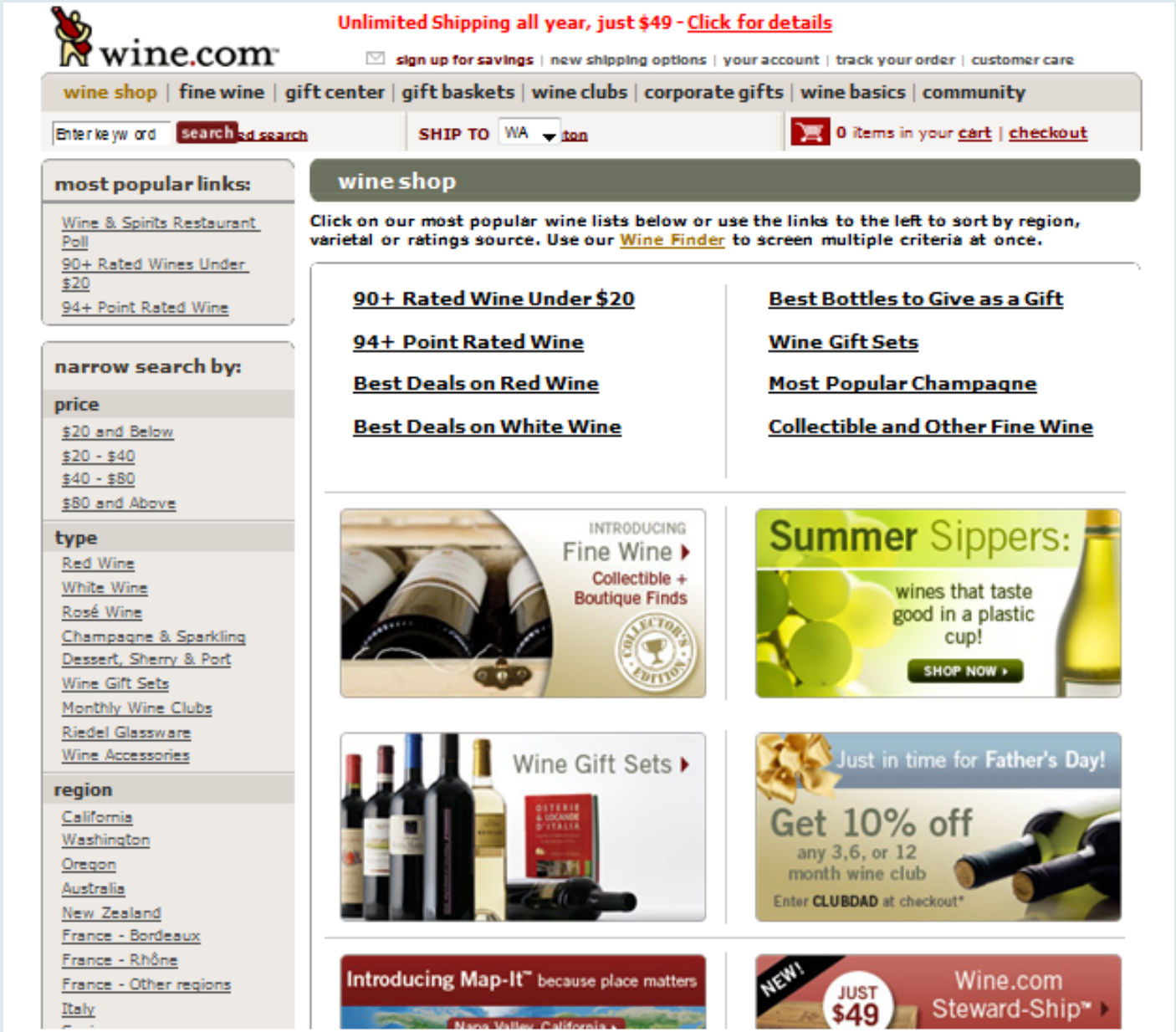
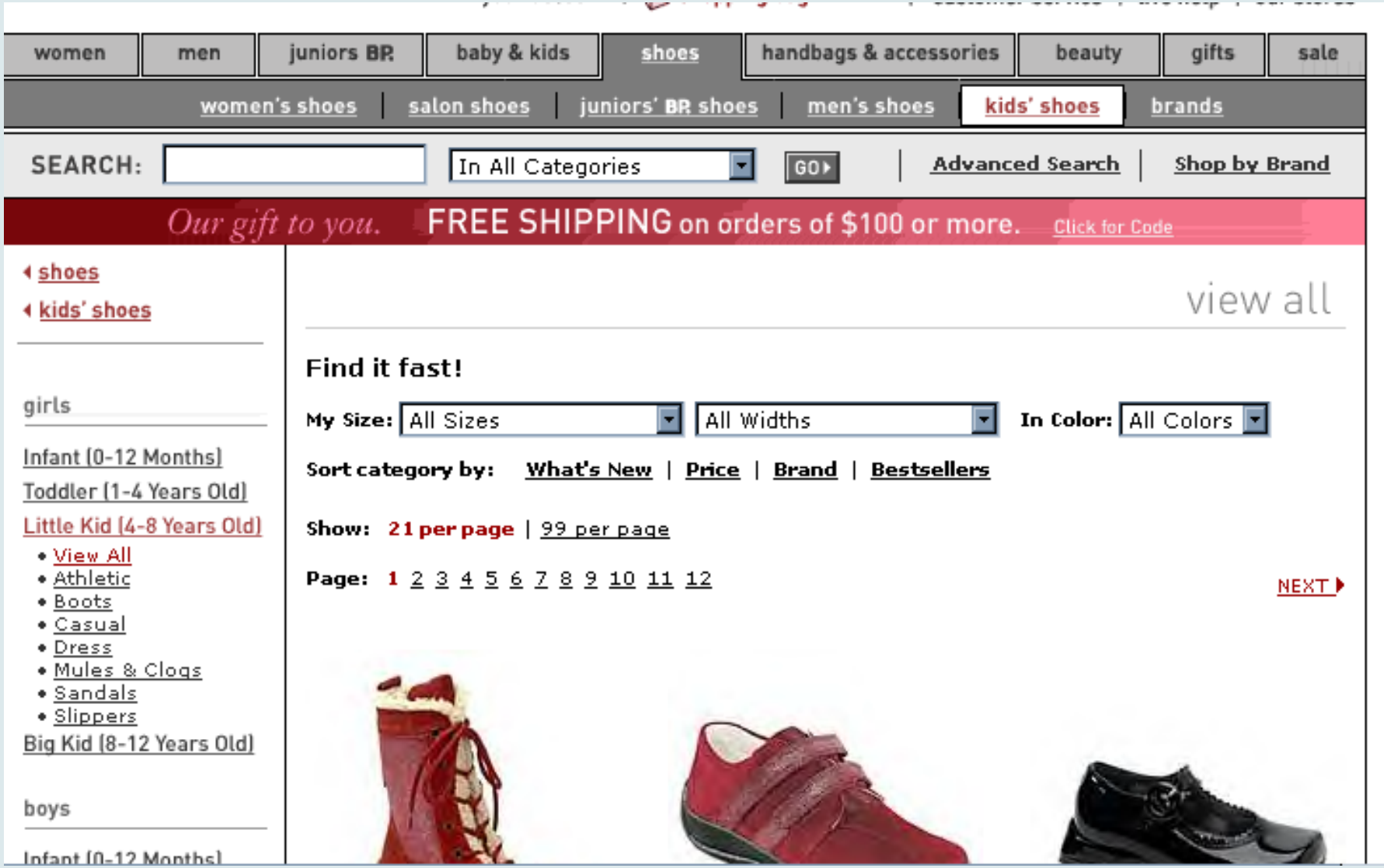
Facet analysis as the theoretical basis of vocabulary tool construction

