

Notes

- 1 www.oxforddictionaries.com/us/definition/american_english/granularity (accessed June 18, 2016)
- 2 From M. Heatter (1982) *Book of great desserts*. New York: Alfred P. Knopf, pp. 65–6.

References

- Anderson, J. D. (1997) *NISO-TR02-1997 Guidelines for indexes and related information retrieval devices*. Bethesda, Md.: NISO Press.
- Kruschwitz, U. (2005) *Intelligent document retrieval*. Dordrecht, Netherlands: Springer, p. 37.
- Mertes, K. (2016) 'Usability testing for indexes.' *Key Words* 24(3), 20.
- Neiberg, M. S. (2016) *The path to war: how the First World War created modern America*. New York: Oxford University Press.
- Piccone, T. (2016) *Five rising democracies and the fate of the international liberal order*. Washington DC: Brookings Institution Press.
- Rosch, E. (1978) 'Principles of categorization.' E. Rosch and B. Lloyd (eds), *Cognition and categorization*. Hillsdale, N.J.: Lawrence Erlbaum.

- Stauber, D. M. (2004) *Facing the text*. Eugene, Ore.: Cedar Row Press.
- Yang, B. (2016) *Heroines of the Qing*. Seattle and London: University of Washington Press.

Fred Leise started his indexing career with the University of Chicago Publishing Program's introduction to indexing course in 1993. Since 1995, he has been indexing nearly continuously (with time out for taxonomy consulting), specializing in scholarly works in international relations, cultural and political history, and landscape studies.

He is a founding member of the Institute of Certified Indexers, an instructor for the University of California, Berkeley Extension Division online indexing course, and has served several terms on ASI's board of directors, of which he is currently immediate past president.

He regularly presents workshops and programs at both national and chapter indexing events and has written a number of articles on various aspects of indexing for both *Key Words* and *The Indexer*.

See also the information on Fred's webinar on page 143.

Multiple entry points: variants and cross-references in indexes and thesauri

Heather Hedden

Indexes point to information, but sometimes multiple pointers are needed for the same information destination. This article reviews the use of See references and double posts in indexes, and then looks at how cross-references are used in thesauri and taxonomies. The focus is on the craft of coming up with synonyms and alternative phrases in indexes and thesauri, how they can be used to reduce scatter, and how they may differ between the two, and between print and online in general.

Indexers may be interested in thesauri for two reasons. First, use of a thesaurus can help ensure consistency of main entries for very large indexing projects that draw out over time and/or involve multiple indexers; and second, creating a thesaurus involves similar tasks and skills to creating indexes. Creating a thesaurus may be for your own large indexing project or for use by other indexers.

The tasks or skills that are similar for book indexing and thesaurus creation include identifying what is important for inclusion, figuring out how to name the concept/topic, determining all the additional ways in which the concept may be identified, creating some structure among the entries/terms, and creating related-concept (*see also*) cross-references. An article I wrote for *Key Words* (Hedden, 2012) covers all of these points in brief. This article goes into depth on techniques for determining the variations for a concept which in an index or thesaurus can be called 'multiple entry points.'

Thesaurus basics

A thesaurus is a kind of controlled vocabulary. A controlled vocabulary is a set of index terms, in which each term stands for an unambiguous concept, and there is control (rules or limits) over the addition of terms to the vocabulary. Indexers use the terms already in the controlled vocabulary. A very small controlled vocabulary might be a simple list without the need for multiple entry points. A thesaurus is a kind of controlled vocabulary that has structure, involving hierarchical and associative (related-term) relationships between its terms, and has multiple entry points. These are both defining features of a thesaurus. If a controlled vocabulary does not have multiple entry points, then it is not a thesaurus. National and international standards explain the proper design of a thesaurus. These include ISO 25964, *Thesauri and interoperability with other vocabularies, Part 1: Thesauri for information retrieval*.

As for thesaurus terminology, entries in a thesaurus are called terms. There are two kinds of term: preferred terms,

which are linked to content, and non-preferred terms, which serve as *see* cross-references pointing to the preferred terms. The cross-references in thesauri traditionally use the word *Use* (often in all caps as USE), instead of *see*, and thus may be called *Use* references instead of *see* references, but they are essentially the same. Meanwhile, what are called *see also* cross-references in a book index are called associative relationships in a thesaurus. Instead of using the designation 'see also', thesauri use 'Related term,' abbreviated as RT, to point from one preferred term to another preferred term. Related term relationships in a thesaurus are used in a similar manner to *see also* cross-references in book indexes.

