

The Matrix Group, Int'l  
 ASI & ISC/SCI 2016  
 Chicago, Illinois

Enter the Matrix

ebook Indexing

17 June 2016

---

---

---

---

---

---

---

---

The Matrix – indexing options



- ▶ With the matrix, you too can discern the way to index outputs in all kinds of formats, from all kinds of tools.
- ▶ In the face of so much information, where's the best access?

---

---

---

---

---

---

---

---

Ebooks, reading



▶ Ebook reading systems:	▶ Ebook formats:
▶ Specialized (Kobos, Nooks)	▶ EPUB 3 (open source)
▶ Personal computers (PCs)	▶ KF8 (Kindle Fire)
▶ Smart phones	▶ Mobi (Kindle)

---

---

---

---

---

---

---

---

### EPUB 3



- Open source ebook standard
- Developed and maintained by IDPF
- Uses existing standards where possible
- International, with strong focus on accessibility

---

---

---

---

---

---

---

---

### EPUB 3



- Uses XHTML
- Much like websites
- Differences relate to packaging: compression (zipped with .epub extension), navigational elements, structural elements, book-wide metadata based on Dublin Core

---

---

---

---

---

---

---

---

### EPUB 3 indexes



- EPUB 3 Indexes Specification (<http://www.idpf.org/epub/idx/>) provides a standard that allows for ebook indexes that do everything print indexes do and more.

---

---

---

---

---

---

---

---

### Ebook indexes



Ebook indexes:

- Active, linked indexes
- Clickable entries or locators linked to unique places in text, or offer other options

---

---

---

---

---

---

---

---

---

---

### EPUB 3 indexes



- Better navigation
- New features: index filtering
- New features: range highlighting
- New features: interactive generic cross-references

---

---

---

---

---

---

---

---

---

---

### Better Navigation

Index groups allow the user to expand and collapse groups...

...or easily access a group by clicking the letter to jump to it

<p>ABCDEF GHIJ KLMN OPQ RSTUVW XYZ</p> <p><b>A</b></p> <p>abbreviations acceptable list, R1:52 in subheadings, R2:18-19 academic theology, R1:14, 18, 22 accents, R1:20 Access Innovations, R2:152 accounting basics of, S:8, 23-26 professional advice for, S:16 software, S:7, 23-24, 26 tracking system, S:49-50</p>	<p>ABCDEF GHIJ KLMN OPQ RSTUVW XYZ</p> <p><b>B</b></p> <p>back-of-the-book indexing. See book indexing back-ups, S:9, 48-50 backslashes, R2:92 Baggiato, Mauri, R2:54 bank accounts, S:8, 24 behavioral science textbooks, R2:57, 59 Bell, Hazel, R1:28, 36, 44, 45; R2:40</p>	<p>ABCDEF GHIJ KLMN OPQR STUVW XYZ</p> <p><b>L</b></p> <p>laser printers, M:34, 47-48; S:7, 8 legal cases, treatment of, R2:75 legal indexes locators in, R2:46 sub-subheadings in, R2:7 textbook, R2:54 legislation, R2:75-76 letter-by-letter alphabetization, R2:61 Levi-Strauss, Claude, R1:29</p>
---	--	--

(Combs)

---

---

---

---

---

---

---

---

---

---

### Index Filtering

**Index\***

A B C D E F G H I J K L M N O P  
Q R S T U V W X Y Z

Filter

...  
Lee, Fitzhugh, [19-20, 21f](#)  
Lee, George Washington, [22-23](#)  
Lee, Robert E., [45-49, 49f, 51f](#)  
Leventhorpe, Collett, [28-30, 32f, 33f](#)  
Logan, Thomas, [39-40, 42](#)  
Long, Armistead, [45](#)  
...

**Index**

A B C D E F G H I J K L M N O P  
Q R S T U V W X Y Z

Unfilter

...  
Lee, Fitzhugh, [21f](#)  
Lee, Robert E., [49f, 51f](#)  
Leventhorpe, Collett, [32f, 33f](#)  
...

Filter by?  
Figures  
Tables

Index filtering allows the user to restrict the view to only figures, or only tables, for quick reference.

