Faceted Classification: Orthogonal Facets and Graphs of Foci?

Martin Frické

SIRLS, The University of Arizona, 1515 E First St., Tucson, USA, AZ85719
<mfricke@u.arizona.edu>

Martin Frické is an Associate Professor in SIRLS, the School of Information Resources and Library Science, at the University of Arizona, United States. His current research interest involves the intersection between the organization of knowledge, logic, and processing by computer.

Frické, Martin. **Faceted Classification: Orthogonal Facets and Graphs of Foci?** *Knowledge Organization, 38(6)*, 491-502. 50 references.

ABSTRACT: Faceted classification is based on the core ideas that there are kinds or categories of concepts, and that compound, or non-elemental, concepts, which are ubiquitous in classification and subject annotation, are to be identified as being constructions of concepts of the different kinds. The categories of concepts are facets, and the individual concepts, which are instances of those facets, are foci. Usually, there are constraints on how the foci can be combined into the compound concepts. What is standard is that any combination of foci is permitted from kind-to-kind across facets, but that the foci within a facet are restricted in their use by virtue of being dependent on each other, either by being exclusive of each other or by bearing some kind of hierarchical relationship to each other. Thus faceted classification is typically considered to be a synthetic classification consisting of orthogonal facets which themselves are composed individually either of exclusive foci or of a hierarchy of foci. This paper addresses in particular this second exclusive-or-hierarchical foci condition. It evaluates the arguments for the condition and finds them not conclusive. It suggests that wider synthetic constructions should be allowed on foci within a facet.

Received 10 November 2010; Revised 1 April 2011; Accepted 26 May 2011

1.0 Two preliminary distinctions

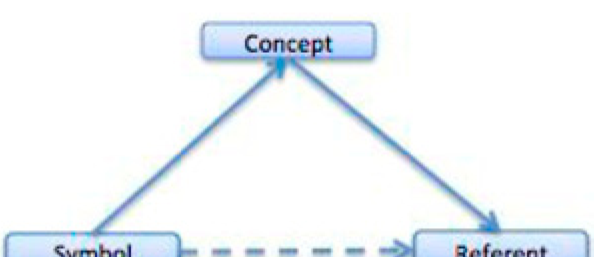
In the realm of knowledge organization, faceting concerns the construction of compound, or complex, or non-elemental, concepts for the purposes of classification. But concepts can play either of two different roles in the organization of knowledge. A classification, i.e., a directed graph of concepts, can either be a classification of things, kinds, processes, and the like, or it can be a classification of subjects, a thematic Baconian 'Tree of Knowledge' (Bacon 1605, 1620). The former is what now would be called an ontology (Smith 2004), and it might be used, for example, for knowledge representation or database design. The latter, in its more general sense, is a directed acyclic graph (DAG) of topics or subjects, and it would be used for annotating or tagging information objects, as, for example, is the practice within librarianship with the *Library of Congress Subject Headings* (*LCSH*) (Broughton 2010b) or the *Medical Subject Headings* (*MeSH*) (Lowe 1994, *MeSH* 2010).

Imagine the owner of an antique shop who also writes books on antiques. The classification scheme that she uses for her antiques might be a synthetic construction from entirely separate individual classification schemes. She may invoke the concept Chair from a classification scheme for Furniture, the period 19^{th} Century from a scheme for Periods, and French from a third scheme for Places, to make the synthesized classification concept '19^{th} Century French Chair.' The resulting fleshed out collection of concepts may well be adequate to represent the knowledge, information, or data about the inventory of her shop. With some simple synthetic classifications, a natural datastructure to represent the classified data is often a table or a relational database. In this case, the different kinds, Furniture, Periods, and Places, ideally would need to be orthogonal or independent (and the resulting database would be Third Normal Form [Date 1977]). There still will be a need to capture and represent any hierarchies, the semantics, in the synthetic classification; for example, the Places kind might have Europe as a Place that encompasses France, Germany, etc.

When the antique dealer writes books on antiques she might use exactly the same collection of classification concepts for antiques to indicate the topics of the books. So, for example, she might write a book on 19th Century French Chairs. The concept now is being used as an annotation. And annotations like these can be used to organize information objects, to organize knowledge. The annotation 19th Century Chairs identifies a broader or more general topic than 19th Century French Chairs, and thus relations like these can establish a DAG of topics.

So, concepts can play double duty: they can organize things, and they can be the basis for topic annotation. One potential point of confusion is that classical librarians often make both uses of concepts, and they often use the latter, topic annotation, to accomplish the former, the organization of things. Brick and mortar libraries certainly have some commonalities with antique shops in as much as there are things, i.e., books, that need to be classified, listed in an inventory, and given physical location. And, of course, librarians have done this. Most actual existing information object (IO) classification schemes use what an IO is about, its subject or topic, to classify what the item is. In the Dewey Decimal Classification, for example, a book on physics is classified differently than a book on chemistry, and the basis of this difference is that the books are about different subjects.

The second preliminary distinction arises from the 'Triangle of Meaning.'



The 'Triangle of Meaning' is a phrase originating with Ogden and Richards' (1972) *The Meaning of Meaning* in the 1920s. Shiyali Ranganathan (1937, 327), the preeminent modern theorist of librarianship, also used the distinction in 1937; he called the Symbol Vertex the 'verbal plane' and the Concept Vertex the 'idea plane.' In fact, the distinction goes back at least to Aristotle (*De Interpretatione*) and it was really focused on by the early 20th Century German philosopher Gottlob Frege (Tichy 1988; see also Almeida, Souza, and Fonseca 2011; Dahlberg 2009; Fugmann 2004).

