INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED'07
28 - 31 AUGUST 2007, CITE DES SCIENCES ET DE L'INDUSTRIE, PARIS, FRANCE

THE USE OF FACETED CLASSIFICATION IN THE ORGANISATION OF ENGINEERING DESIGN DOCUMENTS

M D Giess[1], P J Wild[2] and C A McMahon[1]
[1]University of Bath, UK
[2]University of Cambridge, UK

ABSTRACT
The adoption of a product service or support model in place of product provision has implications upon the strategies employed in the organisation and retrieval of engineering information. In order to ensure that such organisational schemes endure over the extended service time spans, the structure of these schemes must cater for all engineering viewpoints and provide the context necessary to comprehend the information content. Traditional, enumerative classification schemes generally cater for a limited viewpoint or perspective, however faceted classification addresses this by concurrently describing multiple aspects or facets of a given document and allowing retrieval to proceed via browsing of a given facet or combination of facets.
Faceted classification was first formalised within the field of library science in the early 20th century, and whilst notions of a faceted structures have been appended into primarily enumerative schemes there are few implementations of pure faceted models, arguably as within a library it is not possible to reorder physical documents according to a given facet of interest. Electronic documentation has no such physical restriction, and this has fuelled a resurgence in interest in faceted classification. Whilst the concept of faceted classification as a means of browsing is generally understood; the literature specifying how such a scheme should be created for a given domain is sparse, and distributed across different domains and eras. This paper reviews this literature and proposes a means by which the underlying theory may be more readily applied to engineering design documentation.

Keywords: Information management, knowledge management, faceted classification

1 INTRODUCTION
The product-service business model, in which customers receive the *service* of a product for a set period in place of simple product receipt, has many attractive qualities, such as increased opportunities for innovation in service and greater control over operation and intellectual property. Hence, many engineering companies are moving away from traditional product provision models to such a support model. Such a change has implications upon the documentation practices of a company. Whilst retrieval of information is important in any context, the extended timescales over which products must be supported can exceed the careers of engineers, hence documented information must be retrievable and comprehensible to engineers not involved in its creation.
When considering the computer-based retrieval of such document-based information, the use of term-based querying is predicated around the ability of engineers to formulate appropriate queries. Term-based querying is the return of a set of documents which are perceived by a retrieval engine to meet an information need expressed in a set of terms entered by an engineer. The lifespan of documents in the product–service model are such that a requisite understanding of information content and terminology may be lost. This has further implications upon the interpretation of document-based information. Miller [1] notes that information is an expression of a person's understanding which, when conveyed, is subject to both interpretation and misinterpretation. This problem may be increased when considering both the lack of familiarity many engineers may have with the information and the lack of context afforded by term-based querying.
As an alternative to term-based querying, pre-organisation of documentation into meaningful structures provides a means of browsing and of describing the context of the information. The

information content, terminology and context may thus be expressly indicated to the engineer by the scheme itself. It is important, however, to ensure that a scheme does not only reflect the perspectives of the users involved in its creation. Traditional classification schemes, as used in many organisational systems, all induce viewpoint-dependency to a certain extent as the concepts expressed within the scheme and the resultant structure may only be pertinent to a specific subset of engineers. This research is interested in identifying means by which this may be addressed.

1.1 Information Classification

Traditional enumerative schemes (in which domains are partitioned into smaller subsets until the content of a given subset describes a consistent topic) impose a single viewpoint. Enumerative schemes partition by identifying principles of division by which the domain is sub-divided. However, engineering documents generally describe compound subjects (for example 'finite element stress analysis of titanium strut' describes: 1) the objective; 2) method of analysis; 3) the part in question; and 4) its material. The choice of principle of division is significantly viewpoint-dependant. Any choice of primary classification principle to will steer the scheme towards a particular group of users to the detriment of other groups.

Faceted classification was developed in order to more rigorously deal with compound subjects, and is an approach most associated with Ranganathan (e.g., [2]). This approach also addresses viewpoint dependency as it is no longer necessary to give greater emphasis to a single primary classification principle as is the case for enumerative schemes. Faceted classification addresses viewpoint concerns by allowing an entity to be classified according to all concepts which it describes. By associating concepts in different facets, compound subjects such as 'manufacturing statistics' can be concisely depicted, and those interested in either engineering or statistics can ascertain where within their discipline the entity resides.

Previous research by the authors led to the creation of the Waypoint system, a faceted classification software tool. Waypoint allows for rapid browsing of a faceted scheme and of rapid retrieval of all documents of interest. This has been reported in detail elsewhere [3]. This research sought to identify how enduring faceted schemes may be created for use in this tool. This paper focuses upon the creation of faceted classification scheme for the organisation of engineering design documentation.

