      42,48,53,61,75,96,105,108-109,111-112,119,121,200-
201,205,210,214,219,222,231,241-
244,253,275,277,282,290,293,296,301,305,311,319,334,336,340,34
2-343,346,354,358-
359,369,373,378,422,445,462,468,484,486,490,499,514,518,534,54
8
abroad:     141,232,466-467
```

abrupt: 313
abruptly: 281,403

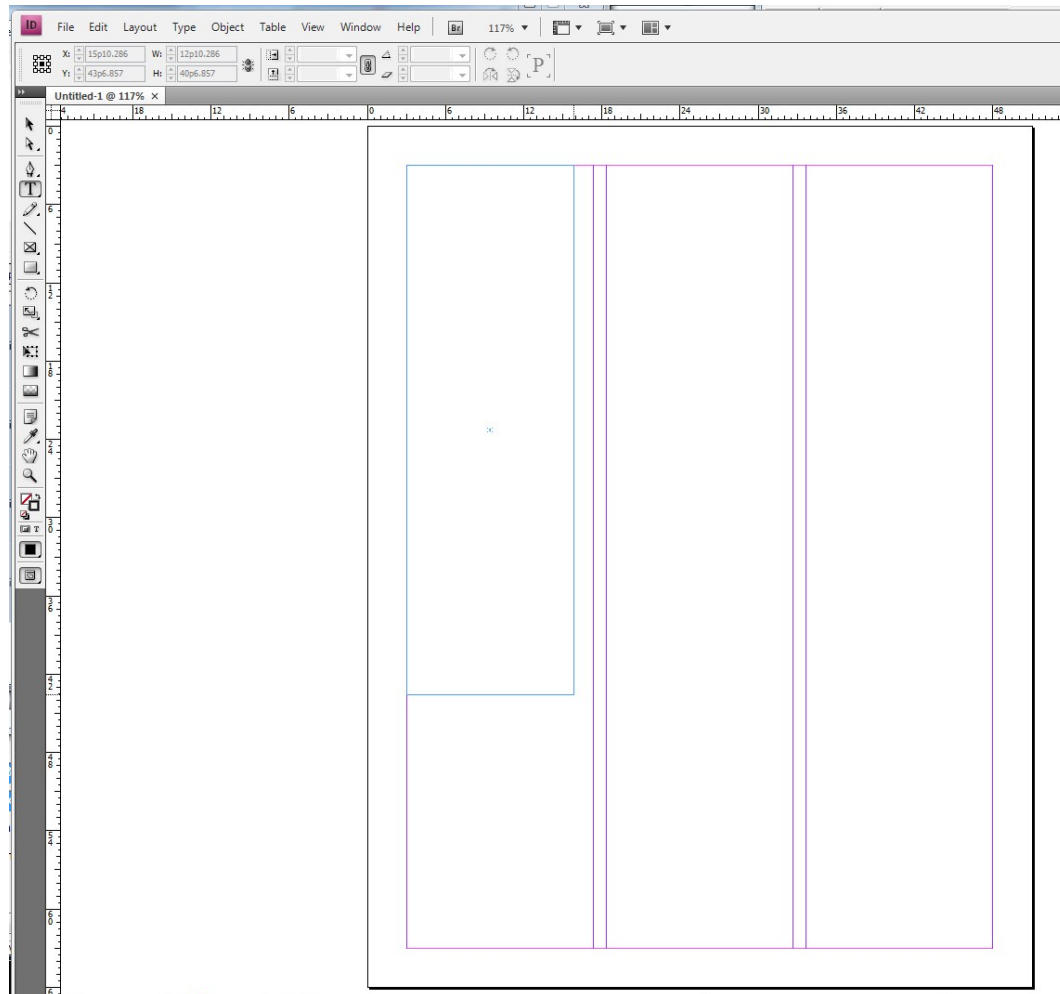
Here you can see the words ABOUT and ABOVE appear on large numbers of pages. It is unlikely the reader will care about these, however the end user may want to find other words such as abruptly. The reader may recall a sentence with that word and wish to find it. You can remove those other common words or leave them, it is a matter of author preference.

You should also look through the skip-list.txt words to see if you do want to include any of them. For example, the word absolute is in the skip list, but if your book includes a product name that includes absolute, you would want to include that word in the index but you still may want to skip absolutely.

