

Working the Embedded Flow: E-book Matrix Flowchart Update 2022

Photo: Sharomka, Shutterstock.com

by Pilar Wyman, Walter Greulich, and Glenda Browne

[This is the second installment of a two-part series. Look for part one in the 2022 winter 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 e-book 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 e-book indexing, along with a consolidation of workflows. A few more tools and utilities are also available for use. In a major change, Amazon’s proprietary MOBI format¹ is gone and Amazon now uses EPUB for Kindle books. Authors publishing for Kindle or to Amazon only need to provide one e-book file.

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

Future changes may come with 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 e-book reading devices respond to page number and other links, sending users unpredictably² to the top, middle, or bottom of content on digital screens. 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, e-book indexing is developing slowly and will likely stay exciting for a while.

Questions to ask

The *who*, *what*, *when*, *where*, *how*, and *why* questions have to be answered for each and every e-book 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 e-books 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/e-book, 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 e-book. How will this indexing and embedding and activating work be done? What software tools and versions will be used for both indexing and publishing?

All these questions need to be addressed when writing static indexes for appending to electronic publications and for activated indexing of electronic publications.

[The first installment of this article, in the fall 2022 issue of Key Words, covered e-book publishing and distribution, EPUB and Kindle file indexes, Word source files, InDesign source files, and FrameMaker source files.]

Working with PDF source files (see figs. 6-9)

More often than not, indexers receive PDF copies of book manuscripts and are asked to provide indexes for them. There are many ways to write indexes for PDF files. With dedicated software like CINDEXTM, Macrex, or Sky, indexes can be written and provided to clients as separate files to be appended to their text. Excel, TExtract, or CMSs can also be used to create such indexes. These index files can then be merged with the rest of the book file and laid out and exported as a larger PDF file. However, such indexes do not provide clickable index content. They are neither active nor linked.

With Index-Manager, a Word version of a PDF source file formatted to match the pagination of the PDF source file can be imported, and an index can be embedded into the Word version of the manuscript. Such a Word file, with an embedded index, can then be activated as detailed above.

Once a PDF file with an inactive index has been generated, it can be *activated* by way of AutoBookmark, Sonar Activate, or by using independent macros. However, the process of creating and enabling links in PDF documents can be tedious.

Calibre can convert PDF files to EPUBs. Typical target anchors in EPUBs are coded as , where the number corresponds to the same page number in the index. Jutoh and

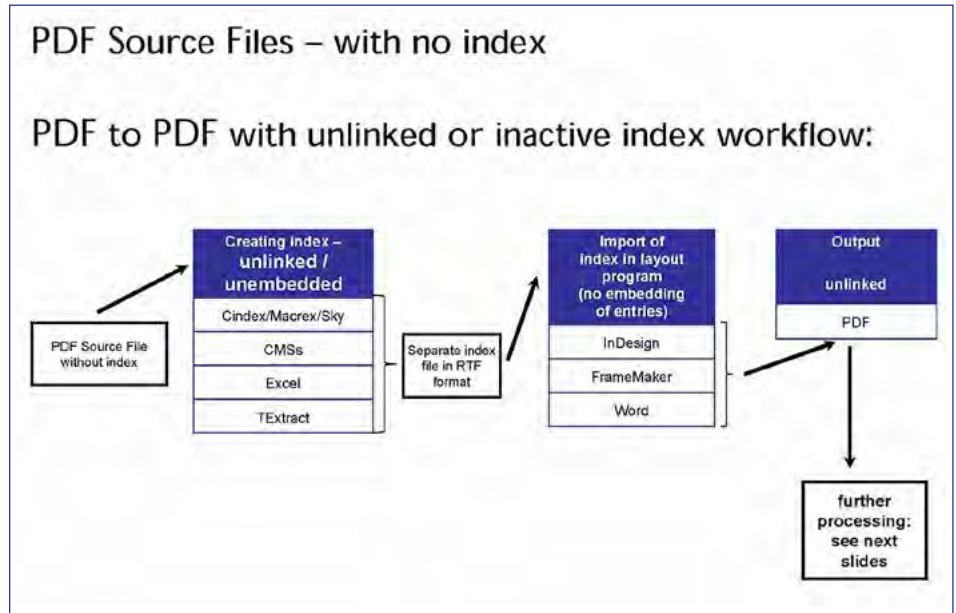


Figure 6: Summary workflow for providing dead or unlinked, inactive indexes for PDFs without indexes.

