Executive Summary: Indexes in eBooks

History

For centuries, printed books have included indexes to provide access to their contents. Since the beginning of the digital age, many content providers have considered indexes extraneous. This first happened with full-text searching on mainframes once space was available to store the whole content of journals and books. Similarly, CD-ROMs and web sites originally elected not to include indexes. Yet with each new technology, users pushed back wanting to have indexes. History is now repeating itself with eBooks. Why this constant attack on indexes? The reasons often cited are: searching replaces indexes; there are no pages in eBooks; and, making indexes functional in eBooks costs money. These reasons are explored below, and some other thoughts on future uses of indexes are given.

Searching *vs.* Indexes

Admittedly search technology has improved over the years, but natural language still befuddles search engines which look for patterns of words and grammar clues, not the meanings behind the written text, to present a set of *possible* results. It is left up to the user to click on each result to review it for usefulness or applicability.

Indexes, on the other hand, are written by information specialists who analyze the content and develop a list of terms representing the topics, their breakdowns, and their relationships. Indexes are crystallizations of access paths to the content's concepts. The terms in an index can be thought of as metadata, but since the index is part of the book, they are really metacontent. Indexes disambiguate homographs, relate synonyms and similar topics, consolidate terminology in multi-authored works, ignore passing mentions, include references to images, guide users to the preferred terms, indicate the span of coverage, breakdown a large topic into subtopics, and more. Indexers often provide alternate terminology making the content accessible to casual readers as well as experts in the field.

Indexes are a familiar access method for users offering them a higher degree of success in finding information. They should be provided as an alternative to searching, and, along with tables of contents, constitute one more research approach. With the advent of the International Digital Publishing Forum's EPUB3 Indexes Specification, indexes can become more than just static sections in an eBook (*see* footnote). At least one study has shown significantly higher success rates using indexes over search as well as speeding the time to finding relevant content. (*see* footnote).

eBooks Don't Have Page Numbers

It is true that many eBooks don't display page numbers. However, the page breaks are often encoded in the eBook and can be destination points for links from an index. Not all publications use page numbers; some use section numbers or other types of locators which are cited by the index. These make suitable candidates to use for hyperlinking as well. In an index, where subheadings have only one locator attached, the text of the subheading can be what activates the link.

The Money Issue

Publishers often pay hundreds or thousands of dollars to have indexes written for their books. Yet they seem ready to toss this investment of human capital aside when creating the eBook. Partly this may be due to the lack of tools for converting legacy books. However, most major conversion houses can create a linked, or active, index during their process. New books produced by one-source publishing can easily yield indexes for both print and electronic publications—possibly with the index locators linking, not to page breaks, but to paragraphs. The tools will only improve over time. Compared to the initial cost of developing this asset, the cost to include an index in the eBook is quite small.

New Functionality

It is expected that the EPUB3 Indexes Specification will allow reading systems to exploit this standard index structure encoding to provide new user functionality for an improved user experience. For instance, upon highlighting a section of the content, the reading system could show the user what index entries point to that text. The user could then jump off to other content based on one of those entries. Indexes and content would become a two-way street to exploration of a book.

Discovery

Books listed for sale on web sites often allow a preview of the book's table of contents and sometimes the index. Indexes can help sell print books. In brick & mortar bookstores, consumers flip to indexes before making a purchase decision. Why shouldn't this same preview be provided to consumers of eBooks? In fact, with the EPUB3 Indexes Specification, for the first time in history, indexes will all be tagged in the same way allowing *index crawlers* to retrieve the index or its headings for use in guiding potential purchasers to the book. For example,

- The index could be displayed to the user for perusal.
- An index's headings could be matched against the headings for other books' indexes which the user bought to suggest another purchase to the user.
- Index entries could be mashed up for a user to browse—linking to the books they came from.
- The main headings from all book indexes could be used as search terms on the site.

These are new ways of using an eBook index as a sales tool—even to support printed books.

Summary

Publishers should give serious thought to retaining indexes in their eBooks. Indexes provide a familiar access method for users to find material; they set the stage for eBook functionality that will come in time as reading systems adopt the EPUB3 Specification; and they provide the publisher and distributor new ways to implement discovery approaches leading to increased sales.