The Triangle of Meaning makes distinctions, first between a concept and an expression or symbol or sign that names or identifies the concept, and then between the concept and the things it applies to or refers to. So, there might be the word 'horse,' the concept of horse, and those particular delightful creatures which fall under that concept, for example, Secretariat, Sea Biscuit, Little Sorel, Trigger, Silver, Black Beauty, etc. In this context, the word 'concept' gets used in pretty well the same way as in ordinary speech and life, and that amounts roughly to 'general notion' or 'general idea' or even 'meaning.' Many describe concepts as being mental or mental constructions; however, it is better to regard them as abstractions or abstract objects. Among other virtues, the Triangle of Meaning gives a transparent account of synonyms and homographs (these are just many-to-one or one-to-many relations between symbols and concepts).

The Triangle of Meaning has significance here and now because, in the areas to be addressed, there typically is much back and forth between the Symbol and Concept vertices. For example, topic annotation is often discussed in terms of strings; so-called 'tagging' is free vocabulary string annotation; *LCSH* and *MeSH* use 'Headings' which are strings; all of this is to use Symbol Vertex in preference to the Concept Vertex. Another example is thesauri. As Jean Aitchison, Alan Gilchrist, and David Bawden (2000, 1) write:

> [a thesaurus is a] vocabulary of a controlled indexing language, formally organized so that *a priori* relationships between concepts are made explicit.

That is to say, a thesaurus gives Symbol Vertex representation of Concept Vertex DAGs of topics (together with vocabulary control).

In sum, there is the need for a flexible awareness of classification and annotation, and of strings and concepts.

2.0 Introduction

Faceted classification is typically considered to be a synthetic classification consisting of orthogonal facets which themselves are composed individually either of exclusive foci or of a hierarchy of foci.

(Broughton 2004, 2006; Buchanan 1979; Gnoli 2008; La Barre 2006, 2010; Ranganathan 1959, 1967; Vickery 1960, 1966, 2008; Wilson 2006). There is a syntax and a semantics to the synthesis. To take a toy example. A syntax might consist of the vocabulary

14th Century [kind=Period],
19th Century [kind=Period],
Renaissance [kind=Period],
French [kind=Place],
German [kind =Place],

and the grammar might deem a well-formed classification label or term to consist either of any single vocabulary item of any kind, or of a single vocabulary item of the kind Period followed by a single vocabulary item of the kind Place. And that would permit the synthesis of well-formed labels like 'German' or '14th Century French,' but not labels like 'French 14th Century' (which would be ill-formed under this grammar, which requires that the Period comes before the Place). There would be a need also for a semantics: that is, a specification of how the labeled classes or types relate to each other (as subtypes, supertypes, instances, and the like). So, for example, the 14th Century type could be deemed to be a subtype of the Renaissance type.

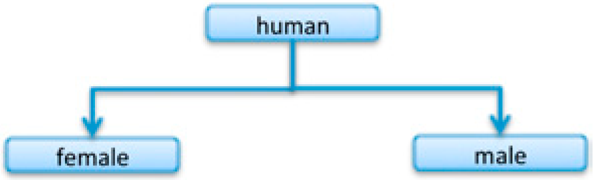
The 'kinds' here are the facets, so the example has a Period facet and a Place facet. And a 'focus' is 'any subject or name or number for it' (to use Ranganathan's [1967, 88] terminology). Then, to move on to the general case, to grammars beyond that of the toy example, the facets are usually required to be orthogonal or independent. This means that, when constructing a synthesized value, the choice of a focus from one facet has no repercussions whatsoever for combination with a focus from another facet. So, for example, the choice of 19th Century from the Period Facet neither compels, nor excludes, a particular choice from the Place Facet—it can be combined with either French or German. Within a facet, though, the foci are not typically assumed to be orthogonal or independent. In fact, they are assumed to be dependent. Choice of one focus precludes or affects choice of others. If, for example, French is chosen from the Place Facet, that choice prevents the additional choice of German; French cannot be combined with German. The foci for a facet are often talked of as being an array, or collection of arrays, of foci, from which one value, or one value from each, needs to be chosen. For example, Vanda Broughton conceives of the foci in a facet as being a collection of separate and individually exclusive arrays, often an enumerated Aristotelian hierarchy, with a choice of no more than one focus from each array (Broughton 2006). And Anthony Foskett and Travis Wilson think something very similar (A. C. Foskett 1996; Wilson 2006).

This paper addresses, in particular, this second exclusive-or-hierarchical foci condition. It evaluates the arguments for the condition and finds them not conclusive. It suggests that wider synthetic constructions should be allowed on foci within a facet.

Faceted classification is widespread nowadays, in the small, so to speak. In the large, there are probably only two examples of traditional classification schemes for Information Objects which are faceted at their core: Ranganathan's Colon Classification (Ranganathan 1960) and the Bliss Bibliographical Classification of Mills, Broughton, and the Classification Research Group (Mills and Broughton 1977). Both these schemes recognize that there are kinds of concepts. Categorizing concepts is also the approach of many others (Austin 1984; Cheti and Paradisi 2008; A. C. Foskett 1996; Lambe 2007; Morville and Rosenfeld 2006; Slavic 2008; Vickery 1960, 1966; Willetts 1975)

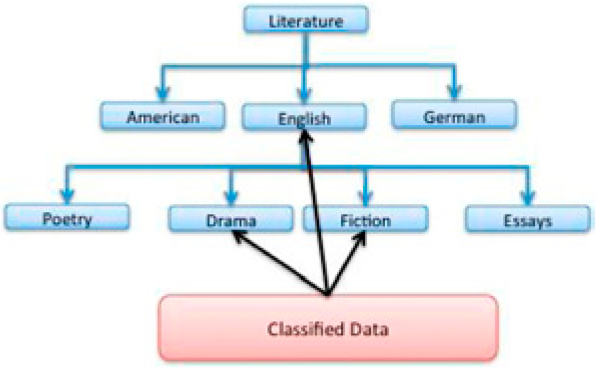
3.0 Some background theory and nomenclature