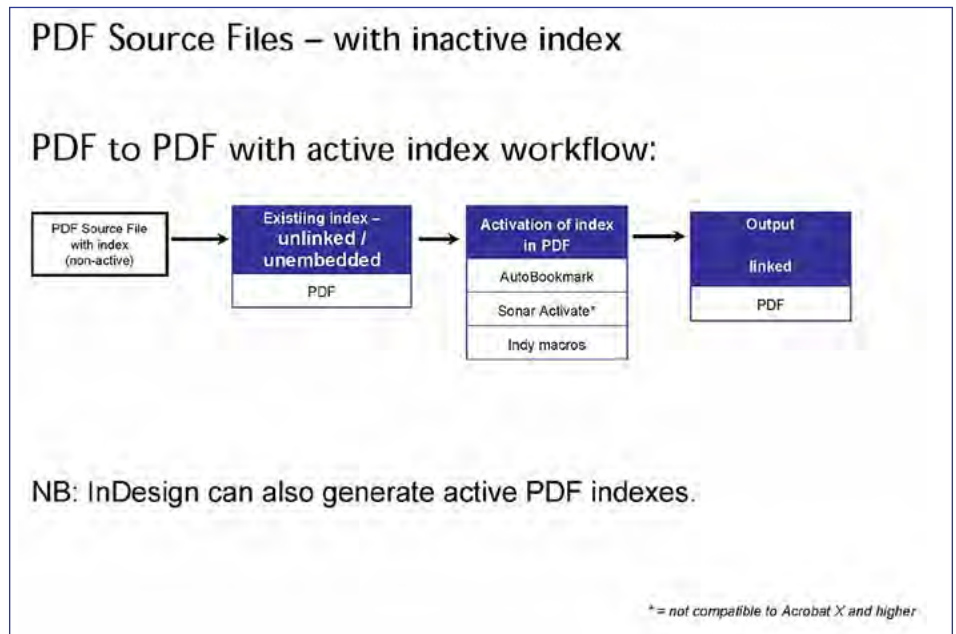


Figure 7: Summary workflow for providing active, linked indexes for PDF files with inactive indexes.

Sigil are similarly useful for exporting EPUBs with linked indexing. Thus, if you have a PDF source file with an index, linked or not, you can convert it to EPUB.

InDesign exports books and book projects as active PDFs and as active EPUBs. If an index is embedded into an InDesign book, an active PDF with linked indexing can be exported (the

page reference locators are clickable, including both starting and ending points of page ranges) as well as an active EPUB with the same functionality. Such EPUBs may require cleanup, and Calibre is great for that.

Recent improvements to TExtract allow for exporting index entries to already existing EPUB files, resulting in EPUBs with active indexes.

