Cataloguers and indexers (except for back-of-the book indexers) use existing classification systems, subject headings, controlled vocabularies, when they assign the classifications, codes, headings, or terms. Cataloguers and indexers may wish to take their skills to the next level to actually design and create controlled vocabularies. Libraries and book publishers benefit, however, by using a standard classification system, subject heading scheme, or set of codes, so there is no need nor desire for new classification systems or knowledge organisation systems for libraries or book publishers. However, there is a much larger world of information management. Companies of all kinds and sizes, nonprofit organizations, and government agencies all have growing collections of digital content for both internal and public access, and customised controlled vocabularies make this content easier to find.

The documents, media (images, audio, video), and other content that organizations own is increasingly in digital form only, and as such, users can access the content with a search engine or a search feature on an organization's website, intranet, content management system, or records management system. However, search alone has weaknesses. The search engine looks for words or phrases in text, not concepts. So, search can miss content described by different synonyms when a user enters a word in the search box, but the text uses a different synonyms for the same concept. On the other hand, a search engine might retrieve document or pages where the searched word is mentioned but where it has a different meaning (as a homograph), or where the word is negated, or where it is a mere passing mention and not the main topic, so false results will be retrieved. Search engines are good for the World Wide Web, because users only want some relevant pages out of the millions. Within an organisation or a single website, however, there may be only one or two pages that the user needs, so the demands on precision and recall in search are much higher. This is where customised controlled vocabularies can help, by being indexed to the content and accessible to the user through browsing and/or searching.

Controlled vocabulary types

Controlled vocabularies comprise term lists, synonym rings, name authority files, taxonomies, and thesauri. At the broader level, controlled vocabularies are kinds of knowledge organisation systems, which also include categorisation systems, classification systems, dictionaries, gazetteers, glossaries, ontologies, semantic networks, subject heading schemes, and terminologies.

Different kinds of controlled vocabularies are suitable for different situations. Terms lists may support individual metadata properties. Synonym rings support search when it's not needed or desired to display terms to the users. Authority files are for named entities, such as person names or organisation names. Large collection of research articles or reports on varied specific subjects are most suited for a thesaurus. Enterprise content, whether on an intranet, content management system, or public website is best served with a taxonomy.

Taxonomies are categorisations or hierarchical arrangement of terms, where terms are linked to each other with broader-term/narrower-term relationships. Increasingly, though, we are seeing faceted taxonomies, with a facet to categorise each aspect or dimension types of terms (such as document types, regions, activities, product types, departments, etc.). So, each facet is similar to a 2-level hierarchy. It is also possible to have a small hierarchy within a facet.

Cataloguers are familiar with classification systems and subject heading schemes, but these are not the same as taxonomies. Classification systems use codes, and taxonomies generally don't. Classification systems provide comprehensive and balanced coverage of a domain, and taxonomies include only the terms needed to index a set of content.

Classification systems are designed to be browsed hierarchically from the top down, but taxonomies can be browsed, searched, or not even fully displayed. Classification systems have a single hierarchy, whereas taxonomies permit a term to appear repeated under more than one broader term (called polyhierarchy). Classification systems provide for limited expansion within the structure, and also include "other" or "not elsewhere classified categories, whereas taxonomies can grow and adapt without limits, and do not have terms for miscellaneous. Classification systems don't usually have alternative labels (nonpreferred terms/synonyms), and taxonomies often do.

Subject heading schemes and taxonomies also differ, because subject headings are not usually arranged in hierarchies. Furthermore, subject headings schemes are characterised by also having a second-level (and sometimes a third level) of subdivisions, such as for different aspects (political aspects, economic aspects, health aspects, etc.), geographic places, dates, or other subtopics, which are similar to the subentries appearing under main entries in back-of-the-book indexes.

Designing taxonomies

Taxonomies tend to be custom-built for a specific implementation of a certain set of content, especially when it comes to an individual organisation's products, services, functions, and activities, how they are particularly organised, and how much content exists in each area. Thus, one of the first steps in creating a taxonomy is surveying the content that will be indexed with the taxonomy terms to identify the main kinds and topics. Identifying potential term from analyzing sample content is an activity is similar to indexing, but without a controlled vocabulary, because it does not exist yet!