In an Aristotelian-Linnaean hierarchy, say



the items being classified are classified by the 'leaves' (Vickery 1975). So everything, every human that is, ends up being female or male (let us not worry about hermaphrodites, etc.). Such leaves have the JEPD property (jointly exclusive pairwise disjoint) i.e., the classification is exhaustive and exclusive. Nothing is classified, or classified directly, by the human node (because it is not a leaf). Certainly, the human node has instances, perhaps Sally, but Sally is classified or 'cataloged' by being female and inherits instantiation of being human thanks to the structure of the hierarchy. The human class and the female class are not exclusive because females are humans, and Sally, for one, is in both classes. So, in a hierarchy, the classes as a whole will not usually be mutually exclusive, but, typically, the leaves, which are the classes that do the cataloging work, will, or should be, mutually exclusive. Distinct from the relations between the classes are the names or labels or notation that the classes have, in this case {'human,' 'female,' 'male'}. These names are all different one from another, but the fact that the names are different, 'unique,' does not mean that the classes that those names signify are exclusive (because, for instance, females are human). In a true Aristotelian-Linnaean hierarchy, the actual classification is done by the leaves, in this case {female, male}, and likely the names or terms will be different and the classes that those names signify, the leaves, will have the JEPD property and be mutually exclusive,

But at least some librarian classification schemes are different from true Aristotelian-Linnaean classification hierarchies, principally in that they use some interior nodes for classification (in addition to the leaves). Here is a fragment of the *Dewey Decimal Classification* (*DDC*), around 820.

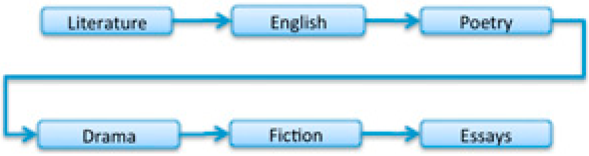


Some books are classified by the leaves (for example, Shakespeare's *Romeo and Juliet* is going to be Literature-English-Drama with classmark notation 822 (and it will gain some other decimal digits in a full classification), and others are classified by the internal nodes (for example, John Keats's *The Works of John Keats [complete Poetry and selected Prose]* will be Literature-English 820).

That there are works, e.g., Keats, that are instances of internal nodes yet not instances of any of that node's children means that the sibling children are not exhaustive as to the contents of their parent. In this case, the *leaves* are not *exhaustive*. And the leaves together with the interior nodes are not *exclusive* one from another, because instances of the children are instances of their parents. The leaves of *DDC* do not have the JEPD property, and neither do the leaves together with the interior nodes. But notice that the names used, or the notation numbers, are different (Literature-English 820, Literature-English-Drama 822). So, talking roughly, the names or terms are 'exclusive,' but the underlying classes are not.

In the setting of classical librarianship, there is the need to produce a systematic, or linear, order from the hierarchy for shelving, bibliographic lists, and the like. This amounts to converting a (classification) hierarchical tree to a list. There are different algorithmic tree traversals that can do this, but typically the children of node are considered ordered, and a shelving traversal can be generated by recursively visiting each node and its children in turn.

So, for example, part of the Dewey fragment above could be 'systematized' to



(omitting American and German literature for simplification and clarity), and here are the corresponding Dewey classification numbers:

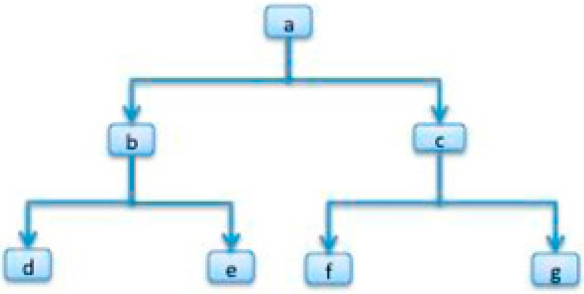
There is the important notion of an array. Unfortunately there is an ambiguity in its use that we need to be clear over. Ordinarily, in this setting, an array just amounts to an ordered list of the children of a node. So, in the first example, the node 'human' is parent to the array {female, male}. In the second example, the node 'Literature-English' is parent to the array {Poetry, Drama, Fiction, Essays}. In this vein, Broughton (2004, 294), for example, offers the definition

> **array:** a group of sub-classes all derived by applying the same principle of division to the containing class;

And Vickery (1975, 14) writes

> Any one level of subdivision gives rise to a group of terms that constitute an *array* (for example, within the class of Metals the array Bismuth, Lithium, Mercury, Potassium, Sodium).

There are no class hierarchies in arrays like these simply because all the values in the array are siblings. If the first level of subdivision is itself further subdivided using a second principle of division, further arrays are generated; in fact, the new ones are what Ranganathan would have called arrays of order 2 (and this process can be continued indefinitely with more principles of division, arrays of order 3, etc.) (Ranganathan 1967). So, for example,



has three arrays {b,c}, {d,e}, and {f,g}. Assuming suitable principles of subdivision have been used, each of these arrays individually is exclusive as to the foci it contains. The arrays individually each have the JEPD property, and the two arrays (of order 2) together, which form the leaves, also collectively have the JEPD property.

But it is also common enough to include parent nodes and even grandchildren and great grandchildren in arrays. For example, Douglas Foskett (2003, 1064), certainly an expert, describes this 'array' from the *DDC*:

370	Education			
370.1		Theory		
370.7		Study		
370.71			Meetings	
370.72			Conferences	
370.73			Teachers' Colleges	
370.732				Courses and programs

Essentially, this is a slice of a systematization of Dewey, and it includes parents, children, grandchildren, etc.

In sum, classification is typically by leaves only, those leaves typically have the JEPD property, and if those leaves are described as being an 'array' or several arrays, those arrays also have the JEPD property so that values of the arrays are exclusive and exhaustive; if, in contrast, classification also includes interior nodes, the classifying nodes do not have the JEPD property, and if those classifying nodes are described as being an 'array,' that array does not have the JEPD property, and its values are not exclusive and exhaustive.

