What's Going On with Embedded Indexing? The Way Forward

BY STEVE INGLE

Steve Ingle is the president and CEO of WordCo Indexing Services (www.wordco.com), located in Norwich, Connecticut. He created his first index (8 lines) at the age of 10. After graduating from Yale University with a degree in German literature, he went on to earn master's degrees in German and Russian Area Studies. In 1988, Steve began freelance indexing part time while also working at the Modern Language Association (MLA) in New York. He began indexing full time in 1991. Steve has served on the national board of ASI. His company now employs a team of indexers and completes about 500 projects annually for a diverse group of clients. His interests include indexing as a business and indexes for digital publications.

Part Two: The Way Forward

In Part One of this article (November 2015), we looked at the reasons why more nonfiction eBooks do not have functional, hyperlinked indexes and why there is not more eBook-index-related work for indexers. Contributing factors discussed include publishers' perceptions that quality digital indexes are prohibitively expensive, ignorance about indexes and the indexing process in general, as well as insufficient advocacy by indexers themselves. The focus of this second, concluding, part of the article will be on what indexers can do to help enrich the rather barren eBook-index landscape. I will provide some suggestions for alternative workflows, as well as providing some tips on how indexers can improve communication with their customers.

There is some confusion, even among indexers, between embedded indexes and eBook (hyperlinked) indexes. Before we discuss workflows, it is important to understand the distinction.

In the case of embedded indexes, it is usual that the entire index is included as tags within the text file(s), and can be generated from these tags. Embedded indexes predate hyperlinked indexes. For example, in DocBook, an XML system for encoding books, each index tag in the text file includes information about the heading and any subheadings, as well as where the discussion begins and ends (spans). The index can be generated via an algorithm which searches for the tags, arranges them alphabetically, and adds page locators based on each tag's location within the file. The idea, in theory, is that this provides flexibility for the publisher. The indexer can add tags to the book while it is still in manuscript; also, several chapters of one book with index tags can be combined with chapters from another tagged book to create a new book, with automatically generated index entries.
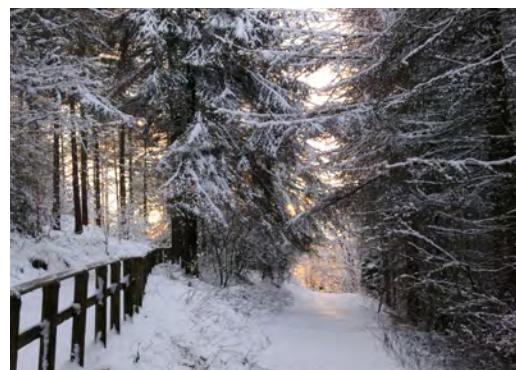
Adobe InDesign's indexing module was created originally with embedded indexes for print books in mind, rather than hyperlinked indexes for eBooks. The fact that in pre-CC versions of InDesign the index was stripped out before the book was exported to ePub indicates that hyperlinked eBook indexes were not the goal. Also, InDesign's menu of page spans for index entries (e.g., "For Next # of Pages") indicates that the focus was on paginated print books, not eBooks. There are certain ways hyperlinks can be added to DocBook or InDesign indexes, but this was not the initial focus of embedded indexes. In general, embedded (tagged) indexing requires that the indexer manually insert tags at specific locations within the file.

Hyperlinked (eBook) indexes, on the other hand, do not necessarily require embedded index tags. All that is needed for a hyperlinked index is a hyperlink from the index entry or index page locator to a specific and unique location (such as an element ID-number) in the text (I will not discuss the more complex case of spanned locations, as this is beyond the scope of this article). For the vast majority of nonfiction eBooks, it is not necessary to have actual index tags embedded in the text files. While it is possible to adapt index tags in the text to produce an index that

is hyperlinked back to that specific location, this requires the addition of anchor tags in the text and hyperlinks from the index. The alternative (creating tags from a hyperlinked index based on element IDs) is easier since it does not involve manual tagging by the indexer. This can be a huge time-saver.