Multiple entry points in general

Multiple entry points means having two or more index entries or thesaurus terms that refer to the identical concept and the identical set of locators, or links to the identical set of text or content. They are 'entry points' for the user to enter the index or the thesaurus to get to the same result. In back-of-the-book indexes, multiple entry points are handled in two different ways: double posts and *see* references. Most indexes will have both kinds of multiple entry points. Printed periodical indexes (to the extent that they still exist) would normally use only *see* references, so as to save space from repeating the citations. Thesauri also use cross-references as their sole method for multiple entry points. Some online implementations of thesauri for the end-users (not the indexers), however, will omit displaying the cross-references and provide links from multiple entry points directly to the content, so it appears more like double posts. Nevertheless, thesaurus creators must always designate a 'preferred' term to which a cross-reference points.

The role of multiple entry points is to capture different ways in which different people might describe or look up the same concept or idea. These include differences of wording between the author and the user/reader, and differences among different users/readers. In the case of a shared thesaurus, this also includes different choices of wording among different indexers. Not every index entry or thesaurus term has multiple entry points, though. A concept may have any number of (multiple) entry points, or it may have only a single entry name.

Multiple entry points, whether in a back-of-the-book index or in a thesaurus, should have a roughly equivalent meaning in the context of the book or body of content. Exact synonyms are only one of many forms of achieving this usable equivalence in context. Other forms include near-synonyms (Careers/Jobs); variant spellings (Defence/Defense); lexical variants (Hair loss/Baldness); foreign language names (German Air Force/Luftwaffe); acronyms/spelled out forms (United Nations/UN); common/scientific names (Cancer/Neoplasms); older and newer terms (Near East/Middle East); phrase variations (in print) (School buses/Buses, school); opposites (in select cases) (Behaviour/Misbehaviour); and broader/more specific concepts (Computers/Laptops).

Multiple entry points in back-of-the-book indexes

In back-of-the-book indexes, the art of creating double posts is no different from that of creating *see* cross-references. The difference between the two methods lies only in the decision when to employ one or the other technique. Typically, double posts are favoured when there are no subentries involved, since space is not taken up by merely repeating a short string of page locators, and it saves the reader the extra step of looking up a referenced entry. However, if subentries are present, which add lines to an index, a cross-reference is preferred to save the space in duplicating the subentries with locators. There are exceptions, though, of using a *see* reference without the involvement of subentries when it is desired to instruct or emphasize the preferred naming of a concept, for example:

Handicapped persons. *see* Persons with disabilities.

Double posts, of course, can be more than double. There could be three or four entries that all point to the same concept with the same locators. Similarly, you may have more than one *see* reference point to the same referenced entry.

Multiple entry points in back-of-the-book indexes differ from multiple entry points in thesauri to the extent that back-of-the-book indexes must serve the user who browses alphabetically. Thus, techniques for creating multiple entry points specifically for back-of-the-book indexes and not for thesauri include:

- substituting the first word of a multi-word phrase with a synonym (e.g. film reviews/motion picture reviews)
- inverting adjective–noun phrases, with a noun–comma–adjective (labour unions/unions, labour)
- exchanging prepositional phrases with adjective–noun phrases (ethics of communication/communication ethics)
- 'flipping' main entries and subentries (if there are going to be subentries anyway) in double posts, with merely the addition of a preposition, as in the example:

African Americans
business ownership, 247–248

Business ownership
by African Americans, 247–248

- 'Semi-flipping,' a double post of a single main entry with a main entry–subentry combination, as in the example:

Design
furniture, 223

Furniture design, 223

While multiple entry points are understood to mean equivalent concepts that point to the same content via the same page locators, it is possible to have a variation, or a 'semi double post,' where one of the double-post entries has a slightly broader meaning than the other, and then actually has additional locators, as in the following example:

body language, 85–87, 118

gestures, 85–87

Body language serves as a double post for gestures on pages 85–87, but on page 118 the topic discussed is the broader body language only and not specifically gestures.

Due to the alphabetical nature of the index, it is not necessary and preferably avoided to create multiple entry points that would lie next to each other or very near, such as a few lines away. For example, an index would not have both Ethnic groups and Ethnic minorities as entries (assuming they were used for the same set of locators), whether as double posts or with one of them as a *see* reference. The same exception, of instruction or emphasis, applies, though, which could result in a *see* reference pointing from one entry to another that lie near each other. For example, the *see* reference Net profit. See Net operating profit in a finance textbook explains to the student that these two terms do indeed refer to the same concept.