4.0 Ersatz faceting and real faceting

As mentioned, faceted classification can either be a classification of 'things,' an ontology, or a subject classification, a thematic 'Tree of Knowledge.'

Chapter 21 of Book 8 of Pliny's *Natural History* has the title (Pliny, 78):

> 21. Of Lynxes, Sphinges, Crocutes, Marmosets, of Indian Oxen, of Leucrocutes, of Eale, of the Ethiopian Bulls, of the Mantichora, the Unicorn, of the Catoblepa, and the Basilisk.

Of these 'land animals that go on foot,' some are Sphinges, and some are Crocutes, and no creature is both a Sphinge and a Crocute. That classification is ontological, it is part of a biological natural kind ontology. However, the subject matter of that chapter is both Sphinges and Crocutes (and some other land animals). Chapter 21 is polytopical. Thematic, topical, or subject, classification accommodates this.

Broughton introduces faceted classification by means of an example involving physical socks (Broughton 2004, 2006). (She does this for pedagogical reasons.) In her example, the socks are items, things, each individually with five different attributes drawn from the 'facets' Color, Pattern, Material, Function, and Length; so there are black-striped-wool-work-ankle socks, white-striped-silk-work-knee socks, etc. Each of the individual facets is (or could be) an Aristotelian exclusive and exhaustive hierarchical classification scheme. And the entire scheme synthesizes the five facets. A similar kind of faceting is also often seen where alternative different principles of subdivision are applied separately to the same underlying class. So, for example, people could be divided up By Age, By Gender, By Occupation, By Religion, By Place of Birth, etc.; and any of the facets, say By Occupation, could itself be a hierarchical scheme of foci or values. This kind of arrangement is commonplace on the Web. Such department stores as Amazon, Target, or Walmart often display their wares essentially as tables or grids or database of orthogonal categories and provide navigation by means of hierarchical facets.

An example that we are all familiar with is the email client. An emailer might contain perhaps 27 messages displayed in a table, and it usually will have the capability of sorting the rows, the email names, By Date, By Subject, and By Sender. The dates, subjects, and senders are column values. When the emails are presented ordered By Date, there are 27 emails; when they are presented by By Subject, there are still 27 emails. There are just different principles of subdivision being applied to the same underlying domain. These principles can be applied one after another; for example, the emails could be sorted by date, and then the result of that could itself be sorted by subject. In effect, the order of a sequence of sorts is a 'citation order,' and each individual column value sorting relation (for example, alphabetical, numerical) is a 'filing order' for the individual column in question. By adjusting the citation order and the filing orders, the individual emails can be grouped and scattered as desired. The natural data-structure here for holding and representing the data is a table. And, in turn, tables are the core of relational databases. The facets are orthogonal—they are mutually independent. Any date can be combined with any subject and any sender. This will likely mean that the associated database table will automatically be in Normal Form, in particular in 3rd Normal Form (which is a good thing for a database table to be) (Date 1977). Here is an example of a less than expert attempt at faceting that is not orthogonal. Imagine a shop that sold pedal-powered personal transportation devices. It might catalog its devices By Kind (with the values {unicycle, bicycle, tricycle}) and By Number of Wheels (with the values {1,2,3}). These 'facets' are not orthogonal because, for example, a unicycle has to have 1 wheel (there cannot be a unicycle with 2 wheels). And, similarly, a relational database table for this would have a transitive column dependency between the Kind of device and the Number of Wheels that it had—the table would not be in 3rd Normal Form.

Although a table is, in some sense, the natural data structure here, in current practice, the faceted data is often alternatively presented in a list or index, often a hierarchical indented list or index. One easy way to do this is to change the column order so that the order reflects the desired hierarchical importance or indentation from left to right (changing column order does not affect the data content of a table), and then to successively sort on the columns from right to left. For example,

Name	Subject	Date
Ann	**Breakfast**	**2010**
Ann	**Breakfast**	**2008**
Beryl	**Dinner**	**2009**
Beryl	**Lunch**	**2008**
Charley	**Breakfast**	**2009**
Charley	**Dinner**	**2008**
Charley	**Lunch**	**2010**

and this can be given the slightly more elegant display

Ann	**Breakfast**	**2010**
		2008
Beryl	**Dinner**	**2009**
	Lunch	**2008**
Charley	**Breakfast**	**2009**
	Dinner	**2008**
	Lunch	**2010**

The tree here is being displayed horizontally with the leaves to the right, instead of the usual inverted vertical display; this type of display is also often used on computers to depict on screen their directory or folder structure. Interestingly enough, many thesauri and indexes collapse the levels in the hierarchy, they reduce the indentation by essentially writing the third column under the left as an alternative (Aitchison Gilchrist, and Bawden 2000; NISO 2005; Zeng 2005). So, for example, the three rows

Ann	**Breakfast**	**2010**
		2008
	Lunch	**2008**

might be depicted

Ann	*(by Subject)*
	Breakfast
	Lunch
	(by Date)
	2010
	2008

The annotations in italics are 'Node labels' and they are indicating the principles of subdivision (Aitchison, Gilchrist, and Bawden 2000; NISO 2005; Zeng 2005). This may or may not be a good thing. It is confusing logically because there are duplicate names, there are four names and only three kinds of rows. And it loses the full hierarchical structure. On the other hand, a faceted scheme might have 10 facets, and 10 levels of indentation would be unusable.

Orthogonal attributes, or principles of subdivision, can provide faceting, 'ersatz' faceting, but it is not the style of faceting, or domain for faceting, envisaged by Julius Kaiser, Paul Otlet, Henri LaFontaine, Shiyali Ranganathan, and the Classification Research Group (CRG) (Classification Research Group 1955; Kaiser 1911; La Barre 2010; Ranganathan 1937, 1951, 1960; UDC 2010).