2 REVIEW OF LITERATURE

2.1 Types of Classification

Broughton [4]and Rowley [5] identified two distinct types of classification scheme, the *enumerative* and the *analytico-synthetic.* Enumerative schemes divide the document corpus into ever smaller classes according to identified principles of division. This may be considered 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 within a classificatory structure. As this approach seeks to elucidate the range of concepts within a corpus before proceeding to develop the classification scheme, this approach may be considered bottom-up as the scheme is constructed from the basic elements of the corpus.

2.1.1 Types of Classification Scheme

Many early classification schemes were developed on an ad-hoc basis, for example the Library of Congress Classification (LCC), which has weak theoretical principles ([4], p. 145), whereas the more frequently implemented Dewey Decimal Classification has greater theoretical rigour The Universal Decimal Classification (UDC) is a development of Dewey, adapting the class structures and adopting the decimal notation, but most significantly it differs from the enumerative Dewey by virtue of its synthetic nature. In DDC only a singular classmark (a codification indicating the relevant class to which a document belongs) can be assigned to a document, whereas UDC allows classmarks to be both combined and for additional, pre-determined classmarks to be appended. Additionally, common class marks are provided as auxiliary tables which allow consistent reference to frequently occurring concept, for example language of publication. This notion that documents may be described and classified by combination of common concept types forms the essence of analytico-synthetic

classification, where each document is described by a compound of a number of different concepts. This has been implemented in the library field as faceted classification, with consistent types of concept being arranged as a facet.
Ranganathan's Colon Classification [6] and the second Bliss Classification (BC2, [7]) are both faceted classifications. Ranganathan's Colon Classification was first published in 1933, and the ideas generated in this work were revisited in the 1960s by the Classification Research Group which led to the generation of BC2 in the late 1970s.

2.2 Creating classification schemes

The specification of a classification scheme is a difficult task, requiring considerable expertise as well as intellectual and manual effort [2, 8-11]. It is not possible to construct an overarching classification scheme that will cater for all requirements, as Broughton [4] states "….it simply isn't possible to create a classification that is truly objective or neutral or absolutely correct" Louie e*t al.* [12], when considering the application of facets within information architecture, recognise the complexity of human knowledge, and when considering document-based representation of knowledge stating "..all attempts to describe or organize [sp] these documents are flawed as well (p.1)."
As noted previously, a faceted classification scheme is generated in principle in a bottom-up manner, as opposed to enumerative approaches, which are top down. The generation of a faceted scheme follows a different approach to an enumerative scheme, and it is was a goal of this research to identify how such a scheme may be created for engineering design documentation.

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 *warrant*. Hulme [13] 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 NISO [14] (defining the enumerative development of a monolingual thesaurus) as "..*Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of frequent requests for information on the concept or free-text searches on the term by users of an information storage and retrieval system.*" Other interpretations of warrant exist. Beghtol [15] 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 Classification Research Group took the position that literary warrant be based more upon the terminology of a field [16]. 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 (either user or scientific warrant).

2.3 Specification of faceted classification scheme

Facet Analysis was proposed by Ranganathan as the mechanism by which faceted classifications can be constructed, however there appears to be little definitive literature describing this method. The following section presents various interpretations upon this theme, from the Library and Information Science domain as well as more practical implementations.
Ranganthan's work is, unfortunately, written in language that makes it somewhat impenetrable. A key part of this work is the Postulate of Five Fundamental Categories, which argues that any subject can be divided into five fundamental categories or types of concept that each individual in general shares and with which any individual could be described. These fundamental categories are described by the acronym PMEST, representing Personality, Matter, Energy, Space and Time. The Classification Research Group [17] revisited this work, suggesting 13 fundamental facets were pertinent. However, Hjorland notes that "*Vickery's expansion of the number of fundamental categories may imply that there is not a fixed set of categories in the world*" [18].

The CRG also revised the prescriptions given by Ranganathan, providing a reduced set of guidelines for the qualities of facets. Recently Spiteri [19] argued that both approaches suffer from practical difficulties, and went on to present a simplified model that unifies the ideas presented by both groups into a rather more pragmatic form. Whilst aimed at Library and Information Science (LIS) students she suggested "*the model could be used by designers of faceted classification systems*."
A number of authors discuss the application of Facet Analysis [4, 20, 21], however such analyses are typically expressed in basic terms. There are several web based discussions about the application of the notion of faceted classification to contexts such as company websites, online catalogues etc, of which Denton [22] and Fast et al [23] are notable examples. In general, each adheres to the stages indicated in Table 1.

