Talking about Classification: Some Vocabulary

SIG/CR Workshop

ASIS&T Annual Meeting

October 27, 2017

Barbara H. Kwaśnik

Parsing a Classification Scheme

- Each classification scheme has the following:
  - A domain – the scope of the classification, or its boundaries
  - Entities – the kinds of concepts being classified
  - Classes – the grouping of similar entities into groups.
  - A structure – the relationships among the classes.

The Domain

- The domain of any classification scheme is the general area of human knowledge or experience the classification represents.
  - It can be very broad and/or general.
  - Or a scheme can have a narrow and specific domain

Understanding Entities as Classes

- The “classes” in a classification represent the various kinds of entities covered by the classification.

Instances

- An instance is a particular example of a member of a class:
  - In the domain of *students*, a person can be an instance of a *graduate student*.
  - At the same time, some can be instances of the class of people called *parents,* and further instances of the subclasses *single parents* or *step parents*.

It’s not a clear-cut definition

- If you are confused at times by this distinction of a class and an instance, don’t be surprised and don’t worry too much about it. It can be quite difficult to draw the line between when something is just an example of a class, and when something actually represents the abstract notion embodied in that class. For example Kleenex and Jell-O are both instances of their respective classes of *tissues* and *gelatin desserts*, but they are often used in the generic sense and represent the class as a whole.

What Makes a “Class”?

- So, what qualifies something to become a “class” in any given classification scheme?
- Classes are clusters of entities that have sufficient similarity to each other to be considered under the same label. As well, they are sufficiently distinct from entities in other classes.

Classification Structure

- When put together, all the classes in a classification form an overall *structure*, which we then call the classification scheme. The structure represents how the classes are related to each other.
- There are several kinds of class structures, representing different class relationships. We'll look at three here:
  - Hierarchies
  - Trees
  - Paradigms (or Matrices)

Some More Concepts

- Before we do that, however, there are a few more concepts that require explanation:
  - Scale
  - Transitivity
  - Inheritance
  - Mutual exclusivity
  - “Necessary and sufficient” criteria
  - Hospitality
  - Expressiveness

Scale

- Scale refers to the conceptual level of granularity for the entities represented in the classification. Granularity means the preciseness or detail of the representation just like in a photograph.
- Most complex phenomena can be represented at various conceptual levels depending on the context and use of the classification.
  - For example, our bodies can be understood using various classifications. A classification of body systems (digestive, respiratory, circulatory, etc.) is at a different scale than is a classification of cells, even though these two are related because they inform each other, and cells are, of course, an integral part of the processes comprising the systems.
  - It's conceptually awkward to have a classification that deals with both aspects at once because a scheme of *body systems* uses a different logic of aggregation and differentiation that does a scheme of *cell types.*

Scale in Application

- Sometimes you'll see a classification that does not represent scale consistently. Commercial stores' websites, such as Lowe's and Home Depot, use such classifications where the classes are at vastly different levels of granularity.
  - E.g., *House and Garden, Plumbing Supplies,*
- This is not a value judgment since in practice a classification of varying scale can be pragmatically the better solution.

Transitivity

- Transitivity refers to the property of a classification in which each of the subclasses inherits defining characteristics from its superclass, and thus inherits all the defining characteristics of the classes above that as well. This can be expressed as follows:
    - If A is a kind of B (a dog is a canine); and
    - B is a kind of C (a canine is a mammal); then
    - A is therefore a kind of C (a dog is a mammal).
- Put another way, information about properties of a class are passed down to the "children" in the subclasses. That is, whatever makes a mammal a mammal by definition, is passed on to canines, and then also passed on to dogs.

Inference

- Classifications that have the property of transitivity are often used to make inferences.
- An inference is a conclusion we make based on incomplete evidence. For example, if we know from the fossil remains that something was probably a *mollusk*, then we can infer a great deal about the creature because most likely it will have had similar properties to other mollusks, even though we can no longer observe the creature itself.