Grey socks are socks, which is to say that the underlying type or universe type of the Color facet is socks. Striped socks are socks, which is to say that the underlying type or universe type of the Pattern facet is (also) socks. So, too, for the other facets. This means that the non-elemental classes like black-striped-wool-work-ankle socks are what Ranganathan would have called superimposed classes (Ranganathan 1960). Grey socks are socks (socks which are also grey), ankle socks are socks (socks which are also ankle length), and grey ankle socks are socks (socks which are also grey, and which are also ankle length).

Contrast this with Kaiser's 'Concretes' and 'Processes' (Kaiser 1911). Examples of concretes are aluminum, iron, and steel; and examples of processes are smelting, welding, and rusting. And, of course, concretes can play a part in processes (or processes can involve concretes)—as in the 'rusting of iron.' But Concretes and Processes do not have the same underlying universe type, the 'rusting of iron' is not a superimposed type.

The story is similar with modern facet analysis (Aitchison, Gilchrist, and Bawden 2000; Buchanan 1979; Gopinath 1992; La Barre 2010; Ranganathan 1960; Spiteri 1998). With real faceting, the actual facets are of different kinds (as opposed to being different attributes, or differently principled divisions, of the same underlying kind).

5.0 The Wilson argument

Travis Wilson (2006) subscribes to the orthogonal-facets-exclusive-foci view for a particular area and kind of analysis (and, to keep the record straight, he is certainly aware that Ranganathan had a different analysis of what probably is a different area). Wilson offers an argument. It first rests on his conception of facet analysis (which addresses mainly ersatz faceting rather than real faceting). Wilson suggests we start with a 'tag soup', say

> Pecan Pie, Chocolate Ice cream, Chocolate Cookie, Cherry Pie, Cherry Ice cream, Pecan Cookie, Chocolate Pie

And we extracting from these the 'atoms' or 'elements,' and that gives, perhaps

> Pecan, Pie, Chocolate, Ice cream, Cookie, Cherry

These are still a 'soup.' There is not considered to be any order or structure here yet. But suppose we wish to extract a structure, in particular a facet structure. One way we can do it is by asking, "Which atoms can be combined with which other atoms?" The ones that can be combined are independent, orthogonal, and belong in different facets. The ones that cannot be combined are dependent and belong as foci in the same facet. So, for example, if we said Pecan can be combined with Pie, Ice cream, and Cookie, but not with Chocolate and Cherry, that would place Pecan is a different facet from {Pie, Ice cream, Cookie} and in the same facet as {Chocolate, Cherry}; if Pie can be combined with Pecan, Chocolate, and Cherry, but not with Ice cream and Cookie, that would place Pie is a different facet from {Pecan, Chocolate, Cherry} and in the same facet as {Ice cream, Cookie}; and if we followed through with this, two facets would be generated:

> Substrates of {Pie, Ice cream, Cookie}
> Flavors of {Pecan, Chocolate, Cherry}.

There is no compulsion that drives to that particular division and combination. We could alternatively have allowed Chocolate to combine with Pecan and Cherry, etc. And that might have led to three facets

> Substrates of {Pie, Ice cream, Cookie}
> Flavors of {Pecan, Cherry}
> Toppings of {Chocolate}

It is up to us what we do, how we do the partition; however the distinction between facets and foci is to be made on the basis of what is independent and what is dependent. And so foci, within a facet, have to be exclusive because they are defined to be exactly that.

Wilson's argument is certainly an argument, and he uses the view it embodies to generate faceted classifications by algorithm. And if the candidate labels were bare, meaningless labels, it would be a reasonable argument. If the atomic tag soup were

> DF2, 27, Km+,*,Wef

And we wanted to establish a faceted scheme of these, what Wilson suggests is presumably exactly right. But in the realistic cases we encounter, the tags in the soup do have meanings, and they do have kinds, independently of what can and cannot be combined. For example, in the soup

> 18th Century, French, German, 19th Century,

two of the labels, or what they signify, are time periods, and the other two are regions or places. And we can use the kinds to do the facet analysis (which is exactly what Ranganathan and the CRG did). So Wilson's argument is not definitive. We can combine or synthesize foci, if we think that desirable. (Wilson's argument is also discussed in Vickery [2008]).

6.0 The (Anthony) Foskett argument

Real faceted classification concerns subjects (or topics or concepts or types or tags), not things (or attributes of things). Its origins are from puzzles concerning the nature of subject nodes in discipline-based Bacon-style Trees of Knowledge (Bacon 1620). (All the traditional library classifications have this style.) Kaiser, Otlet, LaFontaine, Ranganathan, and others, noticed that many, indeed almost all, subject concepts were compound or composite concepts constructed from elemental components. And the elemental components themselves could be of different kinds or categories. And this invites the use of synthetic classification from different categories of elemental, or atomic, component concepts—faceted classification. So the target here is to use faceted classification to produce labels or concepts or annotations for subject classification.

Anthony Foskett (1996, 148) writes:

> The foci within a particular facet should be *mutually exclusive*; that is, we cannot envisage a composite subject which consists of two foci from the same facet. We cannot have the 17th Century 1800s, or German English, or copper aluminum, but we *can* have composite subjects consisting of combinations of foci from different facets: English novels, 17th Century German literature, analysis of copper, heat treatment of aluminium.