Table 1 Activities in Facet Analysis

Stage
Collate representative corpus of documents
Identify discrete concepts that together describe document in entirety
From identified concepts, evolve facets
Structure facets in terms of citation order
Codify facets – develop bookmarks

2.4 Electronic Implementations of Faceted Classification

Where the application of faceted classification in a library context involves some notation such that items may be physically placed, electronic resources impose fewer restrictions on location. They lend 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 when considering physical documents. A number of electronic applications of ostensibly faceted systems have been implemented and will be briefly reviewed below, with their conformance to the theory developed within the library field commented upon.

2.4.1 Faceted Classification Software Tools and their Implementation

A number of software tools facilitate the use of faceted structures, however this review will focus upon tools which cater specifically for faceted classification. Alongside the Waypoint system, the Flamenco system [24] generated at the University of Berkeley allows faceted classifications to be constructed and used for browsing without restricting the nature of the classification scheme itself. This reflects the notion that it is not always possible to decompose a corpus of documents into consistent types of concept. The FacetMap tool takes a notably different approach, where each entity must appear once and only once in each facet. Wilson [25], 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."* The appeal that such an approach is a subscription to Ranganathan's theory is not further justified or discussed, despite the presence of numerous interpretations of Ranganathan's theory within literature. Wilson demonstrates this strict faceted classification upon wine, where facets such as grape and region are identified [25].
Example applications of Flamenco include a classification of Nobel Prize winners and a classification of a Fine Arts museum collection, and literature describing these examples indicate the means by which the scheme itself was generated. When developing the classification of fine arts pieces [as discussed in 26] terms are automatically extracted from free-text descriptions of the content of each piece of art by a curator. These terms are automatically arranged into a hierarchy with some manual refinement [27]. This is done by considering the grammatical relationships between terms as defined in the WordNet lexical database[1], although such arrangement can also be carried out manually.
Such an approach relies upon the suitability of the description of each piece for such purposes, as the description may not be designed to be discriminating or particularly comprehensive. Both the Nobel Prize and Fine Art schemes utilise readily discernable attributes of each entity, nationality, gender,

[1] http://wordnet.princeton.edu/

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 are the constituents of the free-text descriptions used to automatically generate the structure as described previously. Such attributes are useful for retrieving known entities, however it is questionable whether they provide suitable means for retrieval when the domain and constituent entities are unfamiliar. For example, it is not possible to find which recipient received a Nobel Prize for a given achievement, or to find all pieces of art associated with an artistic movement, unless the user has some prior knowledge of the content or context of the entity or entities in question.

Such a consideration is important in an engineering context, where increased emphasis upon product service has acted to increase the lifespans over which engineering information must be maintained and made available. In such a situation it is entirely possible that the information will outlast the careers of those who generated it [28], in which case it must be assumed that those seeking to retrieve it may have limited prior knowledge of either the content of the information resource or the context in which it was created.

2.4.2 Commercial Applications of Faceted Classification

Numerous applications of classifications using a faceted-like structure may be seen on the Internet, of which many (although by no means all) are retail sites, with 69% of eCommerce sites reviewed by Adkisson [29] incorporating some notion of facets. Although faceted-like structures may be seen within many websites, many such sites do not refer to facets by name when describing their organisation. A notable example which specifically claims a faceted structure is eBay Express[2]. As perhaps to be expected of a commerce website, eBay Express classifies physical artefacts as opposed to information entities, which are argued to have more clearly discernable characteristics (or concepts) than the less tangible and more abstract information content of documents.

2.5 Comments upon Approaches to Faceted Classification

A number of issues pertinent to the creation of a faceted classification scheme have been noted by the authors. These issues are not intended as a criticism of the theory underpinning faceted classification, more as a commentary upon issues in the practical application of the theory in areas outside of library science. This section discusses these issues in turn, with the intent of identifying the relevant considerations that must be entertained when generating a faceted scheme for the classification of engineering documentation.

2.5.1 Lack of Methodological Guidance

Ranganathan's Canons, Postulates and Principles and Spiteri's later revisions define the requisite properties that a faceted structure should have, but do not provide a methodology that may be followed in order to arrive at this structure. In essence, many Library Science texts are evaluative as opposed to generative. The methodologies expressed in the more applied texts tend to provide broad overviews, for example Denton [22] suggests that once a representative set of documents have been decomposed into concepts the practitioner should *"examine the resulting terms and see what general, high-level categories appear across all the entities. Study them and narrow them down into a set of mutually exclusive and jointly exhaustive facets…"*. This places great reliance upon the ability of the practitioner to distinguish high-level categories, and it is argued that in complex domains with vast numbers of documents and concepts such as engineering this approach cannot be guaranteed to be successful. As noted in the following section, Denton uses a relatively simple example in which to illustrate this approach, hence the extensibility of the approach to more complex domains is not indicated.

