Describing engineering documents with faceted approaches
Observations and reflections

Peter J. Wild
Institute for Manufacturing, University of Cambridge, Cambridge, UK, and
Matt D. Giess and Chris A. McMahon
Innovative Design and Manufacturing Research Centre, University of Bath, Bath, UK

Abstract
Purpose – The purpose of this paper is to highlight the difficulty of applying faceted classification outside of library contexts and also to indicate that faceted approaches are poorly expressed to non-experts.

Design/methodology/approach – The faceted approach is being applied outside of its “home” community, with mixed results. The approach is based in part on examination of a broad base of literature and in part on results and reflections on a case study applying faceted notions to “real world” engineering documentation.

Findings – The paper comes across a number of pragmatic and theoretical issues namely: differing interpretations of the facet notion; confusion between faceted analysis and faceted classification; lack of methodological guidance; the use of simplistic domains as exemplars; description verses analysis; facet recognition is unproblematic; and is the process purely top-down or bottom-up.

Research limitations/implications – That facet analysis is not inherently associated with a particular epistemology; that greater guidance about the derivation is needed, that greater realism is needed when teaching faceted approaches.

Practical implications – Experiences of applying faceted classifications are presented that can be drawn upon to guide future work in the area.

Originality/value – No previous work has reflected on the actual empirical experience used to create a faceted description, especially with reference to engineering documents.

Keywords Archiving, Information science, Classification

Paper type Case study

1. Introduction
Our concern within this paper is to report on a case study into the development of a faceted classification for engineering documentation and to present our analytic and empirical reflections.

Faceted classification is an approach most associated with Ranganathan (1967), although other notable authors include (Vickery, 1960, 2008; Broughton, 2001, 2002; Broughton and Slavic, 2007; Beghtol, 1986, 1995, 2008). In essence, the approach deals with multiple viewpoints on, and/or features of, entities. Traditional hierarchical classifications require entities to be placed in a single category within a classificatory structure.

This paper contains an extended and revised presentation of some material that appeared in an earlier form in Wild et al. (2006) and Giess et al. (2007).
Its location within the structure indicates how it is related to other concepts/attributes. In cases where an entity describes or embodies multiple concepts such a structure is restrictive, as the entity can only be located according to one specific concept (and hence against one category). This prevents the entity and its attributes from being classified according to all relevant concepts, and introduces problems of viewpoint dependency where the specific “defining” concept may not be common to all users and the relationships within the structure may not reflect the relationships of interest to all users (Rowley, 1992). 

Faceted classification can address viewpoint dependency by allowing an entity to be classified according to all relevant concepts in an open and extensible manner (Ranganathan, 1965). By associating concepts with different facets, compound subjects such as “engineering statistics” can be concisely depicted, and those interested in either engineering or statistics can ascertain where within their discipline the entity resides. In an enumerative schedule, the broader set of “engineering” would be disparate to the “statistic” subset of engineering, and it is also possible that a subset of “engineering” would exist within a broader “statistics” set. The original emphasis of faceted classification was to provide concise and elegant means of classifying such compound subjects. In our context, it provides an elegant approach to covering differing viewpoints on engineering documentation in one structure, differing viewpoints can be introduced by differing roles within the design and manufacturing process (Gunendran and Young, 2006).

Until about 1995 faceted classification received most use and attention in library science. However, increasingly facets are being used outside of library contexts. For example, on the internet, and in localised collections of documents/information resources, such as engineering documentation (Chang and Tsai, 2003; Lowe, 2002; Opdahl, 2003). Here, the library is more a metaphor than a physical and institutional entity. Varying interpretations of the approach can be seen in existence from a more informal use of it in web navigation and knowledge management (Adkisson, 2003, KMConnection), to more formally based approaches (Wilson, 2006b). The divergence in faceted approaches has several roots. One is the difficulty of obtaining the original work. Once obtained there is the difficulty of interpreting Ranganathan’s dense style. Finally, there is a more general tendency to apply notions faceted in different contexts with different philosophical assumptions and pragmatic aims.

Our desire was to follow both the spirit and the methods associated with faceted analysis/classification to deal with the multifaceted nature of engineering documents. However, during our efforts we came across a number of pragmatic and theoretical issues, namely: the divergent interpretations of the facet notion; a lack of pragmatic guidance on how to generate a set of facets; the difference between faceted description and analysis; top-down verses bottom-up construction of facets; differences between user and literary warrant; and whether the approaches (Spiteri, 1998; Vickery, 1960) are truly generative or are criteria aimed at evaluation of generated facets. The development of these empirical and analytic observations is the focus of this paper.

With this in mind, the paper proceeds through another four sections. Section 2 presents relevant background literature, covering library classification schemes in general and focussing on the emergence of faceted classification. Section 3 presents, our case study in applying faceted notions to the description of engineering documents. Section 4 presents our observations and reflections, both analytic and empirical. Finally, Section 5 summarises and concludes the paper.
2. Background literature

In this section, we seek to identify the approaches that are applicable to the organisation of engineering documentation. The review will identify what forms of classification systems exist and where they have been applied. A particular focus will be on faceted classification, and the review will seek to identify guidance on how to construct such a classification scheme. Examples of the application of faceted classification outside of the library domain will be introduced.

2.1 Classification

A number of authors identify two distinct forms of classification, enumerative and analytico-synthetic (Rowley and Fallow, 2000; Buchanan, 1979), although other forms of classification are also noted (as an example, La Barre, 2006, also identifies hierarchical classification). Enumerative schemes divide the corpus into ever smaller classes according to identified principles of division. This can be categorised as a top-down approach, as the ambition is to partition the overall corpus into ever narrower segments until the content of each segment consistently describes the same concept. Analytico-synthetic schemes seek to identify the constituent concepts for each document within a corpus, and from the resultant set of concepts evolve schemes which arrange these concepts in a classificatory structure. Construction of such schemes involves analysis of each document to indicate constituent concepts and subsequent synthesis to arrange these concepts into a classificatory structure. Although considered synonymous with analytico-synthetic classification by a number of authors (including Rowley and Farrow, Buchanan, 1979 La Barre), faceted classification is seen by Broughton (2004) as “tal[ing] the idea of analytico-synthesis to its logical conclusion, [by] 'deconstruct[ing]' the vocabulary into its simplest constituent parts”.

2.1.1 Library classification. In the nineteenth century classification schemes were typically developed on an *ad hoc* basis, focusing upon a single collection. This perhaps reached its greatest expression in the Library of Congress Classification (LCC). It was developed specifically for the contents of the Library of Congress[1] but it has weak theoretical principles (Broughton, 2004, p. 145). The Dewey Decimal Classification (DDC) may be seen to display stronger theoretical principles, and has notable take-up in many libraries. These theoretical principles were introduced from the second edition onwards (Foskett, 1973, identifying the introduction of notation synthesis at this point), and Broughton (2006, p. 56) notes the increasing prevalence of “analytico-synthetic features” in DCC from the 1950s onwards. The DDC has ten main classes, with further sub-divisions via ten sub-classes. The classes proceed from philosophy, through religion to science, mathematics and technology, to geography and history. It is not without fault, being biased strongly towards US perspectives of the early twentieth century (Beghtol, 1986[2]). Whilst both schemes demonstrate different approaches to classification they are both enumerative in nature.