FUNDAMENTAL PURPOSES IN KNOWLEDGE ORGANIZATION

Perceptions of the faceted approach:

- faceted search and retrieval has recently become very fashionable
- faceted search commonly manifests itself in a graphical interface
- terms from different ‘facets’ may be combined using windows or drop down menus
- it offers a structured approach to search that is logical and intuitive for the end-user
- it’s particularly appropriate where the facets consist of various attributes of an entity


wine.com
Unlimited Shipping all year, just $49 - Click for details
sign up for savings | new shipping options | your account | track your order | customer care
wine shop | fine wine | gift center | gift baskets | wine clubs | corporate gifts | wine basics | community
Enter keyword search
SHIP TO WA
0 items in your cart | checkout
most popular links:
Wine & Spirits Restaurant Poll
90+ Rated Wines Under $20
94+ Point Rated Wine
narrow search by:
price
$20 and Below
$20 - $40
$40 - $80
$80 and Above
type
Red Wine
White Wine
Rosé Wine
Champagne & Sparkling
Dessert, Sherry & Port
Wine Gift Sets
Monthly Wine Clubs
Riedel Glassware
Wine Accessories
region
California
Washington
Oregon
Australia
New Zealand
France - Bordeaux
France - Rhône
France - Other regions
Italy
wine shop
Click on our most popular wine lists below or use the links to the left to sort by region, varietal or ratings source. Use our Wine Finder to screen multiple criteria at once.
90+ Rated Wine Under $20
94+ Point Rated Wine
Best Deals on Red Wine
Best Deals on White Wine
Best Bottles to Give as a Gift
Wine Gift Sets
Most Popular Champagne
Collectible and Other Fine Wine
INTRODUCING
Fine Wine
Collectible + Boutique Finds
Summer Sippers:
wines that taste good in a plastic cup!
SHOP NOW
Wine Gift Sets
Just in time for Father's Day!
Get 10% off
any 3,6, or 12 month wine club
Enter CLUBDAD at checkout*
Introducing Map-It™ because place matters
NEW!
JUST $49
Wine.com Steward-Ship™


women
men
juniors BR
baby & kids
shoes
handbags & accessories
beauty
gifts
sale
women's shoes
salon shoes
juniors' BR shoes
men's shoes
kids' shoes
brands
SEARCH:
In All Categories
GO
Advanced Search
Shop by Brand
Our gift to you.
FREE SHIPPING on orders of $100 or more.
Click for Code
shoes
kids' shoes
girls
Infant (0-12 Months)
Toddler (1-4 Years Old)
Little Kid (4-8 Years Old)
View All
Athletic
Boots
Casual
Dress
Mules & Clogs
Sandals
Slippers
Big Kid (8-12 Years Old)
boys
Infant (0-12 Months)
view all
Find it fast!
My Size:
All Sizes
All Widths
In Color:
All Colors
Sort category by:
What's New
Price
Brand
Bestsellers
Show:
21 per page
99 per page
Page:
1 2 3 4 5 6 7 8 9 10 11 12
NEXT

View-based searching:

- this approach originates in the work of Steven Pollitt, who introduced the idea of view-based searching
- view-based searching supports the combination of terms from different fields in a database
- originally developed in the medical field, view-based searching is more sophisticated than most e-commerce sites
- its purpose is query formulation and modification through successive filtering
- it uses a greater range of types, or categories, of concept than just attributes

