This report is based on the case study entitled “A Case Study of a Faceted Approach to Knowledge Organisation and Retrieval in the Cultural Heritage Sector.” The study was performed by Douglas Tudhope and Ceri Binding and investigates the emerging trend within museums and digital heritage institutions to unlock information in their collections by opening up databases to a new range of users. More specifically the case study investigates the use of a faceted approach of classification to structure collections in thesauri. A faceted approach to classification allows the assignment of multiple classifications to an object, allowing searching and browsing of related information through several classes. (Wikipedia, 2005)

The study focused on the use of faceted classification by the AAT, (Art and Architecture Thesaurus) which is split into 7 facets: associated concepts, physical attributes, styles and periods, agents, activities, materials, objects and optional facets for time and place; (Tudhope and Binding, 2004) each facet is a homogenous class of a concept, the members of which differentiate them from members of other classes. By splitting a collection into separate facets the integrity of large collections such as the AAT thesaurus are greatly enhanced. However, what’s the use of dividing collections into facets and hierarchical subdivisions if you cannot search for terms effectively? Using a controlled vocabulary helps searchers locate what they are searching for, even if they are unfamiliar with the terminology in the database. Controlled vocabularies provide a limited set of terms to be used in a search which describe only the documents or objects in a collection. A controlled vocabulary search will result in broader, narrower and related terms being returned, and will also alleviate syntax problems associated with the user; all good things, but there are weaknesses involved with controlled vocabulary use. As explained by Tzitzikas (2002) “…The gains of controlled indexing languages are dependent on the quality, accuracy and consistency of the indexing process.” Hence, if the indexers skills are not of high quality, the quality and integrity of the thesaurus are jeopardised.

It is also more expensive than natural language indexing to implement; there are high input costs due to complexity, there is a strong chance human error will occur during the indexing process, Tzitzikas (2002) backs this hypothesis, stating “…the indexing of objects under a controlled and structured vocabulary is usually done manually. By consequence, if the collection of objects is big, this task is laborious and costly.”

In addition controlled vocabularies lack specificity over natural language indexes and most notably; the searcher must be aware of the correct terms to search with in order to make the search successful. (Harvey, 1999) In my opinion the use of a natural language indexing structure is far more practical than a controlled vocabulary due to not only the obvious cheaper costs and
ease of use, but a collection structured using natural language is far more likely to produce more accurate results for searchers; The most important measure of a search engine is the quality of its search results. (Brin & Page, 2005)

The case study focused on the FACET project, (faceted access to cultural heritage terminology) this project investigated the potential of thesauri in retrieval processes with the added use of an interface; allowing users to choose terms from facets then combine them into a query. The idea behind this concept is that by combining concepts from multiple facets the search becomes more specific. It is in essence a project trying to improve the accuracy of controlled vocabulary searches by using facets. In my opinion the more viable option would be to simply use natural language indexing, but the FACET project managed to determine some positives. On some occasions searchers may not find any exact matches from a search, only similar objects potentially of interest, the solution was developed from the FACET project through the technique known as semantic expansion.

Semantic expansion involves terms being supplemented by additional terms representing similar concepts based on their relative positions and relationships within the thesaurus structure, (Tudhope et. Al (2004) this theory achieves the goals of any organisation using database retrieval methods in that it has the potential to satisfy searcher demands and makes the searching process as easy as possible. This is especially useful to searchers unaware of exact terminology to use when searching because the returned results can include exact matches, close matches and terms related to a particular concept. However, it would still be much simpler for an organisation to implement a natural language structure which is much less complex and resource intensive than a faceted indexing structure utilising controlled vocabularies and semantic expansion.

Whilst developing the FACET project some prototype systems were produced to evaluate the process of user interaction and searching behaviour with the system. This resulted in findings such as design issues and how to best support the controlled vocabulary search process, thus helping developers understand the intricacies of thesaurus making no matter how big the collection is. Most notably the analysis of the prototypes discovered that when using a faceted structure the size of the collection doesn’t matter; both small and large collections are equally intricate and complex to design and maintain. The validity of the evaluation sessions is somewhat questionable however; the sessions involved as few as 23 users and the methods used to accumulate data were as basic as observation notes and transcripts of brainstorming sessions. More appropriate data gathering techniques would have been to either survey users after using the prototypes or to interview them, providing a more detailed description of FACET structured
collections. Despite this, the results seemed to reflect my judgment that FACET collections are far less practical than natural language structured collections, mainly due to the complexity and therefore costs involved with implementing and maintaining.

The introduction of web based information architecture provides the exciting prospect of online museums; this can enable searching across multiple collections as well as ensuring no duplication of effort or data will occur; (Tudhope & Binding, 2004) that is, if protocols are implemented and adhered to. The study discovered that adopting protocols has the added benefit of enabling logical division of effort as well as allowing institutions to perform indexing and create interfaces by utilising resources in different locations, i.e. Indexers can perform their work on a single database from home or different museums; a central location is not necessarily required. This initiative is indeed a very exciting proposition for museums and institutions to take advantage of, in the long term since adopting online collections provides museums with cheaper running costs and increased flexibility for database maintenance activities.

Tudhope and Binding also looked towards the future and how web developments will become more resource intensive and complex; the study found that the main issue facing knowledge organisation systems in online environments was that existing standards are based on the “print world” rather than electronic formats. This lack of standards impedes a wider use of resources in the distributed web environment. Personally I fully agree with Tudhope and Bindings’ findings that if future web developments are to be made in conjunction with the cultural heritage sector, online protocols need to be established. It is obvious that there are many differences between electronic and non-electronic databases, for an online collection to be viable, protocols need to be in place for data input, data manipulation, data retrieval and disposal activities. It is vital protocols exist because the users performing these activities can be working in many different locations around the world, if everyone performs activities the way the want to databases are going to become highly immersed with redundancies and will hence lose all levels of integrity. Protocols will provide users with direction on how to perform activities, ensuring all data activities are performed correctly and in the same fashion as all other users involved with database.

This case study based on faceted approaches to organisation and retrieval presented some interesting results. In my opinion some of the results of this particular case study are definitely inconclusive, the theory of faceted classification does have some advantages for organisations willing to spend the extra money and resources to implement and maintain the database.
However, use of IT technologies such as online thesauri can make use of more practical database structures such as natural language indexing which is far less complex for both indexers and searchers using the database.
Bibliography


