How to Make a Faceted Classification and Put It On the Web

LAST MODIFIED: 28 MARCH 2009

Update February 2011: This has been translated into Dutch: Hoe maak je een facetclassificatie en hoe plaats je haar op het web? (http://iainstitute.org/nl/translations/hoe_maak_je_een_facetclassificatie_en_hoe_plaats_je_haar_op_het_web_.php). Many thanks to Janette Shew and the Information Architecture Institute's Translations Initiative for doing this. Also, How to *Reuse* a Faceted Classification and Put It On the *Semantic* Web (http://eprints.ecs.soton.ac.uk/21488/), by Bene Rodriguez-Castro, Hugh Glaser and Les Carr, takes my example of dishwashing detergents and extends it into ontologies and RDF.

Update February 2007: IA Voice (http://iavoice.typepad.com/) has used this paper as the basis for a series of four podcast episodes! It starts with IA E-Learning: Faceted Classification (1 of 4) (http://iavoice.typepad.com/ia_voice/2007/01/ia_elearning_fa.html).

Denton, William. "How to Make a Faceted Classification and Put It On the Web" Nov. 2003. https://www.miskatonic.org/library/facet-web-howto.html (https://www.miskatonic.org/library/facet-web-howto.html).

This follows Putting Facets on the Web: An Annotated Bibliography (/library/facet-biblio.html), and is the second paper I wrote for Prof. Clare Beghtol of the Faculty of Information Studies (http://www.fis.utoronto.ca/) at the University of Toronto, who led me in a reading course named "Applying Faceted Classification in an Online World."

0. Introduction

Faceted classifications are increasingly common on the World Wide Web, especially on commercial web sites (Adkisson 2003). This is not surprising—facets are a natural way of organizing things. Many web designers have probably rediscovered them independently by asking, "What other ways would people want to view this data? What's another way to slice it?" A survey of the literature on applying facets on the web (Denton 2003) shows that librarians think it a good idea but are unsure how to do it, while the web people who are already doing it are often unaware of S.R. Ranganathan, the Classification Research Group, and the decades of history behind facets.

This paper will attempt to bridge the gap by giving procedures and advice on all the steps involved in making a faceted classification and putting it on the web. Web people will benefit by having a rigorous seven-step process to follow for creating faceted classifications, and librarians will benefit by understanding how to store such a classification on a computer and make it available on the web. The paper is meant for both webmasters and information architects who do not know a lot about library and information science, and librarians who do not know a lot about building databases and web sites. The classifications are meant for small or medium-sized sets of things, meant to go on public or private web sites, when there is a need to organize items for which no existing classification will do. It is certainly not the intent of this paper to show how to build another universal classification, nor to describe how a library that uses a faceted classification scheme can put their catalogue online.

There are four main sections to this paper: when to make a faceted classification, how to make one, how to store it on a computer, and how to make it work on the web. I will concentrate on the middle two sections. The question of when to use facets is not particularly difficult (leaving aside general questions about the purpose and usefulness of classifications). Detailed advice on the design and implementation of a good web site is beyond the scope of this paper and requires a companion web site, with examples, to be best understood (but see Nielsen (2000) for excellent advice). In the final section I offer some guidelines on what to consider when putting facets on the web, but the discussion is not lengthy. The two middle sections about how to make and store a faceted classification receive a much fuller treatment.

What are facets? Consider a common example, wine. Each wine has a certain colour. It comes from a certain place. It is made from a particular kind (or blend) of grape. Its year of vintage is known. It has been guaranteed to be of a certain quality by its country's wine authorities. It comes in a container of a given volume. It has a price. A list could be made of all wines, but it would be enormously long and unwieldy. On the web, it would mean scrolling through screen after screen of endless subdivisions— hard to use, and hard to search. With facets, we can set up a handful of categories that will combine to fully describe the wines: colour, origin, grape, year, appellation, volume, price. Each category is populated with the right terms and organized in an appropriate way. Then each bottle of wine is classified by picking and choosing the right terms from each category. This is a faceted classification: a set of mutually exclusive and jointly exhaustive categories, each made by isolating one perspective on the items (a facet), that combine to completely describe all the objects in question, and which users can use, by searching and browsing, to find what they need.

Facets and the web go very well together. Barbara Kwasnick (1999, 39) said, "The notion of facets rests on the belief that there is more than one way to view the world, and that even those classifications that are viewed as stable are in fact provisional and dynamic. The challenge is to build classifications that are flexible and can accommodate new phenomena." And after they are built, the challenge is to make them easy to use. With hypertext and the web, dynamic views are only the click of a button away. Facets make a multi-dimensional organizational scheme, and web browsers are an easy and familiar tool for navigating many dimensions. All of the benefits of faceted classifications can be realized on the web. Before we can discuss how, though, first we must see when to use facets.

1. When to make a faceted classification

Kwasnick (1999) identifies four classificatory structures: hierarchies, trees, paradigms, and facets. When one of the first three works, use it. If some other organizing principle, such as a timeline or ordering by size, works, use it. The design of the classification must follow its purpose, and different things can be classified in different ways for different purposes, requiring different structures. If the others are insufficient, look to facets.

1.1 When not to make a faceted classification

Hierarchies and trees (imagine indented lists) are best when the entities in question are viewed in such a way that they have one dimension of classification. Hierarchies divide and redivide things into groups where each new group is a sub-species of its parent group; everything that is true of a group is also true of its sub-groups and so on down (Kwasnick 1999, 25). The Linnean taxonomy of living things is the classic example of this. Trees, in contrast, do not have the rules of inheritance (Kwasnick 1999, 30). For example, North America contains Canada, the United States, and Mexico, and Canada contains ten provinces and three territories, but Ontario is not a kind of Canada, and Canada is not a kind of North America.

A paradigm is a two-dimensional classification (imagine a spreadsheet). Use paradigms when there are two independent aspects to consider. Kwasnick (1999, 35-36) uses the example of terms describing kinship relations, which can be organized into a grid, with sex (male/female) along one axis, and relation (parent, sibling, parent's sibling) along the other axis. Comparing English and Polish terms for the various relations shows that "cousin" covers four different Polish terms, which do not ignore the sex of the cousin or which side of the family they are on.

1.2 When to make a faceted classification

Facets will handle three or more dimensions of classification. When, for the purposes of the classification, it is possible to organize the entities by three or more mutually exclusive and jointly exhaustive categories, then facets are probably the appropriate classification. Facets can be used to organize the entire world of knowledge, or the clothes in your cupboard, or anything in between. Ranganathan's Colon Classification and the Bliss Bibliographic Classification, which will be discussed in the next section, are universal classifications. We will see examples of some smaller classifications as well.

Kwasnick (1999, 40-42) lists several things in favour of faceted classifications: they do not require complete knowledge of the entities or their relationships; they are hospitable (can accommodate new entities easily); they are flexible; they are expressive; they can be ad hoc and free-form; and they allow many different perspectives on and approaches to the things classified. She lists three major problems: the difficulty of choosing the right facets; the lack of the ability to express the relationships between them; and the difficulty of visualizing it all. Choosing the right facets is crucial, and requires good knowledge of the items being classified and the users, but that is true of any sort of classification or organization. The lack of the expression of relationships between the facets is a problem, one that this paper will not discuss. We will assume that the items are known well by the users, who can infer relationships from their own knowledge. As to the difficulty of visualization, Kwasnick (1999, 42) notes, "Information technology has promise for new ways of enabling multidimensional visualization and for developing computer-assisted ways of discovering patterns and anomalies that can possibly lead to new knowledge." We will see that the web fulfills that promise.