The Universal Decimal Classification (UDC) is a development of Dewey’s system, adapting the class structures and adopting the decimal notation. But it differs from the DDC by virtue of its synthetic nature. DDC allows only singular classmarks, whereas UDC allows classmarks to be both combined and for additional, pre-determined classmarks to be appended. Combination is provided by the provision of a range of operators, in essence providing a composition of class descriptions. These operators include relationship (i.e. statistics of automotive manufacturing), extension (i.e. all
manufacturing engineering), and addition (i.e. manufacturing and marketing). Additional classmarks are provided as auxiliary tables[3] which provide pre-defined classmarks for concepts that frequently reoccur. For example, indicating that a document is written in a different language to that of the main corpus (where the language in use is not of principal interest).

The description of common concept types (via use of auxiliary tables) and the combination of classes suggests that a document could feasibly be described by a combination of common class types. This approach is taken further in faceted classification. Faceted classification describes, the notion that a given entity is a compound of a number of different concepts. By isolating concepts into consistent facets it is possible to reconstruct any entity from a suitable combination of classes within each facet.

Ranganathan (1965) Colon Classification and the second Bliss Classification (BC2 see Broughton, 2001) are both faceted classifications. Ranganathan’s Colon Classification was first published in 1933, and his ideas were revisited in the 1960s by the Classification Research Group (CRG) leading to the generation of the second Bliss Classification BC2 in the late 1970s (Vickery, 1960). At the time of writing the authors are unaware of any implementations of the Colon scheme within UK libraries[4]. BC2 is in evidence in a small but notable number of libraries throughout the UK (predominantly, but not exclusively, in the Oxbridge college libraries).

Figure 1 shows an approximate timeline for the introduction of different library classification schemes, identifying those that are enumerative and faceted.

2.2 Creating classification schemes
Overall, the specification of a classification scheme is a difficult task, requiring considerable expertise as well as intellectual and manual effort (Kwasnik, 1998; Morville, 2005; Ranganathan, 1967; Star, 1998). The notion that knowledge is simply captured and represented in documents is at best naïve, at worst misleading. Documents act as a representation of concrete or abstract entities in a world. In turn, the properties of knowledge represented in documents can vary in their objectivity or subjectivity[5]. Louie et al. (2003, p. 1) recognised the complexity of knowledge saying that “all attempts to describe or organise these documents are flawed”. In a similar, Broughton (2004, p. 20) states “it simply isn’t possible to create a classification that is truly objective or neutral or absolutely correct”. The idea that a classification scheme is not neutral or objective suggests that any given scheme will be steered in some way by the preconceptions of the constructor, or will be deliberately biased towards a specific anticipated user(s) (Beghtol, 1986).

![Figure 1. Timeline showing approximate introduction of library classification schemes](image-url)
In some respects enumerative classification is more prone to subjectivity than analytico-synthetic schemes as it relies upon identification of criteria for division which may be notably viewpoint-dependent. This gives a hierarchical structure by which the resultant relationships (broader and narrower terms) reflect relationships considered important according to one viewpoint. However, the rigidity of the structure prevents alternative relationships (significant from a different viewpoint) from being expressed within the structure. The description of concepts (particularly compound) within analytico-synthetic schemes potentially mitigates viewpoint-dependency as a greater range of intra-concept relationships can be expressed.

2.2.1 Literary warrant. The method of identifying key concepts, terms, and relationships within documentation and structuring these into a classification scheme is influenced to some extent by the notion of warrant. A classification scheme may be developed around a rationalistic interpretation of the domain and related corpus; or upon the specific literary content of a domain. Hulme (1911) disagreed with rationalistic classification, and coined the term literary warrant to describe the practise of constructing a classification scheme based upon the specific content of literature.

Standards related to thesaurus development refer to both literary and user warrant, with user warrant defined under National Information Standards Organisation (NISO, 1994, p. 6) as “justification for the representation of a concept in an indexing language or for the selection of a preferred term because of its frequent occurrence in the literature.” Other varieties of warrant exist. Beghtol (1986) argues that Bliss identified scientific warrant as fundamental to classification, whereby only the structures defined by experts within the field could yield enduring classification schemes.

In terms of applicability, analytico-synthetic schemes rely upon literary warrant where the concepts which are contained within the document corpus are identified beforehand and the scheme arranged to fit these concepts. The idea of literary warrant espoused by Hulme suggested that book titles be used to provide terminology and concepts upon which to construct a classification scheme, whereas the CRG took the position that literary warrant be based more upon the terminology of a field (Vickery, 1960; Hjørland, 2006). Enumerative schemes can essentially take a range of positions, being based either upon emergent terminology and relationships obtained from literature (literary warrant) or upon a perceived general consensus of the field through user, educational or scientific warrant.

2.3 Specification of faceted classification scheme

The previous discussion on the specification of classification schemes has considered classification in general. Ranganathan proposed an approach to the generation of a faceted classification scheme, which will be discussed in this section. Facet analysis is the method by which a classification scheme is generated, and will be discussed in some detail.

2.3.1 Facet analysis. Facet analysis is the process by which faceted classification schemes are constructed. Documents within a corpus are decomposed into constituent concepts or isolates, and from these concepts a structure can be arranged to collocate similar concepts and distinguish between disparate ones. Ranganathan (1965, 1967) identified three planes to a faceted classification the:

(1) idea plane, which concerns analysing the corpus and decomposing into constituent parts;
A central aspect of the development of a library classification scheme is the need to specify a system of notation. This notation serves two purposes. First, it serves to indicate the proximity or disparity of classes within the overall scheme. Assuming there is a sequential progression to the scheme, and the method of notation reflects this progression, the notation illustrates the relative location of two different classes and from this how conceptually similar they are. Second, the notation provides a coding system which essentially reflects the classification scheme in a more concise manner.

A computer-based archive does not have such a strong requirement for such a system, as it is possible to computationally link a given document into a specific class using methods such as Uniform Resource Indicators or supporting database records (for example the Waypoint system, McMahon et al., 2004) (www.adiuri.com/resource/).

Ranganathan specifies 46 canons, 13 postulates and 22 principles, essentially requirements that the facet selection should meet. A key part of the work is the postulate of five fundamental categories, which asserts that any subject can be divided into five fundamental concept categories. These fundamental categories are personality, matter, energy, space and time (PMEST). Later the CRG (Vickery, 1960) revisited this work, suggesting 13 fundamental facets were pertinent, “Vickery’s expansion of the number of fundamental categories may imply that there is not a fixed set of categories in the world” (Hjørland, 2006). More recently, Vickery (2008, p. 152) asserts:

[...] we must always start from and be guided by the terms and relationships that we find in the subject matter, and that are familiar and helpful to users. But it may well be that a list such as the above can suggest to the classification maker possible aspects of the subject that would be candidates for facets. We should never try to impose such general categories on the subject.

