



Photo: Sharomka, Shutterstock.com

Working the Embedded Flow

Matrix Flowchart Update 2022

by Pilar Wyman, Walter Greulich, and Glenda Browne

[This is the first installment of a two part series. Look for part two in the spring issue of Key Words. Statements regarding technical support for applications are to the best knowledge of the authors and do not reflect the opinions of ASI.]

Introduction

Options for creating embedded and other linked indexes for quality, active information access in electronic books (ebooks or e-books) continue to evolve. There are many tools used in embedded indexing, including traditional publishing software such as Microsoft Word and Adobe InDesign, specialized software such as Index-Manager, DEXembed, or WordEmbed, and spreadsheet tools such as Excel, as well as utilities such as IndexExploit and the KPS Index Utilities. Indexers can also use dedicated indexing software and then output linked files. Just as there are many tools to consider, there are also various production workflows to consider.

In a presentation at the Continental Connections International Indexing Conference, October 18, 2022, **Pilar Wyman, Walter Greulich**, and Glenda Browne shared updated Matrix Flowcharts for producing active, linked ebook indexes. The updated Matrix Flowcharts provide algorithms for navigating various production workflows and selecting which tools to use and when.

The flowcharts are reproduced here and with an overview of the current situation. Some progress has been made since 2018, when the Matrix Flowcharts were last presented (at the Pacific Northwest and Western New York ASI chapters). Thus, we invite the reader to “enter the Matrix” and join us in the universe of electronic publishing and its ever-shifting code.

Summary

There has been a slow but steady increase in ebook indexing, along with a consolidation of workflows. More tools and utilities are also available for use. In a major change, Amazon’s proprietary MOBI format is gone and Amazon now uses EPUB for Kindle books.¹ Authors publishing for Kindle or to Amazon only need to provide one ebook file.

This should offer more opportunity for indexer involvement as EPUBs can include linked locators.

Future changes to the EPUB specification and how we use page numbers, which have proven to be surprisingly resilient. Research by Mary Coe and **Jan Wright** has identified huge variation in the way ebook reading devices respond to page number and other links, sending users unpredictably to the top, middle, or bottom of content on digital screens.² While page numbers provide readers with orientation as to where content is in a book, they may yet be replaced by paragraph or section numbers or something entirely different. In other words, ebook indexing is developing slowly and will likely stay exciting for a while.

Questions to Ask

Who, what, when, where, how, and why questions have to be answered for each and every ebook indexing project, though not necessarily in that order. Who is the indexer—professional, author, or computer? What are the costs or charges? Embedded indexing for ebooks often requires specialized software for activating links and providing truly clickable indexes. For this reason, embedded indexing may take longer than traditional indexing and may cost more than traditional printed-book indexes.

When will the indexing and linking work be done? Be mindful of potential issues when final editing is concurrent with indexing work. Where will index entries link to and from? The choice of software will affect the linking options. Links can be from page number locators, running numbers, or the text of the index entry. Why has the publishing format been chosen, and how will this affect your decisions? Formats include PDF, HTML, digital/ebook, print, and Amazon. Note that HTML is still relevant for help documents and

technical communication because not all online publication is in the *shell* of a digital ebook. How will this indexing and embedding and activating work be done? What software tools and versions will be used for both indexing and publishing?

Ebook Publishing and Distribution

Clients publishing via Kindle can now use linked PDF files to export linked, usable indexes. If they use Word or InDesign, they can export linked EPUB as well as linked PDF files and then publish via Kindle or otherwise.

Kindle Create, which replaced Kindle Textbook Creator in 2017, imports doc/docx and PDF files. It stores files in Kindle Package Format (KPF) and converts doc/docx files to reflowable KF8/EPUB for Kindle readers and PDF files to a special Kindle PDF fixed format. Many self-published authors follow this Kindle workflow. Kindle Direct Publishing (KDP) takes both Word and PDF files as well as EPUB. PDF files with links remain intact in Kindle ebooks.

EPUB and Kindle File Indexes

Depending on the software an indexer uses, their index terms may be embedded or not, and they may be linked or not. There are also variations in how multiple references and page numbers are treated. For links between index entries and text, a distinction must be made between index terms and page numbers. That is, when index terms (heading and subheadings) themselves are linked, page number locators are not displayed. When page number locators are linked, the text of the index terms they accompany is not linked.

Alternatively, linked running numbers can be displayed instead of page number locators, such as when working with EPUB editor programs Sigil and Jutoh. These two open-source programs have simple index functions, so index entries can be inserted directly into EPUB files when editing in the programs. Only indented format is available for these indexes, however, and instead of displaying linked page number locators, clickable running numbers are displayed.