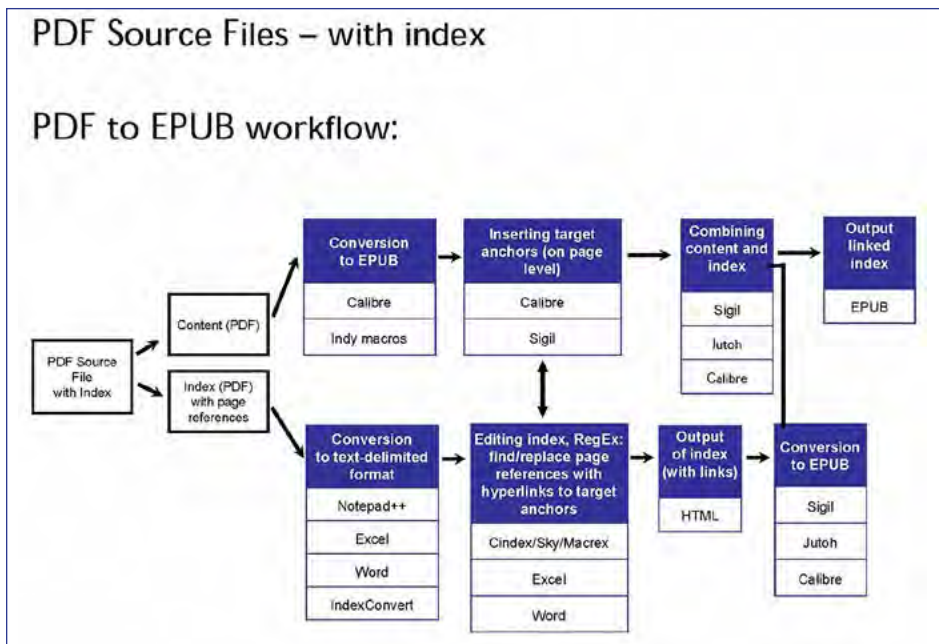


Figure 8: Summary workflow for providing active, linked EPUB versions of PDF files with indexes.

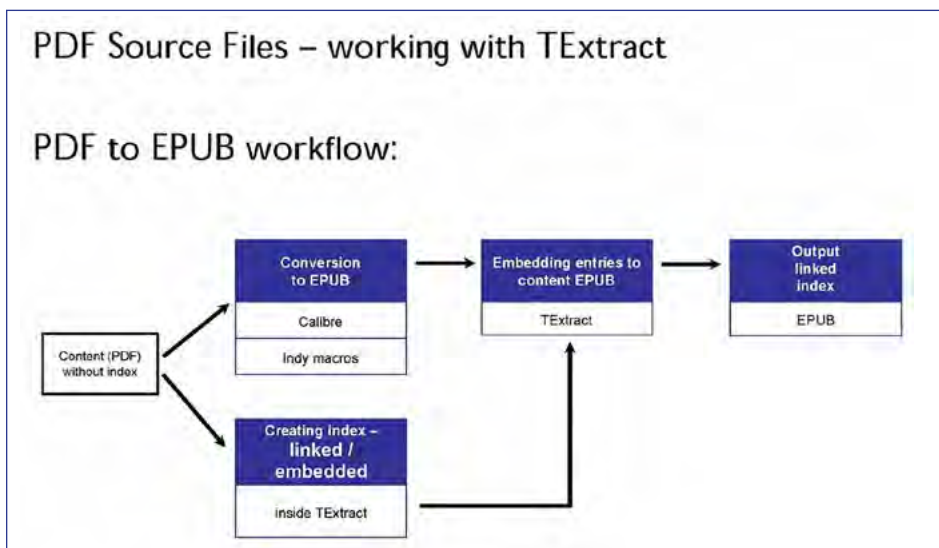


Figure 9: Summary workflow for providing active, linked EPUB versions of PDF files with indexes using TExtract.

Working with XML source files (see fig. 10 on page 25)

DocBook is a popular XML editor, and both DEXembed and IXMLEmbedder create DocBook syntax for index entries and embed them in XML documents opened by Word. In this case, Word is used as a text editor. However, none of these programs activate indexes. While DEXembed is a Word add-in, IXMLEmbedder has two modules, a stand-alone program as well as a Word add-in. There is no support for IXMLEmbedder. Index Manager exports HTML with links by default. It can also work with XML files, after some technical adaptation.

Working with HTML, XHTML, and HTML5 source files (see fig. 11 on page 25)

For XSLT transforms (HTML to XML transformations), it is best to review the XML workflow tools. Spreadsheet tools such as Excel can also be useful here.

Working with LaTeX source files (see fig. 12 on page 26)

LaTeX is still used for scientific and mathematical texts, though other tools work just as well for formulas and such. In addition to all that is detailed in the workflow above, there are several Word to LaTeX converters: LyXConverter (The Editorium), GrindEQ, Word2LaTeX, and Doc2LateX. Doc2LateX also works with Google Docs. Some of these are Word add-ins, others are online. They also convert from LaTeX to Word. LyX also converts index markers to LaTeX \index commands. However, the converter is not installable under Windows 10 and higher.