Inference Can Be Tricky

- Inference is tricky, however, and only as good as the validity of the conceptual structure on which it relies.
  - If someone is classed as a *liberal*, for instance, whose classification of political persuasions do you use to make inferences about that person's beliefs and potential actions? For example:
    - The far Right might class liberals differently than does the far Left, predicting that they might, for example, vote for socialist policies;
    - The far left might class liberals as "fence-sitters" and predict that they would *not* likely vote for socialist policies.

Mutual Exclusivity

- One of the requirements of some classificatory structures, such as hierarchies, is that something can belong to one and only one class. Thus a dog can't be a canine and a feline in the same classification -- one or the other.

Essence

- This principle comes from the theoretical position that a classification has as its goal the expression of the “essence” or essential qualities of the entities. Aristotle posited that once the essence was understood, the classes would divide themselves up “naturally” and the inclusion or exclusion of a member of the class would be based on rules derived from observation and reflection. In other words, the Aristotelian view is that for every entity there is one right class, based on the “essence” of that entity.

Mutual Exclusivity in Practice

- To our modern way of thinking this requirement wreaks havoc in situations where we clearly understand that something can be in more than one class, nor do we believe that any one classification will capture the multidimensional "essence" of most phenomena.
  - For example in a classification of scientists, one could be both a chemist and a biologist.
  - The same city can be a state capital and a county capital.
  - Iodine can be a poison and a necessary element for health.

Tangled Hierarchies

- One of the solutions has been to construct tangled hierarchies. These are two hierarchies that are developed separately but in which certain entities can have two or more parent classes, with the rules of transitivity and inheritance running down separate but “tangled” paths.

“Necessary and Sufficient” Criteria

- This principle refers to the criteria that are used for assigning class membership.
- A necessary criterion is one that you *must* fulfill in order to be a member of a given class. For example, when I lived at Rutgers I lived in student housing. The *necessary* criterion was that you had to be a Rutgers student or a given student’s immediate family in order to qualify.

Sufficient Criteria

- A sufficient criterion is one that pops you into a class the minute you fulfill that criterion. For example, there were several kinds of student housing at Rutgers, but I had to live in married-student housing because I had a child. I was not married, and childless people could also live in married-student housing, but I was automatically placed in that class of student-housing dwellers because having a child was a *sufficient* criterion.
- Put another way, even if you don't fulfill all the possible criteria, if you fulfill the sufficient criteria, you are placed in the class.

An Example of "Necessary and Sufficient"

- Wikipedia has a nice explanation. (Scroll down) http://en.wikipedia.org/wiki/Necessary_and_sufficient

  If a playing card has a single diamond in the middle it is a sufficient criterion for calling it an ace, but it isn't a necessary condition. This is because a card could also have a single spade in the middle and also be an ace. It's *sufficient* to have a single diamond, heart, spade, or club but it's *necessary* to have at least one of these.

Hospitality

- Hospitality refers to the ability of a classification scheme to accept new members.
  - For example, if a new orchid is discovered in Brazil, can it be easily added to the current structure of the orchid family in a plant classification?

Expressiveness

- Expressiveness is the property of a classification scheme that describes whether it has the necessary concepts and relationships to reflect the domain being classified. A scheme that is too general may not accommodate the need for more distinction, or the distinctions that are made are not made using the right criteria.

Requisite Variety

- In general, a classification should have *requisite variety*. This principle means it should match the level of expressiveness that is required for the purpose at hand.
  - For complex and finely differentiated entities you can't effectively use a classification that is too general. For example, dividing music up into *classical* and *popular* definitely doesn't cut it in my book. It's simply too few terms to adequately describe the level of distinction I use in classifying music, and besides, these two terms aren't really the right ones anyway.

Too Much Precision

- Conversely, a minutely specific classification scheme may be overkill for situations that need only a rough division into a "few piles." Aside from taking time, it can lead to a situation where you have many "classes of one."
  - For example, I own only one Polish opera, so a category for just that one item would be too particular for my needs and I put it with some other category.
- Such scattering into many overly fine categories obviates the power of classifications because everything is distributed into many precise but thinly populated and isolated categories.