There is a dearth of literature discussing practical applications and examples of facet analysis (the method of construction of faceted classifications, as discussed in earlier sections), hence how such concerns are addressed in practice is not clear. A number of schema have been published, for example the Art and Architecture Thesaurus [30] uses a faceted structure and many library schema are also faceted (most notably Colon and Bliss/BC2). There has been a widespread failure to explain the empirical or analytical methods employed in the construction of these schemas.

[2] http://www.express.ebay.com/

2.5.2 Nature of the Domain

Those texts that seek to provide some means by which to generate a faceted scheme have, for purposes of illustration, selected relatively simple domains as examples, for example socks [4], wine [31], or detergents [22] There is a tendency to present this work as rationalistic and analytic, whereas much of it is generally a rendition of common sense notions of such simple domains where the boundaries and indeed contents of the sub-domains are clearly ascertainable. Dowell and Long, two leading proponents of domain modelling, note that generally a complete understanding and formalisation of a domain cannot be achieved ([32], p. 130), and similar notions have been expressed by other domain-oriented researchers [33, 34].

In the examples given in literature, those of socks, detergents and wine, it is arguable that the requisite characteristics could be ascertained by cursory examination of the contents of the domain or by a prior understanding of the scope of the domain, feasible for domains of that size but infeasible for domains such as engineering. Not only does the limited size of the domain play a part, this is arguably because each domain has a strong physical presence. In the wine example, the concepts identified were those that are typically currently used by many retailers to arrange wine, for example grape, vintage and region, and did not venture into the more abstract areas such as taste. The remaining examples also utilise concepts that are readily distinguishable and, more importantly, of a character that is already known to exist in other entities in that domain. Such a guarantee ensures that contrastable concepts, of an identical level of granularity, are identified for each entity within the corpus.

In a number of cases it may be argued that the selected facets were those that were readily identifiable, but not those that may be most descriptive or useful. In the case of the fine arts classification facets such as media, artist, shapes, colours and animals were identified and in the case of Nobel Prize winners the nature of the prize, nationality of winner and year were used. Less tangible aspects such as artistic movement and contribution of the Nobel Prize winner were not identified, although they may prove more useful facets upon which to browse the two collections. The use of readily identifiable characteristics of entities contrasts with Star's [35] characterisation of facet analysis as a discovery process akin to grounded theory.

2.5.3 Facet recognition

There appears to be an assumption within the faceted classification community that concept recognition is unproblematic, both empirically and theoretically. This implies that the world is logically ordered and any errors, frailties or problems in depiction in this world are the fault of the practitioner. It is perhaps more accurate to state that the difficulties in depiction are due to the multifaceted nature of the world, with overlapping categories and concerns (see also [36]). Little work seems to challenge the assumption of the faceted classification community although work by Bowker and Star [37], Thellefsen [38] and Margolis and Laurence [39] furnish much theory and evidence on the difficulty of building both local and generic taxonomies.

Bowker and Star [37] point to the situated nature of classification, being a product of a specific time and place, and that a classification is never neutral or objective, despite claims to the contrary. Collections such as Margolis and Laurence [39] consider the complex nature of both the definition of concepts and also the difficulty of recognising them. Documents serve as a representation of events and activities in the world (c.f., [40]), being subject to revised interpretations in the face of improved understanding of the context of their creation, the motivations, knowledge and means of expression of the author. In this respect, facet analysis cannot be guaranteed to identify concepts which are universally accepted definitions of a document, these being subject to the perspective of the practitioner.

2.6 Reflections upon interpretations and applications of faceted classification

In general, implementations such as Flamenco do not adhere to the notions of faceted classification put forward within the library community. In particular, there seems to be little emphasis put upon an inspection of the corpus for purposes of identification of constituent concepts, instead readily identifiable attributes are selected. The use of enumerative classification prior to the implementation of faceted structures at the sub-class level suggests that either the utility of a corpus-wide faceted scheme is deemed unnecessary in certain applications, or that identification of such facets is a non-trivial task.

It is significant that the practical applications of faceted classification outside of the library field take a relatively loose stance, as witnessed in the eBay Express and Flamenco examples, where there is a mix of enumerative and faceted 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.