Before converting to LaTeX, you can convert XE fields to LaTeX \index commands with a Word macro. LaTeX \index commands are embedded in text and can be further processed by the \makeindex command. The index output includes index entries and page numbers per the LaTeX file. In addition, TExtract can import PDF files and export the index, but only the index, in LaTeX format.

While there are numerous LaTeX-to-EPUB converters on the market, the best way to convert LaTeX to EPUB

is to first convert LaTeX to HTML with LaTeXXML and then convert the HTML to EPUB. Good HTML-to-EPUB converters include Sigil, Jutoh, and Calibre.

Working with EPUB source files (see fig. 13 on page 26)

Both Sigil and Jutoh have indexing functions or modules. Sigil processes index entries in its own syntax and does not import XE fields from Word or any other programs. Jutoh can import XE fields from Word and convert them into its own syntax. Both programs also allow editing of existing linked entries. And both programs export to EPUB. Jutoh can also export EPUB to OpenDocument (ODT) format and Word.

Software Resources

The workflows include many tools:

- **AutoBookmark** (Acrobat plugin)—activate links in PDFs
- **Calibre** (e-book manager, open source)—convert linked PDFs to linked EPUBs
- **CINDEX**—dedicated indexing software
- **DEXembed** (MS Word add-in)—embed indexes into Word docs
- **IndexConvert** (MS Word add-in)—embed indexes into Word docs
- **InDesignConverter** (tool)—save MS Word docs as InDesign tag files, import Word index entries into InDesign docs
- **IndexExploit** (MS Word add-in)—embed indexes into Word docs
- **IndexLinker for Word** (MS Word add-in)—create hyperlinks from index page numbers back to Word doc text
- **Index-Manager** (indexing software)—create embedded indexes in Word, InDesign, or XML files
- **IXGen** (FrameMaker add-on)—indexing and marker management
- **Jutoh** (EPUB creation program)—import Word files and automatically activate embedded index markers

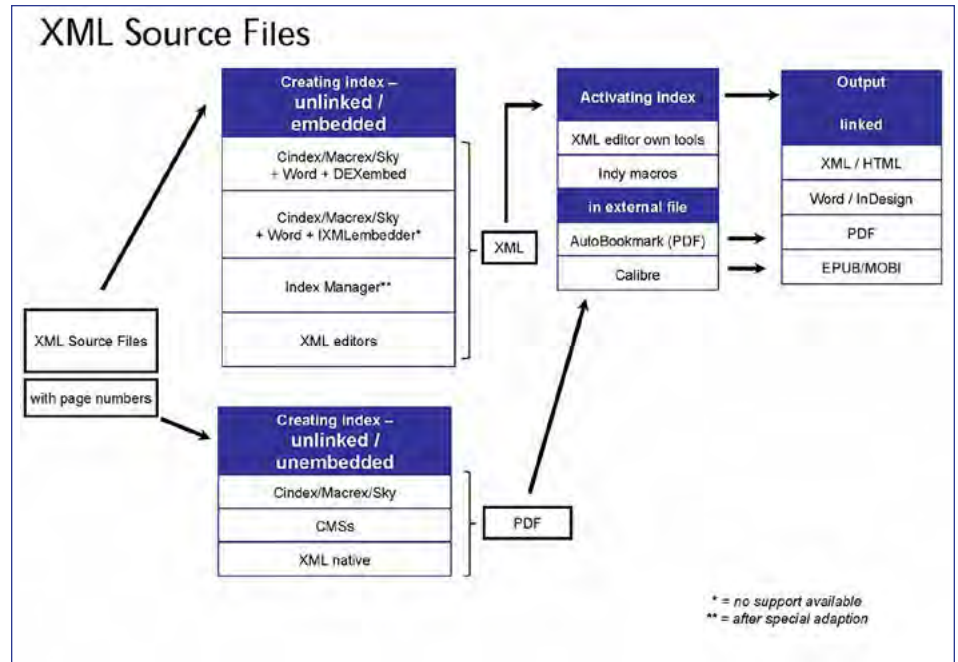


Figure 10: Summary workflow for providing linked indexes for XML files.