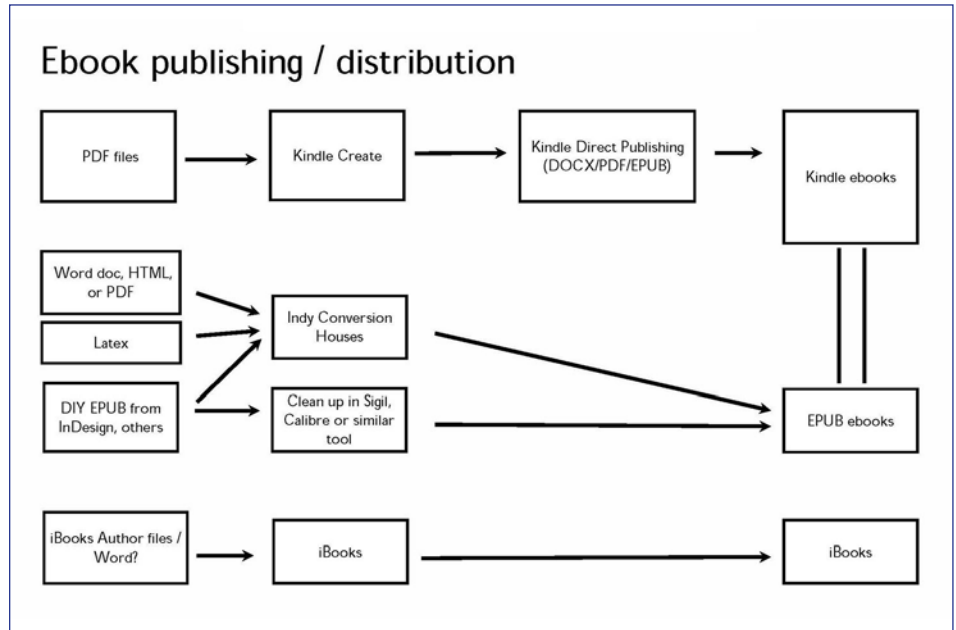


Figure 1: Summary flow for publishing to EPUB ebooks, Kindle ebooks, and iBooks

EPUB/Kindle Files: creation and display matrix

EPUB/Kindle files: Handling of terms, links and page numbers					
Index created where?	Terms in text	Terms in index	Multiple references	Page numbers	Examples for index source programs
Word processing or layout program	not embedded	not linked	shown as printed (page numbers)	displayed, not linked	any program
	embedded	not linked	shown as printed (page numbers)	displayed, linked	InDesign, PDF, Word TExtract
		linked	linked term repetition	not displayed	FrameMaker
		linked	linked running numbers	not displayed	Word (via Jutoh)
EPUB program	embedded	not linked	linked running numbers	not displayed	Sigil
		linked	linked running numbers	not displayed	Jutoh

Designations/synonyms used throughout this presentation:
 terms = index terms
 not embedded = unembedded
 not linked = unlinked = non-active = non-functional
 linked = active = interactive = functional

Figure 2: This table shows how terms, links, and page numbers are handled in EPUB and Kindle files. (Kindle files are now created based on EPUB files.)

With InDesign, PDF, Word, and TExtract, index terms are embedded, multiple references are shown as printed page numbers, and page numbers are displayed and linked. With Jutoh, on the other hand, index terms are embedded, multiple references are shown as linked running numbers, and page numbers are not displayed.

Working with Word Source Files

Indexes embedded into Word documents are NOT automatically linked for hyperactive or clickable functionality. Manual activation of embedded index links can be done in Word or with any HTML/XML tool. Manual activation takes time, however, and is tedious. Unlinked/unembedded (stand-alone) indexes in Word can always be exported and then activated. Figure 3 lists indexing tools available for working with Word files. Dedicated indexing software programs such as CINDEXT, Macrex, SKY, as well as proprietary and other content management systems (CMS) create stand-alone indexes which can be appended to Word documents. However, these indexes are neither linked nor embedded.

In addition, there are quite a few tools available to assist with writing and embedding indexes into Word files (note: Figure 3 includes software for which support is nonexistent or unknown):

- Index Manager, a stand-alone program, embeds indexes into Word, InDesign, and XML files. You can also export HTML files with links.
- IXMLEmbedder is both a stand-alone program and a Word add-in that embeds index markers.
- WordEmbed, DEXembed, and DEXter are Word add-ins that also embed index markers.
- TExtract is a stand-alone program for indexing PDF and Word files. PDF-based indexes can also be exported to EPUB files (see also the later section—Working with PDF Source Files).