This just seems mistaken and a confusion between things and topics. It confuses antiques with books about antiques. An entity, such as a metal spoon made entirely of a single metal, cannot be both made entirely of copper and entirely of aluminium; but a subject (a subject matter, a topic, a concept, a type) presumably can encompass copper and aluminium—isn't 'heat treatment of aluminium and copper' a subject? Isn't '17th and 18th Century German literature' a topic? And isn't 'Sphinges and Crocutes' a topic? (There may just be some lack of clarity of expression here in the text that is being quoted. Foskett would be well aware that the synthetic operations of the Universal Decimal Classification [UDC], especially the '+' operator, in effect permit the forming composites from foci within the same facet [Broughton 2010a; UDC 2010]. Also, there is a connection here to what might be called polytopic reduction. Suppose there is a book with the (honest, accurate, and comprehensive) title 'Heat treatment of aluminium and copper,' and the question is asked, "How many subjects does this book have and what are they?" An answer, favoring polytopicality, is: "Two, and they are {Heat treatment of aluminium, Heat treatment of copper}." Another answer, avoiding polytopicality, is: "One, and it is {Heat treatment of aluminium and copper}." No judgment is passed here on what is most desirable, but the Foskett intuition can be largely retained if polytopicality is the choice.)

7.0 Trying to take Broughton's account a small step further

What Broughton writes seems to be exactly right. However, it may be possible to improve the views it expresses in various ways.

An alternative way of describing a collection of foci, favored by some authors, is to say that there is an array of foci. And it is quite possible to use ersatz faceting as a subfacet of a faceting scheme, in which case there would be arrays of foci for that facet. For example, there could be the 'manufacture of socks,' which could be a combination of a Process and an Entity; then the socks themselves could be ersatz faceted as above, and that would or could generate 'manufacture of white socks,' 'manufacture of grey socks,' etc., and that would subdivide the socks by color, i.e., there would be a By Color array, and it could also give 'manufacture of ankle socks,' 'manufacture of knee socks,' i.e., there could be a By Length array, and so on. And the principles of division could be used sequentially; so, there can be arrays (of foci) of order 1, arrays of order 2, etc., as described earlier.

Broughton writes (2004, 267 and 2004, 54 emphasis in the original): "an important thing to notice about the members of an array [i.e. the foci] is that they are all *mutually exclusive classes*." And (2004, 270 and 2006, 54): "because all the terms within a facet come into the same category … the relationship between them will be those of a hierarchy."

At first glance this does not seem quite right. If the terms are a hierarchy, for example {human, female, male}, they need not be mutually exclusive classes—female and human are not mutually exclusive classes, one is a subtype of the other (a female is a human). But if attention is paid to exactly what is said, it is the arrays that individually have members which are exclusive classes, and that is correct. Notice that the condition exhaustive does not appear (either in conjunction with exclusive or on its own). This typically is an indicator of the use of interior nodes for classification, as opposed to a pure Aristotelian hierarchy with leaves that have the JEPD property.

Moving on to a different point, Broughton favors enumerative or enumerated foci— probably fixed-in-stone hierarchical schemes for the foci—and this is coupled with synthesis between facets. She suggests that the foci for a facet might have a hierarchical arrangement. This hints at an enumeration as opposed to a synthesis. Why? The construction of the classes is top-down. The start is the root. Then a principle of division is applied to produce some children, a second principle of division is applied to produce grandchildren, and so on. So the leaves are not the atoms from which the whole tree is constructed or synthesized bottom up, rather they are the residual fragments after a series of cleavages have been made to the root. It might be thought that this is somehow inessential and that the tree could be synthesized bottom up. But what makes this difficult or awkward in this case is that the arrays are not exhaustive (and this comes from using the interior nodes for classification). If the children were exhaustive of the parents, then the parents could be considered just to be the collection of their children, and attempts could be made to build bottom up.

Broughton (2004, 270) writes: "Where a faceted classification differs most significantly from an enumerative classification is in it potential to combine terms from different facets." Notice "combine terms from different facets" but no mention of "combine terms from within a facet." But why not permit synthesis for everything? Why not permit combining terms from within the *same* facet? Here is an example:

18th Century History
18th Century Geography
19th Century History

are composite subjects synthesized from different facets. But, presumably, we would want also to have the ability to form subjects like

18th and 19th Century History
18th Century History and Geography,

and this requires synthesis within a facet (as well as the synthesis across facets). Neither Broughton nor Wilson would permit this, because, for example, they hold that the choice of the focus 18th Century specifically excludes the choice of 19th Century.

8.0 Muddy waters

Part of what is driving the intuitions here is that concepts as topics for annotation have somewhat different properties to concepts for classification in ontologies. Unfortunately, classification of information objects is not always, and perhaps not even usually, by topic alone.

Earlier it was suggested that the *DDC*, for example, classifies by subject or topic. That is not entirely true. It, in common, with the Library of Congress *Classification* and almost all traditional library classifications, also takes some input from the 'form' of an information object. Form here might include whether the object is a bibliography or whether it is an encyclopedia. There is a good reason for doing this. The whole system is aimed to provide service for the user, and experience has taught that users are often interested in form.

Form has also crept into subject headings (i.e., lists of topics). *LCSH* recognizes about 600 forms of literature. And any of these values are permitted to be components of synthesis to create further subject headings. So, if 'Physics' is a subject or subject heading, so too is 'Physics—Encyclopedias.' Obviously, there could be a book on physics encyclopedias, but a book with the subject heading tag 'Physics—Encyclopedias' is not one of those, rather it is an encyclopedia on physics. This is unfortunate. Subject headings should be, well, subject headings. Information about forms should be separate and separately provided. Other systems are more careful here. *MeSH* will append 'as topic' when required (or use other syntactic devices); so 'Clinical Trial' marks a piece of literature which is a clinical trial and 'Clinical Trial as topic' marks literature about clinical trials. However there is still a mixing of topics and forms at the display level. (*MeSH* is a faceted, or partially faceted, system; it is faceted or 'deconstructed' at the record level; but when it displays to the user it sometimes combines these facets. This is a good approach. The only, mildest of mild, qualification, is that what is displayed to the user is not really a subject or a subject heading, rather it is just a heading, or locator, [which combines subject and form]. There are also initiatives to do facet analysis on LCSH [Chan and O'Neill 2010]).