2.6.1 Differences between Domains

Libraries tend to have uniform collections composed predominantly of books and journals. These entities are formally 'signed off' artefacts, with a clear title and target audience and, essentially, a relatively specific target discipline. There is also limited occurrence of more informal genres such as correspondence, logbooks, notebooks etc., in the actual catalogue.

This contrasts sharply with engineering design information, where it may be that only that held in specific organisational devices such as Product Lifecycle Management (PLM) systems is formally signed off. A significant proportion of engineering documentation is the informal, communicative documentation such as email, and the working documentation such as analysis files. Such files are essential to documenting the engineering activity, as they support the more formal documentation held in PLM type systems. 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.

The case study described in the following section covers the authors' approach to describing engineering design documentation such that a faceted description may be evolved. A key aspect of this approach, when considering the differences between the library and engineering domains, is that a description necessarily considers both content and context as it is this context which will not readily be apparent from the document content itself.

3 DESCRIBING AND PROFILING ENGINEERING DOCUMENTS

The case study described here forms part of a larger research effort into engineering information context, of which aspects have been reported elsewhere [41], all of which share a corpus of documents obtained from an engineering company. The research presented here seeks to provide a means by which the document corpus may be browsed. The company in question has approximately 60 employees and a turnover of £5 million, split between a engineering services division and a design and manufacturing division. Analysis of the document corpus obtained from this company indicates over 250 different distinguishable document types or genres (c.f., [42]), with as few as one or as many as thousands of instances of each type. In order to support browsing of these documents there is 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 as much information about the context and manifestation of the document corpus as possible. Unless recorded, much information about the document will not be self-evident to someone examining the documents in corpus in a few months time, much less over decades as may be necessary.

The method of facet analysis has not been fully espoused to the satisfaction of the authors, much of the library science literature focusing upon evaluation and all practical examples dealing with closed sub-domains. As such, the work described here documents the authors' approach to facet analysis applied to engineering documentation. The evolved approach is based in practice, being grounded in a document corpus and being empirical in nature. In order to perform the analysis two main considerations are key, that the evolution of facets is not exclusively bottom-up (as per analytico-synthetic approaches) and that a relatively broad interpretation of facets and facet analysis is used. The relatively restricted domains used as examples in the more tutorial-based material on faceted classification (e.g., [19, 43]) allows a 'pure' implementation of facet analysis to be followed, with attendant concerns when expanding to the wider engineering domain. In this approach the evolution of facets owes as much to browsing mechanisms (e.g., [44]) as it does to do the varying interpretations of Ranganathan (e.g., [25, 35, 45]).

3.1 Description versus analysis

This research should perhaps be considered as faceted description as opposed to facet analysis, however it is intended to use the descriptions as a basis for evolving a faceted classification. Here we view description as reporting characteristics, in contrast analysis pertains to a resolution into 'simple' elements. We are not making a case for neutral description, a user interested in typographic issues would have a different set of descriptions possibly relating to 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.

3.2 Empiricism and Grounded Theory

Star [35] drew extensive parallels between Ranganathan's work and grounded theory, and whilst this study is influenced by this comparison it is not a grounded theory analysis. The meaning of the documents does not exist purely in the documents themselves, to be generated by a researcher without preconceptions. The work was focussed around describing the qualities of documents, and was heavily 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 in to provide other's insights about documents and the processes (e.g., [42, 46, 47]).

3.3 The Resultant Faceted Scheme

Table 2 presents the facets and illustrates the approach taken in their identification. A number of observations were made during this construction, which to some extent run contrary to the idea that a faceted scheme should be created purely from a bottom-up approach.

3.3.1 Analysis followed by Synthesis

There is often a concern to express classification as being either a top-down or bottom-up process, in essence this is one of the key distinctions between enumerative and analytico-synthetic classification. The development of this work has utilised aspects of both approaches. As understanding of the document corpus has improved facets were moved, omitted, refined and (as new documents are identified) further facets have been added.

In some cases a facet was identified from the corpus, but its range derived from existing theoretical concerns (e.g., [46]). In other cases this situation was reversed. Parallels may be drawn with Carroll and Rosson's [48] characterisation of design as holding the potential to be radically transformational, being bottom-down, top-up. The 'design space' is transformed by the activity. In this context our understanding is the design space, and is transformed by engagement with the documents and pre-existing ideas about them. A classification, faceted or enumerative, can be seen as a hypothesis about the nature of a domain and of what people perceive as relevant by the domain (c.f., [49]).

