

As the name implies a glindex is a blending of a glossary and index. Figure 4 is an example from one of the recent conversions. Other than lengthier main headings than those usually found in an index, the text is straightforward to represent in records in your indexing software.

## Summary

Creating these types of endmatter are two more services you may wish to advertise as work you will perform. They are something different to work on, giving you a break from traditional indexes. Not only will you be an indexer but a glossarist as well!

## Notes

- 1 See Zafran (2001): 51.
- 2 ‘... the “glindex”, a combined glossary and index that is sometimes used as a way of grouping end matter and/or saving space’ (Browne and Jerney, 2007: 38).

- 3 Most often indexes but also tables of contents, bibliographies (for author indexes), tables of cases and other legal references.

## References

Browne, G. and Jerney, J. (2007) *The indexing companion*. Port Melbourne, Vic.: Cambridge University Press.

Zafran, E. (2001) ‘On the table: the problems and challenges of legal tabling,’ in *Indexing specialties: Law*. Medford, N.J.: ASI.

*David K. Ream is Leverage Technologies’ chief consultant for publishers. He has an MS degree in computer science from Case Western Reserve University. He has spent over 40 years working with publishers in the areas of typesetting design and production, database creation, editorial systems, and electronic publication design and production. He has performed thousands of index and related conversions in his career. Email: daveream@levtechinc.com*

# Creating indexes and thesauri: similarities and differences

*Heather Hedden*

*Heather Hedden explores the similarities and differences between back-of-the-book indexing and thesaurus creation, arguing that the similarities are such that a person with interests and skills in one area might consider work in the other, but warning that the differences are such that anyone transitioning from one activity to the other needs to study the other field first and not make unthought-through assumptions.*

A thesaurus can be considered a tool to help indexers index in a consistent manner in cases where the volume of indexable content is very large, and especially when there are multiple indexers working on the same index. However, a thesaurus is usually more than just an aid for the indexer; it also ultimately serves as the index itself. All the terms of the thesaurus, as long as they have locators/links to content, are the terms that are in the displayed index. The structure of the thesaurus also becomes the structure of the resulting index. As such, the creation of the thesaurus has many similarities to the creation of a back-of-the-book index. Thus, those who index books already have the skills for creating thesauri.

## Thesaurus basics

A thesaurus (as a knowledge organization system, not a writer’s type of dictionary) is a specific kind of ‘controlled vocabulary.’ A controlled vocabulary, as its name implies, is a controlled list of terms. Only terms on the list may be used for indexing, so a controlled vocabulary (also called a taxonomy) ensures consistent indexing. The same concept

will then always be indexed with the same term. A controlled vocabulary is particularly useful when indexing a series of volumes or issues of periodicals over time, as it is all too easy to forget what term has been previously used for the same concept. It is also useful when multiple indexers are indexing the same project, because each indexer is likely otherwise to come up with different terms for the same concept. Unless the controlled vocabulary is small enough to fit onto a single page, it typically includes a form of *see* cross-reference terms within it to help the indexer(s) find the approved index term. In contrast to a thesaurus, a controlled vocabulary does not necessarily have a relationship structure between its terms.

A thesaurus is a structured kind of controlled vocabulary, where, in addition to having the *see* references, approved index terms are also related to each other. Terms can be related to each other in one of two ways: first, hierarchically as broader and narrower (also known as generic/specific, parent child, superordinate and subordinate), or second, associatively as merely related, like *see also* terms. A freelance indexer needing to create their own controlled vocabulary for a large project will probably find a thesaurus-type of controlled vocabulary has more structured detail

than necessary, but a thesaurus is quite common to support the indexing of periodical literature, and is essential for operations that index multiple periodicals as part of the same index (e.g. PsychINFO, ERIC, Medline, and the databases of library vendors Gale, EBSCO, and ProQuest). In all these cases, the same thesaurus is used by both the indexers as an aide in consistent indexing and the end-user information seekers, as the resulting periodical ‘index.’