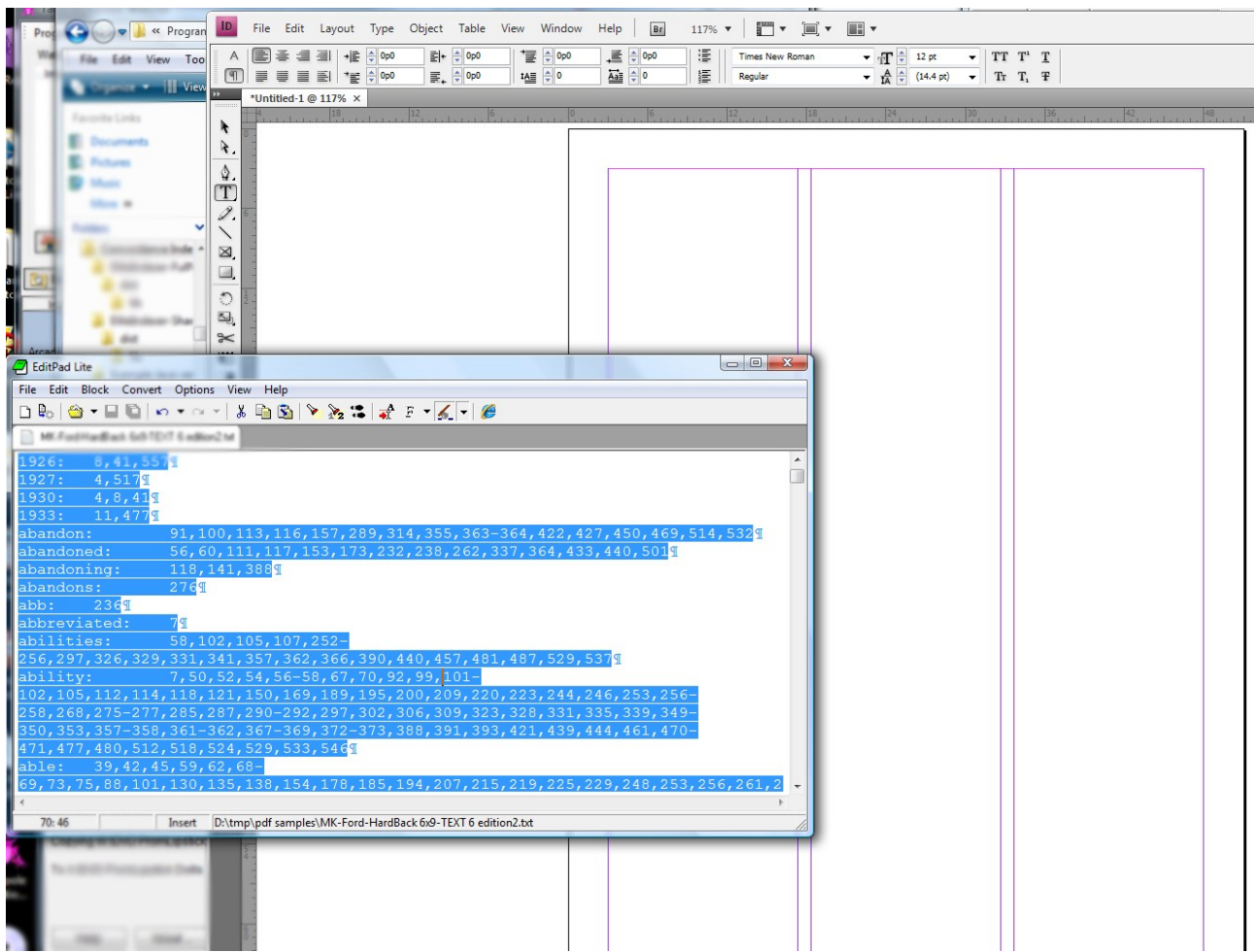
The skip list for each book may be different. You may want a true Concordance which lists all words for some books, for some you may not include common words like "a", "an", "the, and for other books you will want to include words like "anything" but not every book. The ECIC program allows you to customize your skip-list and create different skip lists for different documents.

Adobe Indesign(R)

To import the index into Indesign, I recommend creating a new document, then setting up 2 to 4 columns in your pages. Next, create an empty text box.



Create an empty text box in the first column.



Copy and paste your text into the text box you created in Indesign.

It will autoflow past the end of the box causing a red+ at the bottom. Click the red + at the bottom of the box once, then move your cursor to the top of the next column. Hold the SHIFT key down. This will change the cursor to show an S shaped line which indicates your paste will flow into as many columns and pages as needed. While holding SHIFT, click in the top left corner of the next column. This will paste the text and flow into as many pages as needed.

Another method is to use the Place command, select File/Place and select your text file, then place in the top left corner of your first column.

You can then either append the newly created file to your book, or if your publisher wants a PDF file, you can create a separate PDF and use Adobe's Professional version of Acrobat to combine the files into one PDF. Then you have your book with the newly added index as a final PDF. You can easily update your book, create a new index(since you added all the unwanted words to your skip file) and update the index file separately, then re-combine them for book updates.

If your document appears to be incomplete after pasting or Placing, your lines are too long. Simply do a search and replace to replace ", " with ", " which means replace commas with comma+space. If you do not like the colons after the words, you can also search and replace those with any character or a space. Be sure to look through your index to make sure it is complete. I recommend doing this search and replace operation for commas always.

Distribution

If you are using a print on demand service and want to reduce the number of pages in your book to reduce the cost, you can include a URL link in the back of your book and then provide the index as a PDF download on your website. This is a good way to get people to your website too.

Conclusion

ECIC is designed to be a quick and efficient way to generate a Concordance or an Index for your book. It is by far the easiest method available. Even if you have to edit a long list of words out, it is much easier than going through your book, page by page, and trying to create an index manually.

I hope this program helps you in your writing.

You can find more useful products at

<http://www.elitemindsinc.com/shop>