Were such general information object systems to be approached with a view to performing faceted analysis on them, part of the project, such as pulling out forms, would effectively be ersatz faceting. And the remainder would be real facet analysis of topics (which is the central concern of this paper).

9.0 Conclusion

All concepts can be conceived of as having categories or kinds or types. In particular, the elemental or atomics concepts have kinds. Then the non-elemental or composite or compound concepts are conceived of being constructed or synthesized from the atomic concepts. (These compound concepts also have kinds, in the style of type theory in computer science or categorial grammar in linguistics [van Benthem 1990]. For the most part, this is not of central importance in the context of this paper.)

Elemental or atomic concepts have categories or kinds. Here are some of the kinds they might be. They might be Concretes, Processes, Periods, Places, Things, Kinds, Parts (organs, constituents), Properties, Materials, Operations, Patients (objects of action, raw materials), Products (substances), By-products, Agents, Forms, Genres, and, possibly, other kinds, depending on the discipline or subject matter in question.

Then non-elemental concepts, or labels for such, are constructed or synthesized from values or foci of the kinds in use—there will need to be a syntax or grammar for this.

There can (and should) also be synthesis within a facet. Here is an example scheme. Suppose we decide that the granularity for the kind Period should just come down to centuries, and so any century is permitted as an atom. So in

21st Century Schizoid Man
221st Century Schizoid Man
2021st Century Schizoid Man

The 21st Century, 221st Century, 2021st Century are all good as (elemental) foci. In fact, there are infinitely many elemental foci of the Period kind. Then there can be synthesis to form such periods as '17th and 18th Century.' Synthesis does not have to be restricted to simple (Boolean) additions, the period 'Before Present (BP)' amounts to '*All* centuries before the present.' There can be hierarchically higher-level coverings such as 'Renaissance' for the period 14th to 17th century; 'Paleolithic' to cover the period from a couple of hundred thousand centuries ago up to 100 centuries BP. And these coverings do not have to be exclusive of each other; the Lower Paleolithic and Middle Paleolithic Periods are generally taken to overlap each other.

This is a very important point. If the system allows for overlapping classes, the result is not going to be a hierarchy. (A hierarchy, or tree, is, graph-theoretically, a connected acyclic graph. If classes overlap, they share at least one child, which means, if they also share an ancestor [the root], that there is a cycle, so the structure is not a hierarchy.)

So the overall structure is not really a hierarchy, rather there are infinitely many elemental and synthesized Period foci, and many of these bear subtype-supertype relations to each other. There is a way of representing the semantics, that of directed graphs; and they use, in essence, arrows between the nodes. This convention was used earlier in this paper to illustrate trees and hierarchies; and we are all very familiar with it from links or hyperlinks on the World Wide Web. There can be just links from foci to foci. These links can also be given a different semantics, or several different semantics simultaneously and, possibly, ambiguously. For example, the relations 'is a subtype of,' 'is an instance of,' and 'is a part of' are used foundationally to establish classification schemes and their associated hierarchies or graphs. It is often very important to distinguish these, and to be correct on what they are. For example, classification can support inference; if fish is a subtype of vertebrate, then that supports the inference from Livingstone is a fish to Livingstone is a vertebrate. For this use, it matters whether 'X is a subtype of Y,' 'X is an instance of Y,' or 'X is a part of Y.' But, in the setting of Information Resources, the target is to assist search and to help the Patron to find the relevant Information Objects. Often, in that context, it does not matter what the connection is between X and Y, provided that being guided from X to Y helps in finding Y (or Information Objects labeled with, or given the metadata, Y). Which actual links there are, the semantics, can be established or described in a variety of ways. One such way is symbolic logic and logical inference. With ersatz faceting (and all faceting within a single facet is ersatz), a logic inference engine can produce the links. If one type is socks, and another brown socks, logic has the ability to say that the second is a subtype of the first. Similarly with definitions or other statements, if the Renaissance is defined to span the 14th ,15th , 16th, and 17th Centuries, logic can establish what the links are.

The key to synthetic construction operations on the kind or facet Period is that the result must be a Period. (In other academic literature, there is a logic of periods—that could be invoked to provide assistance here.) And this requirement can be generalized. The results of acceptable synthetic constructions within a facet must themselves be within the facet.

Here is another example: washing is a Process, drying is also a Process and the synthetic constructions 'washing then drying' and 'drying then washing' are both also Processes. (There is also a logic of processes that could provide guidance. The theory of Petri Nets is a somewhat advanced example of such a theory.)

A full generalization would also admit constructors that could have components which themselves were faceted (as Ranganathan described at length); so, for example, there could be a comparison constructor which could be used to produce the class 'Comparisons of the smelting of iron with the smelting of steel.' Indefinitely, many classes can be synthesized, which ones actually are synthesized depends on literary warrant and the needs at hand. (We all have the capability of saying indefinitely many sentences, but, at least in principle, what we actually do say depends on our interests and needs.)

Faceted classification can be fully synthetic classification from different categories of elemental, or atomic, component concepts.

References

Aitchison, Jean, Gilchrist, Alan, and Bawden, David. 2000. Thesaurus construction and use: a practical manual. 4th ed. Chicago: Fitzroy Dearborn.

Almeida, Mauricio, Souza, Renato, and Fonseca, Fred. 2011. Semantics in the Semantic Web: a critical evaluation. *Knowledge Organization 38*: 187-203.

Austin, Derek. 1984. *PRECIS: a manual of concept analysis and subject indexing*. 2nd ed. London: The British Library.

Bacon, Francis. 1605. The Advancement of Learning. Available: http://www.gutenberg.org/ebooks/5500

Bacon, Francis. 1620. *The Great Instauration*. Available at http://www.constitution.org/bacon/instauration.htm

Broughton, Vanda. 2004. *Essential classification*. New York: Neal-Schuman.

Broughton, Vanda. 2006. The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings: New Information Perspectives 58*: 49-72.