As with any index, terms in a thesaurus are listed alphabetically. Under each term is then listed any relationships or information, including *broader term* (BT), *narrower term* (NT), the *see* reference terms (often called ‘nonpreferred terms’ or ‘nondescriptors’) that it is used from (UF), and possibly even a scope note (SN). The nonpreferred terms (*see* reference terms) are interspersed alphabetically in the thesaurus alphabetical display, just like the *see* references with an index, and they are followed by *see* or *USE* and the preferred term that the thesaurus user should use. Figure 1, showing the USDA National Agriculture Library Thesaurus, is an example of an excerpt of the alphabetical display of a thesaurus and a selected term’s details.

### Points of comparison

Back-of-the-book indexing and thesaurus creation are similar, because back-of-the-book indexing is not just ‘indexing’ it is also ‘index creation,’ and the design and creation of an index is not too different from the design and creation of a thesaurus. Both activities center on creating index terms or entries for concepts, arranging

and structuring them, and dealing with their variants/synonymous terms (multiple entry points). This is different from most periodical indexing, which usually relies on the use of pre-existing terms in a controlled vocabulary or thesaurus.

There is some difference in the terminology. What are called *entries* in indexes are called *terms* in thesauri. What are called *cross-references* in indexes are called *relationships* between terms in thesauri. What are called *locators* in indexes are referred to as *references* or *links* in thesauri.

There follows a comparison of the methods to create common features for both for both back-of-the-book indexing and thesaurus creation in four areas:

- style of entries or terms
- multiple points of entry
- hierarchical structure
- indication of related concepts.

#### Style of entries or terms

For both index main entries and thesaurus terms, the editorial style is similar. They should be nouns or noun phrases. They should be concise for easy scanning, yet at the same time clear and unambiguous. It’s easier to skim a list of terms that are one to three words in length, than those that are four or five words long, although proper nouns may need to be longer. Countable nouns should generally be in the plural. Entries or terms comprise both generic concepts and names/proper nouns, and these can be integrated into a single index or thesaurus, or separated out. Capitalization style may vary, and is set by the publisher, but should follow a consistent internal style. However, it is more common to see lower-case entries (when not proper nouns) in book indexes, and it is perhaps more common to have title case (capitalization only for the first word in a title) for terms in thesauri.

There are also differences in style. A book index is designed for alphabetical lookup only. Therefore, the first word of an entry should be a ‘keyword’ that is likely to be looked up. Consequently, phrase terms are often inverted (*noun, adjective*) so that the important word comes first. Thesauri, on the other hand, are designed to be displayed and looked up alphabetically or hierarchically. Furthermore, thesauri are usually accessed electronically, not in print, which means that terms can be searched for and not just browsed. Therefore, the standard practice is *not* to invert the words within preferred terms (although nonpreferred terms may be inverted).

The real difference lies in comparing index subentries with thesaurus terms. Subentries can additionally be prepositional phrases, adjectives, phrases with conjunctions, or other phrases that have no similarity to ‘terms.’ Some of the larger published periodical databases use subentry-like terms, also called subdivisions. They are typically in a list separate from the thesaurus

The screenshot shows a web interface for the USDA National Agriculture Library Thesaurus. At the top, there is a navigation bar with letters A through Z. Below it, a search results window displays a list of terms, with the majority in italics. A 'Next Page' link is visible. A second window, titled 'Jacaranda', provides detailed information for that term, including its RDF/XML format, persistent URI, source, broader term, and narrower terms.

**Search Results: Click on term to display record**

Next Page

- *jaagsiekte*
- Jaagsiekte sheep retrovirus
- Jabiru
- Jabiru mycteria
- *jaboticaba*
- Jacaranda
- Jacaranda copaia
- Jacaranda mimosifolia
- *Jacaranda ovalifolia*
- Jacaratia
- *Jacaratia dodecaphylla*
- Jacaratia dolichaula
- Jacaratia mexicana
- Jacaratia spinosa
- *jacare*
- jaceosidin
- *jack beans*
- *Jack in the bush*
- *jack in the pulpit*