Multiple entry points in thesauri

Thesauri by definition have cross-references (typically worded as *Use*, rather than *see*). This is because the hierarchical structure among terms precludes the inclusion of true double posts in a thesaurus; a hierarchy can only be built out of single terms (preferred terms) for each concept. Furthermore, unlike back-of-the-book indexes, factors do not exist that would lead to deciding on whether to use double posts or cross-references. In a back-of-the-book index, the number of locators per main entry is a factor, but the number of locators in a book is fixed, whereas the number of references linked from a thesaurus term continues to grow over time. In a back-of-the-book index, the presence of subentries is a factor, but thesauri do not have subentries. (Thesauri have hierarchy among terms, but these terms are all akin to main entries. They cannot be compared with subentries, which have a different function.) Thus, to better serve the long-term use of thesauri with a potentially ever-increasing number of reference links (locators), thesauri are designed to make use of cross-references, with the reference pointing from a non-preferred term (NPT) to a preferred term (PT). It is thus the task of the thesaurus creator not only to come up with the various multiple entry points, but also to make the determination of which one is to be the preferred term.

The *Use* cross-reference in thesauri is actually treated as a ‘relationship’ between terms, between a preferred term and a nonpreferred term. Technically, it is referred to as an ‘equivalence’ type of relationship. The other kinds of relationships between terms in a thesaurus are hierarchical (broader term–narrower term) and associative (related term/*see also*). In a thesaurus, all relationships are reciprocal and have indicators in both directions. So, not only do the related-term/*see also* cross-references point in both directions, as they do in an index, but so do the equivalence relationships. The reciprocal of *Use* is UF, which stands for ‘used for’ or ‘used from.’ Thus, for every *Use* reference in one direction there is a UF reference going in the opposite direction. For example, if you have the cross-reference

Public procurement USE Government purchasing, then you will also have Government purchasing UF Public procurement. The display of non-preferred variations at the preferred term is actually helpful to the users, informing them of the meaning and scope of the term.

The issue of using multiple entry points where one entry has a slightly broader meaning than the other plays out differently in thesaurus implementations than in back-of-the-book indexes. Technically it is acceptable to have a narrower-scoped NPT point to a broader-scoped PT, since the preferred term includes the narrower concept within its scope. It is incorrect to go the other way and have a slightly broader concept as a NPT pointing to a slightly narrower PT. For example, Laptops USE Computers is acceptable, whereas Computers USE Laptops is not acceptable in a thesaurus. Computers. See Laptops might be acceptable in a back-of-the-book index, if all mentions of computers in the book were only of laptops. A thesaurus is used for indexing multiple sources and sources that will be indexed in the future that the thesaurus creator cannot see or anticipate. Therefore, the NPT must be created to cover all possible situations, not just the context of one book.

Furthermore, while a narrower NPT pointing to a broader PT is technically correct, it might not always be a desirable cross-reference. If the cross-reference, such as Laptops USE Computers, is fully displayed to the end-user, then such a narrower-to-broader cross-reference is OK. If cross-references are not displayed to the end-users, however, with the NPT of Laptops linked directly to the indexed results for all of Computers, then there could be undesirable results, and this kind of NPT should not be created. For example, in the case of the example of Laptops USE Computers, an indexer might index an article on Supercomputers with the index term Computers. Then an end-user, interested specifically in laptops, looks up the term Laptops and is taken directly to result set of articles indexed with Computers. This result set includes articles on supercomputers and other computers that are not laptops, in addition to articles on laptops. As a consequence, the end-user will think that the indexing is wrong by retrieving articles on other computers besides the selected laptops.

Thesauri also differ from back-of-the-book indexes by appearing almost exclusively online rather than in print. Historically thesauri did exist in print, and so the standards for their basic structure serve both print and online implementations. The current online use of thesauri, however, impacts how the thesaurus functionality is displayed and impacts decisions in creating NPTs.