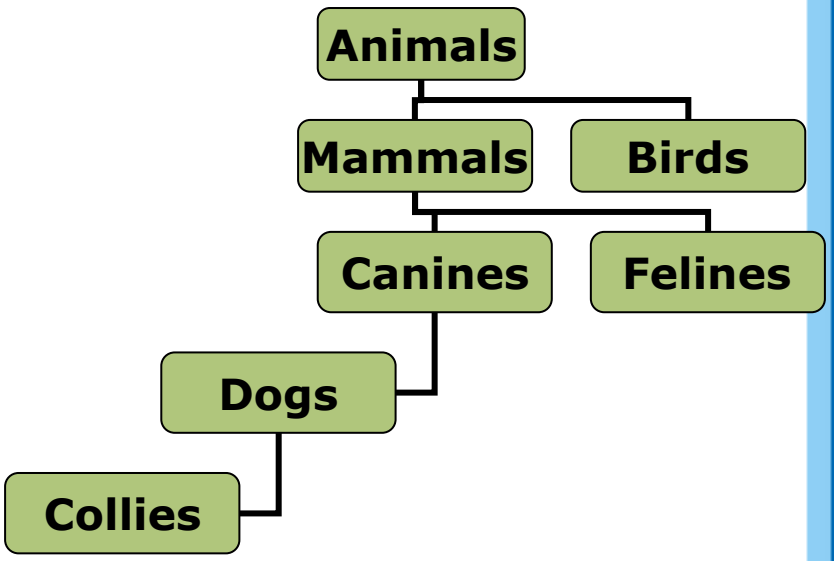
Classification Structures

- I now move to a description of classification structures.
- Each of these structures has properties that can be described in terms of the concepts just outlined in the previous slides.

Hierarchies

- In a hierarchy the classes are linked by the “is-a” or generic relationship.

- Collies are dogs; dogs are canines; canines are mammals, and mammals are animals..



Properties of Hierarchies

- There are systematic and predictable rules for determining the essential necessary and sufficient criteria.
- Each member of a class is therefore as good a representative of the class as any other.
- The rules for aggregation and discrimination are determined in advance, usually guided by a theory or model. Complete knowledge of the domain is therefore a requirement in order to determine the rules

More about Hierarchies

- The classification is a comprehensive outline of the domain – all aspects are included in a structure in which every class can be related to every other class, and ultimately to the top class.
- Classes in hierarchies conform to the principle of transitivity.
- Classes in a hierarchy are mutually exclusive.

Strengths of Hierarchies

- Hierarchies allow inference (because of transitivity);
- A hierarchy contains a great deal of information that is passed down through the hierarchy, and also shared by siblings at each level;
- Hierarchies allow an economy of notation;
- Real definitions (a type of definition that includes similarities and important differences) can be constructed from a hierarchy;
- Hierarchies allow a high-level view and holistic perspective.

Limitations of Hierarchies

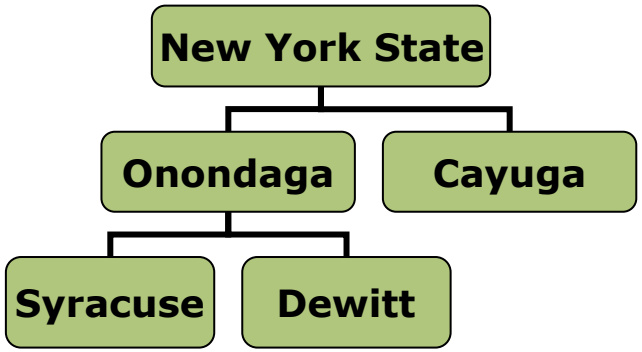
- Mutual exclusivity does not always conform to how we see the world;
- Members of classes might belong to the class more or less and be better or worse examples of that category (e.g. a penguin and a robin);
- Hierarchies require a great deal of comprehensive knowledge beforehand and this is sometimes a problem for emerging or immature domains;

More Hierarchy Limitations

- A hierarchy shows only one perspective (logic), depending on the "first cut" at the top categories and the rules for creating subclasses. This also manifests itself in biases of exclusion;
- Hierarchies can become muddled if scale is not taken into consideration; and
- If they are too rigid or narrow they can be brittle in terms of flexibility and hospitality for new knowledge.