**Jacaranda**

**RDF/XML Format:**  
http://lod.nal.usda.gov/nalt/17129.rdf

**Persistent URI:**  
http://lod.nal.usda.gov/nalt/17129

**Source**  
Germplasm Resources Information Network

**Broader Term**  
Bignoniaceae

**Narrower Term**  
Jacaranda copaia  
Jacaranda mimosifolia

**Figure 1** In the alphabetical display of a thesaurus, cross-reference terms are in italics, in this case the majority. When the user clicks on a term in the thesaurus, such as Jacaranda, the details of that term and its relationships to other terms are displayed in a second window. Source: <https://agclass.nal.usda.gov/>

of terms, which may be referred to as ‘main headings’, like main entries in an index. The use of subdivision supports what is called second-level indexing or structured indexing in periodical/database indexing. It serves a similar role to index subentries, with a controlled vocabulary for the subentries in addition to the thesaurus for the main headings. In any case, the terms in such a controlled vocabulary of subdivisions are still nouns or noun phrases. Examples of controlled vocabulary subdivisions are: Analysis, Cases, Description and travel, Economic aspects, Health aspects, Investigations, and Public opinion. Geographic place names can also serve as subdivisions.

In many cases of thesaurus use, however, subdivisions for second-level indexing are not available because of the complexity of implementing them in a dynamic online environment. Thus, in the absence of subdivisions, thesaurus terms need to be more specific than corresponding index entries, which have subentries available for further specification. For example, where in a back of the book index, a main entry could have subentries for specific aspects, such as:

education  
 administration  
 adult and continuing

in a thesaurus, we would have terms in the thesaurus for each of the following:

education administration  
 education, adult and continuing *USE* adult and continuing education

### Multiple points of entry

The indexer or thesaurus creator (sometimes called ‘taxonomist’ for want of a one-word name) creates multiple terms with the same meaning for a single concept as multiple entry points, to direct the users, who are prone to use various terms that mean the same thing, to the same location in the text. The types of variants are the same for both book indexes and thesauri. They can be synonyms, near synonyms, phrase

inversions (such as digital photography and photography, digital), abbreviations or acronyms and their spelled-out forms, slang versus formal language, and so on.

In back-of-the-book indexing, there are two methods for creating multiple entry points:

- Double-posts, whereby two or more index entries with the same meaning are added to the index, with none designated as preferable to the other, and all have the same locators pointing to the same points of text. (This is the preferred method when the entries do not have subentries.)
- *See* cross-references, whereby additional index entries of the same meaning are added to the index, but they point to a single favored entry, which is the only one with the locators. (This is the preferred method when the entry has subentries.)

In thesaurus construction, on the other hand, there is basically just one method for creating multiple entry points, which is akin to the index *see* reference. There is nothing like double-posts (different terms for the same concept having equal standing) in a thesaurus. The *see* reference terms are called ‘nonpreferred’ terms, and they each direct the user by pointing (whether by the word *see* or *use* or simply by a hyperlink) to the ‘preferred’ term that is used in indexing. Furthermore, while a *see* reference in an index only points in one direction, in a thesaurus this is considered a bidirectional reciprocal relationship. So, in addition to ‘nonpreferred term *Use* preferred term,’ we also have ‘preferred term *Used from* nonpreferred term.’

In a back-of-the-book index, double posts or cross-references are generally not created when the entries would lie close to each other alphabetically, such as starting with the same two or three letters. This is because the index is browsed and the user would easily see entries that are close to each other. (There are exceptions, such as when the indexer believes there is a need to reassure the user that the choice of a preferred entry for a *see* reference was deliberate.) A thesaurus, by contrast, can be searched in addition to (and sometimes instead of) being browsed. Therefore, it

Table 1

#### Book indexes: multiple points of entry

Two different methods:

- Double-posts: both or all of equivalent-meaning entry terms have equal standing.
- *See* references: point the user from an entry term *not* used in the index to one that *is* used in the index.

Indexer decisions:

- When to create double-posts versus *see* references (usually based on presence of subentries).
- If using a *see* reference, then what the preferred term will be.

*See* references are one-directional: *see* (no corresponding ‘seen from’).

Avoid creating cross-references that fall close to each other alphabetically.

#### Thesauri: multiple points of entry

One method only: (nothing like double-posts).

Nonpreferred terms /equivalency relationship: *Use*.

Points the user from an entry term *not* used in the thesaurus to one that *is* used in the thesaurus.

Thesaurus editor decisions:

In all cases, what the preferred term will be.

Equivalency relationships are bidirectional and reciprocal: *use* and *used from* (USE/UF).

Do create nonpreferred terms that would fall close to each other alphabetically.

cannot be assumed that the user will always see the terms that are near each other when arranged alphabetically. The thesaurus creator should include appropriate nonpreferred terms that begin with the same letters or even the same words as their corresponding preferred term.