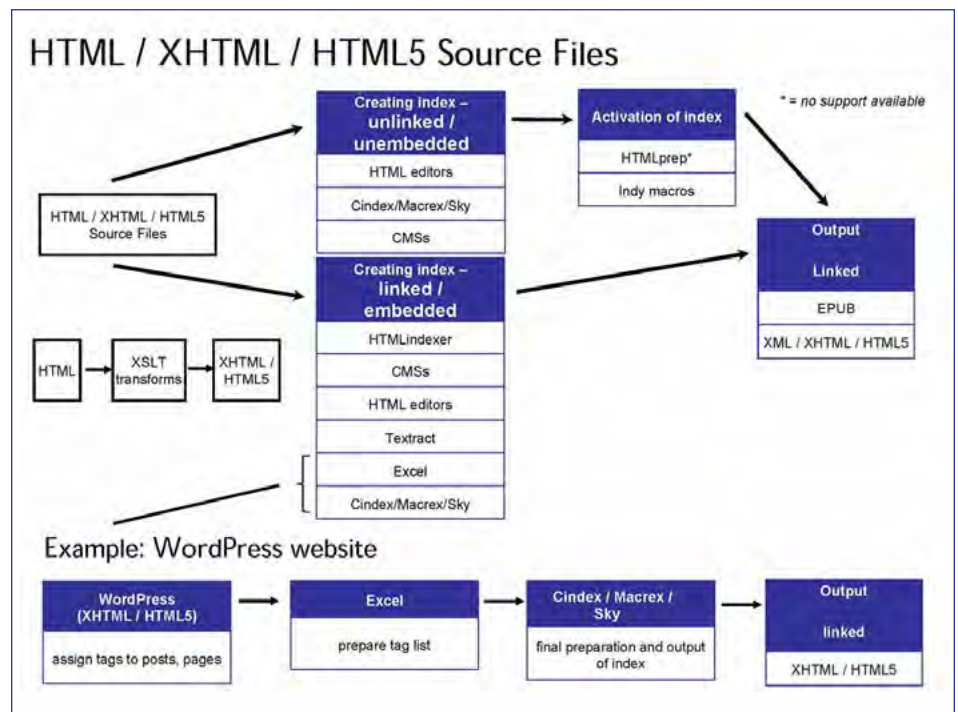


Figure 11: Summary workflow for providing linked indexes for HTML, XHTML, and HTML5 files.

- **Kerntiff Publishing Systems (KPS) IndexUtilities** (InDesign scripts and plugins)—help embed index markers
- **Sigil** (EPUB editor, open source)—perfect for editing EPUBs
- **SKY Index Professional**—dedicated indexing software
- **TExtract** (indexing software)—create embedded indexes for PDF or Word files, including output as EPUB

- **WordEmbed** (MS Word add-in)—embed indexes into Word docs

LaTeX resources:

- **Word to LaTeX converters:** [LyXConverter](#), [GrindEQ](#), [Word2LaTeX](#), and [Docx2LaTeX](#)
- **LaTeX-to-XML/HTML/MathML converter:** [LaTeXML](#) (open source)

For additional resources, including large versions of the graphics and copies of the article installments, please see the ASI Digital Publications Indexing SIG at <https://digital-publications-indexing.org/>

Pilar Wyman, Chief Indexer & Consultant, Wyman Indexing, has been writing indexes and providing related services for over thirty-two years. Currently, she is the leader of the American Society for Indexing (ASI) Digital Publications Indexing (DPI) Special Interest Group (SIG) and Regional Sales Manager, USA/Canada for Index-Manager.

Glenda Browne has been a freelance indexer of books, journals and websites since 1988. She is coauthor of *Website indexing* and *The indexing companion* and author of *The indexing companion workbook: book indexing*.

Walter Greulich, a founding member of the German Network of Indexers (Deutsches Netzwerk der Indexer—DNI), is a physicist and long-time editor and proofreader of scientific journals and books. He has been involved in the professional creation of indexes since the mid-1980s. In addition to the content and organizational side, he has been fascinated by the technology surrounding indexing from the very beginning.