Trees

- In a tree, the relationships among classes are not generic, but may have some other relationship, such as “answers to” or “is a part of”
- Syracuse is *part of* Onondaga County, which is part of NY State.



Several Kinds of Trees

- Trees can express many kinds of relationships:
  - Part/whole in which the top class is the most inclusive, and the subclasses each represent a part of the whole. Rather than being able to say "x is a kind of y" which you do in a hierarchy, you can say "x is a part of y."
  - Other relationships, such as process/product (product x is a result of process y); chain of command (x supervises y; y supervises z); department/function (job x is a function in department y).

Properties of Trees

- The designer of a tree classification devises the structure based on some relationship of interest among the classes.
- You need complete information about the entities but only with respect to the criterion that establishes the relationship being displayed in the tree. For example, you don't need to understand the "essence" of a carburetor if you are creating a classification of parts of a car; you just need to know where in the car a carburetor goes.
- Rules for distinction must be determined in advance.
- The citation order is important. The shape and informational power of a tree is particularly sensitive to what you decide is the first criterion of division, and then in which order the rest of the tree is divided.

Strength of Trees

- Trees can clearly display the relationship of interest (chain of command, part/whole, etc.) in a focused way;
- They can show distribution of classes and the relative numbers of members of each class; and
- They can show the distance of one class from another.

Limitations of Trees

- Trees can be rigid in that they are limited in what is included in the structure. Interesting attributes may have to be excluded for the sake of clarity.
- A tree is a purposely selective view. It can rarely include all aspects of a domain – just the aspect that it was devised to display.
- In a hierarchy there is inheritance of attributes from super-classes to subclasses. In trees there is no inheritance or transitivity so you can't use them for making inferences.
- There is less information inherent in the structure about relationships at the same levels (as there is for the sibling classes of wolves and dogs for instance).

Paradigms

- A paradigm, or matrix, represents classes based on the intersection of two or more attributes at a time.

	Parent	**Parent's Parent**	**Parent's sister**
Male	*father*	*grandfather*	*uncle*
Female	*mother*	*grandmother*	*aunt*

Properties of Paradigms

- The relationship of the cells to the axes is a hierarchical one. Thus, *a father is a parent*, and *a father is a male*.
- This relationship does not apply among the cells. A mother is not a kind of father, although she is a kind of parent.
- Paradigms require the careful choice of attributes on the axes. These are often guided by theory or models. For example in this case, Western kinship models guide making the distinction on the basis of sex and blood ties. In other cultures these might not be the same criteria that are important.
- Cells may be empty or may have more than one member.

Strength of Paradigms

- Paradigms display the intersection of more than one way of discrimination at a time (e.g., sex and blood ties, rather than just sex or just blood ties).
- They are used for studies of naming and what that may reveal about a culture or existing knowledge. This aspect can be helpful in some kinds of conceptual analysis, such as looking at diversity of terms, or lexical holes.
- Paradigms, like trees, are good for discovery because they essentially present a clear and focused picture.
- They are also good for discovery because they can show patterns of similarity and differences when two paradigms are compared (e.g., comparison of “bad deeds” in British or Napoleonic systems of law).

Limitations of Paradigms

- Creating a paradigm requires enough knowledge of the domain to avoid choosing trivial attributes for the axes;
- While paradigms do provide for the intersection of two dimensions, this may still provide a limited view. That is, a paradigm, unlike some richly developed hierarchies, rarely provides a comprehensive picture;
- The paradigm is essentially descriptive and does not provide a great deal of explanatory power as does a theoretically built hierarchy, for instance. The explanatory power comes from the person interpreting or reading the paradigm.

