# Introduction

Adjust this checklist to your needs and make a copy for each conversion if it helps.

Labeling of headings and locators enables the concatenate and export processes to place the text in the correct fields for your indexing program or database. Headings are labeled $H1\_, $H2\_ etc and are placed immediately before the entry, after any indents, tabs or spaces that may be used to define the index structure.

First locators are labeled with a tab followed by $L\_. Cross references are considered locators.

IndexConvert can convert any index to a format suitable for your indexing program and can also convert structured text records to a flat database.

IndexConvert can be extremely fast with a well structured index. If it’s not going as fast as you would like, look for features in the index that can be used with Word Find/Replace to insert labels.

If it’s important to know how long a conversion will take because you’re converting somebody else’s index and need to quote then take a single page of the received index and run through the entire process. Mistakes, tips and tricks learned here can speed full processing.

# Export from PDF

If your index is received as a PDF then it needs to be converted to a Word file. A number of conversion programs and services are available many don’t generate a good facsimile. A good facsimile copy will avoid problems later on.

## Acrobat Pro DC

Set the options ‘Retain flowing text’ and ‘Recognize text if needed’. A docx file is created.

Acrobat Pro DC creates a good facsimile copy. It does not place CRs at ends of lines where they are close to a column boundary and it places unwanted CRs at turnovers. Insert/delete CRs before starting IndexConvert. Make any other changes that may be required.

Indents vary subtly between entries making heading labeling tedious. Use Method C if Method A hasn’t labeled all headings after about 5 tries at each appropriate heading level.

# Export from another indexing program

If possible, get the indexing program to do the labelling. Choose a set-out style. Label level 1 headings with $H1\_, with no following space. Use $H2\_, $H3\_, etc. for lower level headings. Label the first locator in each locator string with {tab}$L\_, where {tab} is the tab character.

Export to RTF/Word as a single column with no headers, footers or continues.

This will save a great deal pf time later but you still need to run through the complete IndexConvert process.

# Word files

## Copy/Paste

Copying and pasting the index to another Word file is not recommended as default settings are likely to disrupt the structure.

## Index Filename

Save as a docm macro enabled file. This enables advanced features required by IndexConvert.

# Quality

IndexConvert contains many features designed to give a very high quality file for import. A remaining weakness is errors made following Audit, which cannot be picked up by another Audit. One outcome is an extra tab in an entry which can be disruptive and difficult to locate. The recommendation is an initial export to Database format. This file can be imported into a spreadsheet such as Excel. An extra column will indicate a problem, which can be easily found and corrected in the Audited file, Select the desired export and continue the process.

# User Interface

A screenshot of a computer

Description automatically generated

# Configuration

## Licence Key

Enter licence key at first use or on receipt of a new licence key.

## Select Export

Details of export formats are listed in *Table 1*.

Database is recommended as an initial export, see Quality above.

Tab delimited exports are available for Cindex, Macrex, Sky Index.

For Macrex you will need to reference the Unicode.ini file.

For Sky Index v7 you may need to reference a copy of your Translation file.

An IXML export is the preferred route for Cindex v3 and v4.

Cindex IXML Expanded places a copy of the locator in a special lowest level subheading.

An IDEX export is the preferred route for Sky Index v8, some early versions will accept an expanded IXML file..

There is an export for Index-Manager

IndexConvert will embed an index into a document. Page numbers in the index must correspond to page numbers in the target document.

Table 1 Export options

| Export | Format | Additional File Required | Note |
| --- | --- | --- | --- |
| Cindex v3+: IXML | UTF-8 (Unicode) file using IXML (XML records). Up to 32 fonts. Styles encoded. |  | Compatible with Cindex 3 and 4.. |
| Cindex v3+: IXML Excpanded | .IXML Unicode file. Up to 32 fonts. Styles encoded. One locator per entry. Locator copied to special subheading. |  |  |
| Cindex V3+: Tab delimited | .txt file. UTF-8 (Unicode) Tab delimited with labels for headings, locators and styles |  |  |
| Database, Unicode | .txt file. UTF-8 (Unicode) Tab delimited with labels for headings, locators and styles. One locator per entry. |  | Use as initial export and import into a spreadsheet. Check for added tabs. |
| Embedded index | Embeds index to Word file using page numbers | Target word file for embedding. | Pagination of the Word file must match the index when XE fields are hidden. |
| IDEX | Idex file for Sky SI8. IDEX. UTF-8 (Unicode) encoding, up to 32 fonts, styles encoded. |  |  |
| Index-Manager | UTF-8 (Unicode) txt file. |  |  |
| Macrex: MBK: ANSI | .mbk file ANSI. | Macrex Unicode file | Keyboard characters from Unicode file for non-ANSI characters |
| SKY IDEX | IDEX file UTF-8 (Unicode). Up to 32 fonts, styles encoded. |  |  |
| Sky SI7: Tab delim. | .txt file. Tab delimited with labels for headings, locators and styles. One locator per entry. | Sky Translation file | Keyboard characters from Translationnfile for non-ANSI characters. |
| Sky SI8 IXML | IXML file UTF-8 (Unicode) |  | Compatible with Sky versions that support IXML. |
| Spreadsheet, Unicode | .txt file. UTF-8 (Unicode) Tab delimited with labels for headings, locators and styles |  |  |
| Word Automark | .docx file containing an automark table. Adjust as required and save as .doc file. | Word Automark file |  |
| Word embed using page nos. | Creates embedded entries using heading and page information | Target file to include embedded index entries | Pagination of index must match the target file for XE fields hidden. |