1.3 Dish detergent example

A running example will help illustrate the processes and principles I will show. I decided to use for the domain (as we will call the set of things to be classified) a small set of commercial products: dishwashing detergents. It serves our purpose for a number of reasons: it is a familiar; it is fairly small (though larger than one might think without having examined supermarket shelves); the language used is somewhat restricted already; and, as we will see later, it will serve as a good example of how to expand a small classification into a larger one.

I visited the largest supermarket in my neighbourhood and wrote down the names of all the dish detergents on their shelves. I have slightly rearranged their order, but they were shelved almost like this—note the classifications the store made when presenting the products to the shopper:

> Dishwasher liquids: Cascade Pure Rinse Formula; Electrasol lemon gel; No Name lemon gel; Palmolive spring blossom gel
>
> Dishwasher powders: Cascade Complete; Cascade, fresh scent; Electrasol Double Action, fresh scent; No Name Premium Formula; No Name Premium Formula, lemon-scented; Sunlight Lemon Fresh
>
> Gelpacs: Electrasol Gelpacs, orange blossom
>
> Tablets: Electrasol tabs; No Name Premium Formula tablets
>
> Washing-by-Hand Liquids: Ivory Classic; No Name Liquid Dish Detergent, lemon-scented; Palmolive Aroma Therapy, Lavender and Ylang Ylang; Palmolive Aroma Therapy, Mandarin and Green Tea Essence; Palmolive Original; Palmolive Spring Sensations, Fresh Green Apple; Palmolive Spring Sensations, Ocean Breeze; Palmolive Spring Sensations, Orchard Fresh; Palmolive, antibacterial; President's Choice Antibacterial Hand Soap & Dishwashing Liquid; President's Choice Invigorating Aroma Therapy, Passion Flower; President's Choice Relaxing Aroma Therapy, Ruby Red Grapefruit; President's Choice Tough on Grease; Sunlight, antibacterial; Sunlight, lemon fresh

For simplicity we will ignore the size and price of the products. Shoppers will be our target audience: chemists and engineers working with soaps and household cleaning products would have different needs requiring a different classification. In the next section, we will take these items and build a classification the supermarket can use to help shoppers find and buy the detergent they want.

2. How to create a faceted classification

This section will describe a seven-step procedure for building a faceted classification, based on the work of B.C. Vickery (1960) and Louise Spiteri (1998). We will begin, however, by looking at some existing faceted classifications to see how they are designed and what ideas they offer for making new classifications.

2.1. Sample faceted classifications

First, let us look at two of the three best-known faceted universal classification systems: the Colon Classification and the second edition of the Bliss Bibliographic Classification (BC2). S.R. Ranganathan's Colon Classification has five facets, now classic (see Ranganathan (1962), among his many books, for an introduction to the facets and how to use them):

- Personality (the something in question, e.g. a person or event in a classification of history, or an animal in a classification of zoology)
- Matter (what something is made of)
- Energy (how something changes, is processed, evolves)
- Space (where something is)
- Time (when it happens)

These five, known as PMEST, may be enough for you. If you need more, look to BC2 for ideas (Broughton 2001, 79):

- thing/entity
- kind
- part
- property
- material
- process
- operation
- patient
- product
- by-product
- agent
- space
- time

Vanda Broughton, one of the editors of BC2, said, "These fundamental thirteen categories have been found to be sufficient for the analysis of vocabulary in almost all areas on knowledge. It is however quite likely that other general categories exist; it is certainly the case that there are some domain specific categories, such as those of form and genre in the field of literature" (2001, 79-80). BC2 makes an excellent starting point for thinking of how to make a faceted classification. Its facets can be renamed and adapted to suit your particular circumstances.

Vickery, a member of the Classification Research Group, had other general suggestions:

> As well as these [general categories], in any scientific classification there may occur a number of terms applicable at several points in the combination formula. For example, any property or process may itself have a general property: rate, variation, and so on. There are general operations on properties (e.g. measurement) and on processes (e.g. initiation, control). There are also a number of operations concerned with apparatus (equipment, instruments) such as design and maintenance. Lastly there are a number of common logical or mental operations: comparison, explanation and so on. [Vickery then quotes other sets of fundamental categories, such as Shera and Egan's] agent, act, tools, object of action, time, space, and product. Barbara Kyle has written of natural phenomena, artefacts, activities, and 'purposes, aims, ideas, and abstracts.' De Grolier suggests the 'constant categories' of time, space, and action, and the 'variables', substance, organ, analytic, synthetic, property, form, and organization. (1960, 23-24)

The smaller the domain, the more specific and detailed the facets can get. There is little or no need to deal with the complications inherent in organizing the world of knowledge, and the system can be as precise as necessary to do what is needed. Here follow some examples of smaller classifications, beginning with the Art & Architecture Thesaurus (http://www.getty.edu/research/conducting_research/vocabularies/aat/) (Petersen 1994, 26), which is not actually a classification scheme, but is indeed faceted. Note how some of the classifications are based on Ranganathan's Personality, Matter, Energy, Space, and Time.

- Associated Concepts (e.g., philosophy)
- Physical Attributes (density)
- Styles and Periods (Simulationist) (similar to Space and Time)
- Agents (People/Organizations) (lighthouse keepers)
- Activities (thinking) (similar to Energy)
- Materials (plywood) (similar to Matter)
- Objects (bunk beds) (similar to Personality)

Epicurious (http://www.epicurious.com/) (n.d.) is a web site about cooking, and organizes its recipes this way:

- Cuisine, i.e. ethnic origin (e.g., Indian) (similar to Space)
- Special considerations (low-fat)
- Meal/Course (soups)
- Main Ingredients (potatoes) (similar to Matter)
- Preparation (grill) (similar to Energy)
- Season/Occasion (Valentine's Day) (similar to Time)

Vickery (1975, 189-192) describes a classification scheme for containers:

- Products (e.g., jam)
- Parts (valves)
- Materials (cork)
- Operations (tooling)
- Miscellaneous Common Sub-divisions (research, information, safety)

There will be times when four or five facets are not enough. Vickery's soil classification (1960, 20-21) needed eighteen:

- Soil, according to constitution (e.g., peat soil)
- Soil, according to origin (granitic soil)
- Soil, according to physiography (desert soil)
- Soil, according to texture (sandy clay)
- Soil, according to climate (arctic soil)
- Physical part of soil (gravel)
- Chemical constituent of soil (nitrogen)
- Structure of soil (profile)
- Layer of soil (horizon)
- Organism in soil (bacteria)
- Parent material of soil (muck)

- Process in soil (mineralization)
- Property of soil (cohesion)
- Measure of property (sticky point)
- Operation on soil (amendment)
- Equipment for operations (plough)
- Substances used in amendment (lime)
- Operations on these substances (placement)

The International Sematech Wafer Services group (http://www.sematech.org/waferservices/find.htm), part of an international consortium of semiconductor manufacturers, uses very specific categories that will make little sense to most people outside the field (International Sematech 2003). These seem to be a mix of special Matter, Space, and Energy facets, but there are no definitions given so it is hard to guess. With a small or expert set of users, there is great freedom to be precise and terse:

- Wafer Size (e.g., 200 mm)
- Application (etch)
- Feature Type (via/contact)
- Feature Size (0.22 m)
- Dielectrics (oxide)
- Fills (tungsten)
- Electrical Test (BTS)

Lillian Vernon Online (http://www.lillianvernon.com/) (n.d.) has a gift finder on its web site that lets the user narrow down the range of possible purchases by selecting from four facets. Depending on the search restrictions, any of a range of products for travel, the bathroom, bedroom, kitchen, garden, or office will be listed.

