EPUB Indexing Scripts, InDesign CS5 (and later)

ePub files and scripts

In versions of InDesign prior to CC, **all embedded index entries are lost in the ePub export.**

To get around the missing entries in ePub problem, Ole Kvern and Jan Wright have developed a set of scripts and a GREP routine to let indexers index outside the files for InDesign books that are going to print and to ePub.

These scripts should only be used when the client needs the following things:

* The client needs to use a conversion service, like eBookArchitects.com, to process the index so that it will work with the ebook.

* The client needs a print index for the same book: the scripts are designed to show the page number the hyperlink is on at the time of indexing. If the publisher wants an epub index that matches the page numbers in the print index, they must freeze their pages and not make any changes to the files, or the index could be off in the print edition. This process is much like standalone indexing in the way files must be frozen.

* If the client needs a PDF index with hyperlinks: the scripts method will only work if you use Sonar Activate on the print index to activate it.

Epub export of a book and index.

These scripts will allow you to get a set of InDesign files, put permanent hyperlinks into each paragraph of the InDesign files, and add a Note at each hyperlink that you can view and copy in Story Editor. You need InDesign during the indexing process to view the marker text, but you don't add index entries to the InDesign files. You index the files in your own indexing software. The locators will contain both page numbers and hyperlink coding.At the end, when you have edited the index, you can place a print index into an InDesign file, and then run the GREP routine to remove all the hyperlinks and produce the print index. Then take your index file, and transform it into .txt. file according to instructions later in this documenr. You then check the links for mistakes using the VerifyIndexLinks.jsx script. It will ask for the txt file, check your links against the book's files, and warn you of any errors so that you can see how to fix your index links.
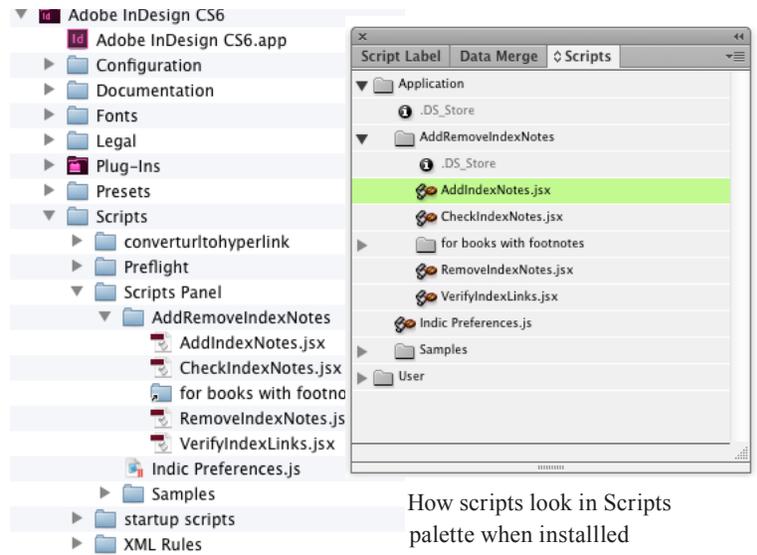
There are no guarantees that the scripts will work with every set of

InDesign files.

Script setup

First, get the scripts from http://www.wrightinformation.com/Indesign%20scripts/Indesignscripts.html.

Close and exit InDesign.

Install the scripts in InDesign: Inside your Mac InDesign application folder you will find a Scripts folder. Put them in the Scripts Panel folder inside that.

How scripts look in Scripts palette when installled

Where the script files live in the directory structure

In Windows, scripts live in Documents and Settings/Owner/Application Data/Adobe/InDesign/Version xxxx/en_GB/Scripts/Scripts panel.