Since taxonomies are built for specific implementations, they should also be designed for the particular sets of users. This could be the employees of an organisation, members of a membership organisation, subscribers or customers of specialised content, students or researchers, or certain market segments of the public. Therefore, user input should be considered when building taxonomies, whether through questionnaires, feedback forms, or even search log reports. While not feedback, this will show what people have been searching for. It's of course easier to gain input from internal users than from external users. For externally used taxonomies, useful input can still be gathered from a variety of internal stakeholders, including those involved in customer support, user experience design, search, etc.

Another kind of user of a controlled vocabulary are the indexers. People who have done or will do the indexing should be interviewed to gain insights on their needs and challenges in indexing the content. Furthermore, after a new taxonomy is created it should also be tested, and at a minimum it should be tested for indexing of sample content by the indexers.

Designing and creating taxonomies is a collaborative effort that brings in experts from different specialties, such as content management and strategy, information architecture, user experience, website design, search engine technology, software systems implementation, etc., and perhaps even subject matter experts. Ideally, however, there is a trained taxonomist leading the taxonomy project and undertaking the actual taxonomy creation. Taxonomy/thesaurus management software can be useful for this task. The final implementation of a taxonomy usually requires the assistance of someone working in information technology.

Best practices and standards for taxonomies

Standards in general are of two kinds, (1) for design consistency and thus ease of use, and (2) technical specifications for interoperability and exchange. For taxonomies, both kinds of standards exist, but are applied to varying degrees. Best practices for controlled vocabulary design are described in the ISO standard, ISO 25964-1:2011 *Information and documentation. Thesauri and interoperability with other vocabularies. Thesauri for information retrieval*. It is also published as a British Standard (BS) with the same number and name.

As the name implies, it is focused on thesauri, rather than taxonomies, but it is the standard most relevant to taxonomies. It provides guidelines for the designation of concepts, format of terms, use of variants, hierarchical relationships, and display formats. Sections on associative relationships (related-terms), which are applicable to thesauri, can simply be ignored for taxonomies which don't have associative relationships. The principles of the ISO 25964:1 standard have been repeated in various publications and courses on taxonomy creation, so it may not be necessary to purchase the original standard.

Examples of best practices standards for taxonomies and thesauri described in ISO 25964:1 include

- unique labels for each unambiguous concept
- broader/narrower term relationships of the generic-specific, generic-instance, or whole-part types
- variant labels (synonyms/nonpreferred terms) that are equivalent or slightly narrower in meaning for the context, and no duplication of variants

There are several technical specifications for taxonomies and other knowledge organization systems, but increasingly adopted is the Simple Knowledge Organization System (SKOS) standard published by the World Wide Web Consortium (W3C). It is described as "a common data model for sharing and linking knowledge organization systems via the web" (www.w3.org/2001/sw/wiki/SKOS). A taxonomy does not have to be built with the SKOS model if it is not to be exchange, but most software dedicated to building and managing taxonomies and other knowledge organization systems now follows the SKOS standards.

Training on taxonomy creation

While taxonomy design comes closest to the fields of cataloging, classification, and thesaurus construction, which are taught in schools of library and information science, taxonomy design is usually not the topic of dedicated university courses. At most it would be covered in a single lesson. That is because taxonomy creation is more of an art, adaptable to specific business cases, than a science. Information professionals gain taxonomy skills primarily from on-the-job training and experience and through continuing education. The latter can be obtained from attending webinars, conferences, conference and corporate training workshops, online tutorials, and online courses. The following are some options for learning more about taxonomy creation:

- Taxonomy Boot Camp London conference, www.taxonomybootcamp.com/London
- SLA conference, www.sla.org/attend
- Hedden Information Management www.hedden-information.com/courses-workshops

Whether working on a single taxonomy project or in a position as a taxonomist, information professionals will likely find this as rewarding work.

References

Hedden, Heather. (2016) *The accidental taxonomist*. 2nd ed. Medford, NJ: Information Today Inc.

ISO 25964-1:2011 Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval, 2011, Geneva, Switzerland. (www.iso.org/standard/53657.html).

SKOS Simple Knowledge Organization System Reference
W3C Recommendation 18 August 2009 (www.w3.org/TR/skos-reference)