- Gift recipient (e.g., her)
- Their interest (entertaining)
- Occasion (birthday)
- Price range ($25-$50)

The site has two more options: one lets the user narrow the results to gifts that can be personalized; the other controls the ordering of the results: by product name or by increasing or decreasing cost. Any commercial classification will need a cost or price facet.

2.2. Spiteri's simplified model for facet analysis

Louise Spiteri (1998) analyzed the complicated sets of rules that Ranganathan and the Classification Research Group laid down for how to make faceted classification systems, and drew up her own simpler set. She did so primarily so as a teaching tool for library and information science students, but notes that "the model could be used by designers of faceted classification systems and IR [information retrieval] thesauri, because these designers might also need to consult a variety of sources to obtain the principles of facet analysis needed for their work" (1998, 4). Spiteri's principles are straightforward and complete and will form the backbone of the procedure set out below.

Spiteri follows Ranganathan and divides classification into three parts: "the Idea Plane, which involves the process of analyzing a subject field into its component part; the Verbal Plane, which involves the process of choosing appropriate terminology to express those component part; and the Notational Plane, which involves the process of expressing these component parts by means of a notational device" (1998, 5). Going from Idea (which Spiteri divides into two parts) to Verbal to Notational Planes take us from idea to word to number, from the general concept of what the entity is about to expressing that concept in a controlled vocabulary to turning those words into notation. The last step of the Notational Plane will be less important than the others for us because of how the web works.

Here are the principles of Spiteri's model for creating faceted classifications:

Idea Plane: Principles for the Choice of Facets

a) **Differentiation**: "when dividing an entity into its component parts, it is important to use characteristics of division (i.e., facets) that will distinguish clearly among these component parts" (Spiteri 1998, 5). For example, dividing humans by sex.

b) **Relevance**: "when choosing facets by which to divide entities, it is important to make sure that the facets reflect the purpose, subject, and scope of the classification system" (1998, 6).

c) **Ascertainability**: "it is important to choose facets that are definite and can be ascertained" (1998, 6).

d) **Permanence**: facets should "represent permanent qualities of the item being divided" (1998, 18).

e) **Homogeneity**: "facets must be homogeneous" (1998, 18).

f) **Mutual Exclusivity**: facets must be "mutually exclusive," "each facet must represent only one characteristic of division" (1998, 18).

g) **Fundamental Categories**: "there exist no categories that are fundamental to all subjects, and ... categories should be derived based upon the nature of the subject being classified" (1998, 18-19).

Idea Plane: Principles for the Citation Order of Facets and Foci

a) **Relevant Succession**: "the citation order of the facets [and foci] should be relevant to the nature, subject, and scope of the classification system" (1998, 21-22). Suggestions: chronological order, alphabetical order, spatial or geometric order, simple to complex, complex to simple, canonical, increasing quantity, or decreasing quantity.

b) **Consistent Succession**: "once a citation order of facets has been established for a classification system, it should not be modified unless there is a change in the purpose, subject, or scope of the system" (1998, 7-8). We'll ignore this if we allow the facets to be rearranged.

Verbal Plane

a) **Context**: "the meaning of an individual term is given its context based upon its position in the classification system" (1998, 11). E.g., London, Ontario, and London, England, can be identified as "London" alone, and which one is meant can be seen by whether that part of the classification is for cities in Ontario or England.

b) **Currency**: "the terminology used in a classification system should reflect current usage in the subject field" (1998, 11). This means the system will require regular attention and revision.

Notational Plane

a) **Synonym**: "each subject can be represented by only one unique class number" (1998, 12).

b) **Homonym**: "each class number can represent only one unique subject" (1998, 12).

c) **Hospitality**: "notation should allow for the addition of new subjects, facets, and foci, at any point in the classification system" (1998, 20).

d) **Filing Order**: "a notational system should reflect the filing order of subjects. This type of notation would reflect the citation order underlying the classification system" (1998, 20).

2.3. Making the faceted classification

Vickery had four steps for making a faceted classification scheme: first, "[t]he essence of facet analysis is the sorting of terms in a given field of knowledge into homogeneous, mutually exclusive facets, each derived from the parent universe by a single characteristic of division" (1960, 12). Following that are three steps: "(i) to assign an order in which the facets will be used in constructing compound subject headings, (ii) to fit the schedules with a *notation* which permits the fully flexible combination of terms that is needed and which throws subjects into a preferred filing order, and (iii) to use the faceted scheme in such a way that both specific reference and the required degree of generic survey are possible" (1960, 13). My procedure for making the faceted classification system rearranges Vickery's steps, and adds to the start and finish to make it complete from beginning to end.

1. **Domain collection**. Collect a representative sample of the entities. In a large domain, get enough to cover all foreseen possibilities. In a small domain, use the entire domain.

2. **Entity listing**. List the entities, breaking down the descriptions into parts and rearranging words. Separate sentences and phrases into their basic concepts and isolate these concepts.

3. **Facet creation**. 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, into which all the terms from the previous step will fit. This is the Idea Plane, so use as guidelines the Principles for the Choice of Facets. Use the Colon Classification's and BC2's facets as starting points, and draw from other classifications for inspiration. As Foskett (2003, 1064) put it:

> In applying facet analysis to a subject, the first step is to examine a representative sample of the literature and enumerate the subject of each article, book, or abstract. It soon becomes clear that the terms encountered may be grouped together in sets according to their relation to the subject and to each other, that they represent, in fact, the various aspects of the subject, each of which can be studied apart from the others, at least conceptually, even though in practice it may be impossible to separate them into groups of static entities apart from the phenomena in which they are observed.

Broughton (2001, 79) says of performing this step with BC2's categories:

> It can be seen that the categorical analysis is on a functional basis in which there are two different types of functionality; linguistic function and operational function. In essence, the categories represent a 'production process', and are particularly suitable for the analysis and organization of terms in technology (where originally they were developed). In analysing a given field one can ask, what is being done (or produced), what are its parts and properties, how this is achieved, by what means and by whom, where and when.

Remember the purpose of the classification and the users. Who will use it? Why? Will they search it, browse it, or both? How well do they know the subject? Always remember it is meant for them to use.