Start up InDesign. You can see if scripts are installed by choosing Window>Utilities>Scripts and open the user scripts folder. There are four scripts, and a subset in a special folder to use if you have a book with footnotes. Use the main scripts for most books, unless you need to index footnotes.
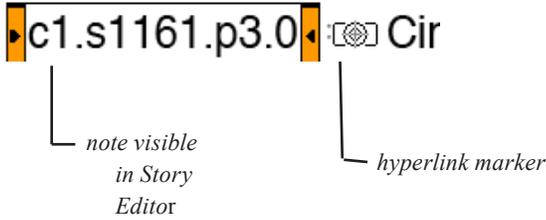
To add the notes and markers to the book:

1. Open InDesign and open the project's book file. You do not need hyperlinks in every file; determine which files need them. For example, the Table of Contents file will not need index anchor codes. Open the first file that needs markers inserted.

2. Choose Window>Utilities>Scripts to open the Scripts palette. Choose AddIndexNotes, and choose Run Script from the flyout menu.

3. It will take time to process. When it is done, you should see sideways orange bowties in the text. These are the notes contaning the inserted anchor codes. You can actually see the marker text if you hover over a marker. If there is a long paragraph, you will get a second marker mid-paragraph to help with indexing closer to the subject.

While never as famous as her husband, Augusta Tabor played a major role–albeit unwittingly–in the history of Oro City and Leadville

[Far right] Denver Public Library Western History Collection X-21986

4. Select the text block, and view it with Edit>Edit in Story Editor. You will see the small hyperlink markers and the note text.

c1.s1161.p3.0 🔘 Cir

note visible in Story Editor

hyperlink marker

5. The note disappears during epub export, but the hyperlink remains. C1 is the note above is the first file in the book, s1161 is the story number, p3 is the page number it was on when the script was run, and 0 is the unique paragraph number.

6. If you have multiple stories (captions, sidebars, etc.) their story numbers will be different. Be sure to open each text block in Story Editor to check the notes. The numbers will be different.

To check notes & markers for page no. accuracy:

If the pages have shifted, or if there is weirdness happening in the pagination, you can check it with the CheckIndexNotes. jsx script. This will tell you when the page number in the note does not match the page it is sitting on. It is good to run this before you start indexing. Try removing (below)and running the AddInDexNotes again.

To remove the notes and markers;

If everything is messed up, and you want to start over, you can remove all the notes and markers by running RemoveIndexNotes. jsx.

To create index entries:

If all is well, you can return control of the InDesign files to the publisher at this point, being sure to retain a copy for yourself. You will need to see the note numbers. But caution your client about any changes that would remove hyperlinks, notes, or cause the text to flow to another page. They should freeze the files, or the pagination will be wrong.

To start indexing, use your usual indexing software, and create entries like you would normally. For the locators, you must create hyperlinks. If you want the page number to display (for a print version) you must include that as well. Here is the anatomy of the hyperlink coding to use as your locator.

This piece makes the "34" show up as a page number

\<a href="c10.s68011.p34.49"\>34\</a\>

note this piece matches the note in InDesign.

Indexing programs all have their own escape characters. This markup is for Cindex. The \ marks are used to keep the codes showing in Cindex. Use whatever your program needs

To verify you have entered hyperlinks correctly:

In order to check your hyperlinks, you must export your index to a tab delimited text file. Export to txt with your locators in one field (See instructions for programs on the following page).

These are the columns you need: All locators must be in one column.

main	subhead	subsubhead	blank tab	locator

In Excel, add a number column to the front. Use Fill>Series to number all of your entries.

1	main	subhead	subsubhead	blank tab	locator

Use Excel's Sort functionality to sort on the locator column. This will get all your cross and generic refs together. Delete those entries from the spreadsheet.

Delete the text in the main head, subhead, and subsubhead columns. We just want the Excel numbers we added and the locators. That's what the script checks. Save the file as text (.txt), and close it.

```
955 →    →    →    →    →    " \<a href=""c21-s2168-p155-55""\>155\</a\>" ¶
956 →    →    →    →    →    " \<a href=""c21-s2168-p155-55""\>155\</a\>" ¶
958 →    →    →    →    →    " \<a href=""c21-s2168-p155-55""\>155\</a\>" ¶
960 →    →    →    →    →    " \<a href=""c21-s2168-p155-55""\>155\</a\>" ¶
949 →    →    →    →    →    " \<a href=""c7-s1831-p5-20""\>5\</a\>" ¶
```

Open the InDesign book, and open all of the files in the book.