# Preparation

| **Instruction** | **Notes** |
| --- | --- |
| Is it a multi-column index?  Check for missing and unwanted CRs at line ends. |  |
| Remove complex group headings, continuation statements and embedded page numbers (i.e. from PDF). |  |
| Is index a legal table containing numbers within headings which could be locators? Manual insertion of locator labels {tab}$L\_ may be required. | If in doubt try things out using a subset of the index. |
| Is index a simple names index? Label headings may not be required if a simple tab delimited export is required. | IXML and IDEX export will not work unless headings are labeled. |
| Define maximum page numbers using Label Locators [ i ]. Set value too high rather than too low. | This is very useful if numbers such as years or specification numbers are included in headings and if locators include chapter numbers.  Don’t set too low. This will cause error symbols to be inserted where a locator is out of range. |
| Are Roman numeral locators present. If so check this selection using Label Locators [ i ] and enter maximum value (Roman or decimal, up to 50) | Roman numeral locators are identified in the second Locators pass.  Minimise the number of Roman locators chosen to control the size of the locator matching library. Copy unwanted terms to the scratchpad. |
| Check cross reference terms. Update using Label Locators [ i ]. | There are no default cross reference terms. Ensure that see and see also or the language equivalents are present.  Minimise the number of cross reference terms chosen to control the size of the locator matching library. Copy unwanted terms to the scratchpad. |
| Check locator prefixes. Enter any using Label Locators [ i ]. | Examples of locators with prefixes are I.20, II.50, A-1, B::-7  Minimise the number of locator prefixes chosen to control the size of the locator matching library. Copy unwanted terms to the scratchpad.  If more than 20 locator prefixes are used then re-run the locator search with a second set of prefixes. |
| Consider how to deal with any locator separators that are not commas.  This affects index expansion. | Cross references separated by semicolons may need duplicated entries. |
| Check for any special characteristics. |  |

# Preprocess

The table below gives some guidance on the dialogs that are presented by IndexConvert when Preprocess is run. Most of the options will only appear if required.

| **Dialog** | **Notes** |
| --- | --- |
| Indented structure | A heading contains a comma or other punctuation, possibly indicating a subheading. This is a reminder to check that the index structure is set-out.  See Label Headings section if you are converting a run-on index. |
| EOF Marker | idxEOF is inserted at the end of the file to support processing. |
| Navigation and preprocessing | A reminder of the navigation options. |
| Convert multi-column index to a single column. | Have you checked the index for missing CRs?  Accept |
| Remove header and footer | Accept (This question is always asked) |
| Form feed | Accept |
| Column break | Accept |
| Section break | Accept |
| Double paragraph marks | Accept |
| Locator string broken by paragraph mark | Accept |
| Index entries run together | Look closely before accepting. A $ErP01\_ label will be inserted at each instance. Optionally alter to $H1\_ or $H2\_ etc. during review and look for other entries still run together. |
| En dash in locator | Indexing programs expect a hyphen. Accept. |

Rerun Preprocess after correcting the errors. Some may have been missed.

# Locators L [ i ]

## Maximum page number

This ensures numbers exceeding this are not considered as page numbers.

If locators are like 30:1-7 including a chapter prefix then set maximum page number to the number of chapters.

## Select Roman numerals

If Roman numerals are to be detected then check the Roman numerals box.

Default is lower case but upper case can be selected. A maximum value can be set as either a decimal or Roman numeral.

## Automatic search (Not recommended)

When selected ‘yes’ is answered to every ‘Is this a…’ question.