The CRG also revised Ranganathan’s prescriptions, providing a reduced set of guidelines for the qualities of facets. More recently Spiteri (1998) argued that both approaches suffer from practical difficulties, and presented a “simplified” model that unifies the ideas presented by both groups into a model she considers more pragmatic.

However, the literature shows no in-depth reports on experiences in applying faceted classification, outside of special classifications. A number of authors discuss the application of facet analysis (Broughton, 2006, 2004; Ellis and Vasconcelos, 1999; Rowley, 1992), however such analyses are typically expressed in basic terms. In general, each covers the following stages:

1. collate representative corpus of documents;
2. identify discrete concepts that together describe document in entirety;
3. from identified concepts, evolve facets;
4. structure facets in terms of citation order; and
5. codify facets (Broughton, 2002, 2004; Broughton and Heather, 2000; Denton, 2003).

There are several web based discussions on the application of the notion of faceted classification in domains such as company web sites, online catalogues, etc. Of these,
the most authoritative is by Denton (2003), who focuses upon creating faceted classifications primarily for web applications, suggesting that the construction of a faceted classification proceeds via the initial construction of a controlled vocabulary.

At this juncture, we can make two observations. The first is a widespread failure to explicate the empirical or analytical methods employed in the construction of these schemas: that is, how did classifiers actually create the scheme? The second is that much of the work appears to the authors to be evaluative as opposed to generative, providing means by which to judge faceted classifications but without pragmatic and procedural methods of actively generating a classification scheme. Both these issues have a bearing when we move outside of library contexts.

2.4 Outside the library: interpretations and implementations of faceted classification

Traditionally, the application of faceted classification in libraries assumes that items will be physically located in just one “place”. Electronic resources can impose fewer location restrictions, lending themselves to faceted structuring; as the actual presentation of the resources can be rearranged and adapted depending upon the facets of interest, something not generally possible with physical documents. A number of applications of faceted systems have been implemented in electronic settings. These will be discussed with focus placed upon the nature of the classification schemes supported and how such schemes are constructed.

2.4.1 Faceted classification software tools. A number of software tools facilitate the use of faceted structures, such as Autonomy, Waypoint, and Flamenco, (www.autonomy.com; www.adiuri.com; http://flamenco.berkeley.edu/index.html) this section will focus upon tools which cater specifically for faceted classification. Waypoint (McMahon et al., 2004) and Flamenco (Hearst et al., 2006) are similar in that both allow faceted classifications to be constructed and used for browsing without restricting the nature of the classification scheme itself. Example applications of Flamenco include a classification of Nobel Prize winners and a classification of a fine arts museum collection.

When developing a faceted classification for images of fine art pieces using Flamenco (Yee et al., 2003) terms were automatically extracted from free-text descriptions of the content of each piece of art[7]. These terms were then automatically arranged into a hierarchy with some manual refinement (Stoica and Hearst, 2004). This was done by considering the grammatical relationships between terms as defined in the WordNet (http://wordnet.princeton.edu/) lexical database, although such arrangement can also be carried out manually. This approach can be considered to be literary warrant in terms of constructing the classificatory structure from the identified concepts, however there is significant dependence upon the free-text descriptions adequately and consistently representing the underlying pieces. The analysis of entities relies upon an accurate free-text description of the piece as decided by curators of the museum and not by the creator of the classification scheme. As the curators may have some further knowledge of the context and content of the pieces that would not necessarily be included within the textual description, and the specific location of the piece may provide additional context (as it is located in a specific area of a specific gallery) we argue that this approach may unduly impact upon the utility and expressiveness of the classification scheme.

Neither tool imposes a restriction upon the classification scheme itself, therefore it is possible to locate an entity an arbitrary number of times in a given facet range.
This arguably reflects the notion that it is not always possible to decompose a group of entities into a consistent number and form of concepts. FacetMap (Wilson, 2005) takes a notably different approach, each entity must appear once and only once in each facet. Wilson (2006b), the creator of FacetMap, argues that, when trying to locate a given entity more than once within a facet:

[...] a strict faceted classification model forbids you to assign both those headings, and with good reason. This is counterintuitive, controversial, and if you subscribe to S.R. Ranganathan’s original facet theory, heretical.

It is difficult to compare with any analytic depth the effectiveness of both Wilson and Hearst as each author covers different domains. However, the contrasting approaches demonstrate that differing notions of faceted classification are in evidence.

2.4.2 Applications of faceted classification. Numerous applications of classifications using a faceted-like structure may be seen on the internet, of which many are retail sites. Adkisson’s (2003) survey of 75 sites noted that 69 per cent incorporate some form of faceted classification. An approach commonly used in retail sites is the enumerative-faceted. Here, the complete domain is described under an enumerative structure, and when this enumerative structure is traversed a faceted structure is utilised in sub-classes containing similar content. The enumerative structure is retained until the members of a given class are of a consistent enough nature that facets are readily distinguishable. The facets evident in such structures are bespoke for that particular sub-class within the overall enumerative structure, and are not common to the domain as a whole.

Although facet-like structures may be seen within many web sites, for example Merholz (2001) and Broughton (2006) note that a wine retailing web site (wine.com) takes advantage of the inherently faceted structure of the wine domain, the only example to the authors’ knowledge specifically to claim a faceted structure is eBay Express (as indicated by the collaboration between eBay and the UC Berkeley School of Information, Hearst et al., 2006). As can be expected of a retail web site, eBay express classifies physical artefacts and not information entities. As was the case for Flamenco, where concepts were identified based upon either physical manifestation or tangible descriptions of each entity, physical traits are argued to be notably more tangible and hence distinguishable than the rather more abstract and intangible information content of documents.

The use of an enumerative structure prior to facets, and the resultant limitation in scope of the facets, signifies the difficulty inherent in identifying facets relevant to an entire domain covering disparate entities. eBay Express is required to cater for a potentially unlimited breadth of entities, as there are little restrictions upon what may be hosted for auction. This, in conjunction with the continual variation in entities under auction, does not readily allow a bottom-up approach to the identification of facets.

2.5 Reflections upon interpretations and applications of faceted classification
In general, the applications described previously do not adhere to the notions of faceted classification put forward within the library community. Hearst et al. (2006) do not seek to classify entities according to their constituent concepts, more upon those attributes
that can be readily identified. This is argued to have some adverse effects upon the scheme itself where the user has no prior understanding of the domain in question.

In domains with wide content, such as eBay Express, facets extending across the whole domain are not identified. Instead enumerative division was used to identify narrower groupings within which facets could be generated from prevalent physical (or readily distinguishable) properties of each entity.