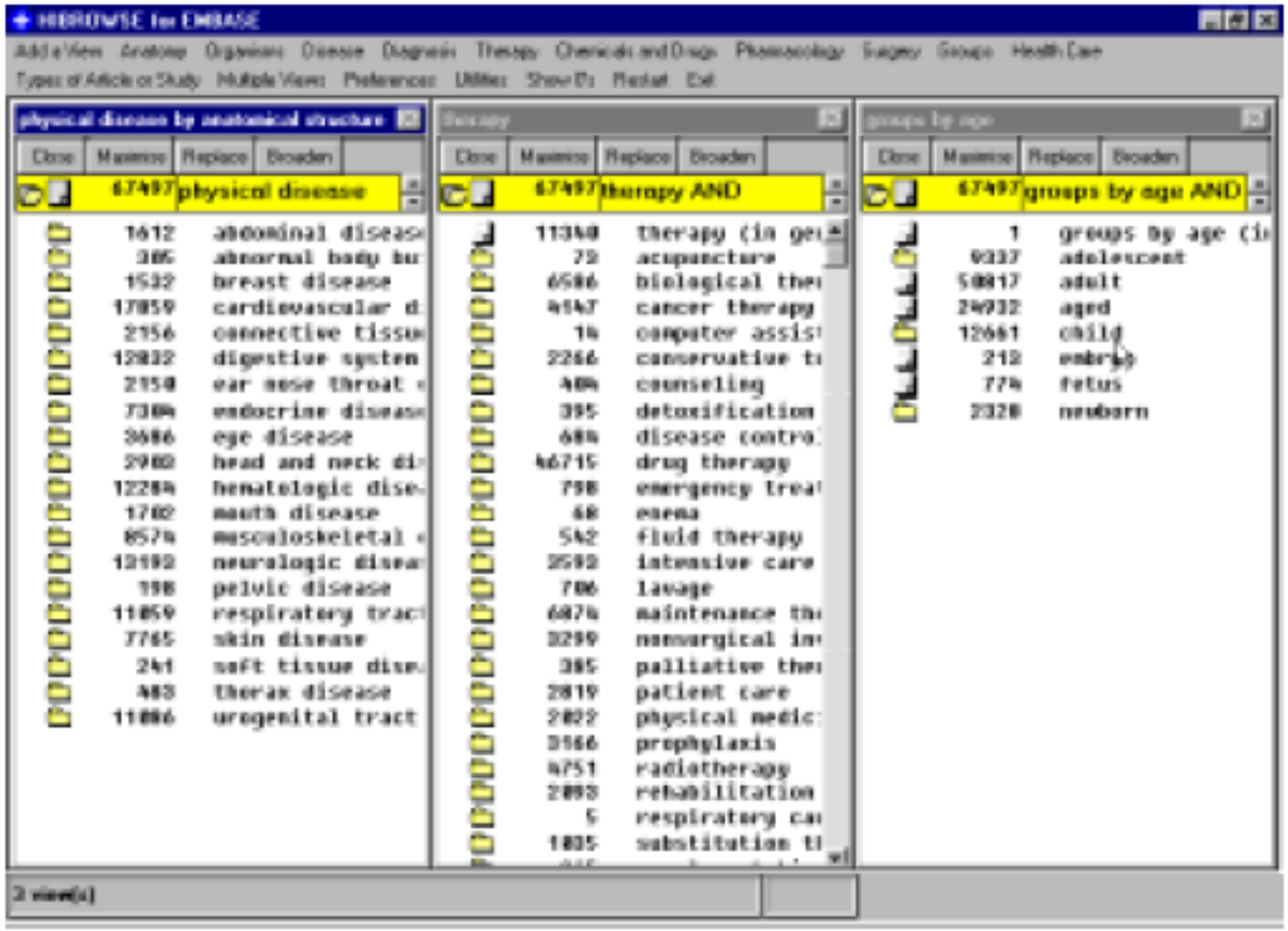


Figure 1. Views of Disease by Site, Therapy and Groups by Age

The key role of classification and indexing in view based searching.
http://www.ifla.org.sg/IV/ifla63/63polst.pdf

Databases and faceted structures:

- view-based searching arises from the idea that the faceted approach is particularly appropriate for information held in databases
- this was first discussed in the 1980s
- it acknowledges the structured nature of the data and metadata
- it is much closer to the original concept of faceted systems than the e-commerce applications

The origins of faceted classification:

- faceted classification was originally devised as a means of creating tools for indexing and for physical organization
- it centrally addresses the problem of complexity in subject content and provides rules for dealing with it
- hence it is concerned with the relationships between concepts in a subject domain
- it attempts to discern, and to impose, regularity of structure on the domain as represented by its constituent concepts
- in a managed environment this regularity and predictability improves retrieval performance