Table 1 summarizes the differences in how multiple points of entry are treated in a book index and in a thesaurus.

Another feature of a thesaurus, polyhierarchy, while not exactly another ‘point of entry,’ serves as an additional way for users to navigate to the specific term and content they want. This is explained in more detail in the next section on hierarchical structure.

### **Hierarchical structure**

Both indexes and thesauri have some form of hierarchical structure between terms/entries to guide users from more generic concepts to more precise topics. If a term has too many locators/references, it needs to be broken out by creating multiple corresponding subordinate concept entries/terms. In indexes these are called subentries, and in thesauri these are called narrower terms (or sometimes ‘subtopics’). In both situations, indexing may be done using the main entry/broader term for a general treatment of the subject and also with its subentries/narrower terms for specific aspects or sub-topics.

Despite their similarities, though, subentries in an index and narrower terms or subtopics in a thesaurus are *not* the same and do not function identically. Index subentries serve as subdivisions or more specific aspects of the main entry. The options for subentries are limitless. For example, the main entry of a country name, such as Russia, could have subentries of periods of history, relations with other countries, politics, economics, ethnic groups, religions, etc. By contrast, narrower terms in a thesaurus are limited to, one, specific examples or instances of a more generic term, and two, parts of a whole. In the case of a term of Russia, the only narrower terms would be geographic subentries, such as its subregions, republics (states) or cities, for example, Siberia or Moscow.

Table 2 lists the differences subentries in a book index and narrower terms in a thesaurus.

Just the fact that both an index subentry displays indented

under a main entry and a thesaurus narrower term displays indented under a broader term does not mean they are the same thing. As an *aspect* of a main entry, index subentries must be related to the main entry and can be mere adjectives or prepositional phrases. Narrower terms in a thesaurus, on the other hand, are standard terms themselves which can stand on their own, so must be nouns or noun phrases, as all thesaurus terms are. In fact, ‘narrower term’ is only relative to another term; ‘narrower term’ and its reciprocal ‘broader term’ designate *relationships* between terms and not types of terms. The fact that a main entry and a subentry in an index can be interchanged or ‘flipped’ is indicative of the fact that it is their mere *pairing* that is significant rather than any intrinsic hierarchical (parent-child) relationship. You cannot switch the position of a narrower term and broader term in a thesaurus.

Index subentries and thesaurus narrower terms lend hierarchical structure that is evident when the user skims the entire index or thesaurus, but the hierarchy is usually deeper in a thesaurus. In an index, a single sublevel of subentries is most common for trade books, up to two levels of subentries is common for scholarly, and only specialized subjects, such as law, would have more levels. In non-specialized fields, indexers avoid creating deep levels of sub-entries. In a thesaurus, however, hierarchical levels to the depth of seven are not uncommon, nor are they a problem, and it may not even be necessary to keep track of the levels.

What is especially interesting about a thesaurus structure is that it can be generated in different displays: hierarchy for each top-level term (as is done in an index), hierarchy for *every* term listed alphabetically, or the immediate narrower term and broader term indicated for a selected term but not the full hierarchy. Finally, a thesaurus may even have terms with such broad meaning, such as *Countries*, that they are not even used for indexing, but just for grouping narrower terms. All terms in an index, by contrast, are used for indexing.

As mentioned previously, in some cases in a thesaurus, a specific narrower term may have more than one broader term, a feature called ‘polyhierarchy.’ This helps users who are browsing the thesaurus hierarchically and may start from a different broader term. The feature of polyhierarchy is *not* the same as the reuse of the same subentries under different

**Table 2**

#### **Book indexes: subentries**

As subdivisions:  
 specific aspects of the main entry  
 any additional concept in combination with the main entry.

Must be related to main entry.

Can be prepositional phrases, gerunds, adjectives, etc.

‘Flips’ of main entry/subentry may have same meaning.

Hierarchy is usually 2 or 3 levels, and only in specialized fields more.

Indicated by indentation or run-in following colon and semicolons.

Narrower concepts may be subentries or other main entries.

#### **Thesauri: narrower terms**

Terms that are:  
 specific kinds or members of a class  
 named instances of a generic term  
 parts of a whole.