The online environment permits hypertext, so the text of each thesaurus term can link directly to a page with a list of relevant citations/references to the indexed content. Even though a thesaurus is designed with cross-references which are utilized by the indexers, if a thesaurus along with a body of indexed content is displayed to end-users online, it may be decided that there is no need to waste the user’s time with the redirection of a PT to an NPT. Rather, the NPTs may also hyperlink directly to the page of relevant citations/references that are indexed with the PT. On the other hand, similar to back-of-the-book indexes, the display of cross-references, otherwise unnecessary, may be displayed

to the end-user for purposes of instruction or emphasis regarding the ‘correct’ term. Unlike a back-of-the-book index, though, where the decision as to whether to present a cross-reference or double-posts is made on a case-by-case, term-by-term basis within the same index, the decision to display cross-references (displaying the NPT *Use* PT), has to be made for the implementation of the thesaurus as a whole, not term by term. The tendency is for scholarly, technical or medical thesauri, or internally used thesauri with repeat users, to display the cross-references to the end-users, whereas thesauri that are more general for general public use do not. What this means for the thesaurus creator is that if cross-references will display to the end-users, care should be taken not to create too many cross-references that might clutter up the thesaurus.

The online implementation of thesauri means that thesauri can be ‘searched’ and not just browsed. As a consequence, thesauri may often be presented to users as searchable by default instead of browsable. Thus, the thesaurus creator should not assume that the users will browse the thesaurus alphabetically the same way they might browse a back-of-the-book index. Whereas a back-of-the-book index would not have double posts that lie alphabetically near each other, such as Ethnic groups and Ethnic minorities, a thesaurus should include them both (one as the NPT and one as the PT), because they will probably not both be seen in the same view.

The sophistication of the search technology in use will impact the number and choice of multiple entry points created. Typically, search software will make matches on words within terms, so word order does not matter. That means that phrase inversions, which are typical for back-of-the-book indexes, generally can and should be omitted from thesauri. This would include omitting both noun–comma–adjective inversions and variations of prepositional phrases. Search software may also support ‘stemming’ so that variations of plural/singular, verbal noun endings, and other standard suffixes do not need to be included as NPTs. Including them clutters up the search results. For example, State education standards *USE* Educational standards might be desired in a thesaurus that is searched and not browsed, but the sophistication of the search engine would result in Educational standards being retrieved anyway if a user were to enter State education standards, with an exact match on one word and a stemmed match the other word.

It’s not simply a matter of the search technology but also the user interface display of thesaurus terms that can impact the decision in creating certain NPTs. Typically, there are different user interface displays for the indexers and for the end-users. Examples of different user interfaces for viewing the subject thesaurus of Gale, a part of Cengage Learning, show what happens when different users look up ‘education standards’ which is an NPT for Educational standards. Figure 1 is a screenshot from the thesaurus management software used by the Gale thesaurus editors with the selections of ‘Smart’ search (allowing searches for words within terms). (Alphabetical/Begins search is also an

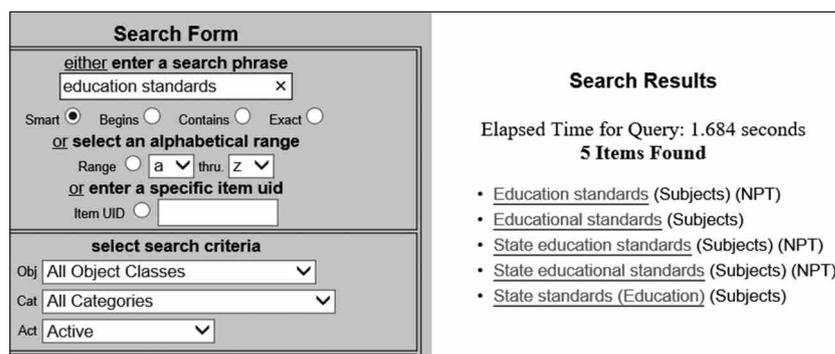


Figure 1 Screenshot from Gale thesaurus management software

option.) Notice that three NPTs and two different PTs are retrieved. The three NPTs point to one or the other of these PTs, so there is some redundancy in the retrieval. Figure 2 is a screenshot from the indexing software used by Gale database indexers with the selections of ‘Smart’ search. Figure 3 is a screenshot from the Gale database product InfoTrac’s user interface as used by the customers, with the results upon entering ‘education standards’ in the search box of the Subject Guide. Notice that both the desired cross-reference and other terms with the words (including names, not just Subjects) are retrieved.

Comparison summary

Multiple entry points in back-of-the-book indexes and in thesauri have the same goal: to direct various users, who use various terms that mean the same thing, to the same content location. For back-of-the-book indexes, the users are the different readers. For thesauri, users are more varied, and comprise multiple indexers along with multiple end-users. Multiple entry points can also include various designations of the same concept in the text. Within a book an author might refer to a concept with different designations. Within the content indexed with a thesaurus, on the other hand, inconsistent terminology is more widespread, since thesauri are typically used to index content from multiple authors.