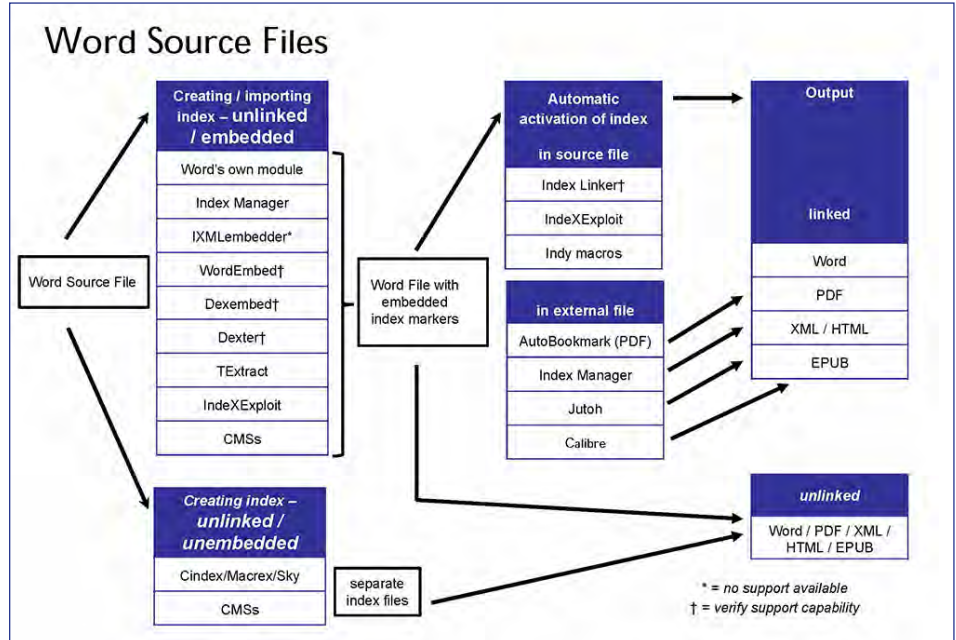


Figure 3: Summary workflow for providing indexes in electronic books that start out as Word manuscripts

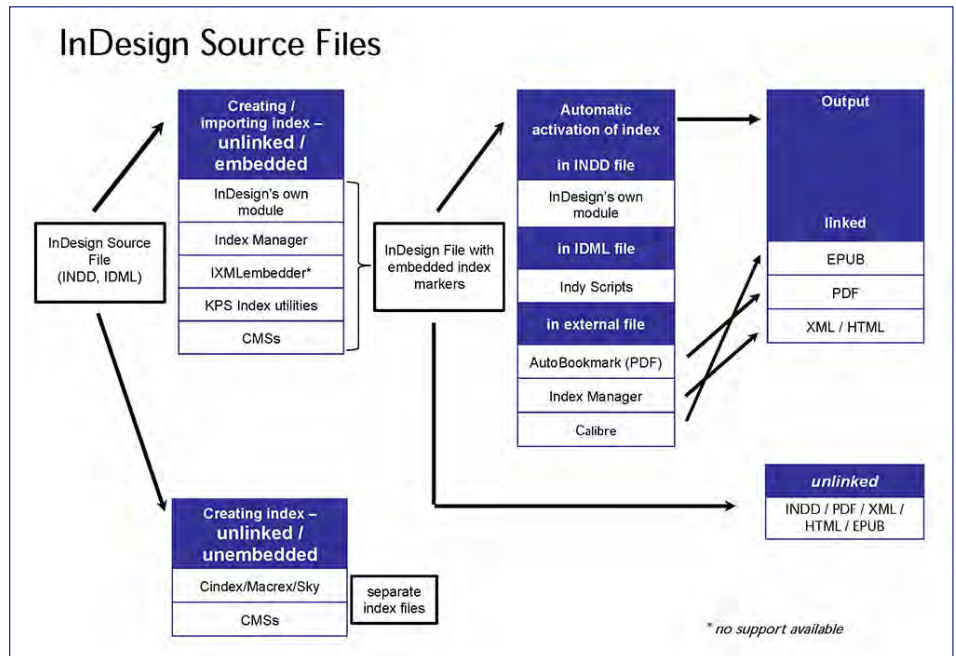


Figure 4: Summary workflow for providing indexes in electronic books that start out in InDesign

- IndexExploit is a Word add-in that adds page number links in PDF file copies of Word files.
- IndexLinker is a Word add-in that links index entries but is not suitable for larger, complex indexes (automatic linking in Word is not recognized).