Broughton, Vanda. 2010a. Concepts and terms in the faceted classification: the case of UDC. *Knowledge Organization 37*: 270-279.

Broughton, Vanda. 2010b. *Essential Library of Congress Subject Headings*. London: Facet Publishing.

Buchanan, Brian. 1979. *Theory of library classification*. London: Clive Bingley.

Chan, Lois Mai, and O'Neill, Edward T. 2010. *FAST: Faceted Application of Subject Terminology: principles and application*. Santa Barbara, Calif.: Libraries Unlimited.

Cheti, Alberto, and Paradisi, Federica. 2008. Facet analysis in the development of a general controlled vocabulary. *Axiomathes 18*: 223-241.

Classification Research Group. 1955. The need for a faceted classification as the basis of all methods for information retrieval. *Library Association Record* 57: 262-68.

Dahlberg, Ingetraut. 2009. Brief communication: concepts and terms – ISKO's major challenge. *Knowledge organization* 36: 169-77.

Date, C.J. 1977. *An Introduction to Database Systems* (2 ed.). Reading, MA: Addison-Wesley.

Foskett, Anthony C. 1996. *Subject approach to information.* 5th ed. London: Facet Publishing.

Foskett, Douglas J. 2003. Facet analysis. In Drake, Miriam A. ed., *Encyclopedia of library and information science*. 2nd ed. New York: Marcel Dekker, pp. 1063-67.

Fugmann, Robert. 2004. Learning the lessons of the past. In Rayward, W. Boyd, and Bowden, Mary Ellen eds., *The history and heritage of scientific and technical information systems: Proceedings of the 2002 Conference, Chemical Heritage Foundation*. Medford, NJ: Information Today, pp. 168-81.

Gnoli, Claudio. 2008. Facets: a fruitful notion in many domains. *Axiomathes* 18: 127–30.

Gopinath, M. A. 1992. Ranganathan's theory of facet analysis and knowledge representation. *DESIDOC Bulletin of Information Technology* 12n5: 16-20.

Kaiser, Julius Otto. 1911. *Systematic Indexing*. London: Pitman.

La Barre, Kathryn. 2006. *The use of faceted analytico-synthetic theory as revealed in the practice of website construction and design.* Indiana University, Bloomington.

La Barre, Kathryn. 2010. Facet Analysis. In Blaise Cronin (Ed.), *Annual review of information science and technology* 44. Medford, NJ: Information Today, Inc., pp. 43-86.

Lambe, Patrick. 2007. *Organising knowledge: taxonomies, knowledge and organisational effectiveness*. Oxford, England: Chandos Publishing.

Lowe, Henry J. 1994. Understanding and Using the *Medical Subject Headings* (*MeSH*) Vocabulary to Perform Literature Searches. *Journal of the American Medical Association* 271: 1103-08.

MeSH. 2010. *Medical Subject Headings - Home Page*. Available at http://www.nlm.nih.gov/mesh/

Mills, Jack, and Broughton, Vanda. 1977. *Bliss bibliographic classification. second edition. introduction and auxiliary schedules*. London: Butterworths.

Morville, Peter, and Rosenfeld, Louis. 2006. *Information architecture for the World Wide Web*. Sebastapol, CA: O'Reilly.

NISO. 2005. *NISO Standards: Z39.19*, Available at http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a

Ogden, Charles Kay, and Richards, Ivor Armstrong 1972. *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism*. New York: Harcourt, & Brace.

Pliny, the Elder. 78. The Natural History. Available at http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.02.0137

Ranganathan, Shiyali R. 1937. *Prolegomena to library classification* (3rd ed. 1967; 1st ed. 1937 ed.). Madras: The Madras Library Association.

Ranganathan, Shiyali R. 1951. *Philosophy of library classification*. Copenhagen: Munksgaard.

Ranganathan, Shiyali R. 1959. *Elements of library classification* 2 ed. London: Association of Assistant Librarians.

Ranganathan, Shiyali R. 1960. *Colon classification* (6 ed.). London: Asia Pub. House.

Ranganathan, Shiyali R. 1967. *Prolegomena to library classification* 3d ed. Available at http://dlist.sir.arizona.edu

Slavic, Aida. 2008. Faceted Classification: Management and Use. *Axiomathes* 18: 257-71.

Smith, Barry. 2004. *Beyond Concepts: Ontology as Reality Representation.* Paper presented at the Proceedings of FOIS 2004. International Conference on Formal Ontology and Information Systems, Turin.

Spiteri, L. 1998. A simplified model for facet analysis: Ranganathan 101. *Canadian journal of information and library science* 23: 1-30.

Tichy, Pavel. 1988. *The foundations of Frege's logic / Pavel Tichy*. Berlin: de Gruyter.

UDC. 2010. *UDC Consortium Home Page* Available at http://www.udcc.org/

van Benthem, Johan. 1990. Categorial Grammar and Type Theory. *Journal of philosophical logic* 19: 115-68.

Vickery, Brian C. 1960. *Faceted classification: a guide to construction and use of special schemes*. London: Aslib.

Vickery, Brian C. 1966. Faceted classification schemes. In S. Artandi (Ed.), *Rutgers series on systems for the intellectual organization of information* 5. New Brunswick, NJ: Graduate School of Library Science at Rutgers University.

Vickery, Brian C. 1975. *Classification and indexing in science* (3 ed.). London: Butterworths.

Vickery, Brian C. 2008. Faceted classification for the Web. *Axiomathes* 18: 145-60.

Willetts, Margaret. 1975. An investigation of the nature of the relation between terms in thesauri. *Journal of documentation* 31: 158-84.

Wilson, Travis. 2006. *The strict faceted classification model*, Available at http://facetmap.com/pub/

Zeng, Marcia Lei. 2005. Construction of Controlled Vocabularies, A Primer (based on Z39.19) Available at http://www.slis.kent.edu/~mzeng/Z3919/index.htm