## Define locator cross references

There are no default values. Ensure that your see and see also terms are defined. Case sensitive so it’s good to have see, see also, See, See also. See specific advice for Sky Index SI8.

## Define additional locator prefixes.

Non-numeric prefixes. Punctuation and spaces should be included. Search is case sensitive.

p. , f. ,Chapter , etc.

# Locators L

IndexConvert looks for the end of headings, punctuation and the start of locator strings and highlights a candidate. For heading examples include us www. us http: ng, 3 ng, www. ng, http: ng; 5 ng; www. ng; http: ng: 1 ng: www. ng: http: g) 8

Take great care answering ‘Is this a first locator’. Use the Not Sure or No option if uncertain.

$Er labels are added for ambiguous locators. Adjust before Label Headings.

The first pass does not identify Roman numeral locators. It does identify run-on entries and entries run together during PDF export.

# Level 1 [i]

Select [ i ] for Label Headings options.

## Options

The Options tab allows the labeling method to be selected.

## Method A

Set Multiple choice. This defines the number of searches to be carried out before labeling begins. The recommended value is 3.

Show search detail is intended for software diagnostics.

## Method B

Method B is a legacy process and is not recommended.

## Method C

There are no options for Method C. Recommended if the index has been exported from PDF. The user assigns a heading level to each entry.

# Label Headings

If the index has been exported by an indexing program and received as a RTF file then labeling of heading should run smoothly.

A manually created or edited Word file should be fairly straightforward.

If the starting point was a PDF file then it may take a lot longer.

For run-on subheadings, first label all level 1 headings. Identify the level 2 heading separator. Using Word find/replace, replace each to give an indented new line including a desired number of spaces or tabs and a $H2\_ before the entry. This should convert the run-on index to set-out and label the entries. If necessary adjust the process and repeat for lower level headings.

## Method A (Standard)

This option uses many parameters including indentation, tabs, spaces, character size and others to identify heading levels.

If possible select the entire document and set character size to a single value. This may reduce the number of iterations required.

An [ i ] option allows multiple selections, up to 50, before processing takes place. 3 is recommended.

## Method B (Indents Only)

This uses less information than Method A which can save time sometimes. It may not work so well at other times. A mix of Method A and Method C is normally the most reliable.

## Method C (Interactive)

Method C is interactive. It is very useful as an alternative to Method A and for labeling the last few remaining entries. It is recommended for scanned PDFs where it is necessary to review every entry for smudges and wrongly identified characters.

# Locators M

Selection of Roman numerals and a large number of locator cross references and locator prefixes will slow locator search. Use only the options necessary for the current index.

Take great care answering ‘Is this a first locator’. Use the Not Sure option if uncertain or Quit and start again..

$Er labels are added for ambiguous locators. Adjust before Audit.

If Roman numerals and prefixes are not defined, likely cases will be labeled as errors during Locator Labeling and Audit. This may save time compared to setting an excessive Roman numeral range.

# Audit

Audit inserts $Er labels where possible errors exist. Adjust and remove error labels before moving on. Headings are audited first. Locators will be audited when heading error labels have been addressed. Commas in locator strings are set to normal text to prevent unpaired style labels in locators.

Audit is now interactive, if U (Up) or D (Down) is used be sure to select R (Return) to label the right entry.

Audit expects every heading to include a locator. For headings correctly containing no locator choose ignore to prevent labeling.

## Audit Report

The Audit Report is used for review and for user support. It collects together all entries containing $Er labels into a separate document.

Special audit reports are generated during Concatenate if the Macrex Unicode.ini file or the Sky Index translation file needs updating.

# Save File

Saving the file at this point provides the opportunity for recovery or multiple concatenations/exports. Regular saves are recommended.

# Select Export

If you’ve selected Database, Unicode as the export to enable format checking in a spreadsheet then select the final export now.

# Elided Locators

**Run Elided here**

Elided 123-4 expands ‘pure’ numeric locator ranges to their un-elided form.

# Concatenate

Not applicable for IXML, IDEX, Index-Manager

If Audit required additional Unicode characters for Sky Index or Macrex then make sure that the file is correctly referenced in the Configuration dialog before Concatenate.

For Sky Index v7 export the Translation file following any additions.

# Label Styles

Not applicable for IXML, IDEX, Index-Manager

Automatically labels bold, italic, underline, small-caps, subscript and superscript. The labels are used to recover the styles in the target program.

# Export

Tab delimited files are exported using a save command. IXML files are saved using an IXML generator, IDEX files are exported using an IDEX generator. Index-Manager files are exported using a structure compatible with Index-Manager.

# Import to Indexing Program

See the instructions for your indexing program.