There is definitely a place for embedded (reflowable) indexes, such as in annual publications like tax guides that don't change much from year to year, but most eBooks require a hyperlinked index only. Even if index tags are required, there are ways (discussed below) to add them to the book file(s) automatically at a later stage. Thus, hyperlinked indexes can be made into embedded indexes. Also, creating hyperlinked indexes allows a more familiar indexing workflow, in contrast to manually adding index tags. For example, by using element ID numbers, we can index *to* the element number, rather than having to insert tags. As indexing professionals, we need to be focused on indexing rather than tagging.

Especially because there are no standard workflows for creating hyperlinked indexes, we need to understand what our customers are trying to achieve, and we need to be able to help them do it. Aside from knowledge, this requires communication, and good communication begins with listening. Knowledge comes from the experience of coming up with solutions to our customers' eBook-index needs.

A good place to start is to engage our customers by asking them questions: What are you trying to achieve with a hyperlinked index? Do you want hyperlinks from index headings or from page locators? Do you want hyperlinking to top-of-the-page, element (paragraph),

or exact location? Will you be working in InDesign or some other platform? Will there be a print version, or is this book going directly to digital?

These are just a few of the questions you might consider bringing up with a customer when they have asked you to create an eBook index. They may not know the answers themselves, but are looking to you for guidance. For example, questions they have for you may include the following: Is there any advantage to hyperlinking from page locators versus from headings? Is it faster and/or cheaper to hyperlink to the page versus the exact location in the text? By having a conversation with your customer, you can ensure that the project is (1) doable, (2) profitable, and (3) is going to be satisfactory for the client.

I have found that publishers and authors inquiring about hyperlinked indexes have vastly different goals and levels of knowledge. There are no standards for how to get the job done (i.e., techniques and workflows), since every customer has unique needs. Here are a few examples of customer eBook-index requests I have received in the past several years:

- An author: "I have my book as a PDF, can you index it and add hyperlinks to the index entries?"
- A small publisher: "We want an index for the print version, but we also want hyperlinked index that will go right to the location in the eBook version. Can you do this?"
- A medium-sized publisher: "We need you to index using our online authoring platform. The platform will let you add tags without having to manually type out the tags. Can you do this?"
- A large publisher: "We have a book that's going directly to ePub from Word. It may be released as a print book later. It's 700 pages in Word and there's a two week turnaround. What do you suggest we do about the index?"

I started to realize that my customers knew a lot less about indexes and indexing than I did. In each case, we at WordCo had to come up with a workable (and worthwhile) solution. For example:

- I learned the hard way that adding hyperlinks to a PDF document is extremely labor-intensive and tedious. The author would have been better off if we had hyperlinked the index in the original InDesign and then re-ouput to PDF.
- The small publisher may not have been willing to pay for precise hyperlinks, but they were OK with element- or even page-level linking. We gave them both options, and they chose to go with page-level linking. We could offer this service at a nominal cost.
- The medium-sized publisher with the authoring platform had a method that "worked," and was better than direct tagging of files, but was still terribly slow. We invented a workflow that allowed us to use our indexing software to build and edit the index, and then insert the tags automatically at the final stage. This new method took less time and got better results.
- The large publisher with the book going directly to digital needed to have an idea of what the final product would look like, as well as a workflow for achieving this. With input from the comp, we succeeded in developing yet another workflow: we added element numbers to the Word files and then indexed using "dummy" page numbers and created a hyperlinked index in the ePub. Perhaps not the most elegant solution, but it was one that worked and that the customer considered satisfactory given time and budgetary constraints.