It is significant that the practical applications of faceted classification outside of the library field take a looser stance, with a mix of enumerative and faceted schemes and where the concepts used for classification do not always depict the specific content of the domain. It is arguably the case that differences in domain influence the formality and rigour of the approach. The following section will consider the differences between the library domain and other domains such as engineering, in the hope of identifying issues influencing the application of faceted classification to engineering documentation.

2.6 Contrasting the warrant of library and engineering documents
Our concern now shifts to contrasting library and engineering document collections and practices.

Libraries have their own discipline, training courses and degree programmes. Their role in many organisations and communities is acknowledged and considered essential to the academic, social, and economic life of the communities they serve. In contrast, concern with information, knowledge, and document management in engineering remains a community within a discipline (engineering design), albeit with links into library, information and knowledge management.

Libraries also have more uniform collections composed predominantly of books, journals, conference proceedings and official statistics[8]. In terms of the maturity of library contents, items are formally “signed off” artefacts, with a clear title and target audience. In contrast, the variety of documents in engineering design is huge with much occurrence of more informal genres such as correspondence, logbooks and notebooks (Wild et al., 2006).

Library collections and engineering document collections have different temporal cycles, concerning when items join the collection and when they are used. In the former case these are tied to a publishing cycles, and user needs. In contrast, engineering documentation is based on a temporal cycle relating to the design and manufacturing processes. A significant proportion of engineering documentation is the informal, communicative documentation such as e-mail, and “working” documentation such as analysis files. These files are essential to documenting the engineering activity, as they support the more formal documentation. These are not written as a formal record but as a means of communication or of carrying out work. In such a case interpretation of these documents requires a degree of tacit knowledge and understanding of the context of the documents.

Libraries have dedicated systems for classification (e.g. Dewey, LCC and BC2) and “off-the-shelf” database catalogues. In contrast, engineers are flooded with a number of tools and file formats, which are interacting, competing and changing. Owing to cost constraints smaller engineering concerns have less sophisticated file management tools at their disposal (Wild et al., 2006). In contrast, larger engineering companies may utilise multiple document management systems.
3. Case study: describing and profiling engineering documents
This case study is situated in understanding information and knowledge in engineering design contexts. As part of these research efforts, we have obtained a significant document corpus. Aspects of this work have been reported at a high level in other venues (Wild et al., 2006), our concern here is to delve into the process by which we came about the set of facets for profiling and describing document collections in engineering domains.

Our initial modest aim was to be able to setup a structure with which an interested party could browse the document corpus. Hence, the reader would be able to browse all documents pertaining to X or Y; the intersection of X and Y. The “meaning” of X and Y would relate to a range of issues including document purpose, manifestation, design phase, functional division, genre, source and creation mechanism. This work is situated within a wider research theme concerning information and knowledge usage in engineering contexts, related projects concern annotation, logbooks (McAlpine et al., 2006), and engineering information ontologies (Darlington and Culley, 2008). The document corpus serves as a source repository of exemplars for all these efforts. The more refined and restricted context of the home project’s focus on decomposition schemes should not discount rich descriptions of the document’s manifestation and context. Hence, we found that could not discount a priori what is, and what is not important to record about the document corpus.

As we delved into the documents more, we started to consider how we could compare the documentation “profile” of different corpora. At a general level, we would be able consider differences between design and manufacturing documents; between organisations working with and without quality systems; and between different types of activity and manifestation. Initially, we aimed to use the Waypoint functionality (McMahon et al., 2004) to support both aspects of this vision (browsing a document corpus and comparing document corpuses), however, as the profiling notion took hold Waypoint was dropped because it is non-zero match functionality did not allow us to see the “holes” in a document corpus, which are as important to us as the documents that are present. In the end MS access was used to develop the database that stored records of some of the document types we encountered.

From these two goals (browsing and profiling) emerged a need to be able to define a set of facets that would adequately describe the document corpus. An aim of this was to represent information about the context and manifestation of the document. Unless recorded, much information about the document will not be evident to someone examining the documents in corpus in a few months time.

3.1 The body of documents
The bulk of the document corpus relates to work we have undertaken in an engineering company referred to as “TrollCo” (Wild et al., 2006). TrollCo is an Engineering Design and Manufacturing company based in Wiltshire, with around 60 employees and a turnover of £5 million. They have an engineering services division that undertakes general engineering and manufacturing work, however they also design and manufacture their own product line. Across the two divisions our analysis showed over 250 different distinguishable document types or genres, with the number of instances of each type running from one through to thousands for documents such as purchase
orders and design documents. Table I illustrates a number of document types from the TrollCo case study.

TrollCo was an opportunity for us to deal with the order and mess of real world engineering documents. Our first attempts at “typing” the documents resulted in a cumbersome hierarchy which when applied to the corpus seemed to classify most documents as “general”. Hence, one researcher on the project undertook the task of describing relevant features of the document corpus. Our main concern was therefore with issues surrounding context and manifestation. In part, this is because a related project strand was concerned with the decomposition of document structure and content (Darlington, 2005; Liu et al., 2006). Furthermore, as an organisation undertaking general engineering, gaining substantial and traceable information about the definitive semantic content of the documents would be difficult (Section 2.2.1).

3.2 Qualities of the endeavour
There were a number of issues we came across as we developed the facets. This has caused considerable soul searching and discomfort. We have never been able to avoid a nagging feeling that we were missing some key document pertaining to how to carry out the generation of facets for a document corpus. We have been unable to find such as body of work that is generative, focusing upon the generation of facets. We could not even find a definitive statement on fundamental facets (i.e. PMEST or CRG). However, we did see a number of qualities in this endeavour, these being: the application of a general notion of facets; the approach being descriptive rather than analytic; the approach being empirical and grounded in a real document corpus; the richness of the domain; and the process was both top-down and bottom-up.

3.2.1 A general notion of facets was applied. We started by making reference to tutorial format material on faceted classification (Denton, 2003; Spiteri, 1998). However, this material chooses restricted domains. In wider domains such as engineering documents, the purity of the facet model is not immediately applicable. Hence, a general notion of a facet was applied. This owes as much to browsing mechanisms as it does to the varying interpretations of Ranganathan (Hjørland, 2005; Star, 1998; Wilson, 2006b).

3.2.2 Descriptive rather than analytic. This research should be seen to be faceted description rather than analysis. Here, we view description as reporting characteristics; in contrast, analysis pertains to a resolution into “simple” elements (WordNet). We are not making a case for neutral description. Someone interested in typographic issues would have a different set of descriptions (e.g. font size, font style, page formatting and layout). In contrast, the characteristics we are concerned with are the contextual and higher level manifestation issues that surround engineering documents. Hence, Table II shows facets concerning contextual issues such as: manifestation (i.e. electronic or physical); grouping status (single or grouped) and mechanism (electronic such as email attachment or folder, and physical such as staple, folder), annotation and template status.