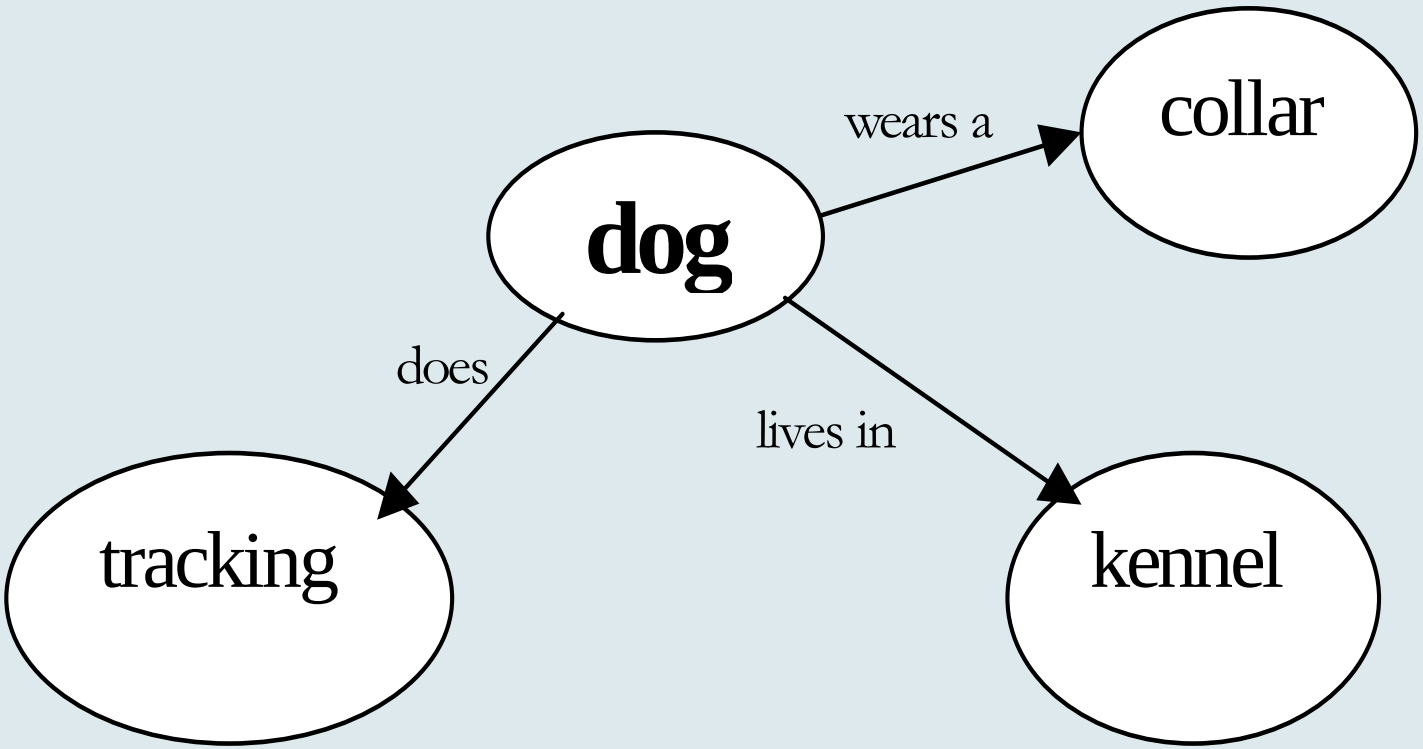
Why is faceted classification important?

- structured vocabularies allow us to create structure in the metadata that corresponds to structure in the data i.e. structure in the subject domain
- structure here is taken to mean the relationships both between concepts and in complex combinations of concepts
- structured vocabularies provide a model of the subject domain that can simply act as a surrogate for browsing and navigation or can, in combination with the system syntax, support query formulation and modification, and a degree of automatic classification
- standards for document or object representation usually provide excellent structure for *description* of the item concerned
- it's much less likely that the provision for *subject* in the same standards will be equally well structured

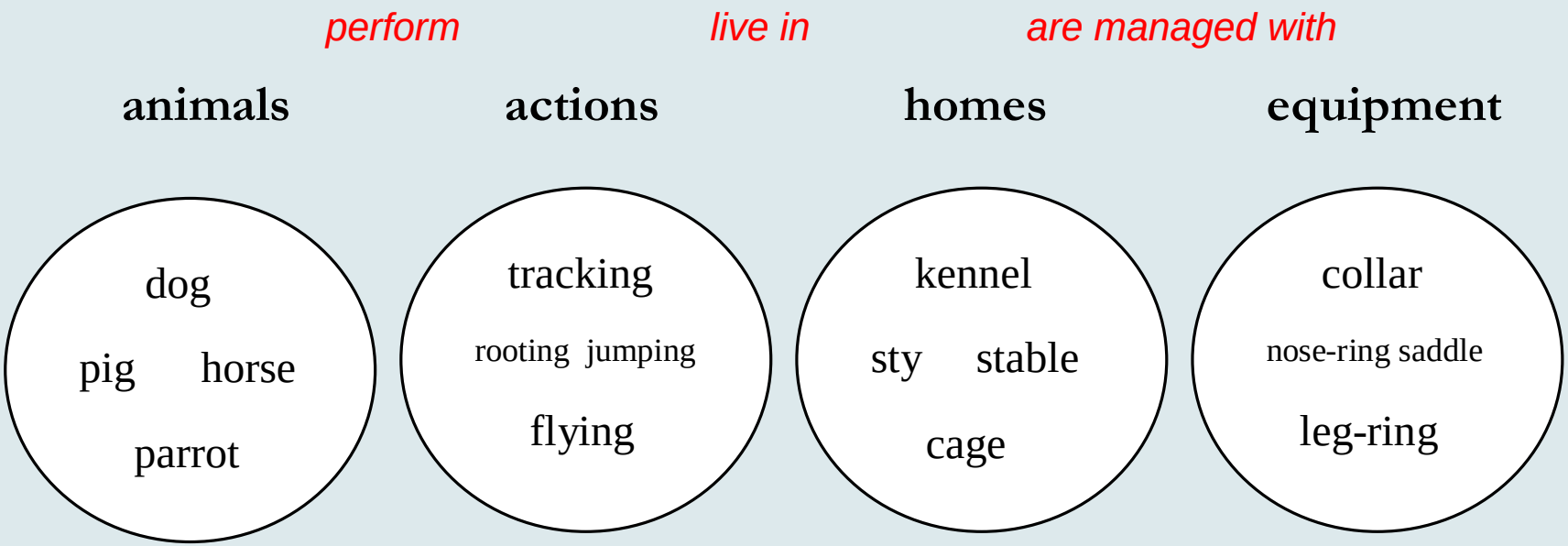
Structure of a faceted classification:

- the broad structure of a faceted system is achieved by analysing concepts, and assigning them to a series of fundamental categories
- Ranganathan identified only five of these, but the UK Classification Research Group developed his methodology , and worked with a set of thirteen categories
- using categories allows general rules to be proposed about the relationships in the domain (as opposed to the relationships between particular pairs of concepts).

‘Ontological’ relationships:



Categorical relationships:

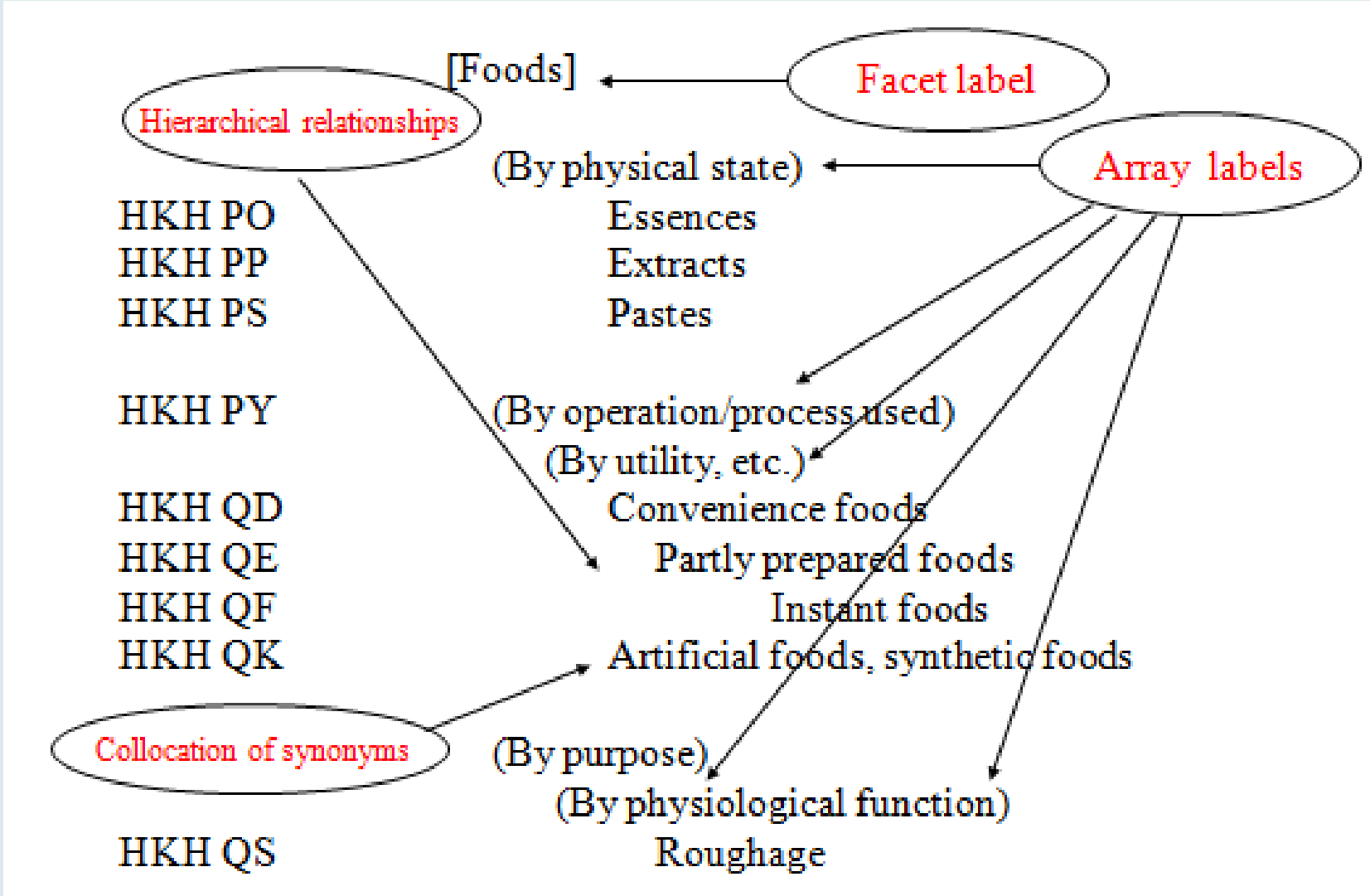


Concepts grouped in categories

Classification based on relationships:

- the rules of the classification govern the way in which the categories are ordered, and how combination is controlled between facets, including the positioning of compounds
- this system syntax is also a part of faceted classification theory alongside the structural elements
- a very sophisticated structure can be built up when the 'core' classification structure is populated by examples of compound topics
- in BC2, the inclusion of a great number of compound classes enriches the vocabulary, particularly where unique *terms* are generated by combination of *concepts*

‘Core’ classification showing structural elements:



Classification populated by examples of compound topics:

HUQ W	Thymus gland
	(Physiology)
HUQ WH	(Pathology)
	(Hyperplasia)
HUQ WMD V	Lymphatism, status lymphaticus
	(Causal agents)
	(Symptoms)
	(Treatment)
	(Neoplasms)
HUQ WME	Thymomas
	(Products)
HUQ X	Thymus hormones
	(Molecular structure)
HUQ XS	Thymopoietins

[Compund terms pre-synthesized and added to published schedule]
[Examples of potential synthesized compounds]

The significance of relationships:

- the centrality and significance of these relationships is demonstrated by the fact that they hold good in different types of controlled vocabulary
- current work on the Bliss Bibliographic Classification 2nd edition (BC2) shows how a thesaurus (as well as an alphabetical subject index) can be derived from the input data for the systematic schedules
- this is enabled by the fact that the relationships of sequence, hierarchy, and facet status can be encoded
- this encoded data can then be automatically manipulated by the classification software to generate the different formats.

```
CGHGKYG         07Cumulated bond compounds
CGHGKYH         07Isolated bond compounds
CGHGKYL         07Covalent bonds special to a given context ]I
                     *   For example, CJT_GKY_M Metal bonds.
CGHGL           06Ionic bond compounds, electrovalent compounds
                     *RT Coordination compounds = CIL
CGHGLN          07Resonance bonded compounds, resonance hybrids
                     *   For aromatic compounds, see CRA.
CGHGLT          06Hydrogen bond compounds
CGHGLU          06Bonds special to a given class ]I
                     *   The classmark -LU is reserved for qualifying
                         particular classes; eg, CIA_GLU Binary acids
CGHGO           05(Kinds by molecular structure)
                     *   Add to CGH_G letters O/W following CA; eg,
CGHGOQ          06Chain structures
CGHGOS          07Linear chains, straight chains, open chains
                     *RT  For aliphatics (organic straight chains): = CP
CGHGOU          07Branched chains
CGHGOV          08Crosslinked chains
                     *RT polymers = CGV
CGHGOW          06Ring compounds, closed chains
                     *RT inorganic ring structures = CIP_GOW
                     *RT organic rings = CQ
                     *RT chelate rings = CIM_GOW
CGHGOWJ         07Separately linked ring compounds
CGHGOWL         07Fused rings, condensed nuclei (rings)
CGHGPS          06Symmetrical compounds
CGHGPU          06Asymmetrical compounds
@               05(Kinds by variations in structure) ]IT
CGHGQ           06Polymorphic substances, polymorphs, allotropes
                     **  classes CGH_GQH_P/CGH_GX are all added (moved
                         from original CAQ/CAW and CGQ/CGV).
                     *   The term Allotropes is often used when
                         allotropes of elements is meant; in such
                         cases, use CGE_CQ.
CGHGQH          07Polymorphism, allotropy, allotriomorphism
                     *   Assumption by a substance of two or more
                         different structures which are most frequently
                         stable in different temperature ranges.
                     *   Note at CGC_Q on loose use of terms applies
                         also to allotropy.
CGHGQHP         08Stable allotropes
CGHGQHQ         08Metastable allotropes
CGHGQHR         08Dynamic allotropes
CGHGQJ          08Dimorphs
```

BC2 encoding:

- the BC2 code is relatively crude ☺
- it identifies:
  - the position in the linear sequence
  - the position in the hierarchy
  - non-classes (e.g. node labels, facet names, etc. other aspects to do with index entry
- from this the software can compute the hierarchical layout for classification purposes, as well as appropriate fonts and headings
- the programs 'understand' the rules for structuring and ordering concepts
- therefore the software can also infer BT/NT and sibling relationships, as well as equivalence relationships, for the purposes of the thesaurus

W3E	.	Art product, art object, works of art
W3H	..	(Properties) Content, meaning.. Emotions.. Comedy..
W3I L	..	(Elements) Composition & design.. Form.. Colour..
	..	*Kinds of arts defined by earlier facets*
W3R	..	*Kinds by movements, schools, styles*
		* For original styles see place/time of origin (eg, WK9 FLR I Impressionist painting).
W3S	...	Universal styles, A/Z.. Classicism.. Realism..
W3T	...	Historical styles, influences, revivals
W6A	..	*Kinds of arts by special categories of persons*
F	...	Folk arts.. Ethnic.. Religious..
W77	..	(Kinds of arts by period) Ancient.. Modern..
W9	..	(Kinds of arts by place) Western.. Non-Western..
	..	*Kinds of art by special categories of artists*
WBA M	...	Children's arts.. Disadvantaged persons' arts..
	..	(By representation) Abstract.. Figurative..
	..	*Artforms by medium*
WC	...	Visual arts
WC3 7K	....	(Operations) Manufacturing.. Aesthetic..
BV	....	(Materials) Substrates.. Marking materials..
WC9	....	*Kinds of visual arts by place*
		* Notation modified to give shorter classmarks for large literature on art history by place/tir (as W9C/Z), WDM/Z is Non-Western (as WAM/Z).
WEC M	....	(By special categories of persons) Children's arts
WEE	....	(By representation) Abstract.. Figurative (subject)..
WF	....	Applied arts, design arts
WFG	.....	Graphic design
		* For graphic fine arts, see WIT Y.
WFI	.....	Calligraphy.. Illuminated manuscripts..
WFK	.....	Print design.. Illustrations.. Books.. Press..
WFQ	.....	Commercial art.. Advertising.. Maps.. Signs..
WGB	.....	Decorative arts
WGC	.....	Arts & crafts, craftwork, handicrafts

Thesaurus

Africa south of the Sahara *(contd.)*
NT Tribes in African art WDV E9
NT West Africa WDV FB
Africa visual arts WDV
BT Visual arts WC
NT Africa south of the Sahara WDV E
NT Ancient period WDV 77
NT Ethnic groups WDV 6BE
NT Linguistic groups WDV 6BC
NT Medieval & modern periods WDV 8AY
NT North of the Sahara WDV 9X
NT Religious groups WDV 6BJ
African WC5 V
BT Non-Western styles revived WC5 M
African WFA V
BT Design product WF3 E
African art styles (influences, revivals) WCC 5V
BT Non-Western styles revivals WCC 5M
African arts in the Western tradition W9V
BT Western arts W9C
African style influences WCD 5V
BT Historical movements WCD 3V
African styles (influences) W5V
BT Non-Western styles revived W5M
African Western art WCV Q
BT Western visual art WCC
After-image WC3 JML
BT Illusion WC3 JMI
Aged persons WEH S
UF Old age
BT Human person WEH B
Ageing WC3 BJ
USE Degeneration (of art objects)

Aircraft WEV FI
UF Spacecraft
BT Transport vehicles WEV AI
Aircraft (toys) WHJ SR
BT Transport vehicle models WHJ SB
Aisles WHT KF
USE Corridors
Aisles WHU XGK F
USE Naves
Ajanta caves WDQ AMJ 72A J
BT Sites WDQ AMJ 72
Akamba WDV UKQ
UF Kamba
BT Kenya WDV UK
Akbar period WDQ HL
BT 16th century WDQ H
Akbar period WKA QHL
BT Indian paintings WKA Q
Akkad WC7 CK
UF Akkadian art
BT Mesopotamia WC7 CD
Akkadian art WC7 CK
USE Akkad
Aknetaten WC7 NJQ
UF Amenhotep IV
UF Ikhnaten
BT Eighteenth dynasty WC7 NJN
Alabaster WLK P
BT Stone sculpture WLK
Alabaster (onyx marble) WLK O
USE Onyx marble
Aleatory art W3P JMV
BT The Arts W

Enhancing encoding:

- the FATKS project (www.ucl.ac.uk/fatks) examined how we might represent a faceted classification in a database format
- the classification data was encoded with:
  - a notation representing linear order
  - indent codes indicating hierarchical position
  - a code for the subject area
  - a code representing the facet (*entity*, *operation*, *process*, etc.)
- this encoding supports some degree of automatic classmark synthesis, since:
  - a measure of vocabulary control can be exercised
  - from the encoding the order of combination can be inferred by the software
- this type of encoding when added to the BC2 mark-up could facilitate the identification of inter-facet combinations and the existence of associative relationships (RT) as well as BT/NT

Building compound and complex numbers

a) Combining concepts within the same facet

Notation:	Description:	Facet:
J15	Marriage and family	J Religious activities. Practice
J1477	Abstinence. Celibacy	J Religious activities. Practice
J15J1477	Abstinence in marriage	

b) Combining concepts between facets

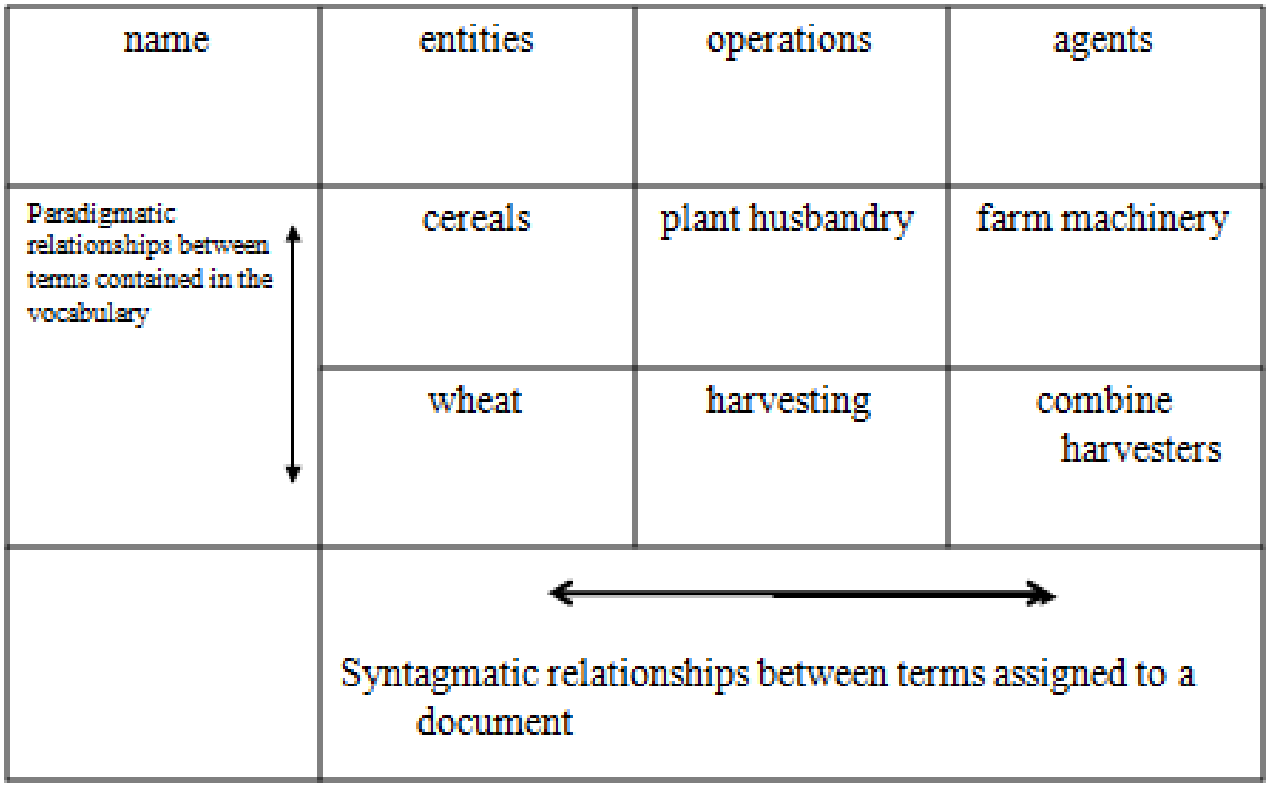
Notation:	Description	Facet:
5904	Buddhism	590 Religions and Faiths
E31	Originator, founder	E Agents. Subfacet: Persons as agents
A443	Physical form, appearance	A: Theory and Philosophy. Subfacet of God.
5904E31A443	Trikaya. Doctrine of the three bodies in Buddhism	

Notation:	Description:	Facet:
59033	Hinduism	590 Religion and faiths
5904	Buddhism	590 Religions and Faiths
5907	Christianity	590 Religion and faiths
J14247	Abstinence. Fasting. Prohibition	J Religious activities. Practice
59033J14247	Upavasa. Fasting in Hinduism	
5904J14247	Abstinence. Fasting in Buddhism	
5907J14247	Fasting in Christianity	

Relationships in faceted schemes:

- relationships between concepts *within* a category (intra-facet relationships) are by definition semantic (hierarchical, paradigmatic) relationships
- relationships between concepts in different categories (inter-facet relationships) are by definition syntactic (syntagmatic) relationships
- relationships of the latter type are, by definition *associative* relationships or RTs
- in combination with the BT/NT inference, this ensures that the theory holds good for thesaural structures as well

Paradigmatic and syntagmatic relationships



The next step:

- current work is focusing on the representation of BC2 on the web
- while it could be disseminated as an electronic text, it was thought more useful to investigate ways of properly representing the structure and relationships within the terminologies
- various web representation languages were investigated but were found incapable of the required level of complexity

SKOS (Simple Knowledge Organization System):

- SKOS is a good attempt to represent the kind of information about concepts that is required
- however, it covers only a limited range of relationships, and has no potential to expand on these
- a major difficulty is the lack of any way to encode individual concepts in ways other than with a simple identifier
- as a result SKOS is not appropriate for more complex KOSs
- in particular it cannot deal with complex subjects
  - "... the SKOS vocabulary itself does not provide any mechanism for expressing that a given concept consists of a pre-coordination of other concepts."

Text encoding:

- text encoding of the kind used in the TEI appears closer in nature to the encoding used in BC2 and FATKS
- TEI encoding allows the mark-up of individual concepts to support machine manipulation
- the functional elements of TEI are closer in essence to the fundamental categories of facet analysis

Module corpus: Corpus texts Elements defined: activity channel constitution derivation domain factuality interaction locale particDesc preparedness purpose setting settingDesc textDesc

```
<settingDesc>
Â <settingÂ who="#p1 #p2">
Â Â <nameÂ type="city">Bedford</name>
Â Â <nameÂ type="region">UK: South East</name>
Â Â <date>early spring, 1989</date>
Â Â <locale>rug of a suburban home</locale>
Â Â <activity>playing</activity>
Â </setting>
Â <settingÂ who="#p3">
Â Â <nameÂ type="city">Bedford</name>
Â Â <nameÂ type="region">UK: South East</name>
Â Â <date>early spring, 1989</date>
Â Â <locale>at the sink</locale>
Â Â <activity>washing-up</activity>
Â </setting>
Â <settingÂ who="#p4">
Â Â <nameÂ type="place">London, UK</name>
Â Â <time>unknown</time>
Â Â <locale>broadcasting studio</locale>
Â Â <activity>radio performance</activity>
Â </setting>
</settingDesc>
```

Collaborative work:

- current collaboration with humanities computing researchers will, we hope, produce a solution
- it is expected that some software can be written which will incorporate more of the features of the faceted KOS than do currently available systems
- at present discussions are at the planning stage, but there is every cause for optimism.

Summary:

- faceted structures are more than a convenient way to manage search interfaces
- faceted schemes are built on sound conceptual principles of domain analysis and are logical and regular in structure
- they can be easily understood by machines because of this logical nature
- a variety of relationships are clearly expressed and can be encoded for machine manipulation
- the faceted structure, when populated, acts as a model of the subject domain
- automatic generation of systematic schedules and thesaural formats can be achieved by such means
- encoded data provides a basis for automatic classification
- this suggests that the underlying principles are robust, and hold true for different kinds of KOS
- web representation must be able to cope with the semantic richness and complexity of the faceted system, but currently there are no tools that can do so.

References:

Broughton, V. (2008) “A faceted classification as the basis of a faceted terminology” *Axiomathes* 18(2) 193-210

Broughton, V; Slavic, A. (2007) “Building a classification for the humanities: principles and procedures” *Journal of documentation* 63(5)

Facet analytical theory in knowledge structures www.ucl.ac.uk/fatks

Pollitt, A. S. (1997) *The key role of classification and indexing in view based searching*. http://www.ifla.org.sg/IV/ifla63/63polst.pdf

SKOS Primer at: http://www.w3.org/TR/skos-primer/

SKOS Reference and definitions at: http://www.w3.org/TR/skos-reference/

Text Encoding Initiative (TEI) at: http://www.tei-c.org/release/doc/tei-p5-doc/html/CC.html

Wine.com at: http://www.wine.com/Default.asp