Each customer is unique and has different needs, and we need to do our best to understand what they are looking for. As we come up with solutions, we have found that a particular solution or a piece of that solution may be applicable to a future challenge. We are beginning to develop a few workflows that apply to most scenarios and
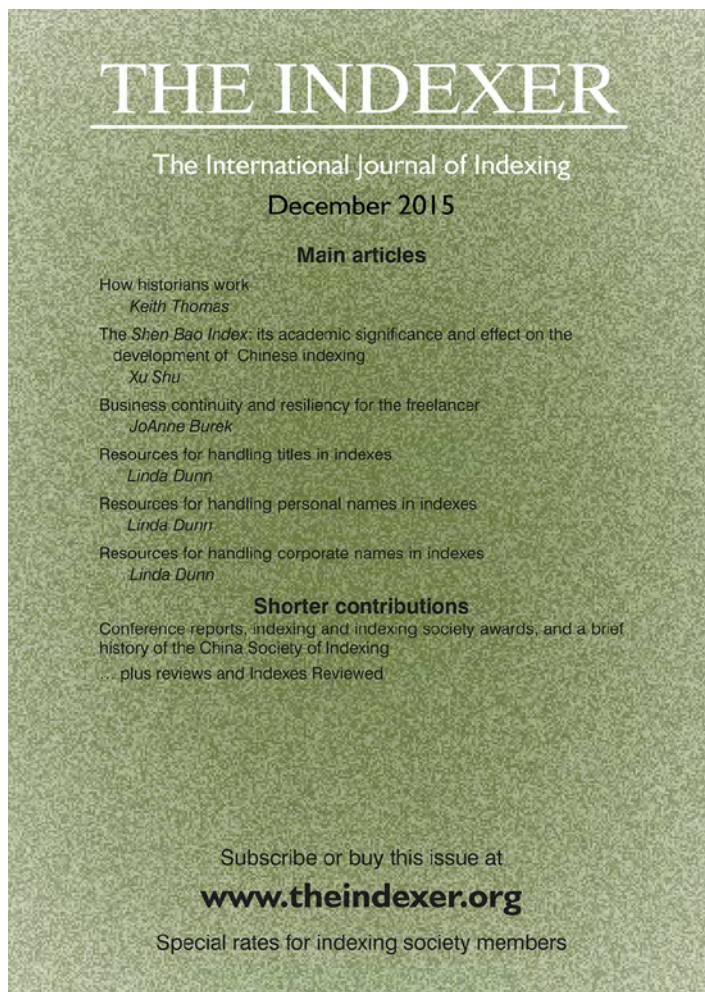
could potentially be adopted by other indexers. In the meantime, I can offer the following advice to indexers looking to do more eBook indexing:

- Make sure you know what you can and cannot do when it comes to hyperlinked indexes. Don't promise more than you can deliver.
- Work with individual authors and small publishers as your eBook-index "laboratories." It's bad to lose a customer because of failure, but you don't want it to be one of your bread-and-butter clients. Moreover, small publishers not encumbered with top-heavy bureaucracy tend to be more flexible and willing to try something new.
- Avoid direct tagging whenever feasible. If you have to tag, use software tools (for example, WordEmbed) that allow you to work using your regular indexing software.
- Use technology to your advantage. There are plenty of tools out there, some better than others, and some indexing-specific and some not. If a tool can save you time without compromising quality, consider adding it to your toolbox.
- Element-level IDs seem to be a good compromise between overly vague (e.g., page-level) linking and exact (word-level) linking. Element IDs can be added and removed automatically, freeing the indexer from the tedium of tagging. Also, index-referenced element-level IDs have the advantage of enabling an association between index-entry records in your indexing software and the IDs in the text file. This approach is powerful because each element-ID in the text file that is referenced in the index record is associated with particular fields in the index record

(i.e., heading and subheadings). As a result, embedded index tags can be added automatically by searching for IDs in the text and replacing them with properly formatted index tags containing all of the field information from the index record. This may require some creative scripting, but it can be done (tip: get to know some programmers!).

- We need to index only **once** if possible, not twice. In other words, we need to incorporate the digital index into our workflow from the very start. This is an ideal, and often not the reality in a print/digital world. A majority of publishers still follow the Manuscript → InDesign → PDF → ePub workflow. But if the eBook is to be more than just the "digital version" of the print book, we need a new approach.

We are indexing professionals, not tagging technicians. Most publishers don't understand that creating an index is an organic process. The index evolves. We indexers use computer-assisted indexing software to facilitate this process, but the process of index creation is not linear. We often find ourselves changing index entries from the beginning of the book once we get to the end of the book and understand the context. The index is rough until it gets "polished" at the very end.

The challenge for us is to create the very highest-quality indexes for eBooks as well as print books and sell this value to our customers. Once we can do that dependably and affordably, and quality eBook indexes become more prevalent (as they should be!), the next step will be to explore creative ways of using/displaying eBook indexes. The future of eBook indexes promises to be exciting, as long as we indexers are up to the challenge! ●