3.2.3 Empirical, qualitative, and grounded in a document corpus. Hjørland’s (2005) assertion that Ranganathan does not deal with empirical realities does not preclude the empirical use of facets. Star (1998) drew extensive parallels between Ranganathan’s work and Grounded theory. Our work takes influence from this comparison and has also been qualitative in orientation. Whilst, we acknowledge the influence of various forms of
<table>
<thead>
<tr>
<th>Product demonstration/order</th>
<th>Build</th>
<th>Test/assess quality</th>
<th>Delivery</th>
<th>Finance</th>
<th>Service/maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brochure</td>
<td>Contract</td>
<td>Final inspection check</td>
<td>Advice</td>
<td>Invoice</td>
<td>Routine service</td>
</tr>
<tr>
<td>Demonstration-request</td>
<td>Review</td>
<td>Certificate of conformity</td>
<td>Note</td>
<td>Maps</td>
<td>Customer feedback form</td>
</tr>
<tr>
<td>Draft order</td>
<td>Production programme</td>
<td></td>
<td></td>
<td></td>
<td>Parts order form</td>
</tr>
<tr>
<td>Draft quote</td>
<td>Product stocking sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final quote</td>
<td>Planning sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final order</td>
<td>Call off sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order-acknowledgement</td>
<td>Welding build sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sub assembly build sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time and temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Electrical options</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credit checks</td>
<td>Internal reject note, external reject note, supplier quality approval record, corrective action request form and vendor assessment form</td>
<td></td>
<td>Second invoice</td>
<td>Legal proceedings</td>
<td>Breakdown checklist</td>
</tr>
<tr>
<td></td>
<td>Exceptions/options</td>
<td></td>
<td></td>
<td></td>
<td>Call out service</td>
</tr>
</tbody>
</table>

**Source:** Wild *et al.* (2006)
Grounded theory (Seldén, 2005), we do not claim to have undertaken a Grounded theory analysis furthermore, we eschewed Grounded theory’s positivist roots, and subscribe to something that is more influenced by Charmaz’s (2000, 2006) work. Namely, we consider that the meaning of the documents does not exist purely in the documents themselves (Crotty, 1998), to be generated by a researcher without preconceptions (Glaser and Strauss, 1967). Our concern was to both address our preconceptions and to see what features emerged out of detailed examination of a document corpus. Classifications are not neutral, objective or context independent (Bowker and Star, 1999; Broughton, 2004; Medin et al., 1997; Medin et al., 2006; Worsley, 1997; Yeh and Barsalou, 2006). We were concerned to mirror Crotty’s (1998, p. 48) concern to bring “objectivity and subjectivity together and hold them together”. The range of facet descriptors for the document corpus ranges from harder aspects of the documents to softer, more situated and interpretative aspects. In the former we include its manifestation; grouping status; and grouping and manifestation mechanisms. In relation to softer more interpretative aspects descriptors concerning genre and purpose are relevant. Descriptors such as design phase, functional division, ISO type, etc. involve an element of discrimination from either research partners or researcher. The work was iterative in nature. Facets were generated, and then compared with further documents in the collection, documents were coded, compared against the emerging set of facets and then the facets were re-examined. Further work was drawn into utilise other’s insights about documents and the processes they are used in (Malone et al., 2003; Sellen and Harper, 2002; Yoshioka et al., 2001).

3.2.4 Richness of the domain. Documents are rich and diverse in their nature. They exist in both virtual and physical forms. Attempts to define what form they are in both before and after digital technologies has been difficult (Buckland, 1997; Windfeld Lund, 2003). They are both an artefact in themselves and the result of a documenting process (Buckland, 1997). Documents are embedded in a context and contain representations of both concrete and intangible aspects of the world. They have multiple formats, and with technology such as printers, digital pens and scanners and optical character recognition; they can be transferred between physical and digital forms with relative ease. Therefore, their fluidity is debated (Levy, 1994) and exploited (Henderson, 1990; Zellweger, 2000)[9].

3.2.5 The process was both top-down and bottom-up. There is often a concern to express classification (indeed, many representations of a domain) as being a cleanly top-down or bottom-up process (Cheti and Paradisi, 2008; Szostak, 2008). The development of this work has been both and more, as our engagement with TrollCo’s

<table>
<thead>
<tr>
<th>Issue</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Versions</td>
<td>One, two, three</td>
</tr>
<tr>
<td>Scrapped</td>
<td>Version two saw the removal of a “contain” facet that referred to contents such as geometry, words, and numbers</td>
</tr>
<tr>
<td>Refined</td>
<td>Manifestation, refined from single facet to two covering electronic and physical. The use of the MIT process handbook functional concerns rather than local division names</td>
</tr>
<tr>
<td>Additional sources</td>
<td>Additional sources such as the ISO9001 standard (ISO9001, 2000), the MIT process handbook (Malone et al., 2003; Laitinen’s, 1992) classification were added</td>
</tr>
</tbody>
</table>

Table II. Examples of iterations to facets
documents has evolved, facets have been, scrapped and refined and as new work has come to our attention additional facets have been added. Some of the facets were derived from the document corpus and this corpus itself was the acid test of whether preconceptions and conceptions held up against the data. Table II provides illustrations of changes to facets as we proceeded through three versions of the faceted classification. Table III presents the facets and illustrates their status as being derived top down, bottom-up or a mixture.

4. Reflections on faceted classification
This section serves to summarise a number of issues and reflections based on our experiences with applying faceted notions. We cover the following issues: that there are differing interpretations of the facet notion; that there is confusion between faceted analysis and faceted classification; that there is lack of methodological guidance; that there is use of simplistic domains as exemplars; the difference between description and analysis; the view that facet recognition is unproblematic; and whether the process purely top-down or bottom-up.

4.1 Differing interpretations of the facet notion are in existence
We raise two observations here. The first is that there appear to be a variety of interpretations of the notion of facet in existence. The second is that faceted classification does not have to be inherently associated with a particular epistemological standpoint.

Ranganathan’s work appears subject to significantly varying interpretation. This comes in part from the texts being out of print and people relying on second hand knowledge, in part because of the density and obscurity of his writings. Yet all who refer to his works seem to consider themselves true to the spirit of Ranganathan. A number of positions can be found within the facet community. We cover that of Wilson (2006b), the general model espoused by those such as the KMConnection (KMConnection), and Star (1998).

Wilson’s (2006b) paper is concerned with answering the question “Why can’t I assign multiple headings, from a single facet, to one of my resources?” Wilson appears to go further than Ranganathan in formality, insisting that a strict faceted classification allows assignment to one and only one element from a given facet. This is a rationalistic and non-empirical position[10]. The use of arrays allows a given facet to be decomposed, should a clear principle of division be identified with which to proceed, thus individuating the entities and promoting clarity of structure. This, however, is not considered within Wilson’s work, where the generation of additional facets is proposed as means of ensuring an entity is treated only once in each facet. Wilson is of course, entitled to take a strict view, and he also has implementation issues to address, however, more concrete evidence is needed to confirm whether Wilson is truly within Ranganathan’s worldview.