4. **Facet arrangement**. First, do a test ordering of all of the terms (now to be known as "foci", or individually, "a focus") under the facets. Use the Principles for the Citation Order of Facets and Foci as guidelines. You may have a few marginal concepts that do not have a good home, and we will discuss that problem later. When the foci are set out, do a test classification and make sure that all the original entities can be described by picking foci from the facets. If something was missed, go back and reanalyze or rearrange the terms.

Second, after the rough draft of the classification has been shown to work, do the final facet arrangement, using the Principles for the Citation Order of Facets and Foci. At this point you are working on both the Idea Plane and Verbal Plane, so mind the relevant rules. Kwasnick (1999, 39-40) says of this step that "[e]ach facet can be developed/expanded using its own logic and warrant and its own classificatory structure. For example, [in the Art and Architecture Thesaurus] the Period facet can be developed as a timeline; the Materials facet can be a hierarchy; the Place facet a part/whole tree, and so on." You are free to use whatever is best.

Now you must decide on a controlled vocabulary, if you have not already naturally been using one. Foci are official words and phrases that will always used for the concepts or things they represent. Other terms must be translated into existing foci. For example, if you are classifying tinned vegetables, now is the time when you choose between "chick peas" and "garbanzo beans." Authority and vocabulary control like this is a complex field of its own and is beyond the scope of this paper, but however you handle it, users and other classifiers must be led from their own words to your chosen terminology. Without that, the system will work for no-one but you.

5. **Citation order**. This is the Notational Plane. Kwasnick (1999, 40), says, "[i]n organizing the classified objects, choose a primary facet that will determine the main attribute and a citation order for the other facets. This step is not required and applies only in those situations where a physical (rather than a purely intellectual) organization is desired." She means, for example, library shelves or printed bibliographies. On the web, we can make the order as changeable as necessary if we wish. However, decide on a standard citation order that will be the default way things are ordered on the web site. It should be sensible, understandable, and useful to all users. Experienced or curious users can rearrange the order if they wish and are allowed to do so, but most people will use the basic view provided.

6. **Classification**. The classification system is now finished. Use it to classify everything in the domain. Analyze the entities using the facets, and apply the proper foci to describe each entity.

7. **Revision, testing, and maintenance**. If there were any problems in step 6, go back as many steps as necessary to correct the problem. You may find that the arrangement of foci in a facet needs changing, or that some foci are missing, or perhaps even that your choice of facets is inaccuruate or incorrect. Go back through the steps until you are able to classify everything to your complete satisfaction. Test it on users (how to do so is unfortunately outside the scope of this work) and make any more necessary changes. Finally, prepare to do regular maintenance on the classification: update terminology as it changes; verify that any new

entities can be clearly and accurately classified; and make sure that the facets and foci are complete enough to handle the domain as it changes over time. You may need to add or rearrange foci, add a new facet, or someday even make a whole new classification.

2.4. Classifying dish detergents

Now we will apply the above seven rules to the domain of dish detergents, as listed in §1.3. Step 1, Domain Collection, is done. Step 2, Entity Listing, is done too: the product names are already broken down into constituent concepts such as brand name, product name, scent, etc. This makes step 3, Facet Creation, easier.

In Facet Creation we want to begin by seeing what general, high-level categories appear across all the entities. Some stand out: Brand Name (Ivory, Palmolive, Cascade); Form (dishwasher liquid, dishwasher power, handwashing liquids); Scent (lemon, ocean breeze). The classification is meant for shoppers, so Brand Name is the right name for that facet, but if the classification was meant for stores, we might want to list Manufacturer or Distributor. Is that Form category workable? On closer examination, we can see it violates the Principle of Homogeneity. There are actually two things there: the physical form of the dishwashing soap (liquid, gel, powder), and how the cleaning is done (by machine or by hand). Let us call the first category Form, and the second (from BC2) Agent. Ultimately, of course, a person is behind the cleaning, but the Agent will be how they perform the cleaning—by machine or by themselves.

Brand Name, Form, Agent, and Scent take care of almost all the words and phrases in the entity listing, but we are left with antibacterial Sunlight and the aroma therapy products. The antibacterial property does not fit in with any of the categories we have already made, and is quite different from them, so it will need one of its own. We will not use a yes/no Antibacterial category, because facets should be more than on/off switches. The property is related to, perhaps, a special coffee-stain-fighting property of a detergent, or something that makes white dishes whiter. Special Cleaning Property? The entire classification is about cleaning. Special Property? That is a bit awkward, but will do. For the aroma therapy detergents, we can see that this property is different from any other because it affects (or at least purports to) the user of the detergents, not the dishes. This is similar to the inclusion of skin lotion to prevent chapping. It does not fit any of the general BC2 categories, but Effect on User would handle it—but only if the user is a person. What if the detergent helps the longevity of a dishwashing machine? That's basically the same idea, so Effect on Agent is a better term.

Do these facets obey Spiteri's principles? The Principle of Differentiation holds: the facets will make clear distinctions about the different aspects of the entities. The Principle of Relevance holds: these are all important things to know about when buying a dish detergent. The Principle of Ascertainability holds: we can determine everything we need just by reading the label. The Principle of Permanence holds: the detergents will not change from liquid to powder by themselves, nor will an orchard fresh scent turn into the heady odour of ylang ylang. The Principle of Homogeneity holds, as does the Principle of Mutual Exclusivity. Finally, the Principle of Fundamental Categories holds: we have derived the facets (all the ones possible, in fact) from the nature and purpose of the detergents. Spiteri's principles are all fulfilled.

Nevertheless, the facets can be challenged. It can be argued, for example, that Scent is really an effect on the ultimate agent, the human doing the dishes or running the machine. Philosophers would say that while the Form of the detergent is a primary quality, inseparable from the object, Scent is a secondary quality,

existing in the person smelling it. As to the particular foci in Scent, what exactly are "ocean breeze" and "orchard fresh"? How do ruby red grapefruits smell different from other grapefruits? But those are the terms used, so we will stick with them, ridiculous though they may be. As well, perhaps the aroma therapy detergents should be bundled in with Scent in a more encompassing facet, since they are all about an effect on the human using the product. Or perhaps not. Our facets do seem appropriate for the needs of the classification, so we will leave them as they are.

Going into Step 4, Facet Arrangement, we have Brand Name, Form, Agent, aEffect on Agent, Scent, and Special Property. Now we must arrange the foci under those categories.

Brand Name: Cascade, Electrasol, Ivory, No Name, Palmolive, President's Choice, Sunlight

Form: gel, gelpac, liquid, powder, tablet

Agent: dishwasher, person

Effect on Agent: aroma therapy (with subdivisions or sub-foci: invigorating, relaxing)

Scent: green apple, lavender and ylang ylang, lemon, mandarin and green tea, ocean breeze, orange blossom, orchard fresh, passion flower, ruby red grapefruit

Special Property: antibacterial

A check will show that all the detergents can be classified with these facets and foci. Arranging the foci in alphabetical order is simplest and most appropriate, because there is no other sensible way to order them (national origin of manufacturer, for example, or botanical classification of the source of the scent) that would help our users, the shoppers. This follows the Principle of Relevant Succession. The only unusual case here is the two kinds of aroma therapy in Effect on Agent. I subdivided it into invigorating and relaxing, making one focus with two sub-foci. In Scent, we will have to break down the combination scents ("lavender and ylang ylang") into their components, and then put them back together later when the entities are classified. It is perfectly legal for two foci in the same facet to apply to one entity. If several apply, they can be handled invididually, or perhaps grouped together into a larger class if there are many. We could group "mandarin" and "orange blossom" together somehow under "oranges," but for simplicity we will leave them as they are.