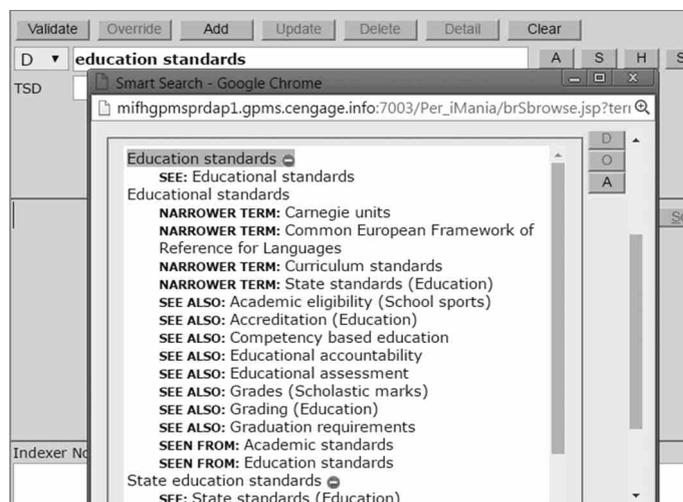


Figure 2 Screenshot from Gale database indexing software

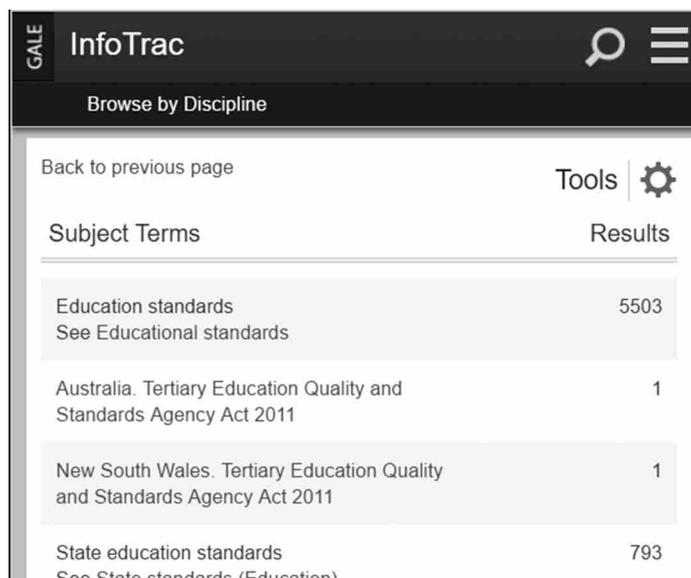


Figure 3 Screenshot from Gale InfoTrac database product

Multiple entry points in back-of-the-book indexes and in thesauri follow the same basic approach of utilizing synonyms, near synonyms, slang or jargon, abbreviations or acronyms and spelled out forms, former and current names, pseudonyms, phrase variations and inversions, and so on.

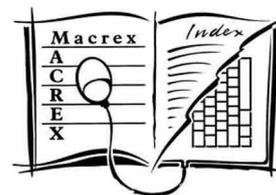
Back-of-the-book indexers have two methods to choose from for creating multiple entry points, double posts and *see* references, whereas thesaurus creators have the single method of creating NPTs for PTs. So, while back-of-the-book indexers have two levels of decisions to make, the majority of their multiple entry posts will likely be straightforward double posts. Thesaurus creators, on the other hand, need to decide in every case which of the variant terms will be the PT.

Back-of-the-book indexers create multiple entry points to serve the needs of alphabetical browsing. Thesaurus creators create NPTs that will facilitate searching the thesaurus first and browsing it second.

A note: multiple entry points in SKOS vocabularies

Software for the development and management of controlled vocabularies may support the thesaurus standard (ISO 25964 Part 1) or it may support other vocabulary standards, or both. An increasingly common standard for online vocabularies is SKOS, which stands for Simply Knowledge Organisation System. It's a recommendation by the World Wide Web Consortium for the application of the RDF (Resource Description Framework) interoperability format. While the end result of an ISO-compliant thesaurus or SKOS vocabulary is essentially the same, the underlying structural model is different. In SKOS vocabularies, instead of PTs and NPTs linked by an equivalence relationship, there are 'concepts' which each have 'preferred labels' (just one in each language), and any number of 'alternative labels.' Alternative labels are treated as a concept's attributes and are not considered as distinct 'terms.' These