\*“f” following a locator indicates a figure (Comba)

---

---

---

---

---

---

---

---

---

---

---

---

### Generic cross references

**Index**

...  
Battles. See [names of specific battles](#)  
...

Term categories (nav doc)

Battles  
[Antietam](#)  
[Chattanooga](#)  
[First Manassas](#)  
[Fort Pulaski](#)  
[Harper's Ferry](#)  
[Lexington](#)  
[Pea Ridge](#)  
[Second Manassas](#)  
[Shiloh](#)

**Index**

...  
Antietam, [65](#)  
...  
Battles. See [names of specific battles](#)  
...  
Chattanooga, [56](#)  
...  
First Manassas, [32](#)  
...  
Fort Pulaski, [54](#)  
...  
Harper's Ferry, [62](#)  
...  
Lexington, [40](#)  
...  
Pea Ridge, [45](#)  
...  
Second Manassas, [58](#)  
...  
Shiloh, [51](#)  
...

Fully-functional generic cross references using term categories prevent the user having to guess at relevant terms. See also the Index to the IDPF EPUB 3.0.1 Indexes Specification 1.0:  
<http://www.idpf.org/epub/idx/epub-indexes-index.html>

(Comba)

---

---

---

---

---

---

---

---

---

---

---

---

### Range Highlighting

**Index**

Buford, John, [21](#)  
Gettysburg, [20-22](#)  
...  
Lee, Robert E., [21-23](#)

**Body of book**

¶20 After his success at Chancellorsville in Virginia in May 1863, the Confederate Army marched through the Shenandoah Valley to begin their second invasion of the North—the Gettysburg Campaign...

¶21 Elements of the two armies initially collided at Gettysburg on July 1, 1863, as Lee urgently concentrated his forces against Brig. Gen. John Buford...

¶22 Lee was forced to change his plans. Longstreet would command Pickett's Virginia division of his own First Corps, plus six brigades from Hill's ...

¶23 Around 1 p.m., from 150 to 170 Lee ordered an artillery bombardment...

Range highlighting helps the user quickly identify where coverage of a topic begins and ends

(Comba)

---

---

---

---

---

---

---

---

---

---

---

---

### Software tools for ebook indexing



---

---

---

---

---

---

---

---

### Ebook indexes: overview



- |                |            |         |         |
|----------------|------------|---------|---------|
| ▶ index inputs | ▶ outputs  |         |         |
| ▶ standalone   | ▶ Frame    | ▶ print | ▶ PDF   |
| ▶ MSWord       | ▶ InDesign | ▶ HTML  | ▶ ebook |
| ▶ XML eds      | ▶ other    | ▶ XML   | ▶ Apps  |

---

---

---

---

---

---

---

---

### Work flow issues



- ▶ Linking to entries vs embedding entries in content files

---

---

---

---

---

---

---

---

### Software options

#### Content that is inserted into files:

Linked entries	Embedded entries
<p>XHTML anchors or unique IDs.</p> <p>Unique IDs must be convertible by a compiler into XHTML anchors or other anchor format as required for links to work.</p>	<p>Index entries are inserted directly into content as fields, XML elements, or using the program's own unique marking system.</p>

---

---

---

---

---

---

---

---

### Software options

#### Granularity:

Linked entries	Embedded entries
<p>Yes. But links only go to where unique IDs are located.</p> <p>NOTE: Display of content varies depending on reading device size and view settings.</p> <p>NOTE: Indexers should consider carefully which IDs to use as locators.</p>	<p>Links can go to any level.</p> <p>NOTE: Different workflows dictate what is displayed as locators. Indexes must be recompiled as content changes.</p> <p>NOTE: Indexers should consider carefully where to embed markers.</p>

---

---

---

---

---

---

---

---

### Work flow issues

#### Single sourcing capabilities:

Linked entries	Embedded entries
<p>Yes.</p> <p>NOTE: If content is later edited, unique IDs may be lost.</p> <p>NOTE: Most workflows will rely on an outside compiler to generate different outputs, and to generate the index and create links to unique IDs. If the publishing software doesn't natively provide good output, publisher may have to tweak manually.</p>	<p>Yes.</p> <p>NOTE: InDesign CC outputs EPUB with active indexes, and active indexes in HTML if the book is one file.</p> <p>NOTE: Indesign and Frame can output active PDF indexes.</p> <p>NOTE: Word can output print indexes.</p> <p>NOTE: XML requires XSL style sheet or other automated process.</p>