Must stand on their own as terms.

Must be nouns or noun-phrases, just like main heading terms.

Broader terms and narrower terms cannot be ‘flipped’.

Hierarchy is deep, usually 3–7 levels, and often more.

Indicated by reciprocal hierarchical relationships of broader term/narrower term (BT/NT); often displayed by indentation.

Narrower concepts *must* be assigned NT relationships.



main entries in an index. A subentry has a specific meaning in relationship to the main entry it falls under and restricts. A thesaurus term, even as a narrower term to another term, is completely independent in meaning from the narrower term. The same term that appears under more than one broader term has the exact same meaning and thus points to the same content in both cases.

### Indication of related concepts

A third feature of both book indexes and thesauri is to have a method to indicate *related* topics of possible interest to the user. In an index, the designation is *see also*, and in a thesaurus it is a relationship called *related term* (RT). They function in much the same way, and it tends to be a judgment call of the indexer or thesaurus creator as to when creating such a reference would be helpful without having so many that they get in the way of the ease of use of the index/thesaurus.

The situations in which the related term relationship can be created in a thesaurus, however, are more limited than when the *see also* reference can be created in an index. As mentioned previously, there are strict rules about when the hierarchical (narrower term/broader term) relationship can be created in a thesaurus, as the narrower term must be a specific type or instance of a broader term or an integral part of a whole broader term. By extension, if two terms in a thesaurus have a broader–narrower relationship, then they *must* be arranged in such a hierarchy and consequently cannot be related terms. For example, in a book index, you could have cancer *see also* melanoma (a kind of cancer). In a thesaurus, however, your only option is to have cancer NT (narrower term) melanoma. Cancer RT (related term) melanoma would be incorrect, no matter if it was in addition to or instead of the other, hierarchical relationship.

Another difference between *see also* references in an index and related terms in a thesaurus, is that the *see also* reference is not necessarily always reciprocal, whereas the related term always is. A common reason to have a *see also* reference point in just one direction is that the locators for both entries overlap, and if all of the locators of one entry are included among the locators for the second entry, which has additional locators, then the *see also* reference only needs to direct from the entry that has fewer locators to the entry that has those same locators and more. In a thesaurus where related term relationships are used, it is mandatory to have the relationship at both terms pointing in both directions, as the links to content are not known.

**Table 3**

Book indexes: <i>see also</i>	Thesauri: related term (RT)
<i>See also</i> is often two-way, indicated at both pairs of terms, but not necessarily always.	RT is <i>always</i> bidirectional reciprocal, indicated at both pairs of terms.
Don't create <i>see also</i> cross-references between entries that fall close to each other alphabetically.	<i>Do</i> create RT relationships between terms that would fall close to each other alphabetically, because a thesaurus is online and might be searched.
If pointing to a subentry, the corresponding main entry needs to be named: <i>see also under</i> [main entry]	RT relationships may point to terms at any level in the hierarchy without distinction.
May refer to a group of terms at once: <i>see also specific...</i> [class of terms]	Must refer to an individual term only.

Table 3 summarizes the differences in how indication of related concepts are treated in a book index and in a thesaurus.

## Skills and processes comparison

The skills used by book indexers and by thesaurus creators are similar. Both involve content analysis and term creation. Indexing requires perhaps greater specific content analysis, since the indexer must read the entire text and index to specific passages, to identify topics and names that are important, whereas thesaurus creators only look at sample texts. Thesaurus creators, on the other hand, must do analysis that is more broad-based. They often need to examine other existing thesauri on the subject. Thesaurus construction also requires more study of the 'audience' or users, who will use the thesaurus, and consider how it will be used, something that book indexers almost never do.

Thesaurus creation *may* also demand more subject-specific knowledge than does book indexing. Book indexers usually require subject expertise only for technical and scholarly subjects. Generally, book indexers can let the book be their guide for terminology and structure. The thesaurus creator, on the other hand, does not have a single work to rely on, so must consult multiple sources. These multiple sources may cover the subject area unevenly or even conflict with each other, so that is where subject-matter knowledge can be helpful.