Step 5 is Citation Order. Is the order given above (Brand Name, Form, Agent, Effect on Agent, Scent, Special Property) satisfactory? We will reapply the Principle of Relevant Succession, this time to the facets. When people look for a dish detergent, the first thing on their mind is what Agent will do the cleaning (a person, or a dishwasher) and what Form they need. After that, they probably look by Brand Name, because they have a favourite brand and they stick with it. After that, Scent is probably the next most interesting facet, and then Effect on Agent and Special Property can follow. That will be the default citation order, but when it goes on the web we can let the users reorder the facets to their liking. (This will allow users to navigate and browse the classification, the equivalent of scanning the shelves. Most shoppers will just grab the same detergent they always buy, which is the equivalent of a known- item search.)

Here is the final form of the classification scheme:

```
Agent: dishwasher, person

Form: gel, gelpac, liquid, powder, tablet

Brand Name: Cascade, Electrasol, Ivory, No Name, Palmolive, President's Choice, Sunlight

Scent: green apple, green tea, lavender, lemon, mandarin, ocean breeze, orange blossom, orchard fresh,
passion flower, ruby red grapefruit, ylang ylang

Effect on Agent: aroma therapy (subdivisions: invigorating, relaxing)

Special Property: antibacterial
```

Step 6 is Classification. For example, "President's Choice Antibacterial Hand Soap & Dishwashing Liquid" would be:

```
Agent: person

Form:  liquid

Brand Name: President's Choice

Scent: (none)

Effect on Agent: (none)

Special Property: antibacterial
```

"Palmolive Aroma Therapy, Lavender and Ylang Ylang," would be

```
Agent: person

Form: liquid

Brand Name: Palmolive

Scent: lavender, ylang ylang

Effect on Agent: aroma therapy

Special Property: (none)
```

We will skip Step 7, Revision, Testing, and Maintenance, since we have no users to consult.

2.5 Marginal subjects and expanding the classification

Eventually every limited classification will have to handle something that does not quite fit. Vickery (1960, 16-19) offers a few ways of handling these situations. If some uncommon but related terms appear, one solution is to draw on another classification that handles them fully. Vickery's example, from the soil classification, involves chemical substances. Some are listed in the classification, but others, as needed, can subdivide an "Other" focus by how they are handled in another classification. "Other" is always a convenient place to drop anything that does not have a proper home, but not an easy place to find things.

If the mismatch is more than just a missing term or two, the entity can be classified as something that is related to it, even if the relation is tenuous and only exists in the way the users associate things. Vickery's example is of an insurance classification where a book on the textile industry is put with "Fire insurance:

textile industry." Even though the book is not about insurance, fire insurance is why the users would need to read about the textile industry.

In our dish detergent example, a potential marginal entity would be the products that prevent streaks from appearing on glasses. Sometimes this is made part of the detergent, in which case we would file it under Special Property. A stand-alone anti-streak agent would fit into our classification if we ignore the fact it is not a dish detergent and isolate the anti-streak aspect in a note or comment. If two such products are handled the same way, there would be no way to connect their shared property. That is when we would need to think about adding a new facet.

If too many things are filed under "Other," or need to have important aspects of their nature ignored so they can be made to fit, re-examine the classification and see if it needs rearrangement or a new facet. Follow the rules above to see what the facet should be. When added, it should mean that reclassifying all the marginal entities gives them a new, accurate home and makes them easy to find. In the dish detergent example, because of the facets we have chosen, it would be easy to expand the classification to include other cleaning products. The classification is for dish detergents, and each of those words could be the beginning of a new facet: the dish is the object being cleaned, and a detergent is a special sort of cleaning agent, chemically different from bleach, ammonia, and even soap. (It might even be true that we are mixing soaps and detergents in our domain, without acknowledging the difference, but it does not matter for our purposes.) Adding, for example, Object and Cleanser Type would mean we could classify hair shampoos, rug shampoos (which would need a new Agent focus: rug cleaners), face soaps (possible new Effect on Agent: acne reduction), Lysol, Vim, Ajax, Comet, Windex, and every other cleaning product. Two facets, neither of which would have many foci, would mean we could handle an enormous range and number of new entities. It would not be the end of our problems, though: eventually there would be more marginal entities to consider. Does mouthwash fit? Furniture polish? Hair conditioners? Products to prevent static cling?

3. How to store the faceted system in a computer

Two ways to store the faceted classification system on a computer are to use XFML or a relational database. I will examine both with reference to the dish detergents. For both, it is important to note that it will be a lot of work to change the classification after it has been implemented. In the last section we saw how powerful the addition of two facets would be. What took a paragraph to explain would take much longer to actually make work. When designing a system to store a classification on a computer, make it as easy as possible to handle changes, but do everything to prevent that from ever being necessary. The hospitality and flexibility that Kwasnick (1999, 40) listed as good features of facets can be found in the web interface, but are not often matched by similar qualities in software.

3.1 XFML

XFML is a markup language written in XML, and hence looks similar to HTML. It is used to put faceted classifications into a standard machine- and human-readable form that is easy to store, transmit, and manipulate. The specification (Van Dick 2003) has the complete rules on how to use it, and we will just cover the basics here.

There are two main elements in XFML: `facet` and `topic`. The `facet` element defines the top-level facets. It has only one attribute, `id`, which will be the name used internally to identify the facet. It can be an abbreviation or code number, but here we will use the full name:

```
<facet id="agent">Agent</facet>
<facet id="form">Form</facet>
<facet id="brand_name">Brand Name</facet>
<facet id="scent">Scent</facet>


<facet id="effect_on_agent">Effect on Agent</facet>
<facet id="special_properties">Special Properties</facet>
```

That is all that is needed to define the facets. Defining the foci in each one is done similarly, with the `topic` element (as XFML calls foci). Each `topic` refers back to its parent facet. For example, the foci in the Brand Name facet are defined this way:

```
<topic id="cascade"    facet_id="brand_name"><name>Cascade</name></topic>
<topic id="electrasol" facet_id="brand_name"><name>Electrasol</name></topic>


<topic id="ivory"      facet_id="brand_name"><name>Ivory</name></topic>
<topic id="no_name"    facet_id="brand_name"><name>No Name</name></topic>
<topic id="palmolive"  facet_id="brand_name"><name>Palmolive</name></topic>


<topic id="presidents_choice"
                facet_id="brand_name"><name>President's Choice</name></topic>
```

Again, the `id` attribute is an internal identifier. `facet_id` points to the facet where the focus belongs, and inside the `name` tag is the actual name of the focus.