---

---

---

---

---

---

---

---

Work flow issues

File Management:

Linked entries	Embedded entries
<p>Indexer has index files. Publisher has content files (indexer gets copy). Publisher needs to notify indexer when/if content changes.</p> <p>NOTE: Publisher OR indexer can run scripts to insert anchors.</p>	<p>Indexer has copy of content files.</p> <p>NOTE: (1) Indexer works directly in content files, OR (2) Indexer works outside files with interim unique IDs as locators, client or indexer later runs macro to merge entries into files, OR (3) Indexer works in content file copy which will later be merged via software workflow with the master copy.</p> <p>NOTE: Workflow should include error checking.</p>

---

---

---

---

---

---

---

---

Work flow issues

Content Changes:

Linked entries	Embedded entries
<p>This is only a problem if using numbered locators such as page numbers, or paragraph or section numbers. If so, index must be reviewed carefully and changes made as appropriate.</p>	<p>Index must be recompiled after each content change and the index reviewed for accuracy or errors. Every time.</p>

---

---

---

---

---

---

---

---

Work flow issues

Translation/Localization issues:

Linked entries	Embedded entries
<p>Translation of entries: Index is translated as a complete document, along with the text. OR publisher may choose to request a complete re-index of the new translation.</p> <p>Locators: If using page numbers as locators, pagination may change.</p>	<p>Translation of entries: Embedded entries must be translated along with the text.</p> <p>Locators: Locators will be generated when the index is regenerated.</p>

---

---

---

---

---

---

---

---

Work flow issues

---

Chunking/Customizing content:

Linked entries	Embedded entries
Yes. Potential issues include: orphaned subheads, missing cross-references, and broken links.	Yes. Potential issues include: orphaned subheads, missing cross-references, and broken links.

---

---

---

---

---

---

---

---

Work flow issues

---

Updates and Revisions:

Linked entries	Embedded entries
If original entries can be sorted consecutively (in order of appearance in the content), the indexer can easily review the index for changes.  NOTE: Consider planning for this capability if later editions are anticipated.	It depends ... on software and production processes.  NOTE: View index entries in context is helpful, as is a tool that highlights differences in versions of files.

---

---

---

---

---

---

---

---

Work flow issues

---

Legacy Book Index Conversions:

Linked entries	Embedded entries
(1) Obtain content files including index; (2) If necessary, insert unique anchors into content files; (3) Sort existing index into locator order; (4) Enter new locators that point to anchors rather than to page numbers; (5) Output the index in appropriate format to include in the eBook.	If legacy book is in a layout/desktop publishing software, and if the software has a later version that supports active ebook export, convert files to the later version.

---

---

---

---

---

---

---

---

Break / Q&A




---

---

---

---

---

---

---

---

Software Options for Active Indexes

Standalone indexing:

Print	PDF	HTML/Web	e-Book (XHTML)
Yes.  NOTE: Unique IDs must be available to use as locators.	Yes, with Sonar Activate  NOTE: Locators must match Sonar requirements. NOTE: Acrobat 10 doesn't work with Sonar at this time. Use Sonar Activate 6 for Acrobat Pro 9. Use Sonar Activate 5 for Acrobat Pro 8.	Yes, with locators that include anchor IDs (<a id="xx"> tags)  NOTE: HTML/Prep + CINDEXT or Sky Index → links + appropriate files. NOTE: Macrex + style sheets → links + appropriate files.	Yes, with locators that include anchor IDs (<a id="xx"> tags) or Canonical Fragment IDs (CFIs)  NOTE: HTML/Prep + CINDEXT or SKY Index → links + appropriate files. NOTE: TExtract supports anchor IDs

---

---

---

---

---

---

---

---

Software Options for Active Indexes

Microsoft Word:

Print	PDF	HTML/Web	e-Book (XHTML)
Yes.  NOTE: No décor, specialized locators, or cross reference checking. Watch for multiple targets. Regenerate index before final printing.	Yes, with Sonar Activate  NOTE: Cross references will be static, not active.	Yes, with third-party macros  NOTE: Requires specialized macros to insert anchor tags (<a id="xx">) and links NOTE: MSWord links to page level.	Yes, with third-party macros  <i>Same notes as for HTML/Web, PLUS:</i> NOTE: TExtract may be of use. NOTE: While Amazon Create Space and Kindle Direct accept Word docs, indexes are NOT active.