From our experiences with engineering documents this strong position cannot account for the existence of boundary objects as well as the general complexity of the world outside of restricted or special domains. The notion of boundary objects emerged from Star’s (1989) work and has found common currency in a number of contexts (Bucciarelli, 2002; Henderson, 1990; Mambrey and Robinson, 1997; Morville, 2005). Boundary objects are “those objects that both inhabit several communities of practice
<table>
<thead>
<tr>
<th>Facet name</th>
<th>Facet</th>
<th>Range</th>
<th>No.</th>
<th>Notes and examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>BU</td>
<td>BU</td>
<td></td>
<td>E.G., TrollCo</td>
</tr>
<tr>
<td>Purpose</td>
<td>TD &amp; BU</td>
<td>TD</td>
<td>6</td>
<td>Genre theory, earlier discussion with research partner, stated purposes in ISO documentation and other internal guides to documents</td>
</tr>
<tr>
<td>ISO9000 type</td>
<td>TD &amp; BU</td>
<td>TD</td>
<td>7</td>
<td>BU because it was ISO accreditation at TrollCo</td>
</tr>
<tr>
<td>Product phase (after BS 7000)</td>
<td>TD</td>
<td>TD &amp; BU</td>
<td>9</td>
<td>One of the range came from TrollCo the remainder were from the BSI documentation</td>
</tr>
<tr>
<td>Document status</td>
<td>BU</td>
<td>BU</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Review status</td>
<td>BU</td>
<td>BU</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Distribution status</td>
<td>BU</td>
<td>BU</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Main manifestation</td>
<td>BU</td>
<td>BU</td>
<td>4</td>
<td>Physical, electronic, or more of one than another</td>
</tr>
<tr>
<td>Physical manifestation mechanism</td>
<td>BU</td>
<td>BU</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Electronic manifestation mechanism</td>
<td>BU</td>
<td>BU &amp; TD</td>
<td>10</td>
<td>Contents was in part from the tools at TrollCo but in part from other tools that are available</td>
</tr>
<tr>
<td>Grouping status</td>
<td>BU</td>
<td>BU</td>
<td>2</td>
<td>Single or grouped</td>
</tr>
<tr>
<td>Physical grouping mechanism</td>
<td>BU</td>
<td>BU &amp; TD</td>
<td>10</td>
<td>Some of the elements were taken from TrollCo, some from catalogue, we expect the number to increase</td>
</tr>
<tr>
<td>Electronic grouping mechanism</td>
<td>TD &amp; BU</td>
<td>BU</td>
<td>4</td>
<td>Once the physical grouping element concept was come upon, the electronic version came into being</td>
</tr>
<tr>
<td>Template status</td>
<td>BU &amp; TD</td>
<td>BU</td>
<td>6</td>
<td>Colleague asked what about templates […]</td>
</tr>
<tr>
<td>Annotation</td>
<td>BU</td>
<td>BU</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Paper interface</td>
<td>TD &amp; BU</td>
<td></td>
<td>3</td>
<td>Work on genres and the nature of the forms in existence, novel concept, but one that builds on notion of UI</td>
</tr>
<tr>
<td>Descriptions – main descriptions</td>
<td>BU</td>
<td>TD</td>
<td>5</td>
<td>The notion of documents being tied to the quality process was derived from the document corpus. The specific facets and ranges were adopted from an existing paper on software quality documentation (Laitinen, 1992), as they are generic enough to describe key design documents</td>
</tr>
<tr>
<td>Descriptions – description</td>
<td>BU</td>
<td>TD</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>appendices</td>
<td>BU</td>
<td>TD</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Utilisation documents</td>
<td>BU</td>
<td>TD</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Development plans</td>
<td>BU</td>
<td>TD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality control documents</td>
<td></td>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Administrative documents</td>
<td>BU</td>
<td>TD</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Table III. Top-down and bottom-up derivation of facets

(continued)
and satisfy the informational requirements of each of them” (Bowker and Star, 1999, p. 16). Some objects exist across domains, and to reflect this in a faceted classification needs several elements from a range to be applicable.

It is also possible to generate explicit links between different facet schemas, however this moves us into the territory of ontologies rather than faceted classification, along with its own debates and idiosyncrasies (Darlington and Culley, 2008; Sim and Duffy, 2003). In our experiences documents frequently acted as boundaries between different design activities and functional concerns.

There is also in existence a looser interpretation, much more concerned with browsing of information entities. Adkisson (2003) observed that 69 per cent of the web sites she assessed used facets for browsing product categories. Waypoint (McMahon et al., 2004) does not enforce a “puritanical model” of facets, that is it allows more than one element in the facet’s range to match a document. The KMConnection note several differences between the model put forwards by the library community and the knowledge management community (KMConnection, 2005)[11].

Hjørland (2002b, 2005) has stressed his view that faceted classification embodies a rationalist epistemology. Whilst Hjørland should be thanked for highlighting epistemological considerations in relation to information, knowledge and documents (Hjørland, 2002a, b; Hjørland, 2005), we disagree with his immediate connection between facets and a specific epistemology. Another interpretation is that construction of a faceted classification is an open-ended discovery process akin to qualitative data analysis. Star (1998) draws extensive parallels between Ranganathan and

<table>
<thead>
<tr>
<th>Facet name</th>
<th>Facet</th>
<th>Range</th>
<th>No.</th>
<th>Notes and examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional concern</td>
<td>BU</td>
<td>TD</td>
<td>13</td>
<td>The general idea for the facet was developed bottom-up. Some documents can be perceived as boundary objects (Star, 1989). The facets are derived from the MIT process-modelling handbook (Malone et al., 2003), which serves as a repository of knowledge about organisations and provides case examples expressed in a common format and framework. They “abstract” across the naming idiosyncrasies of organisation’s different functional divisions, and by representing them as non-discrete facets, we can (a) account for documents that are embedded in multiple functional divisions, and (b) profile differences in document types between organisations.</td>
</tr>
</tbody>
</table>

Notes: TD refers to top-down; BU refers to bottom-up; most ranges contain two options for not applicable and not known, these are excluded from the count.

Table III.
Grounded theory, the latter being a qualitative research method that aims to systematically generate theory and concepts. Positivist versions and visions of Grounded theory do exist[12], but it is not clear that Star could easily be pigeon-holed as a positivist, her work with Bowker and Star (1999) points to the sociality of the construction of classifications and the citation of Latour in her work (Bowker and Star, 1999), do not appear to reinforce a positivist epistemology.

Despite Hjørland’s (2005, p. 144) continued assertion that Ranganathan is “is a position that does not consider the empirical basis of systems very much” there are approaches to facets that are empirical. Our case study is a further example. Beghtol (1995, p. 207) notes “that empirical data collection and analysis often precede(s) the final determination of appropriate facets”.