The Effect on Agent facet demonstrates how to arrange foci and sub-foci. Remember that the only focus in this facet is "aroma therapy," which is divided into "invigorating" and "relaxing." The facet and its contents can be described thusly, first by defining the facet, then the focus, and then connecting the focus to the two sub-foci `parentTopicid` attribute:

```
<facet id="effect_on_agent">Effect on Agent</facet>
<topic id="aroma_therapy" facet_id="effect_on_agent"><name>aroma therapy</name></topic>


<topic id ="invigorating" facet_id="effect_on_agent" parentTopicid="aroma_therapy">
<name>invigorating</name></topic>
<topicid="relaxing" facet_id="effect_on_agent" parentTopicid="aroma_therapy"><name>relaxing</name>
</topic>
```

When all the facets and topics (foci) have been defined, we can define the entities. We would create a web page for each one (a description of the entity) and then locate it in the classification. For example, President's Choice Antibacterial Hand Soap & Dishwashing Liquid would look like this:

```
<page url="http://www.example.com/dishdetergents/pc/ahsdl.html">

    <title>President's Choice Antibacterial Hand Soap & Dishwashing Liquid</title>

    <occurrence topicid="person" />            <!-- Agent facet -->


    <occurrence topicid="liquid" />            <!-- Form facet -->

    <occurrence topicid="presidents_choice" /> <!-- Brand Name facet -->

    <occurrence topicid="antibacterial" />     <!-- Special Property facet -->
</page>
```

That matches up the entity with its facets. There is more to XFML than this, but not a lot. As markup languages go, XFML is fairly simple.

Once the whole classification has been rendered in XFML, there remains the problem of how to use it. There seems to be little open source software for XFML available on the Internet (http://www.xfml.org/software.html), and there are no full-featured XFML libraries under continuing maintenance for any major programming language. However, because XFML is written in XML, any of the many XML libraries can be used to handle it. A programmer can build custom XFML-handling code without much trouble. There is one commercial product that handles XFML, Facetmap (http://www.facetmap.com/).

3.2 Relational databases

The other option is to store the classification system in a relational database. I will show first how to design the database, and then how to use it for searching and navigation. The design is based on an entity-relationship model (Chen 1976). ("Entities" in the title means something slightly different from the way we have been using it, and to avoid confusion we will not use it in that sense again.)

We begin by seeing how each facet relates to the entities: is it a one-to-many or many-to-many relationship? In our example, most are one-to-many: for example, one Brand Name at a time can be used by many detergents. One Form may be the form of many detergents—there are several liquids and several powders—but each detergent can have only one form. Such one-to-many relationships result in simple database table structures. For example, the Brand Name facet becomes the BRAND_NAME_T table. Each focus has its own row and unique primary key:

PK	BRAND_NAME
1	Cascade
2	Electrasol
3	Ivory
4	No Name

5	Palmolive
6	President's Choice

For the sake of the example, EFFECT_ON_AGENT_T will be much simplified. Effect on Agent is a hierarchical facet and that should be shown in the database design. As we have seen, this is easy to handle in XFML, but here we will just treat it as a list so as not to get bogged down in details:

PK	EFFECT_ON_AGENT
1	aroma therapy
2	aroma therapy—invigorating
3	aroma therapy—relaxing

The other facets are all similar, except for Scent. It is different because a detergent can have more than one scent at a time, for example "mandarin and green tea." This requires a many-to-many relationship, which requires a special table to join together detergents and scents. We will leave that to last. First, as with the other facets, we list Scent in the SCENT_T table:

PK	SCENT
1	green apple
2	green tea
3	lavender
4	lemon
5	mandarin
6	ocean breeze
7	orange blossom
8	orchard fresh
9	passion flower
10	ruby red grapefruit
11	ylang ylang

Now we build the table that will pull almost everything together and describe the detergents: ENTITY_T. It will have a field for the name of the detergents, and other fields to handle all the one-to- many relationships. Here are how the two sample classifications from §2.4 will look. For readability the usual row and column view of a database will be shown with one line per column. The numbers here are the primary keys of the relevant rows from the other tables, e.g., row 6 in BRAND_NAME_T is President's Choice. If the table was not given in full above, row numbers are inferred from the facet listings in §2.4.

```
PK: 1
NAME: President's Choice Antibacterial Hand Soap & Dishwashing Liquid
AGENT: 2
FORM: 3
BRAND_NAME: 6
EFFECT_ON_AGENT: (null)
SPECIAL_PROPERTY: 1


PK: 2
NAME: Palmolive Aroma Therapy, Lavender and Ylang Ylang
AGENT: 3
FORM: 3
BRAND_NAME: 2
EFFECT_ON_AGENT: 1
SPECIAL_PROPERTY: (null)
```

Notice that there is no mention of Scent. Last, we build HAS_SCENT_T, where we associate detergents and scents using the primary keys from the existing tables:

PK	ENTITY	SCENT
1	2	3
2	2	11

This table says that the entity with primary key 2 in ENTITY_T (i.e., Palmolive Aroma Therapy, Lavender and Ylang Ylang) is associated with scents 3 and 11 in SCENT_T (i.e., lavender and ylang ylang, respectively).

The database structure is now complete. We have AGENT_T, FORM_T, BRAND_NAME_T, SCENT_T, EFFECT_ON_AGENT_T, and SPECIAL_PROPERTY_T, one table for each facet. We have ENTITY_T to hold the names of the entities and handle the one-to-many relations, and HAS_SCENT_T to match up the many-to-many relationships of entities and Scent. In general, for classifications like this, the database will require (# of facets + # of one-to-many relations + 1) tables.

Once the tables are populated and the entities are stored in the database, we can use SQL to generate navigation and to answer user search queries. For example, if the web site had a "Browse by Brand Name" option, the list of manufacturers would be pulled from the database by running:

```
select PK, BRAND_NAME from BRAND_NAME_T;
```

(The user would not see the primary key, but the system would need to remember it.) The user could choose Palmolive, then choose a final "Browse by Form" option to see what powders, gels, and liquids Palmolive makes, regardless of where they are to be used, their scent, or any special properties. This query will list the keys of the forms used by Palmolive entities:

```
select FORM from ENTITY_T where BRAND_NAME = '5';
```

If the user wants to see all information about all detergents made by Palmolive, this query will find it:

```
select e.*, s.     SCENT from ENTITY_T e, SCENT_T s, HAS_SCENT h
where e.FORM =      '5'
and h.ENTITY =      e.PK
and h.SCENT =       s.PK;
```

(The many-to-many relationship for Scent makes the SQL statements longer than they would be otherwise.)

Searching is easy to handle. If the user is shown a menu listing each facet and its foci, and given the ability to pick and choose what particular elements he or she wants, then an SQL statement is easy to build. For example, if the user asks to see all lemon-scented powders (where "powder" is fourth in the FORM_T):

```
select e.*, s.     SCENT from ENTITY_T e, SCENT_T s, HAS_SCENT_T h
where e.form =      '4'
and h.scent =       '4'
and h.entity =      e.pk
and h.scent =       s.pk;
```

3.3 Which is better?

XFML offers all the benefits of XML, while relational databases offer SQL. Whoever is in charge of designing the implementation of a web-based faceted classification system will need to choose whichever is best given his or her particular circumstances: which will run faster, the expertise of the programmers, technical requirements and limitations in the organization, the size of the classification, how the classification will relate to the entities it organizes, and future plans. In XFML's favour is that its files are plain text and can be created and edited by hand without any special tools, and we saw how easy it is to handle hierarchical facet listings. However, while using a database adds some complexity, SQL and relational databases are not only very powerful but familiar to most programmers. There are many SQL libraries available for all major programming languages.