Facetted Classification

- The faceted approach provides a method for creating multidimensional classifications that then provide a way of conceptually organizing phenomena according to a number of aspects in the same classification scheme.
- The advantage of this approach is that each entity is seen as complex – that is, it can be viewed from more than one perspective and we are not bound to any one logic of classifying. A faceted scheme can accommodate the notion that a dog can be construed as a mammal and at the same time as a domesticated pet

Two Parts to the Faceted Approach

- The design of a facetted classification comprises two processes:
  - **Facet analysis**. The identification and definition of the facets and the *analysis* of the phenomenon into these facets. Facet analysis in this sense means the "breaking down" into the important aspects of the phenomenon.
  - **Synthesis.** Once the phenomenon has been broken down into important aspects, a faceted classification will outline a way of putting the pieces (facets) back together again into a meaningful and useful whole.

Facet Analysis Is not Componential Analysis

- Please note, facet analysis doesn't mean breaking a concept down into components, but rather into *aspects*.
  - For example, an *apple pie* broken down into components (ingredients) would yield: flour, shortening, salt, apples, spices, sugar, etc. This is *not* facet analysis.
  - Facet analysis would identify important aspects of the concept *apple pie*: nutritional aspects, role in a menu, cooking methods, things made with apples, symbol of American values, price, etc. – whatever seems important for the context in which the classification is being used.

Same Object – Different Views

- When you break something down into components, each component has its own identity – flour, sugar, apples.
- When you facet analyze something, it remains the same object, but you are given a view from different angles. Thus, the pie remains a pie – but you can represent its different aspects.

Synthesis and Citation Order

- Synthesis means arranging the facets into a meaningful and useful order.
- Once a domain has been facet analyzed, and the objects classified according to the facets, then the scheme must be organized so it can be used.
- Faceted schemes are most gracefully used for indexing artefacts and documents. The process of doing this involves creating indexing-term strings and possibly deciding on a citation order.

Strings

- A class in a faceted scheme usually consists of a *string of terms,* each term chosen from a separate facet.
  - A "15th Century, Japanese, carved, wood, screen"
- One could then think of the *family of classes* to which this key belongs, each one representing a different aspect of the key:
  - All screens
  - All wood things
  - All carved things
  - All things Japanese
  - All things made in 15th C.

Mix and Match

- Each of these aspects can be represented by a different classificatory design that best expresses it, and then they can be mixed and matched.

- People often refer to faceted schemes as being like Legos – the interchangeable individual parts coming together to form a whole in a flexible way -- but I think the analogy breaks down a little in the sense that Legos are components of the whole, not aspects, but you get the point.

Designed Dynamically to Grow

- The very impressive advantage, though, is that because the "parts" are separate, you can imagine putting them together in novel ways.
- For example, if screens began to be made out of some new material, using some new technology to form them, a faceted scheme could accommodate a new material and new fabrication process. Thus, you don't have to anticipate every eventuality. You merely set up your facets in such a way that they will be sufficient into the future.

Applications of Faceted Schemes.

- Faceted schemes are useful when the entities being classified have complex and diverse characteristics, each of which could potentially be important.
- They are also useful when the theory in a field is not mature enough to require or support a particular set of relationships.
- A faceted scheme is good when the scope and nature of the entities is in flux, expected to grow in number and kind, and so on. If a robust basic set of facets is selected then the scheme can be hospitable to many new classes and combinations of classes.

Limitations

- Faceted schemes are exceptionally flexible, expandable, and expressive, especially if you take care to choose the basic facets.
- What is often missing is some guidance as to how to make connections *among* the facets. For this reason faceted schemes can be thought of as primarily descriptive rather than theoretical.
- While they are able to provide a rich, multifaceted view, they are sometimes very difficult to present in visual form, especially if each facet is developed using a different structure or logic. It's sometimes impossible to show all the facets at once.