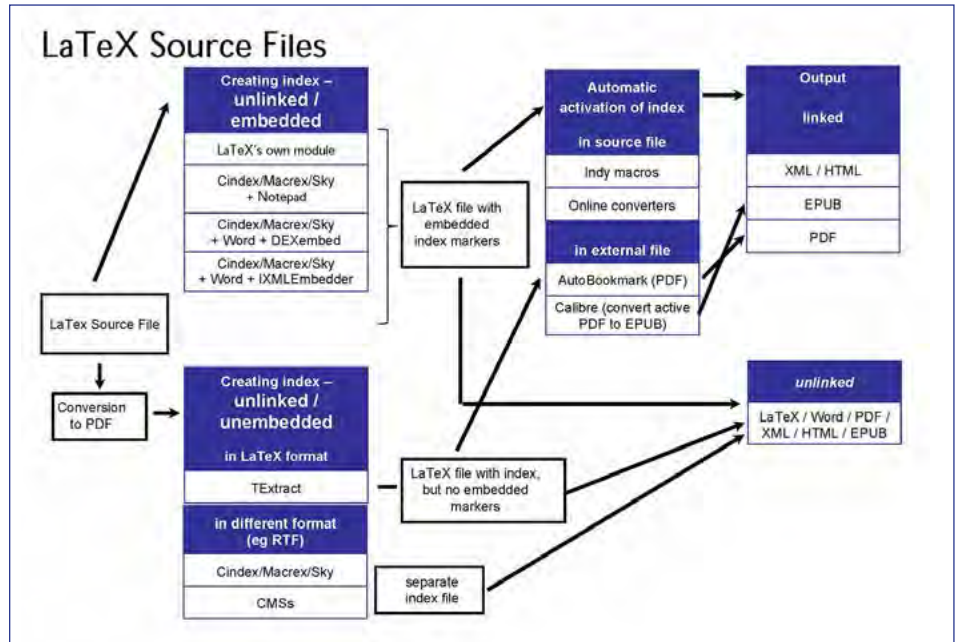


Figure 12: Summary workflow for providing linked indexes for LaTeX book files.

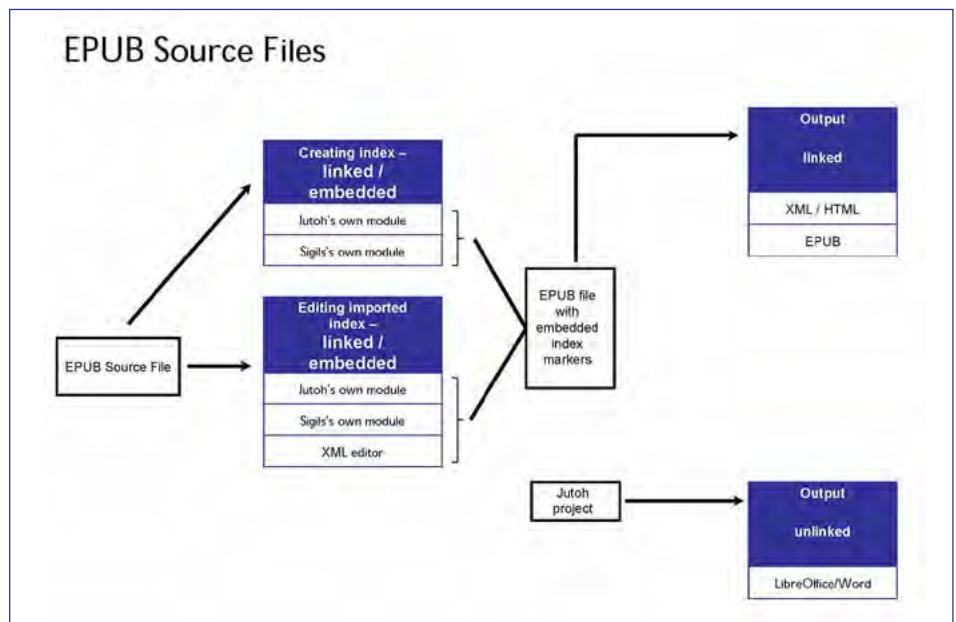


Figure 13: Summary workflow for providing indexes for EPUB source files.

1. Jason Hamilton, "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.