---

---

---

---

---

---

---

---

### Software Options for Active Indexes

---

XML editors:

Print	PDF	HTML/Web	e-Book (XHTML)
Yes.	Yes	Yes	Yes
NOTE: Requires a style sheet to output desired format.	NOTE: Requires a style sheet to output desired format.	NOTE: Requires a style sheet to output desired format and insert anchors and links.  NOTE: Can link to word level.	NOTE: Requires a style sheet to output desired format and insert anchors and links.  NOTE: Can link to word level.

---

---

---

---

---

---

---

---

---

---

### Software Options for Active Indexes

---

Adobe InDesign:

Print	PDF	HTML/Web	e-Book (XHTML)
Yes.	Yes	Yes, in InDesign CC	Yes, with InDesign CC EPUB export.
NOTE: Regenerate before final printing.  NOTE: To single-source in older versions, use scripts and standalone indexing software.	NOTE: Indexing plug-ins from Rich Bines activate print indexes for InDesign PDFs.	NOTE: InDesign CC indexes link only if you create books as one file. NOTE: Older versions strip out index entries upon export. Anchors must be inserted and then used as locators.	NOTE: Kindle formats are not supported. NOTE: Older versions strip out index entries. Use scripts or plug-ins, or standalone index with inserted HTML anchors.

---

---

---

---

---

---

---

---

---

---

### Scenarios

---



---

---

---

---

---

---

---

---

---

---

Scenarios

---

▶ We are in a rush and must get a book to press ASAP! – What’s the quickest way to get the index in? What tools should we use?

---

---

---

---

---

---

---

---

---

Scenarios

---

▶ We are not in a rush, and we would like to publish both digital and print editions. – Where do we start? What tools should we use?

---

---

---

---

---

---

---

---

---

Scenarios

---

▶ We are only interested in publishing electronically. – We can do the indexing any time, right?

---

---

---

---

---

---

---

---

---

Scenarios

---

▶ We just want to convert a print edition of a book we already published in digital format – that’s easy, no? Can’t we just work off the old PDF and the index we already got?

---

---

---

---

---

---

---

---

---

Scenarios

---

▶ We do everything in InDesign/Word/Oxygen (name your poison), but haven’t upgraded to the latest version. And now we want to include an index. – What do we do?

---

---

---

---

---

---

---

---

---

Scenarios

---

▶ We are about to release a new edition. The only difference is a new chapter 25. We’re going to do both print and ebook editions this time, though. – Can you get us a new index? How quickly?

---

---

---

---

---

---

---

---

---

Scenarios

▶ I'm self-publishing my book, and it's going to be the greatest thing ever! I'm going to do both print and ebook versions, of course. – Can you get me an index for it? I have no idea how to do that, or what I'm doing, really.



---

---

---

---

---

---

---

---

Resources



- ▶ ANZSI Indexing ebooks: <http://anzsi.org/resources/reading-lists/indexing-ebooks/>
- ▶ ASI Digital Trends Task Force (DTTF) <http://www.asindexing.org/about-indexing/digital-trends-task-force/>
- ▶ ASI Matrix Resources for Active Indexing: <http://www.asindexing.org/wp-content/uploads/2013/08/matrixresourceslist.pdf>
- ▶ The Indexer March 2012: <http://www.theindexer.org/>
- ▶ SI Publishing Technology Group: <http://www.ptg-indexers.org.uk/about/ebooks.htm>



---

---

---

---

---

---

---

---

Resources

Contact us!

- ▶ ASI DTTF [dttf@asindexing.org](mailto:dttf@asindexing.org)
- ▶ The Matrix Group, Int'l: Glenda Browne, Michele Combs, Dave K. Ream, Jan Wright, and Pilar Wyman
- ▶ Glenda Browne thanks Copyright Agency Limited (Australia) for sponsorship to attend this conference.



---

---

---

---

---

---

---

---