MACREX



A

Enthusiastically used by professional indexers worldwide

Version 9 available now

With a host of new features adding to the best qualities of the earlier versions

C

A Few Features of MACREX

- View the index in real time as it is compiled
- Auto-complete option ensures accuracy of added headings and subheadings
- Sophisticated error prevention for entries and page numbers
- Instant selective display of specific words or page ranges to aid editing and checking
- Ability to propagate a changed heading including the option to duplicate headings from one topic and place them under another
- Support for Unicode character set including Cyrillic and Arabic (under Windows 7 or later)
- Exhaustive volume number/page number features – a boon for multivolume works or journal indexes

R

E



Contacts:
 Head office: dcalvert@macrex.com
 North America only
macrexna@gmail.com
 Australasia only
max.mcmaster@masterindexing.com

X

www.macrex.com

differences do not have much impact on the intellectual work of the thesaurus or controlled vocabulary editor, but familiarity with the different nomenclature is important if using different software.

Acknowledgment

This article is based on a presentation of the same name given by Heather Hedden and John Magee at the joint ASI/ISC conference in Chicago on 17 June 2016.

Reference

Hedden, H. (2012) 'Back of the book indexing and thesaurus creation: a comparison.' *KeyWords* 20(2), 49–51.

Heather Hedden is a senior vocabulary editor at Cengage Learning, where she edits the subject thesaurus for Gale research database products and develops discipline taxonomies for Cengage's learning products. Previously she did freelance back-of-the-book indexing and taxonomy consulting through Hedden Information Management. Heather is also author of The accidental taxonomist, Indexing specialties: web sites, and chapters in Indexing names and Index it right, Volume 2. She is a past president of the ASI New England Chapter, a past manager of the ASI Web & Electronic Indexing SIG and the Taxonomies & Controlled Vocabularies SIG, and is currently a member of the ASI national board. Email: heather@hedden.net

Ebook indexing update

Glenda Browne

The future of ebook indexing will depend on specific developments relating to ebooks as well as trends in publishing overall. Book metadata is a related and important field. Change is slow. This article reports on software support for ebook indexing along with developments in the EPUB standard and in publishing in general.

Introduction

Some years ago there was a major fire at Warragamba, NSW, Australia, about seven days burning away from our home in the Blue Mountains. The situation was not urgent, but as there was nothing but bush separating us from the fire, we believed it would arrive at our home eventually.¹ The situation with ebook indexing is similar. Change is inevitable, but not rapid, so we are in limbo waiting for developments that we expect, but don't control.

This article provides an update on some of the changes that have occurred specifically in ebook indexing and also in the general publishing environment. It assumes knowledge about ebook indexing and the EPUB standard and its Indexes specification. For background information see Browne (2013), Combs and Ream (2014) and the Matrix documents (linked to from www.asindexing.org/about-indexing/digital-trends-task-force/).

Publishing environment

As indexers, when we think of ebook indexes, we are often focused on the digital equivalent of a print book: that is, a stand-alone ebook for a single work, in EPUB or Kindle format. We have also considered the role of indexes for multiple individual ebooks, as seen with the ASI ebook mashup² and in the EPUB 3 Indexes Charter.³

Beyond this, some digital 'book' content is accessed in HTML format through subscription platforms, where books are just one of a number of available resource types along with periodicals, images and databases.

For example, at the ASI/ISC conference, Ted Trautmann from Wolters Kluwer said that they had 'de-emphasized' traditional publication-centric indexes in favour of more comprehensive ways to aid online search. He also noted that they had tried publishing ebooks, but they were a dud for their market as they didn't fit into the attorney workflow. Users did not welcome another channel in which to read content, and ebooks didn't allow for cross-publication search.

A similar theme emerged at the Law via the Internet conference 2015, where Stephanie Booker spoke about the replacement of the print version of the Northern Territory Law Handbook with a web-based version.⁴ Other speakers considered HTML to be the easiest standard to work with and said that EPUB was used primarily to access distribution channels. HTML is considered to be acceptable for a university textbook as students are a captive audience.

Because of the importance of HTML, the proposed merger of the IDPF with the W3C (that is, of EPUB and web standards) could be of great importance to indexers. EPUB is discussed further below.

When a large collection of content is made available through a centralized platform, navigation may be through a standardized taxonomy or a controlled vocabulary as well as, or instead of, through indexes handcrafted for one-off publications. The principles of collection indexing may be applicable alongside those of book indexing.

Automated indexing

While to my knowledge nobody has been able to create software that effectively indexes a stand-alone book, auto-