

Taxonomy Design Best Practices

Heather Hedden

Taxonomy Consultant

Project Performance Corporation

October 2011

Introduction: What Makes a Good Taxonomy?

Designing a taxonomy involves a careful balance of both general taxonomy best practices and the specific use cases involved. Since the purpose of a taxonomy is to help users, taxonomy design best practices are flexible. There are industry “guidelines” (ANSI/NISO), rather than “standards.” This white paper looks at the best practices guidelines for various aspects of taxonomy design, including overall structure, term inclusion and wording, hierarchy, facets other relationships, and variant terms.

Overall Structure

There is more to the initial design of a taxonomy than identifying the top-level, broadest terms. The overall structure needs to be determined. Is the taxonomy going to be primarily hierarchical, faceted, a combination of hierarchy and facets, a thesaurus, or an ontology? The choice depends on the content, the users, and the purpose for content retrieval. Choosing the wrong kind of taxonomy structure for your content or users can result in a less useful taxonomy.

A **hierarchical taxonomy** is suitable for a limited subject scope, such as geographic places, industries, detailed products or parts, academic disciplines, fields of science (e.g., plants and animals), or engineering, etc. There is no presumption on the search goals of the users.

A **faceted taxonomy** is most suitable when the content or documents being retrieved are somewhat uniform in that they share common attributes, as common fields for a database record. The decision to use facets considers the search goals of the users, in that they will likely need to combine several attributes/parameters into a single search.

A taxonomy may also **combine hierarchical and faceted structure**, but one format needs to be dominant. It could be primarily faceted, but there could be limited hierarchies within one or more of the facets, such as in a topical subject facet. Or it could be primarily a hierarchical taxonomy with the introduction of facets to narrow the choices further at the lowest level of the hierarchy, where there is more uniformity in the kinds of content.

A **traditional thesaurus-type of taxonomy** is a practical approach when the subject scope is too broad and varied to be neatly categorized into a few large hierarchies, and the content is too varied to be described in facets or most everything is a “subject” or “topic.” From the user perspective, a thesaurus is most suitable if there are trained or experienced manual indexers. Subject expert end-users also appreciate thesauri.

Finally there is the **ontology type of taxonomy**, which is structured around “classes” of terms (which may also serve as facets or hierarchies), and the classes have customized, semantic relationships between each other. An ontology also needs to be clearly defined, and the content needs some uniformity. Most importantly, an ontology is highly complex, so is only suitable when users already have some knowledge of the subject area and there is a use case for semantic relationships.



Term Inclusion and Wording

A taxonomy is a kind of “controlled vocabulary” after all, so it is important to give attention to the inclusion and wording of each of the terms. Rather than thinking first of words and phrases, the best approach is to start out considering the concepts that should be included. Concepts need to be within the subject area scope of the taxonomy, relevant to the content so that there is sufficient information within the content on the topic, sufficiently important to be likely looked up by users, and something that user would want or expect.

The most important issue to consider in the wording of terms is that the term name be both clear and concise and also likely to be looked up by the intended users. Standard language that is neither too technical nor too colloquial (or jargon) is recommended. Additional factors to consider in term wording include enforcing organizational or enterprise controlled vocabulary, conforming to academic or professional standards, applying consistency in style throughout the taxonomy, and following the wording of concepts within the documents and content indexed.

If the taxonomy will be browsed alphabetically, then the first of the term should be significant and likely to be looked up. Term inversions of the format noun-comma-adjective, however, should be avoided as they are less likely to be picked up in search engines. It is better to avoid prepositional phrases, but they may be suitable for standard expressions, such as Power of Attorney.

Hierarchy Design

For taxonomies that contain hierarchies (which are the majority) a well-designed hierarchy is crucial, because this is often the primary way users navigate through the taxonomy. A hierarchy comprises multiple terms, each linked together in hierarchical relationships. Therefore, a good hierarchical taxonomy is based on sound hierarchical relationships between individual terms. Hierarchical relationships should be based on one of three types: generic > specific, generic > instance (proper noun), or whole > part. Other kinds of merely “related” relationships should not be constructed as hierarchical.

Hierarchy depth and breadth are important design considerations. Users may lose interest if they have to click through too many levels of a deep hierarchy, and too many terms at the same level can take too much time to browse through. A depth of three levels is recommended, but if the subject area is intuitively hierarchical (such as product or part types), and/or the users have some familiarity with the subject (as repeat site visitors or internal organization members), then four or five levels may be acceptable.

The number of terms that are at the same level (i.e., share the same broader term) should be at least two, and in many hierarchically-based taxonomies, a minimum of three or more is considered a better design. The preferred maximum number of terms at the same level will depend on how varied the terms are. A miscellaneous list of terms (such as the top level menu of a web site navigation) should be limited to approximately seven for ease of browsing, but a list of similar items that can be listed alphabetically (such as types of jobs or occupations) could be much more, as long as they are all displayable in one screen view to the user without scrolling.