The process of book indexing and the process of thesaurus creation may be similar depending on the approach taken. Depending on the book and the indexer, back-of-the-book indexers may take one of either two approaches: after a cursory review of the table of contents, read and index page by page from the beginning, or skim the book and write down common themes, topics, and names, as likely index terms, considering even which will be broad topics with many subentries, and then go back and begin indexing. Thesaurus creation is similar to the second book indexing approach, but then additional work is put into refining the terms, and the indexing stage is not necessarily even performed by the thesaurus creator.

In conclusion, due to the number of similarities between back-of-the-book indexing and thesaurus creation, a person with interests and skills in one area might consider work in the other area. Because of the number of differences between back-of-the-book indexing and thesaurus creation, though, anyone transitioning from one activity to the

other needs to study the other field first and not make assumptions about any apparent similarities.

## Acknowledgment

This article is an expanded version of 'Back of the book indexing and thesaurus creation: a comparison,' which appeared in *Key Words* 20(2), April–June, 2012.

Heather Hedden is a senior vocabulary editor at Gale/Cengage Learning. Previously she did freelance back-of-the-book indexing and taxonomy consulting through Hedden Information Management, and she continues to offer online training in taxonomy and thesaurus creation. Heather is also author of *The accidental taxonomist, 2nd edition*, *Indexing specialties: web sites, and chapters in Indexing names and Index it right, Volume 2*. She is currently a member of the ASI board.

Email: heather@hedden.net

---

# Awards roundup

---

## 2017 ISC/SCI Ewart-Daveluy Indexing Award

The 2017 Ewart-Daveluy Indexing Award was presented to Judy Dunlop at the awards banquet of the ISC/SCI in Montréal on 2 June 2017. The award honours Judy's indexing of *One child reading: my auto-bibliography* by Margaret Mackey, published by the University of Alberta Press.

The ISC/SCI Ewart-Daveluy Indexing Award is presented each year to an individual who has created an index that demonstrates outstanding expertise, the ability to analyze complex text and the ability to design an index that significantly enhances reader use of the text.

*One child reading* is a unique and fascinating examination of reading and literacy development. Author Margaret Mackey revisits the things she read, viewed, listened to and wrote as she grew up in the 1950s and 1960s in Newfoundland. Her reading included school texts, knitting patterns, musical scores, games, church bulletins, family magazines and hundreds of books. In *One child reading*, Mackey weaves her growing literacy and social consciousness with the books of her childhood and youth and the history of the time and place.

The indexer's challenge was to combine in one comprehensive, cohesive index the three aspects of the book: the author's memories, the theoretical discussion and the analysis of specific texts. In addition to standard terminology to cover the biographical details, the indexer had to incorporate the sometimes unique terms the author created for the textual criticism and social analysis. As one judge noted, 'This is an indexer who's not afraid to directly express the language of the text ... and also to use some ingenuity in handling sections like the distinction between a subject in theory vs. its relation to the author's life.' Said another, 'The index is wonderfully fulsome and narrative, and brief and concise – quite a feat.' 'There are some lovely discoverables in this index,' said the third. The author herself was 'awestruck' by the 'sensitivity of [the indexer's] reading.'

In responding to the Award, Judy spoke as follows.

The author, Margaret Mackey, is a library science professor at the University of Alberta, who specializes in children's and adolescent literature and literacy. Her book, *One child reading: my auto-bibliography*, is about her childhood in St. John's, Newfoundland. She examines her experiences reading the word/the world from birth to early teens.

Margaret's 500-page book combines three major threads. First, it is a memoir of a child's life in St. John's, including her school days, reading memories, and family experiences. It is also a 'bibliography' of all of the texts she 'read,' including *Dick and Jane* basic readers, *Roy Rogers* television programs, series books, *Jack and Jill* children's magazine, comics, maps, games, sheet music, and even knitting patterns. Second, it is a critical approach to those texts, including commentary on topics such as stereotypes and gender issues. Finally, it is an application and development of reading theory applied to her own literacy experiences, including her memories and her time spent as an adult re-reading childhood texts. Some of the main entries for this area were the most challenging. For example, 'the Murk' is the term she uses for the period from birth to her earliest memory. 'Folding and unfolding' refers to the moment of change in the brain that occurs while reading, including the reader,



Awarding the Ewart-Daveluy Indexing Award: left to right, Mary Newberry, Judy Dunlop, Noeline Bridge and Heather Ebbs