The reading focuses on methods of using and managing information in various knowledge organization systems to enhance the effectiveness of information retrieval. Tudhope and Binding explore the idea of the thesaurus and a faceted approach to classification as a potentially effective mechanism to organize and classify large volumes of information in digital heritage institutions such as museums. The writers discuss the issues that surface when knowledge organization systems are used in the Web environment. The reading also contains an analysis of “The FACET Project” which was a study designed to investigate the potential of the thesaurus for information retrieval purposes.

The digital age has been the stepping stone for the availability of large amounts of data and information on the World Wide Web. Tudhope and Binding describe the trend within museums and digital heritage institutions making information such as exhibits and objects in their collections available to the public who may wish to seek specific information. Of course, when dealing with such large amounts of information, “There is a need for tools to help formulate and refine searches and navigate through the information space of concepts that have been used to index the collection” (Tudhope & Binding, 2004).

This is an obvious statement. It all comes down to determining the best method of organizing information and enabling it to be retrieved efficiently. The first idea presented in the reading is the notion of “controlled vocabulary”, a mechanism used to index information that involves technical terms. The writers claim that “if both searchers and indexers draw on the same standard set of words then the synonym mismatch problems common with web search engines can be avoided” (Tudhope & Binding, 2004).
opinion, the writers are correct to an extent. Their statement is simply an objective that needs to be achieved in a practical sense.

The use of controlled vocabulary in indexing systems is widely used today. The problem is further enhancing it to cater for the many different types of users who possess different knowledge and information seeking methods. We have to also consider the fact that information is of a constant changing nature. New terms may be created from scratch or even derived from old or existing terms, and existing terms may no longer be used for some reason. The indexing of large volumes of information may undergo constant refining. On the other hand, we have to question whether the new terms are common amongst the different types of users. This sends us directly back to the idea that the searcher and indexer must draw upon standard sets of words to avoid synonym mismatches.

This leads us into controlled vocabulary driven indexing tools such as the thesaurus. A thesaurus can be viewed as a hierarchical structure of broad terms, narrow terms, and related terms that are organized in an interlinked manner to distinguish the context in which terms are used. The writers agree to the idea that thesauri to classify all sorts of information should be developed in a faceted manner. The faceted approach to classification involves the grouping of terms in subdivisions known as facets which by definition encompasses terms with similar concepts, attributes, subjects, and topic areas.

The faceted approach is quite useful because it is more specific when users seek information such as searching web pages or simply generating queries from databases. One of the important characteristics of faceted classification is its integration of controlled vocabulary as an indexing tool. In another article regarding the same topic area, Ruben Prieto-Diaz states that “Software products have certain characteristics that make controlled vocabulary a more attractive approach over free-text analysis” (Prieto-Diaz, 1991). This may be very true because the software focuses on searching within subdivisions and pre-defined terms. Free-text searching is more dependent on more
complex functions such as statistical and positional properties of words in large amounts of text.

The reading details findings of a study called The FACET Project where facet structure and retrieval were put into practice to investigate its potential. It involved setting up an application where users can select terms from multiple facets and use them in a query. “However, in many cases it is unlikely that exactly the same combination of terms will have been used in indexing” (Tudhope & Binding, 2004). This finding directly reflects back to the point made earlier about the searcher and indexer having to use a standard set of terms for better results. In the case of The FACET Project, terms may have been omitted; the searcher may have wanted something more specific, or simply no specific results returned. My point here is simply to suggest that this method of classification won’t be fully integrated any time soon.

The study incorporated a technique known as semantic expansion to counter for synonym mismatch. The concept of semantic expansion has a lot to do with the indexing side of information organization. It involves including additional terms that represent a similar concept e.g. guitar, bass guitar, acoustic guitar, ukulele etc. The aim of semantic expansion is to account for the range of different terms that users may adopt when searching for particular information. This is to say that a searcher can use different terms that still enable the retrieval of the main result(s) which are associated with the pre-defined search terms.

From the analysis of The FACET Project, the writers form a conclusion in saying that common standards need to be adopted for thesauri style classification tools to be used on the broader scale such as the idea of having a global museum where major museums are interconnected, enabling users to search on a broader scale. The adoption of standards in this case will help in removing duplicate data. The writers suggest that semantic expansion of query terms is an effective approach to assist in information retrieval as it has been proven in web based searching applications. However, it all comes down to the user. There are many different search techniques that users may not be aware
of. Some of these techniques are effective and some are not. I think user awareness should be addressed in a sense where free-text searching and controlled vocabulary systems can both be enhanced to maximize information retrieval.

In conclusion I believe that faceted approaches to knowledge organization such as the thesaurus have no definite future in the way we classify. Using the thesaurus as a standard for all future classification systems is a difficult concept. As humans, there is only so much knowledge that we can gain. Each individual possesses a different level of knowledge. The information that our knowledge is built upon is of a constant growing nature. The evolution of language and culture are also direct influences on information organization. To account for these important factors, knowledge organization systems require constant updating and refining to keep up with information growth. This is also very much dependent on the knowledge of the users.
References:


