Ambiguities in representing thesauri using extended SKOS

Examples from real life

Jutta Lindenthal

TPDL – 2012-09-27 – NKOS workshop

- Concept hierarchies
    - hierarchical relationships between concepts vs. grouping concepts
- Node Labels and Top Terms
    - node labels in the strict sense: denoting arrays of sibling concepts
    - nodes representing concepts not used for indexing
    - nodes representing classification-like elements not used for indexing
    - top terms used to cluster concepts
- Groups
    - classification-like structures, often called themes, fields, subject areas, subject categories, subject groups, micro-thesauri, sub-thesauri, and the like
    - often used for systematic display and browsing
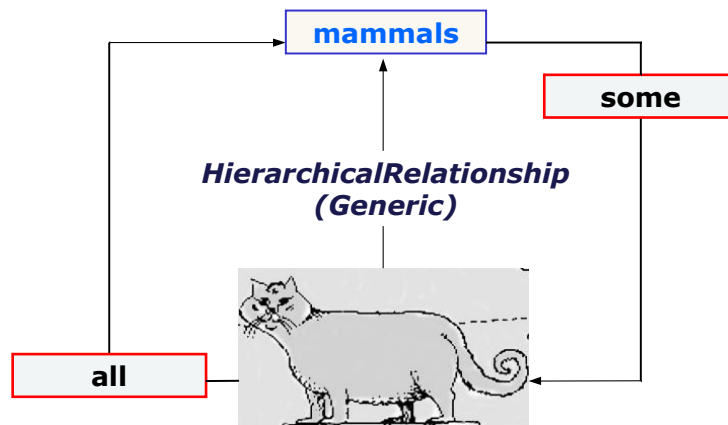- Displays and user interfaces
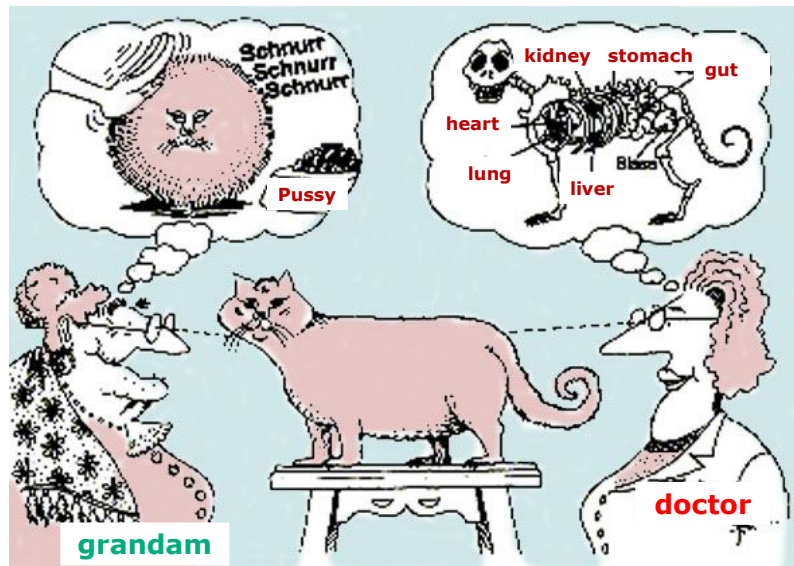
Beauty is in the eye of the beholder ...

Different persons can have different views on the same thing. While a cat is a pet to one person, another one may regard it as a biological object.
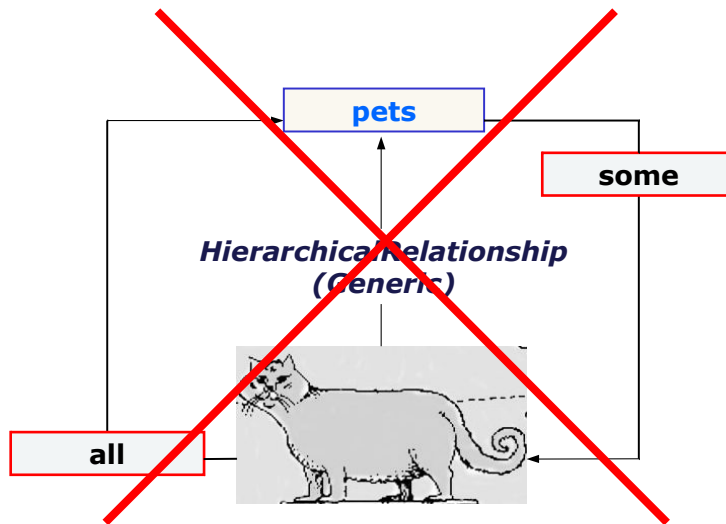
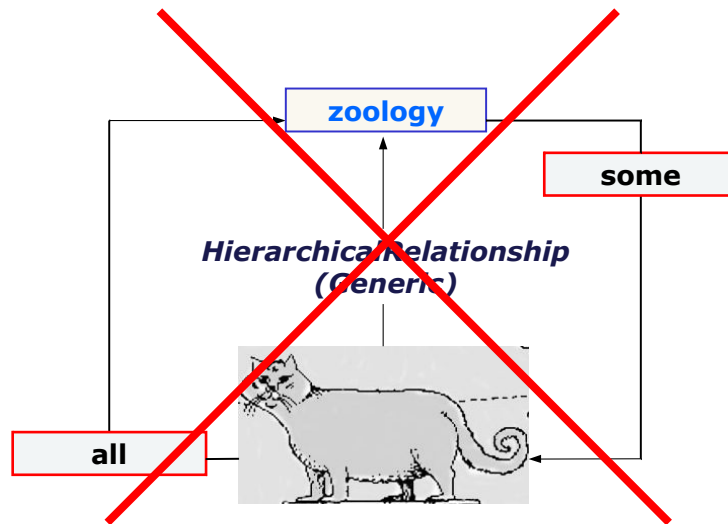Concept hierarchies