A secondary point to be made relates to the collation of a representative corpus of documents prior to undertaking facet analysis. As the acquisition of documentation, and perhaps more pertinently the understandings of the exact meaning and context of such documentation, takes place over an extended period, the number and form of identified concepts tends to adapt and increase over time. This incremental change suggests that it is not feasible to take a completely bottom-up approach, as the dynamic nature of the document corpus means that the range of extracted concepts only hold for a given point in time. Only by revisiting and updating the scheme can new documentation (and resultant new concepts) be introduced.

3.3.2 Subtle differences with user and literary warrant

Literary warrant dictates that a classification scheme is developed via consideration and analysis of the documents within the corpus. It is argued that such analysis does not proceed without some understanding of the needs of the users, and as such there is a degree of user warrant involved. In essence, it is not considered feasible to consider evolving facets from observations of the document corpus as exclusively literary warrant, as the preconceptions of the practitioner play a large part in identifying salient features of each document (this point was discussed previously, however it is useful to reinforce it here). This supports the notion, as expressed by Broughton (2004), that it is not possible to create a classification scheme that is absolutely correct, and that the application of facet analysis cannot be guaranteed to give repeatable or optimal results.

Table 2 Top-Down (TD) and Bottom-Up (BU) derivation of Facets

	Facet	Range	No [3]	NOTES AND EXAMPLES
Source	BU	BU		E.G., TrollCo
Purpose	TD & BU	TD	6	Genre theory, earlier discussion with research partner, stated purposes in ISO documentation and other internal guides to documents.
ISO 9000 Type	TD & BU	TD	7	BU because it was ISO accreditation at TrollCo, TD because the categorisation existed in ISO 9001, that is the range from ISO
Product Phase (after BS 7000)	TD	TD & BU	9	One of the range came from TrollCo the remainder were from the BSI documentation
Document Status	BU	BU	6	
Review Status	BU	BU	2	
Distribution Status	BU	BU	4	
Main Manifestation	BU	BU	4	Physical, Electronic, or more of one than another.
Physical Manifestation Mechanism	BU	BU	6	
Electronic Manifestation Mechanism	BU	BU & TD	10	Contents was in part from the tools at TrollCo but in part from other tools that are available
Grouping Status	BU	BU	2	Single or grouped.
Physical Grouping Mechanism	BU	BU & TD	10	Some of the elements were taken from TrollCo, some from catalogue, we expect the number to increase
Electronic Grouping Mechanism	TD & BU	BU	4	Once the physical grouping element concept was come upon, the electronic version came into being
Template Status	BU & TD	BU	6	
Annotation	BU	BU	2	
Paper Interface	TD & BU		3	Work on genres and the nature of the forms in existence, novel concept, but one that builds on notion of UI
Descriptions - Main Descriptions	BU	TD	5	The general notion of there being documents that are 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 [50], as they are generic enough to describe key design documents
Descriptions - Description Appendices	BU	TD	5	
Utilisation documents	BU	TD	8	
Development Plans	BU	TD	12	
Quality Control Documents	BU	TD	11	
Administrative Documents	BU	TD	6	
Functional Concern	BU	TD	13	The general idea for the facet was developed bottom up. Some documents can be perceived as boundary objects ([51]), or cross-role documents. The facets are derived from the MIT process-modelling handbook[46] - serves as a repository of knowledge about organisations and provides case examples in a common format and framework. These 'abstract' across naming idiosyncrasies of organisation's different functional divisions. By representing as non-discrete facets, we can a) account for documents embedded in multiple functional divisions, and b) profile differences in types between organisations.
Class Related Facets				

[3] Most range contains 2 options for Not Applicable and Not Known, these are excluded from the count.

CONCLUSIONS

Facets are a powerful way of expressing the complexity of information. 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 and electronic means of organisation, faceted techniques have gained wider exposure. However, our experiences of developing facets outside of the library science domain suggest that the application of faceted techniques is not without problems, some of which stem from inadequate descriptions of the methods employed in generating such a scheme and others from the innate differences between the domains. This paper has sought to identify where the theory of facet analysis is insufficient for application in an engineering context, and further to this has proposed a means of performing facet analysis upon an engineering document corpus. As the theory of facet analysis and faceted classification has received little attention in this domain, the authors have detailed their reflections upon both theory and practice in this regard.

ACKNOWLEDGEMENTS

The work reported in this paper has been supported by the Engineering and Physical Sciences Research Council (EPSRC) under grant reference EP/C534220 and also as part of the EPSRC Innovative Manufacturing Research Centre at the University of Bath (grant reference EP/E00184X/1). It has also been supported by a number of industrial companies. The authors gratefully acknowledge this support.