Faceted analysis/classification can be interpreted and reinterpreted within different epistemologies and theoretical standpoints. The differing views outlined above are evidence of this. There is no inherent and absolute commitment to a positivist epistemology when applying faceted analysis/classification. For an analogy consider the empirical milieu. Interviews come in variety of formats, ranging from open ended, participant driven affairs, through to tightly controlled questions. Some authors class them as solely qualitative in orientation, whereas others demand that they be quantitative and fully structured. A method can be used in different methodologies, and thus be driven by different theoretical perspectives and epistemologies[13]. Methods can be used in a number of methodologies, sometimes without great adaptation, in other cases they take on the assumptions of the methodology or overarching theoretical and epistemological frameworks. It does not follow that because an approach has been used within a particular epistemology it embodies that epistemology and cannot be used with different assumptions (Crotty, 1998).

4.2 Confusion between facet analysis and faceted classification
The differentiation between facet analysis and faceted classification is not always clear. Broughton (2004, p. 70) states that “the methodology of faceted classification [is] what we shall call facet analysis”, suggesting that faceted classification is a tangible product of the process of facet analysis. Broughton (2004, p. 70) then proceeds, on the same page, to state that “faceted classification is not a particular scheme or system, but rather a method for making a classification”. This implies that faceted classification is itself the process by which the product comes into being. It is not clear, from these two statements, exactly what is meant by the two terms as the definitions appear contradictory. For the purposes of this paper facet analysis has been taken to be the process by which the tangible faceted classification is generated, however this example indicates how confusion arises when entering into such an endeavour.

La Barre (2004) suggested that a distinction should be made between faceted analysis and faceted classification by examination of them in relation to Ranganathan’s notion of idea, verbal and notational plane. Facet analysis covers the idea and verbal planes, whilst faceted classification covers all three. We however suggest that the distinction should be based around process and product. Facet analysis is the process by which a faceted classification is developed. As such a faceted classification is a product, whose manifestation reflects the idea and verbal planes. The notational form can manifest itself as a traditional library notation, a database, or a browsing mechanism.
4.3 *Lack of methodological guidance*

There is nothing than can be considered a step-by-step method to the creation of a faceted classification. There are a set of criteria that are:

- debated/altered (Spiteri, 1998); and
- have inconsistent and varying definitions, contrast Wilson (2006b) and Star (1998) for example.

Evidence of this lack of methodological guidance can be found in the lack of consistent stages for facet analysis; lack of specification of expected or needed iteration for activities and a lack of agreement about the number and usage of fundamental facet types (e.g. PMEST vs CRG).

There is a tendency to present the work on facets as being rationalistic and analytic, whereas much of it is actually rendition of common sense notions of a simple domain (e.g. socks, detergents). In turn, there is a tendency towards evaluation rather than generation. As if the researcher is asking the question “How do we sit in judgement of faceted classifications” rather than “How do we support people in constructing faceted classifications”. In turn, the empirical basis for faceted classifications is limited, both through the absence of empirical data and examples, but also by restricting itself purely to libraries. For us our concern was to identify what we could “break” whilst still considering ourselves to be undertaking a faceted classification. It is not clear that we have found an answer to this question, especially when we consider that historically the faceted notion has emerged in more than one context (Beghtol, 1995) and has been used by others without explicit reference to either the work of Guttman or Ranganathan nor Broughton or Vickery (Cheung et al., 2005; Karlson et al., 2006; Priss and Jacob, 1999; Sawyer et al., 2005).

A common distinction in design is the difference between generative and evaluative methods. Whilst the boundaries are not always clean-cut between them, they are formalised activities in design contexts. Generative methods concern the generation of artefacts; evaluative with some form of evaluation, whether summative or formative. Design walkthroughs and usability testing are example of evaluative approaches. Task analysis, brain-storming, focus groups, are generative methods/activities. When we consider the use of the notion of facets, the models proposed by Ranganathan, the CRG (Vickery, 1960) and Spiteri (1998) are in practice evaluative as opposed to generative. They state what the nature of the identified facets should be, but are less useful in identifying how they may be identified.

4.4 *Simplistic domains as exemplars*

Within this section, we make the observation that contemporary domains used to illustrate faceted concepts are simplistic (e.g. wine and confectionary ingredients, Wilson, 2006a, b, dishwasher powder, Denton, 2003).

Dowell and Long (1998, p. 130), two leading proponents of domain modelling, note that a complete understanding and formalisation of a domain cannot always be achieved. Similar notions have been expressed by other domain-oriented researchers (Booch, 1994; Pejtersen and Rasmussen, 1997)[14]. Domains are not generally such small closed entities, however, the demonstrations of faceted analysis have focuses around small close domains, such as socks (Broughton, 2004) and detergents (Denton, 2003).
These domains all use concepts that are already divisions which are to a great extent obvious under cursory examination – colour of socks chosen to match clothes, type dependant upon purpose (i.e. sports, formal). If the idea of facet analysis is that documents should be treated as the source of concepts, as opposed to the idea that concepts are identified according to a pre-defined classification scheme, the example domains are not indicative of this approach. It is recognised that the use of such “simple” domains is to a great extent intentional; as they usefully serve for purposes of illustration, however by nature they do not fully explore the potential or difficulty of facet analysis. For example, a cursory examination of socks would indicate that different colours exist, and hence the colour of a sock would be used as a descriptive attribute in the knowledge that this distinguishes it from other socks. The principle of division (in effect the condition used to create a hierarchy) is therefore identified prior to extracting concepts.

Furthermore, these domains have a strong physical presence. Their tangibility is more apparent and readily suggests the elements for a facet classification. In turn it is often the more objective properties of things (e.g. colour, region) than the more interpretative aspects that are considered in faceted classifications (taste of wine). In the case of the classification of fine arts and of Nobel Prize winners as conducted within the Flamenco environment (Hearst, 2006) both schemes utilise readily discernable attributes of each entity, nationality, gender, affiliation, year of award and award type in the case of Noble Prize winners and attributes such as media, artist, location in the fine arts classification. These aspects are the principal constituents of the free-text descriptions (such as curator’s labels) used to automatically generate the structure as described in previous sections. Such attributes are useful for retrieving known entities, however they do not distinguish why a piece of art is significant, or to what movement the art belongs (such as renaissance, modern, medieval, etc.) or what contribution a Nobel Prize winner was commended for. For example, the Mona Lisa would be classified as a painting (media) of a woman in a mountainous setting (heaven and earth) in Europe (location) by Leonardo Da Vinci (artist). In contrast, it is posited that an expert on art may have identified (amongst others) the concepts of “Sfumato” technique[15] or “Renaissance” as the period in which it was created. These aspects, whilst arguably a more detailed treatment of the piece, are not readily ascertainable and it is argued are likely to be of great diversity across the entire collection.