4. How to put the classification on the web

Facets and the web go well together. It is easy show the user a menu of facet listings and let him or her pick and choose what is of interest. The user can make quick choices, thinking, "I'd like to see something with that, and that, and a bit of that, and I don't care about the rest," then click a button and see the results. Such systems are becoming more common and users will be ever more comfortable with them. Complete coverage of all the issues involved in putting facets onto the web is beyond the scope of this paper, however, touching as it does on information visualization, human-computer interaction, design and use of OPACs, and hypertext and/or graphical display of information. I will discuss some of the basic issues involved in putting facets on the web and suggest how to handle them with simple HTML and Javascript. For deeper analysis, the reader should begin with Rosenfeld and Morville (2002), who cover everything about information architecture and the

web, and Baeza-Yates and Ribeiro-Neto (1999), who discuss all aspects of information retrieval including interface design. As well, IFLA's (2003) guidelines for OPAC displays will be of great use even to people outside the library world.

There are two basic ways to make a faceted classification usable on the web: keyword searching or facet-based navigation.

4.1 Keyword searching

In keyword searches, the user types in one or more words to see if they match anything. The system will search facets, foci, and entity descriptions. The results display will depend on many things: the number and location of any keyword matches, the size and nature of the classification, the nature of the entities and their descriptions, the users, and the intent of the web site. Whatever is done, it should help the user find the most relevant results in the shortest time. If the search terms match entity descriptions, they should be shown—but there is the problem of how to handle too many results. Matching facets and foci should have precedence on the page: there will be fewer of them, they are of more importance in the classification, and they can serve as starting points for navigation. In the dish detergent classification, for example, searching on "lemon" should return all the lemon-scented products, each listed under its full classification. The user could look at details of the products, use the listed classifications as starting points for further browsing, or refine the search with more terms.

Efficient keyword searching like this requires a well-managed controlled vocabulary, as mentioned in the Facet Arrangement step in §2.4. Such a system would mean that a user searching for citrus-scented detergents would see the lemon-, grapefruit-, and orange-scented products. "Citrus" does not appear in any product names, but the classification creator could arrange that anyone searching for that word would be led to the two known related words. Aside from such a controlled vocabulary, if the entities are text (for example, articles or books), then full-text searching is possible. A full treatment of these searches is beyond the scope of this essay (but see Rosenfeld and Morville (2002) for a good web-focused treatment of the topic), and the rest of the discussion will focus on searches where the user's options are controlled. The key point is that any search should show the user what matches and where it sits in the classification, so that the user can either navigate directly to the entity they want, or browse for related items.

4.2 Facet-based navigation: three questions and four principles

There are three questions to ask when planning how to build the web site. Finding and blending the appropriate answers will give you a good starting point for building navigational tools for the site.

1. Do you want to focus on free navigation or on navigation by selection? The first lets the user move from page to page through a list of hypertext links; the second has the user navigate by choosing options in forms (single or multiple select menus, radio buttons, or checkboxes) and clicking a submit button. Form elements allow for more interaction. Single and multiple select menus differ in their size on the page and because more than one option can be selected simultaneously in the latter. Checkboxes and radio buttons show a list of choices to the user, who must make a selection by activating a button beside a term. Only one radio button can be active at a time, but with checkboxes more than one can be chosen. When forcing foci selection in a menu, use single select menus or radio buttons for one-to- many relationships, and multiple selects or checkboxes for many-to-many relationships.

2. What are the facets like? How many are there? Which are the most important? How long are the foci listings? Do they vary widely? Do the entities all use only one focus from each facet, or are there some many-to-many relationships? Are they trees, hierarchies, timelines, lists, or some other ordering? Simple alphabetical or numeric lists, or timelines, can be displayed as they are. Hierarchies and trees will need to be specially formatted for clarity, perhaps into indented lists. Collapsible lists (as found in most graphical file manager utilities) are understood well by users, but require Javascript.

3. How much control over the facet ordering (the citation order) will users have? Will you restrict them to a particular ordering, or can they rearrange the facets as they please? Facets can be reordered by giving each an ordinal ("second") or cardinal ("2") number and letting the user choose their own order, or by letting the user drag and drop the form elements on the page (perhaps with Javascript and dynamic HTML). Allowing users to rearrange facets will give them a power over the classification impossible to realize with paper. Forcing a defined order on the user will be easier for the webmaster, but it prevents the user from using one of the most powerful benefits of facets.

There are four general principles to remember:

1. The user should not be able to form a query that is known to have no results. Links and form elements can be changed on the fly so that as the user selects a focus from one facets, the foci listings in other facets adjust themselves so that the only ones shown are the ones leading to possible matches. This will save the user's time. For example, in the dish detergents, Electrasol makes no liquids for hand washing. If the user chooses to see all of Electrasol's detergents, the Form and Agent facet listings should be limited. Of course, the user must be able to see what has happened and know why the choices are unavailable.

2. Users must always know where they are in the classification. Always show them the facets and foci they have chosen, and make every point in the classification a hypertext link. Every facet and focus is a pivot point around which the user can rearrange the classification. Every hyperlink is a branch point leading from one dimension of the classification into another. The users should be able to travel through the facets and foci however they please, and whenever they please, they should be able to adjust the path they took to get there, or the path they intend to take to get out. This, combined with the power to rearrange the ordering of the facets, is what makes it possible to realize the full potential of faceted classifications on the web.

3. Users must always be able to refine their query or adjust their navigation to see what is nearby in the classification. Imagine a series of dials, one for each facet, all in a row, each like the tuning dial on a radio. When the user has tuned in the classification they want, they should be able to twiddle a dial back and forth to see what is nearby. Adjusting the Brand Name dial would run through all the other brand names. Adjusting the Form dial would run through powders, gel, and so on. This interface cannot be built with simple HTML, but some sort of related browsing feature is needed. One of the purposes of a classification is to show what is like a given thing, and with facets there are many possible ways for things to be alike.

4. The URL is the notation for the classification. It should be compact but comprehensible, and editable. When a knowledgeable user examines it, he or she should be able to understand how it is built and how editing it could lead to other entities.

These questions and principles cover all the issues involved in putting facets on the web. With them, all of the features of a faceted classification can be realized on the web. They can undoubtedly be improved with testing and research, but they demonstrate that the web is a perfect home for facets.

4.3 Suggestions for web page design

If your classification is small, in number and length of facets, it is probably best to use controlled searching. Array all the facets across the page in the preferred order (see http://www.sematech.org/waferservices/find.htm (http://www.sematech.org/waferservices/find.htm) for an example of this). Using checkboxes or single select menus (or multiple if necessary), allow the user to choose from each facet or to specify a wildcard indicating they want to see everything matching. Keep the form at the top of the page when the results are displayed so the user can redefine the search. Let the users reorder the facets if you can.