On the Scripts panel, choose VerifyIndexLinks.jsx. It will check each hyperlink in your index against the links in the book. If all goes well, it will give you a report saying all links matched "Yay!" If not, it will give you a list of broken links.

If VerifyIndexlinks finds mistakes, open your index file, and find the matching problem hyperlink: You can just search the page field and find the page number easily. Then look at the corresponding page in the InDesign file and see what the correct link should read. Fix your index file entries.

Start from the beginning again and re-export the list. Run VerifyIndexLinks again until you get no errors.

To remove all coding from the laid-out hyperlinked index so that it can be a plain print index:

This is a GREP routine, based on the A href standard anchor coding. In Cindex, we use escape characters in front of < and > characters. Other programs may not. When you export the index for print, be sure that the resulting Word file is cleanly coded, without escape characters, in order for the GREP to work.

34 is a clean entry. Once the Word file is placed in InDesign, if it doesn't work with your encoding, try doing a search and replace to make your links look like the clean example link above, and then try GREPping again.

To run the GREP routine:

Place your index in InDesign as you would a normal index. Close all the other InDesign files.

Place an insertion point in the beginning of the index story.

Open the Edit>Find/Change dialog box, and choose the GREP panel. Make sure it says "story" only.

In the Find field, type in   (?i)(?m)<a.*?>(\d*?)

In the Change to field, type in $1

Click Change All. Your story should have all hyperlink coding removed, and just page numbers left. It will be ready for print.

Cindex Export instructions:

For the VerifyIndexLinks script, you must export your index file and manipulat it in Excel. To export it so that you can easily open and manipulate it:

1. Choose File>Save To and choose the Delimited Records option.
2. Click the Options button next to the file type selector.
3. Choose All Records In View.
4. Set Minimum Number of Fields to at least 5. If you have more than 4 layers of subheads, choose a number of fields that is at least one higher than that. This gets all your locators into one column.
5. Save the file.
6. Open it with Excel, and tell Excel it is a tab delimited file. Then perform the manipulations described on p. 2.

Sky Export Instructions:

1. Choose File>Export
2. Change the Output Type to: Delimited Text File
3. Verify that "Single character delimiter" is set to "Tab" (the default)
4. Uncheck the "Include field names" checkbox.
5. Verify that the "Enclose fields within quotes" checkbox is NOT checked (the default)
6. Check or uncheck the "Include Formatting" checkbox as desired.
7. Click OK
8. In the Export As dialog, type a name for the exported index.
9. Click Export.

Macrex Export Instructions:

1. Put index in PNO.
2. Go to Output Subprogram.
3. Load InDesign for Scripts (Excel; no ranges).LAY
4. Load InDesign for Scripts (Excel).MWP
5. Make word processor file without any description; it should have the extension of TXT.
6. Open <output>.txt with Excel and delete initial rows with x-refs.
7. Sort all remaining entries so that locators can be moved to the last column. Sorting procedures will vary according to the structure of each index. Add columns so that the locators are in column F.
8. Fill Column A with sequential (1…) numbers
9. Save the changes.
10. Open file with TextPad or other text editor and remove final hard return so that <Ctrl><End> takes you to the line with the last entry. Save the file. Proceed with Jan's documentation for verifying index links.

Contact Macrex Support Office, North America, for definition files and for documentation on Print and Hyperlink-coded RTF outputs.

Contact Gale Rhoades at galerhoades@wisebytes.biz

Please report any bugs in these scripts to Jan Wright, jancw@wrightinformation.com. Olav Kvern and Jan Wright aren't responsible for anything that happens: use your scripts wisely!