REFERENCES

[1] Miller, F.J. I=0 (Information has no intrinsic meaning). *Information Research*, 2002, 8(1).
[2] Ranganathan, S.R. *Prolegomena to Library Classification*. Asia Publishing House, New York, 1967.
[3] McMahon, C., Lowe, A., Culley, S., Corderoy, M., Crossland, R., Shah, T. and Stewart, D. Waypoint: An Integrated Search and Retrieval System for Engineering Documents. *Journal of Computing and Information Science in Engineering*, 2004, 4(4), 329-338.
[4] Broughton, V. *Essential Classification*. Facet Publishing, London, 2004.
[5] Rowley, J.E. *Organising Knowledge*. Gower, Aldershot, Hants, UK, 1992.
[6] Ranganathan, S.R. *The colon classification by S.R. Ranganathan*. Graduate School of Library Service, New Jersey, 1965.
[7] Broughton, V. Faceted Classification as a Basis for Knowledge Organization in a Digital Environment; the Bliss Bibliographic Classification as a Model for Vocabulary Management and the Creation of Multi-Dimensional Knowledge Structures. *The New Review of Hypermedia and Multimedia*, 2001, 7, 67-102.
[8] Kwasnik, B.H. The Role of Classification in Knowledge Representation and Discovery - 1. *Library Trends*, 1998, 48(1), 22-47.
[9] Lowe, A. Studies of information use by engineering designers and the development of strategies to aid in its classification and retrieval. Thesis (Ph.D.). Department of Mechanical Engineering, University of Bristol 2002
[10] Morville, P. Ambient Findability. O'Rielly, Sebastopol, 2005.
[11] Star, S.L. Grounded Classification: Grounded Theory and Faceted Classification. In *Information Systems and Qualitative Research Conference.* Philidelphia, May, 1997.
[12] Louie, A.J., Maddox, E.L. and Washington, W. Using Faceted Analysis to Provide Structure for Information Architecture. In *ASIS&T Information Architecture Summit.* Portland, Oregon, March 21-23.
[13] Hulme, E.W. Principles of Book Classification. *Library Association Record*, 1911, 13(October), 354-358.
[14] *ANSI/NISO Z39.19-1993, Guidelines for the construction, format and management of monolingual thesauri*, 1994
[15] Beghtol, C. Semantic validity: Concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, 1986, 30(2), 109-125.

