

# Facets for Enterprise Content

Faceted classification is relatively new and unlike taxonomies and so many other intellectual and philosophical ideas, it was not invented by Aristotle and some of the other ancient Greeks. It was first developed in the 1960's by S.R. Ranganathan and the Classification Research Group.

Neither group thought small, they both were trying to develop a classification system that would exhaustively classify everything in the universe and they defined facets and faceted classifications very precisely. Facets are orthogonal dimensions (mutually exclusive) and every object in the universe should have one value and only one value in each facet. In fact, an object was defined by the combination of its facet values in every dimension.

Facets are most often used for objects while taxonomies are used for topics or subjects found in documents. As we shall see later, this difference is changing, but explains why facets have been mostly associated with eCommerce sites selling things (wine, computers, hardware, etc.) rather than within the Enterprise where finding documents is the dominant interaction.

## Faceted Navigation and Browsing

Users often experience faceted navigation as a browse experience, which is one of the reasons that it is so successful. However, faceted navigation is actually the application of one or more filters to a set of objects or documents. Within each facet, the experience is the same as a browse, selecting from lower nodes to zero in on your target. However, unlike a traditional browse taxonomy where you follow a single trail through the hierarchy, individual facets are designed to work in conjunction with other facets in the system.

[Figure 1 – Flamenco – Marti Hearst]

In this example, someone might know the general period for the particular work, so they start with making a selection within periods, say 18<sup>th</sup> century. They also know that the work was created in Western Europe so they select that within the Location facet. With these two simple choices, they have reduced the search space or size of the search result set dramatically. They can then either zero in further by selecting the second half of the 18<sup>th</sup> century or picking a specific country within Western Europe.

With a few simple selections you can go from a catalog of all art works in the entire world to a few paintings by a particular artist. Contrast that with the traditional experience of browsing through a taxonomy.

If it is a true formal taxonomy organized by types of art, it will either have to stop at a level in which the total number of items was 1,000's or even millions (how many oil paintings are there in the world?), or introduce elements like region at a level in the

taxonomy that will either require the user to make decisions that they might not know or else duplicate large chunks of the taxonomy.

### Advantages of faceted navigation

First, faceted navigation is a dynamic interface that allows for multiple perspectives. Instead of starting at the top of a hierarchy that was designed by a taxonomist, users can select which facet to start with depending on what they already know. In addition, they can make any combination of selections – they can browse down a single facet for 2 or 3 levels and then switch to another facet and repeat or take one selection from facet A, one from facet B, and one from facet C.

Second, the output from a set of selections is equivalent to putting together an advanced query, Select all the wines that are color = red, price = medium (\$15-\$25), location = Napa Valley in California in the United States. The difference (and it makes all the difference in the world as those of us who have tried to get users to use advanced search know) is that the experience is one of browsing through a number of easy to use facets rather than trying to construct a complex query from scratch.

Third, as noted above, facets are easy to use. They are more intuitive because each facet is relatively simple (the complexity is constructed) and because they are mutually exclusive, it is easier to guess what is behind each door, that is, what to expect within each facet.

A fourth advantage is that it takes much fewer elements to achieve the same specificity as a taxonomy. Since you can map each element within each facet to any other, if you have 4 facets each with only 10 nodes, it is the equivalent of a 10,000 node taxonomy (10x10x10x10).

### Are Taxonomies obsolete?

Of course not. Facets are extremely powerful but they have one major limitation, taxonomies can model much richer knowledge representations that incorporate much greater conceptual depth. There is just no substitute for representing conceptual domains such as physics, genetics, business concepts, and any deep subject area.

In addition, documents are much more complex entities than a bottle of wine or even a work of art and taxonomies are much more suitable for categorizing documents. This is because taxonomies have the conceptual complexity needed and because unlike facets, taxonomies are designed for multiple values per entity. A bottle of wine only comes from one place and has one price and one color, but a document can be, and almost always is, about multiple topics.

### Combining Facets and Taxonomies

The answer to using facets and taxonomies is not which one is better but how do you combine the strengths of each. There are many ways in which facets and taxonomies can be combined. What is discussed below is based on a project we did earlier this year, creating what we called a faceted taxonomy.

The first thing we realized is that there really is no clear dividing line between facets and taxonomies except in how they are used and the relative depth of each. For example, an individual facet can be organized as a traditional hierarchical taxonomy.

A faceted taxonomy differs from a traditional taxonomy in that the anticipated use is to make choices from multiple dimensions and achieve small result sets by combining choices in each facet rather than browse down to a small result set within a single dimension. The design implication of this projected use is to favor a larger set of shallow and relatively simple “taxonomies”, rarely more than 3 or 4 levels deep.

Some other design considerations were finding the right balance of elements – do you have a set of equal facets or one dominant facet and several secondary ones. In many faceted navigation implementations, it makes sense to have a dominant facet. For example, for hotels, location is clearly a special facet in that people don’t normally select a vacation spot based on the hotel and some of it’s features, but rather they select location first and then want to see hotels in that location.

We found that when combining a large taxonomy with a number of facets, that it made worked best to have the subject taxonomy the dominant element with facets that could be applied to the taxonomy at any point in a browse.

### Implementation Issues

One additional issue deserves special mention. Facets require a lot of metadata since every document needs to be tagged with a value from each facet. In one project, we defined 7 facets as well as a subject matter taxonomy. That means a minimum of 8 tags per document that someone needs to add. This is in addition to other metadata fields like title, author, description, and others. And we all know how easy it is to get users to add metadata to their documents.

What we found was that there are a number of ways to overcome the implementation cost, but the best approach was a combination of governance / work flow, some dedicated librarian / metadata specialists, communities of interest, and finally some software components – specifically content management and text analytics.

The basic model was to publish into the CM application with an entry page that included all the metadata elements that had to be entered (and access to optional and local metadata fields). What saved the experience was the text analytics software that provided default values for the taxonomy and each of the facets. The software would categorize each document based on the taxonomy and categorization rules, and it would extract entities to populate each facet.

What was asked of users then was a task that was cognitively simpler by orders of magnitude. Instead of trying to come up with the right terms to categorize the content of a document with no supporting context, all they had to do was look at a suggested term and agree or disagree with the categorization. The amount of cognitive energy we are asking users to expend is minimal and usability studies show that users can and will do this.

In addition, there is a highly structured controlled vocabulary to add to if users disagree with the suggested term and a community of interested people, including metadata specialists and taxonomists that are working to improve the taxonomy and categorization rules

### Conclusion

Faceted navigation is a powerful new way to organize and provide access to information for Enterprise content and it can lead to a whole range of new applications that combine facets, search, taxonomies, ontologies, and visualization.

[Figure 2 – Seamark – Inxight ]