If you have many facets, or they are long, it may be better to have users navigate the site facet by facet. Have users work their way into the classification step by step and narrow down the range of entities until reaching a point in the citation order where the detail is exact enough that they want to see a list of everything matching (see http://www.cmsreview.com/Directory.html (http://www.cmsreview.com/Directory.html) for an example of something like this). For this, text links or single select menus work best. Make the user choose a focus from the first facet, then show a list of all the possible choices from the second facet in the citation order, leaving out any foci that lead to dead ends. The user may go through as many facets as desired, then see a listing of all entities that match so far. The chosen facets can be listed horizontally along the top of the page, or vertically along the left. Let the users reorder the facets if you can. The end result of this will be the same as when the classification is small enough to be shown all at once, except that here they are making their choices sequentially instead of simultaneously.

These are merely two suggestions on possible ways to choose from the basic principles. Each can be adjusted, and new combinations can be made. Good design fundamentals (see Nielsen (2000)) must underly any web site, and the site must be tested on users: if they cannot make sense of it, the classification is useless. Adkisson (2003) examined 75 e-commerce web sites to see how they used faceted classification. She found that 69% of them used facets. Of those, 77% offered facet-based navigation, 6% offered facet-based searching, and 17% offered both. Sixty-seven per cent of the sites with faceted navigation did not make full use of it: they let the user choose one facet, to begin navigating the site from a particular point of reference, but then never showed facet options again. The user's first choice was the only one he or she was permitted to make. Of the other sites, 28% let the user select other facets as he or she navigates an increasingly restricted set of possibilities and results. Four per cent let the user browse with a search-type interface, making many choices at once with a search- like array of pop-up menus. These results show that e-commerce sites make fairly limited use of facets. Further study on how facets are used across the web, and what interfaces are the most usable, is deserved. Indeed, more research on all aspects of how facets relate to human-computer interaction and information visualization is warranted.

5. Conclusion

I have covered all aspects of using faceted classification systems on the web: when to use one, how to make it, how to store it on a computer, and how the web interface should work. I have given a seven-step model for the creation of a faceted classification, and five questions to ask and four principles to follow when building a facet-based web interface. Faceted systems are very powerful, and their increasing popularity on the web is no surprise. They will only become more common, so it is important to design and deploy them well. All the benefits of a faceted classification can be fully realized on the web, giving users power that they have not had with simpler web-based systems or with faceted systems on paper.

There is much about all this still to be studied. Throughout the paper, we have had to cast aside or ignore important issues from many fields of study to keep the discussion manageable: the philosophy of classification and knowledge organization; details about hierarchies, trees, and paradigms; how to test classification systems with users; whether the model described here for the creation of a faceted classification system is acceptable, and how to test that; implementation details about using XFML and relational databases to store the classification system, for example, how to handle hierarchies in a database; most aspects of how to design a good web site, from the HTML that makes it work to the principles of human-computer interaction and usability that make it good, especially including what user interfaces are best for facets; and whether the five questions and four principles about web interfaces are valid and how they can be tested. Some of subjects are amply discussed elsewhere, but the particulars of my model for classification creation need testing, and the use of facets on the web needs investigation.

References

Adkisson, Heidi P. 2003. Use of faceted classification. http://www.webdesignpractices.com/navigation/facets.htm () (accessed 2 November 2003).

Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*. New York: ACM Press.

Broughton, Vanda. 2001. 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: 67-102.

Chen, Pin-Shan. 1976. The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems* 1 (1) (March 1976): 9-36.

Denton, William. 2003. Putting facets on the web: an annotated bibliography. http://www.miskatonic.org/library/facet-biblio.html (http://www.miskatonic.org/library/facet-biblio.html).

Epicurious. n.d. Recipe search. http://eat.epicurious.com/recipes/enhanced_search/ (http://eat.epicurious.com/recipes/enhanced_search/) (accessed 15 November 2003).

Foskett, Douglas J. 2003. Facet analysis. In *Encyclopedia of Library and Information Science*, 2nd ed., ed. Miriam A. Drake (New York: Marcel Dekker), 1063-1067.

IFLA Task Force on Guidelines for OPAC Displays. 2003. *Guidelines for online public access catalogue (OPAC) displays* [draft]. http://www.ifla.org/VII/s13/guide/opacguide03.pdf (http://www.ifla.org/VII/s13/guide/opacguide03.pdf) (accessed 17 November 2003).

International Sematech. 2003. Wafer processing services: find a patterned product. http://www.sematech.org/waferservices/find.htm (http://www.sematech.org/waferservices/find.htm) (accessed 2 November 2003).

Kwasnick, Barbara H. 1999. The role of classification in knowledge representation and discovery. *Library Trends* 48 (1): 22-47.

Lillian Vernon Online. n.d. Gift finder. http://www.lillianvernon.com/cgi-bin/giftfinder.pl (http://www.lillianvernon.com/cgi-bin/giftfinder.pl) (accessed 4 November 2003).

Nielsen, Jakob. 2000. *Designing web usability: the practice of simplicity*. Indianapolis, IN: New Riders.

Petersen, Toni. 1994. *Art & architecture thesaurus. Vol. 1, Part I: Introduction, Part II: Hierarchical displays*. New York: Oxford University Press.

Ranganathan, S.R. 1962. *Elements of library classification*. New York: Asia Publishing House.

Rosenfeld, Louis, and Peter Morville. 2002. *Information architecture for the World Wide Web*. 2nd ed. Sebastopol, CA: O'Reilly.

Spiteri, Louise. 1998. A simplified model for facet analysis: Ranganathan 101. *Canadian Journal of Information and Library Science* 23 (1/2) (April-July): 1-30.

Van Dijck, Peter. 2003. XFML Core eXchangeable Faceted Metadata Language. http://www.xfml.org/spec/1.0.html (http://www.xfml.org/spec/1.0.html) (accessed 15 October 2003).

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

——— 1975. *Classification and indexing in science*. 3rd ed. London: Butterworths.

William Denton <wtd@pobox.com (mailto:wtd@pobox.com)> Toronto, Ontario

(https://github.com/wdenton) (https://orcid.org/0000-0003-3184-599X) (/feed/all.xml)

"Legendo autem et scribendo vitam procudito." — Marcus Terentius Varro (116–27 BCE)

SEARCH

Search site at DuckDuckGo

CONTENTS

- Library Science (/library/)
- Publications and Talks (/publications/)
- RARA-AVIS (/rara-avis/)
- Old Stuff (/old/)

Projects

- Conforguration (https://github.com/wdenton/conforguration): configure things (servers and dotfiles) in Org.
- GHG.EARTH (https://ghg.earth/): a sonification of climate change.
- *Listening to Art* (https://listeningtoart.org/): field recordings of visual art.
- STAPLR (https://staplr.org/): Sounds in Time Actively Performing Library Reference.
- *France Audio Montage* on *framework radio*: episodes #693 () and #691 (http://www.frameworkradio.net/2019/10/691-2019-10-20/).

See also

- Kady MacDonald Denton (https://www.kadymacdonalddenton.ca/), children's book author and illustrator
- The Arts and Letters Club of Toronto (https://www.artsandlettersclub.ca/)

CO$_2$

Atmospheric CO$_2$ at Mauna Loa (ppm) in October over the last 20 years.

Sources: data (http://www.esrl.noaa.gov/gmd/ccgg/trends/), code (https://github.com/wdenton/jekyll-co2).