[16] Hjørland, B. Literary Warrant [online]. 2006. Available from: www.db.dk/bh/Lifeboat_KO/CONCEPTS/literary_warrant.htm [Accessed October 2006]
[17] Vickery, B.C. *Faceted Classification: A Guide to Construction and Use of Special Schemes*. ASLIB, London, 1960.
[18] Hjørland, B. Facet, facet analysis and the facet-analytic paradigm in knowledge organization (KO) [online]. 2006. Available from: http://www.db.dk/bh/facet_and_facet_analysis.htm [Accessed October 2005]
[19] Spiteri, L. A Simplified Model for Facet Analysis. *Canadian Journal of Information and Library Science*, 1998, 23, 1-30.
[20] Rowley, J. *Organising Knowledge*. Gower, Aldershot, England, 1992.
[21] Ellis, D. and Vasconcelos, A. Ranganathan and the Net: using facet analysis to search and organise the World Wide Web. *Aslib Proceedings*, 1999, 51(1), 3-10.
[22] Denton, W. How to Make a Faceted Classification and Put It On the Web [online]. Miskatonic University Press. 2003. Available from: http://www.miskatonic.org/library/facet-web-howto.html [Accessed 1 April 2006]
[23] Liese, F., Fast, K. and Steckel, M. All About Facets & Controlled Vocabularies [online]. boxesandarrows. 2002. Available from: http://www.boxesandarrows.com/view/all_about_facets_controlled_vocabularies [Accessed 1 June 2006]
[24] Hearst, M.A., Smalley, P. and Chandler, C. Faceted Metadata for Information Architecture and Search - CHI 2006 Course. In *CHI 2006.* Montreal, Canada, 22-27 April 2006.
[25] Wilson, T. The strict faceted classification model: an effective alternative to free-form tagging. In *Information Architecture Summit.* Vancouver. (March 23-27)
[26] Yee, K.-P., Swearingen, K., Li, K. and Hearst, M. Faceted Metadata for Image Search and Browsing. In *CHI 2003.* Fort Lauderdale, Florida, USA, 5-10 April 2003.
[27] Stoica, E. and Hearst, M.A. Nearly-Automated Metadata Hierarchy Creation. In *HLT-NAACL'04.* Boston, USA, 2-7 May 2004.
[28] McMahon, C.A., Giess, M.D. and Culley, S.J. Information management for through life product support: the curation of digital engineering data. *Int. J. Product Lifecycle Management*, 2005, 1(1), 26-42.
[29] Adkisson, H.P. Use of Faceted Classification [online]. 2003. Available from: http://www.webdesignpractices.com/navigation/facets.html [Accessed
[30] Trust, J.P.G. *Art and Architecture Thesaurus*. Oxford University Press, New York, 1994.
[31] Wilson, T. Facetmap Gold Demo Server [online]. 2006. Available from: http://www.facetmap.com/browse/ [May 2006] [Accessed
[32] Dowell, J. and Long, J.B. Conception of the Cognitive Engineering Design Problem. *Ergonomics*, 1998, 41(2), 126-139.
[33] Pejtersen, A.M. and Rasmussen, J. Ecological Information Systems and Support of Learning: Coupling Work Domain Information to User Characteristics. In Helander, M.G., Landauer, T.K. and Prabhu, P.V., eds. *Handbook of Human-Computer Interaction*, pp. 315-346 North-Holland, Amsterdam, 1997.
[34] Booch, G. *Object-Oriented Analysis and Design with Applications*. Benjamin Cummings, California, 1994.
[35] Star, S.L. Grounded Classification: Grounded Theory and Faceted Classification. *Library Trends*, 1998, 47, 218-252.
[36] Law, J. *After Method*. Routledge, 2004.
[37] Bowker, G.C. and Star, S.L. *Sorting Things Out: Classification and its Consequences*. MIT Press, Cambridge MA, 1999.
[38] Thellefsen, M. Concepts and terminology reflected from a LIS perspective. - How do we reflect meanings of concepts? In *Proceedings of the 12th Nordic conference for information and documentation.* Aalborg, September 1st-3rd. 68-75
[39] Margolis, E. and Laurence, S., eds. *Concepts: Core Readings*. (MIT press, Bradford, 1999).
[40] Palmer, S.E. Fundamental Aspects of Cognitive Representation. In Rosch, E. and Lloyd, B., eds. *Cognition and Categorisation* Lawrence Erlbaum and Associates, Hillsdale, 1978.
[41] Wild, P.J., Culley, S., McMahon, C., Darlington, M.J. and Liu, S. Towards a Method for Profiling Engineering Documentation. In *Proceedings of the 9th International Design*

Conference DESIGN 2006. Dubrovnik, May 15-18th. 1309-1318
[42] Yoshioka, T., Herman, G., Yates, J. and Orlikowski, W. Genre taxonomy: A knowledge repository of communicative actions. *ACM Transactions on Office Information Systems*, 2001, 19(4), 431-456.
[43] Denton, W. How to Make a Faceted Classification and Put It On the Web [online]. 2003. Available from: http://www.miskatonic.org/library/facet-web-howto.html [Accessed
[44] Hearst, M. Clustering versus Faceted Categories for Information Exploration. *Communications of the ACM*, 2006, 49(4).
[45] Hjørland, B. Empiricism, rationalism and positivism in library and information science. *Journal of Documentation*, 2005, 61(1), 130-155.
[46] Malone, T., W., Crowston, K. and Herman, G., A., eds. *Organizing Business Knowledge: The MIT Process Handbook*. (MIT Press, Cambridge, MA, 2003).
[47] Sellen, A. and Harper, R.H.R. *The Myth of the Paperless Office*. MIT Press, Cambridge MA., 2002.
[48] Carroll, J.M. and Rosson, M.B. Usability Specifications as a Tool in Iterative Development. In Hartson, H., ed. *Advances in Human-Computer Interaction*, pp. 1-28 Ablex, Norwood, 1986.
[49] Woods, D.D. Designs Are Hypothesis About how Artefacts Shape Cognition and Collaboration. *Ergonomics*, 1998, 41(2), 168-171.
[50] Laitinen, K. Document Classification for Software Quality Systems. *SIGSOFT Software Engineering Notes*, 1992, 17(4), 32-39.
[51] Star, S.L. The Structure of Ill-structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In Huhs, M. and Gasser, L., eds. *Readings in distributed artificial intelligence*, pp. 37-54 Kaufmann, Menlo Park, 1989.

Contact: M D Giess
University of Bath
Department of Mechanical Engineering
Bath BA2 7AY
UK
+44 1225 386131
m.d.giess@bath.ac.uk