There are also a few tools available for activating embedded indexes so that index entry links are *active* and clickable:

- AutoBookmark is a new Acrobat plug-in replacing Sonar Activate.
- Jutoh is a stand-alone EPUB creation program that imports Word files and activates their embedded indexes. EPUBs exported from Jutoh show linked page numbers.
- Calibre is a stand-alone ebook manager that converts Word files to EPUBs (without active links) and PDF files to EPUBs (with active indexes).

Working with InDesign Source Files (see fig. 4 on previous page)

InDesign's index module creates linked page number locators for PDFs, both for page ranges (start and ending page numbers) as well as for single pages. InDesign can export EPUBs and PDFs with active, linked indexes. All that is needed is to embed and generate the indexes into the InDesign file. In addition to the tools listed above (IXMLEmbedder, for which there is currently no support, and KPS index utilities), Index-Manager does this seamlessly.

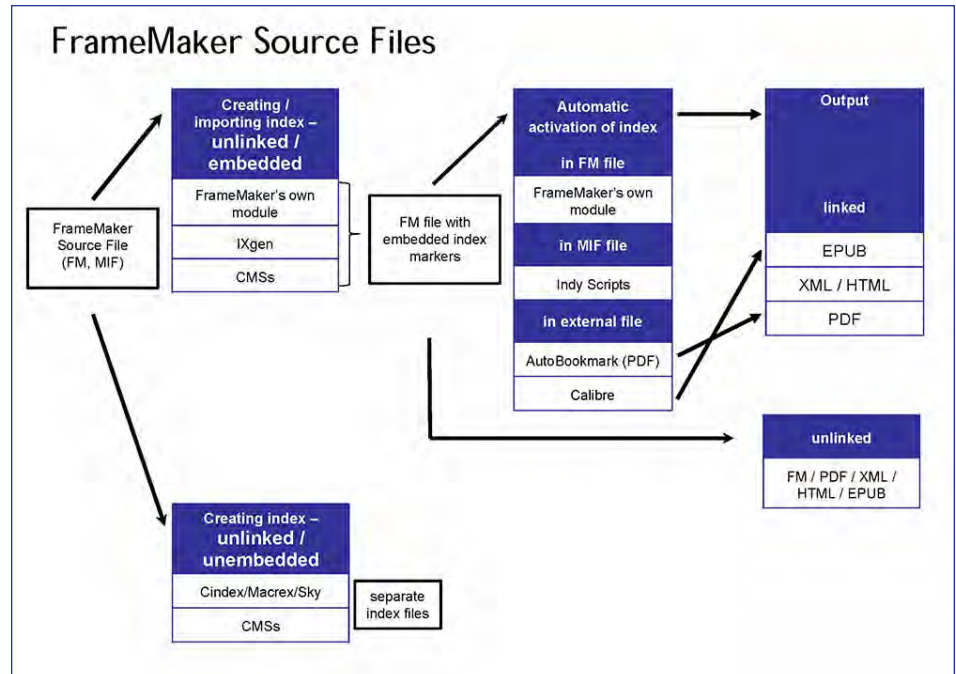


Figure 5: Summary workflow for providing indexes in electronic books that start out in FrameMaker

Working with FrameMaker Source Files

FrameMaker is still used in technical communication, with many technical editorial offices creating indexes in-house with FrameMaker. FrameMaker exports linked indexes to EPUB and XML/HTML formats. However, page number reference locators in indexes are not displayed in EPUB files exported from FrameMaker. FrameMaker-created EPUB indexes link directly from index headings and terms; in cases of multiple references, terms are repeated. IXgen is still supported (2022). ■

[Please see installment two of this update in the spring issue of *Key Words*. The spring installment will address several other types of document publication source files from PDF environments to various markup languages and LaTeX.]

Pilar Wyman is the North American representative for Index-Manager and a professional freelance medical indexer and indexing consultant. Pilar is an ICI-certified indexer and has been involved in indexing since late 1983, writing indexes since 1990. She also provides consulting and expert witness services for professional associations, organizations, and individuals. When her schedule permits, she gives training workshops and presentations on indexing and related topics.

¹ Hamilton, Jason, "The MOBI File is Dead: Long Live the EPUB," *Kindlepreneur*, August 16, 2022.

² Coe and Wright, "Looking for needles in a haystack: how do ebook reader applications handle active indexes? Part 3 - tablet devices," *The Indexer: The International Journal of Indexing* (2020), 38, (3), 271-289.