Conclusion

- If you're in a position of creating a classification for:
  - an emerging and dynamic set of entities,
  - where a guiding conceptual framework (such as Atomic Theory for the Periodic Table) is non-existent or immature, or people have not reached consensus on it,
  - and where the relationships among the various aspects of the entities have yet to be determined,
- Then a faceted scheme is probably a wise place to start. If nothing else it will provide a clear and systematic description, and that often provides an opportunity for further insights and theorizing.

What Is Warrant?
In Beghtol's Own Words

- "…the warrant of a classification system can be thought of as the authority a classificationist invokes first to justify and subsequently to verify decisions about what class/concepts should appear in the schedules, what units class/concepts are divided into, how far subdivision should proceed, how much and where synthesis is available, whether citation orders are static or variable and similar question. Warrant covers conscious or unconscious assumptions and decisions about what kinds and what units of analysis are appropriate to embody….The semantic warrant of a system thus provides the principal authorization for supposing that some class or concept or notational device will be helpful and meaningful to classifiers and ultimately to the users of documents (p. 110-11).

Assessing the Semantic Aspects of a Classification

- To assess warrant, one might ask:
  - *Why* are these concepts here?
  - *Who says* this is the way to divide them?
  - *Who thought this classification up* and *under what circumstances?*
  - What is the *conceptual framework or theory* that determines the choice of concepts, the rules of division and the relationships?
  - What are the *assumptions* both hidden and explicit?
  - Is this classification *extensible*? That is, given that this classification was devised to serve X, can it be repurposed and now do a good job of serving Y?

Aspects of Warrant

- So how do you use the these notions of warrant to analyze the quality and effectiveness of a classification with respect to its intended meaning and utility *in the world*?
- Beghtol suggests looking at them through several different aspects of warrant. These are:
  - Literary Warrant
  - Educational/Philosophical Warrant
  - Scientific Warrant and
  - Cultural Warrant

Literary Warrant

- A classification built using ***literary warrant*** is based on the collection, that is, the "literature" in the sense of a body of works and the concepts embodied in it.
  - Following literary warrant, we first look to see what literature exists on a given topic or in the collection, and then decide what entities should be included as subjects in the classification.
  - With respect to assessment, then, we ask "Can this classification accommodate and reflect the body of works in our collection on this topic?"

Issues with Literary Warrant

- One of the problems is that a scheme that starts out following the literary warrant of the time and context in which it was created, may not adequately change as the "literature" and the context changes.
  - For example, the Library of Congress Classification was designed to organize the collections of works that served the U.S. Congress. Over time it has expanded to serve as a *de facto* national and even international system. Do the same criteria of literary warrant still apply?

Scientific/Philosophical Warrant

- A classification following scientific/philosophical warrant is consistent with scientific and scholarly consensus and relies on the authority of scholarship and research.
  - In academic libraries, literary and scientific warrant may coincide. That is, the collections are compiled in order to reflect and support scientific warrant and thus the classification for one will map onto the other.

Whose Consensus

- Another problem is that scientific warrant often assumes that we can create a permanent classification built on enduring "meanings."
  - Consider, though, the different discourse of Western and complementary medicines;
  - Or, think of the various theoretical debates in biology on how to name and classify the living world.

- Thus, we can well ask, whose consensus, or whose science or philosophy and at what time is that consensus captured?

Educational Warrant

- A classification following educational warrant is responsive to the pedagogical needs or the specialized needs of an institution.
- This notion can also be extended to mission-specific schemes that organize knowledge in a way most conducive to the goals of the organization.

In Summary

- Assessing a classification can be difficult and subjective to some extent. After all, a classification is a way of communicating ideas. Here are some considerations:
  - How well does the classification live up to its warrant?
  - Is the classification hospitable, expressive and have requisite variety?
  - Is it memorable so it can be easily used?
  - If its aim is to represent a domain, is it heuristic and fruitful in terms of representing knowledge and generating new knowledge?
  - Is it well designed, parsimonious, elegant, consistent, comprehensive, and coherent?