The schemes developed in these two examples use harder more objective aspects such as country of origin (for both Art and Nobel Prize winners), media (art), affiliation (Nobel Prize), and award type (Peace, Chemistry). These are entirely legitimate aspects upon which to classify, however they are clearly ascertainable[16]. Facets that may provide some illumination of the subject, and which may provide more useful clues in retrieval, are perhaps less readily identifiable. This contrasts with Star’s (1998) comparison of both facet analysis and Grounded theory as discovery processes. The utility of the resultant classification schemes is subject to opinion[17], we merely wish to reinforce the point that the generation of faceted-like structures in practice tend to move away from rigid adherence to facet analysis as proposed by the library community.

4.5 The impression that facet recognition is unproblematic!
Someone new to the domain could be forgiven that there is an assumption within the facet community that concept recognition is unproblematic, both empirically and
theoretically. A corollary of this is the assumption the world is logically ordered and coherent and that mess is due to lack of rigour in the processing or consideration of the facets. It is possible that the mess may be due to the multifaceted nature of the world, with overlapping categories and concerns (Kent, 1978; Law, 2004). Little work seems to challenge this although work by Bowker and Star (1999), Thellefsen (2004), Margolis and Laurence (1999) and Worsley (1997), provide theory and empirical evidence on the difficulty of building both local and generic taxonomies.

Bowker and Star (1999) point out to the situated nature of classification. That is to say that they are contextual, and that a classification is never neutral or objective, despite claims by some (arguably those with vested interests) that it is. Collections such as Margolis and Laurence (1999) consider the complex nature of both the definition of concepts and also the difficulty of recognising them.

4.6 Is the process purely bottom-up or top-down?

Within Section 3, we noted that the faceted description of engineering documentation emerged from top-down and bottom-up concerns. In some cases the facet was identified from the document corpus, but its range was derived from existing theoretical concerns (Malone et al., 2003). In other cases this situation was reversed. We draw parallels with Carroll and Rosson’s (1986) characterisation of design as holding the potential to be radically transformational, being bottom-down, top-up. The “space” is transformed by the activity. In our context, our understanding is transformed by our engagement with the documents and pre-existing ideas about them. A classification, whether faceted or not, can be seen to be a hypothesis about the nature of both a domain, and what people perceive as relevant by the domain (Woods, 1998[18]).

5. Conclusions

Facets are a powerful way of expressing the complexity of knowledge. Their emergence in multiple contexts (Beghtol, 1995) is testament to this. However, as with many others concepts and methods, it can be used in different ways with different implicit and explicit philosophical assumptions.

With the rise of the internet faceted techniques have gained wider exposure. There is the view that existing classifications such as BC2 would be a powerful addition to the internet (Broughton, 2001, 2002). For better or worse we do not have any indication that this is being used. People are applying faceted notions in a range of contexts. However, our experiences of developing facets outside of the library suggest that the application of faceted techniques is not without problems (Broughton and Slavic, 2007). Our case study generated observations and reflections on the application of faceted classification; namely that differing interpretations of the facet notion exist; there is confusion between faceted analysis and faceted classification; there is a lack of methodological guidance; the use of simplistic domains as exemplars; the assumption that facet recognition is unproblematic; and that the process is both top-down or bottom-up. We make a more general call for similar reflections in other “documentation” domains that have endeavoured to generate faceted classifications.

Notes

1. Although developed for the Library of Congress collection the LCC is used extensively in many research libraries.
2. For example, the religion category originally assigned classmarks 220-289 to Christianity and only 292-299 to all other religions combined (although recent revisions have addressed this to some extent).

3. Auxiliary tables exist as systematic, covering all documents, or specialist, covering only a specific subset or subject.

4. The last UK library to use Colon switched to an enumerative scheme in 2001 (Broughton, 2001).

5. Objective properties concern knowledge’s invariance, distinctiveness and controllability. Subjective properties concern knowledge’s individual utility; simplicity; coherence; and novelty. Intersubjective properties of knowledge concern publicity, expressivity, formality, collective utility, conformity and authority (Heylighen, 1997).

6. Dewey introduced the idea of relative location (Broughton, 2004). Whereas previous notation systems indicated where a book was to be physically located, Dewey’s notation indicates where it fits within the overall scheme. In this manner books can be relocated or additional books included without unduly influencing the notation, and the notation itself serves to indicate the degree of conceptual relativity of two books.

7. Standard Information retrieval techniques, such as stopword removal (the deletion of articles, for example) and the removal of any inflection (for example plural terms), were used.

8. Exceptions would be within the “personal” document collections of esteemed figures.

9. Although Ranganathan did in some way predict this “621: Canon of Classics: A Scheme of Book Classification should have a device to bring together all the editions, translations, and adaptations of a classic, and next to them all the editions, etc. of the different commentaries on it, the editions, etc. of a particular commentary all coming together, and next to each commentary all the editions, etc. of the commentaries on itself in a similar manner (commentaries of the second order), and so on”.

10. An additional problem with Wilson’s work is that there is no quotation or reference to the original work of Ranganathan.

11. Their arguments can summarised as: immediate effectiveness of the scheme over exhaustiveness; they have more rapid feedback mechanisms; their domains are more dynamic and driven by business objectives; information-seekers are looking for answers, not documents; speed of information response; and people are indexed too.

12. Three versions of Grounded theory can be discerned. The original, still championed by Glaser, which is general classed as Positivist, Strauss’s work suggests a number of conceptual tools with which to analyse data, this is still grouped under the label positivist but should probably be more fairly classed as post positivist. More recently Charmaz (2000, 2006), has put forwards a constructivist “version” of Grounded theory (Seldén, 2005).

13. Methods are the techniques or procedures used to gather and analyse data related to some research question or hypothesis. In contrast, a methodology is the strategy, plan of action, process, or design lying behind the choice and use of particular methods and linking the choice and use of methods to the desired outcomes. (after Crotty, 1998).

14. Hjørland (2002a) notes that domain analysis with computer science tends to be restricted, but even within their relatively naïve view Booch (1994, p. 145) notes “there is no golden path to classification”. and devotes an entire chapter to classification.

15. The painting technique perfected by Da Vinci, of which the Mona Lisa is the most widely known example (Capra, 2007).

16. The fine arts classification includes some aspects of the subject matter of a piece of art, for example what objects, animals or structures the piece of art depicts.
17. In terms of the fine art classification, it is argued that the generated schemes are useful to someone with a reasonable understanding of the field, whereas schemes which seek to provide greater insight might be more useful to those unfamiliar with the domain.

18. Since this paper was submitted Hjørland (2008) has noted that classification involves inductive, deductive historic and pragmatic processes. However, we see his categorisation as idealised, much empirical work can be pragmatically and historically, for example action research and ethnography and blend elements of deduction and induction.

References


Morville, P. (2005), Ambient Findability, O’Rielly, Sebastopol, CA.


Ranganathan, S.R. (1965), The Colon Classification by S.R. Ranganathan, Graduate School of Library Service, New Brunswick, NJ.


**Corresponding author**

Peter J. Wild can be contacted at: pw308@cam.ac.uk

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints