Rigorous facet analysis as the basis for constructing all types of knowledge organization scheme

Leonard Will

"Concepts are our building bricks"

Steps in building a knowledge organization scheme

     1. Concepts

     2. Facets

     3. Thesaurus

     4. Classification scheme

Concepts

1. Gather concepts ("units of thought")

2. Define the scope and meaning of each concept
   What is its parent and what distinguishes it from its parent?
   Write scope notes where necessary

3. Split or retain compound concepts
   coal mining = coal + mining

4. Choose a term to label each concept
   One preferred term, any number of non-preferred terms

5. Sort concepts into facets based on "fundamental categories"
   "is-a" relationship requires both to be In the same category

6. Build concepts into hierarchies within each facet

iterative process

inductive and deductive

top-down and bottom-up

Fundamental categories – "facets"

Distinct and mutually exclusive groups, e.g.

- Things - *ships, shoes, cabbages, power stations, heating systems*
- Activities, processes, disciplines - *cutting, thinking, dancing, rusting, physics*
- Abstract concepts - *love, war, costs, benefits, access, rights*
- Places - *continents, mountains, countries, political groupings, rivers, Europe*
- Times - *21st century, pre-war, 2012-12-25, mediaeval*
- Materials - *sealing wax, water, dirt, adhesives, polymers*
- Properties - *size, colour, intelligence, plasticity*
- People and organizations - *kings, children, hospitals, ISKO*
- Events - *battles, conferences, wars, investigations, festivals*

The choice of "fundamental" categories is not absolutely objective, but should be consistent within a knowledge organization scheme. Hierarchies within a facet must be genus/species = "is-a" relationship.

Hierarchical structure

- Genus/species relationship ("is-a")
- As used between classes in ontologies
- Properties which define a concept (or class) are inherited
  - ♦ If it is a defining property of a bird that it lays eggs, then if a cuckoo is a bird it must lay eggs.
  - ♦ Laying eggs is a necessary but not a sufficient condition for being a bird.
  - ♦ Being able to fly is not a defining property of a bird, so an ostrich can still be a bird even though it cannot fly.

Facets – different meanings to different people

- Fundamental facets (as in ISO 25964-1 and AAT)
  - *based on fundamental categories*

- Not "facets of a document"
  - *metadata elements or fields of catalogue record*

- Not "facets of a subject"
  - *related to something chosen as "personality"*

- Not restricted by discipline
  - *discipline can be one facet among others*

Non-disciplinary facets – "phenomena"

- Traditionally facet analysis has been applied within disciplines
- General schemes then are a set of disciplinary schemes
- Some facets have been recognized as interdisciplinary
  – *UDC now has common tables for properties, materials, relations, processes, activities and persons*
- Bliss classification has "phenomena" classes
- *Integrative Levels Classification (ILC)* has a single list of concepts, including disciplines
- B. C. Vickery suggested two sequences: "phenomena" and "activities"

Three possible facets – subassemblies of a KOS

people
- *<people by age>*
- babies
- children
- adults
- old people
- *<people by occupation>*
- cooks
- farmers
- information scientists
- librarians
- students

activities
- artistic activities
  - drawing
  - painting
- cooking
  - *<cooking by process>*
  - boiling
  - braising
  - frying
  - steaming
- information processing
  - cataloguing
  - indexing
- manufacturing activities
  - boatbuilding
  - knitting
- recreational activities
  - boating
  - swimming

materials
- *<materials by use>*
- building materials
- cosmetics
- foods
  - *<foods by origin>*
  - animal foods
  - vegetable foods
    - *<vegetable foods by part of plant>*
    - grains, seeds
    - leaf vegetables
      - cabbages
      - lettuces
    - root vegetables
      - carrots
      - parsnips
      - turnips
    - *<vegetable foods by colour>*
    - green vegetables
      - cabbages
      - lettuces
    - orange vegetables
      - carrots
      - pumpkins

Node labels showing characteristics of division

8

To complete a thesaurus

- Record synonyms as non-preferred terms for each concept
  – *chefs USE cooks*

- Record compound terms that have been split
  – *baby foods USE babies + foods*

- Record associative relationships between concepts
  – *cooking RT foods*

- Possibly create subsets as "concept groups" of related terms and sub-trees
  – *microthesauri, themes*

- Other presentation and management issues
  – *displays, metadata*

Post-coordinate indexing and searching

● Concepts are linked separately to documents being indexed

● Concepts are combined (coordinated) at the time of search, not at the time of indexing

● Searchers can specify concepts in any order

● Searchers cannot usually specify relationships between concepts

From a thesaurus to a classification

- Add a mechanism for combining concepts to express compound subjects

- Choose a citation order
  - *to ensure consistency of expressing compounds*
  - *to provide a helpful sequence of classified resources*

- Ranganathan's sequence
  - *PMEST*

- CRG sequence
  - *Thing – kind – part – property – material – process – operation – patient – product – by-product – agent – space – time*

- These sequences depend on roles as well as on facets

Combining concepts

- Add a mechanism for combining concepts to express compound subjects

- In a hierarchical display, introduce "facet labels" to show when a new facet is being introduced and if necessary the role that that facet plays in relation to the preceding concept

(people)	*(activities)*	*(materials)*
people	activities	materials
<people by age>	cooking	*<materials by use>*
babies	*<cooking by process>*	building materials
children	boiling	cosmetics
adults	braising	foods
old people	roasting	*<foods by origin>*
<people by occupation>	steaming	animal foods
cooks	artistic activities	vegetable foods
farmers	drawing	*<vegetable foods by part of plant>*
information scientists	painting	
librarians	information processing	grains, seeds
students	cataloguing	leaf vegetables
	(agents : people)	cabbages
	librarians	*(activities)*
	indexing	cooking
	manufacturing activities	*<cooking by process>*
	knitting	steaming
	recreational activities	*(agents : people)*
	boating	people
		<people by occupation>
		cooks

Facet labels showing facets and roles

Concepts in colour "brought down" from earlier facets

13

Distinguish between node labels and facet labels

- **Node labels** introduce arrays, showing the criterion by which the following concepts have been selected or arranged
  – the label <*cooking by process*> introduces an array listing kinds of *cooking*, such as *steaming* and *roasting*. These three concepts are necessarily in the same facet and
  are hierarchically related.

- **Facet labels** introduce concepts from a new facet, to be combined with the previous one.
  – *cooking* and *cabbages* are **not hierarchically related.**
  The concepts before and after the facet label are **combined** to express the compound concept *cooking cabbages.*

Notations and symbols

- When a compound subject is expressed verbally, the components may be linked by words or symbols to indicate their relationships.
  – *cabbages : cooking*
  – *cabbages : cooking [by] children*

- British Technology Index used punctuation symbols and words
  LASERS; Carbon dioxide—Nitrous oxide,, Chemically driven : Nitrous oxide : Reaction with sodium vapour

- PRECIS used quite complex codes which controlled some aspects of the display and typography of verbal strings.

- When a notation is used, this information is given by symbolic "facet indicators" as part of the notation
  – UDC punctuation symbols, DDC "0"

15

Integrative levels classification – facet indicators

0 under aspect

1 at time

2 in place

3 through process

4 made of element

5 with organ

6 from origin

7 to destination

8 like pattern

9 of kind

The specific interpretation of these relationships may vary according to the nature of the concepts which they link.

Hierarchies

● **Polyhierarchy,** If a concept occurs in more than one array, choose one location as the primary one, to use in compounds.
  – *normally the "place of unique definition"*
  – *a carrot is inherently a plant*
  – *a carrot is incidentally an orange vegetable*

Hierarchies

- **Instances.** "Classes of one".
  – *proper names: London, Loch Ness, Microsoft Windows, Nelson Mandela, Titanic, ISKO.*

- Can have broader concepts, denoting the class(es) to which they belong

- Cannot have lower-level species

Hierarchies

● Narrower concepts of synthesised compounds such as *joints : inflammation* may be enumerated under the most appropriate element of the compound.

diseases
  inflammation
    <inflammation by body part>
     arthritis
      osteo-arthritis
      rheumatoid arthritis
     appendicitis
     cystitis
     dermatitis

body parts
  bones
  joints
    (diseases)
     inflammation
      arthritis
       osteo-arthritis
       rheumatoid arthritis
  skin
    (diseases)
     inflammation
      dermatitis

Hierarchical whole/part relationship

- **Avoid.** *Wheels* are not necessarily parts of *bicycles.*

- Even if specified as *bicycle wheels, they* do **not** inherit the attributes of *bicycles.*

- Use *bicycle components*
  NT *bicycle wheels*

- Should the whole/part relationship be considered as hierarchical or associative?

- In a pre-coordinate scheme we can link two concepts with the whole/part relationship by using a suitable role indicator. This does not indicate a hierarchical relationship, e.g.

  bicycles
  (parts)
  chains
  pedals
  wheels

Whole/part relationship

- ISO25964-1 lists four special cases where the partitive relationship is acceptable as hierarchical

(1) **Body parts**. Why is this different from *bicycle parts?*
– AAT puts them under *animal or human components.*

(2) **Geographical locations.** These are all instances, identified by proper names, rather than classes. Attributes are not inherited. Misleading "non-transitive" relationships if they are mixed with generic relationships. Keep them separate.

(3) **Disciplines or fields of discourse.** Consider these as having generic relationships rather than partitive ones. *Physics* is a kind of *science, painting* is a kind of *art,* not a part.

(4) **Hierarchical social structures.** Not always valid. *Regiments* are not always parts of *battalions.* Same as *wheels* and *bicycles.* Specific named units are instances, like geographical locations.

Conclusion

Build knowledge organization schemes step by step

      1. Concepts

      2. Facets

      3. Thesaurus

      4. Classification

Several iterations may be needed.

But it is not a mechanical process – skill and judgement are essential.