Assigning a term more than one broader term (a structure called “polyhierarchy”) can be useful in providing more than one navigational path to a complex concept, such as Car Audio under both Car Parts and Audio Components. Too many polyhierarchies, however, can compromise the hierarchical structure of a taxonomy.

Facet Design

In designing a faceted taxonomy, the biggest decision is determining its facets. The facets need to reflect both the nature of the content and the purpose, business process, or use case for end-user retrieval. Facets for an image database could be along the lines of: who, what, where, when, etc. Facets for

products could be: size, brand, price range, color, features, etc. Facets for an expert person database could be: job title, academic degree, subject specialization, skill, location, etc.

It is important, however, not to create more facets than needed. The more facets there are, the more difficult and time-consuming it becomes to index or tag the content, and those responsible for tagging may prefer to take shortcuts and not tag accurately or sufficiently. The presence of too many facets also becomes an obstacle to ease of searching for the end-user. Four to six facets is a good number to strive for, although expert users who frequently use the system may accept more.

Associative Relationships

While most commercial or business taxonomies tend to have only hierarchical relationships, the value of other, non-hierarchical relationships, also known as “associative” or “related term” relationships should not be overlooked. For example, Memory Cards in a Camera Accessories hierarchy may have a related link to Flash Drives in a Computer Accessories hierarchy.

Creating associative relationships, however, is not as intuitive as creating hierarchical relationships, and thus this practice is usually left to taxonomy experts. Related term links should be created sparingly, as needed, otherwise the taxonomy can become full of too many unnecessary inter-term links. The basic idea is that the two concepts should be related in all circumstances, not merely possibly related under certain conditions for some users. The key is to anticipate user needs without overburdening users with all possible options.

Variant Terms

Taxonomies are not merely browsed. As long as they are online they are also searched, whether through sophisticated search engines or simple database search tools. Therefore, to leverage the most use of a taxonomy, it is a good idea to create multiple equivalent variants for each taxonomy term. Sometimes called “synonyms,” the correct name for these alternative entry terms is non-preferred terms, because they do not have to be exact synonyms. For example, a term elementary education may have non-preferred terms of: elementary schools, primary education, grade school, and K-5 schools, even if these are not all exact equivalents.

An important factor in designating non-preferred terms is that they are equivalent within the context of the content repository. It is also okay to have narrower concepts serve as non-preferred terms, but broader concepts cannot be non-preferred terms. While a greater number of non-preferred terms aid users in finding the preferred term, too many can be counter-productive because a non-preferred term can usually point to only a single preferred term. When a term comprises multiple words, multiple preferred terms can be created by varying each word in turn. Common acronyms can also serve as non-preferred terms.

Conclusion

Taxonomies can be indispensable tools in helping users locate desired content within a repository. However, taxonomies are only useful when they are well designed. A well designed taxonomy needs to have a structure that is planned out to serve both the specific content and the target users. There is no “one size fits all.” The taxonomy design also needs to follow best practices for hierarchical relationships and facet design, as relevant, to be most useful. Finally, additional features such as associative relationships and non-preferred terms should be considered and not left out merely due to lack of time or in-house expertise. Taxonomy consultants at PPC are experts in all areas of taxonomy design and implementation, so they can create the right taxonomy for its purpose.



About Project Performance Corporation

Project Performance Corporation is part of the AEA group, a 1,200-person, multi-disciplinary team of information technology and knowledge management professionals, project management experts, scientists, technologists, and regulatory specialists. With more than 20 years of experience in information management and enterprise services implementation, we specialize in delivering effective, reliable answers to the most complex challenges. At the heart of it all are our people - innovative thinkers, well versed in business processes and drivers, with unequalled focus on delivering quality products and services.

About the Author

Ms. Heather Hedden is a taxonomy consultant with Project Performance Corporation. She also teaches continuing education workshops in taxonomy creation through Simmons College Graduate School of Library and Information Science. She has previously worked as a taxonomist at First Wind, Viziant Corporation, Earley & Associates, and Gale. Heather is the founder and current manager of the Taxonomies & Controlled Vocabularies SIG of ASI, a past manager of the Web Indexing SIG, and past president of the New England Chapter of ASI. She is the author of *Indexing Specialties: Web Sites*, a chapter in *Index It Right: Advice from the Experts*, Volume 2, and *The Accidental Taxonomist* (May 2010).

For more information contact:

Heather Hedden

Project Performance Corporation
1760 Old Meadow Rd., McLean, Virginia 22102
p: 703.748.7000
e: heather.hedden@ppc.com