Concept hierarchy – All-and-some-test

Nevertheless, most of us would agree in that a cat **is a** type of animal, more specifically a mammal, on ontological and/or epistemological grounds. This type of hierarchical relationship, known as the generic relationship, is amenable to a logical „all-and-some-test". We can say „**all** cats are mammals, and **some** mammals are cats".
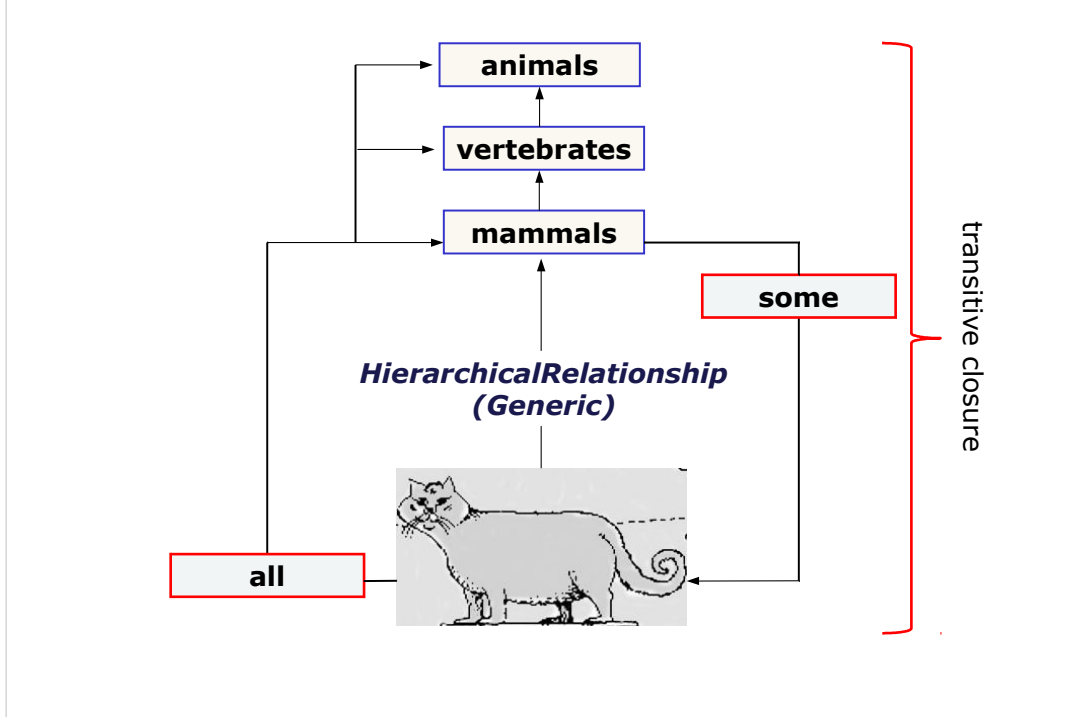
But we can't say „**all** cats are pets", since only **some** cats are pets,

and the all-and-some-test neither holds for a statement like „all cats **are** (a kind of) zoology".

Nevertheless, these are contexts in which cats do occur and in which indexers and users of information systems wish to **find** them. How can we satisfy this demand?

Concept hierarchy – transitive closure

Should we give up the idea of a strictly logical hierarchy and thus of a prerequisite for advanced retrieval options, e.g. by exploiting the transitive closure for enhanced recall or query refinement? Or should we rather create **additional structural elements** to do justice to these different views and aspects?

Concept hierarchy and thesaurus arrays

Generic Hierarchy

.... Agents Facet
.... Living Organisms
.... Eukaryota (domain)
.... Animalia (kingdom)
.... Chordata (phylum)
.... Vertebrata (subphylum)
.... Mammalia (class)
.... Carnivora (order)
.... Felidae (family)
.... Felis (genus)
.... Felis domesticus (species)
....
....
....
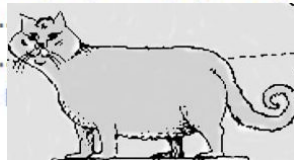
Source: Art & Architecture Thesaurus

ThesaurusArray

cats
. *(cats by degree of domestication)*
.. wild cats
.. semi-feral cats ⎤ **array** <ordered>
.. domestic cats ⎦
. *(cats by usage)* ← **node label**
.. laboratory cats ⎤ **array**
.. pet cats ⎦

Indeed, we can retain our generic hierarchy and use additional elements to group sibling concepts corresponding to different aspects. These characteristics of division are introduced through **node labels**, the concepts thus forming a **thesaurus array** according to the ISO standard.

In this example we have a compound concept, „pet cats" which can be linked poly-hierarchically to the broader concepts „pets" and „cats". (Alternatively, the compound „pet cats" could be split so that the concept is represented in post-coordinate fashion using a combination of „pets" and „cats".)

ThesaurusArray and NodeLabel according to ISO 25964-1
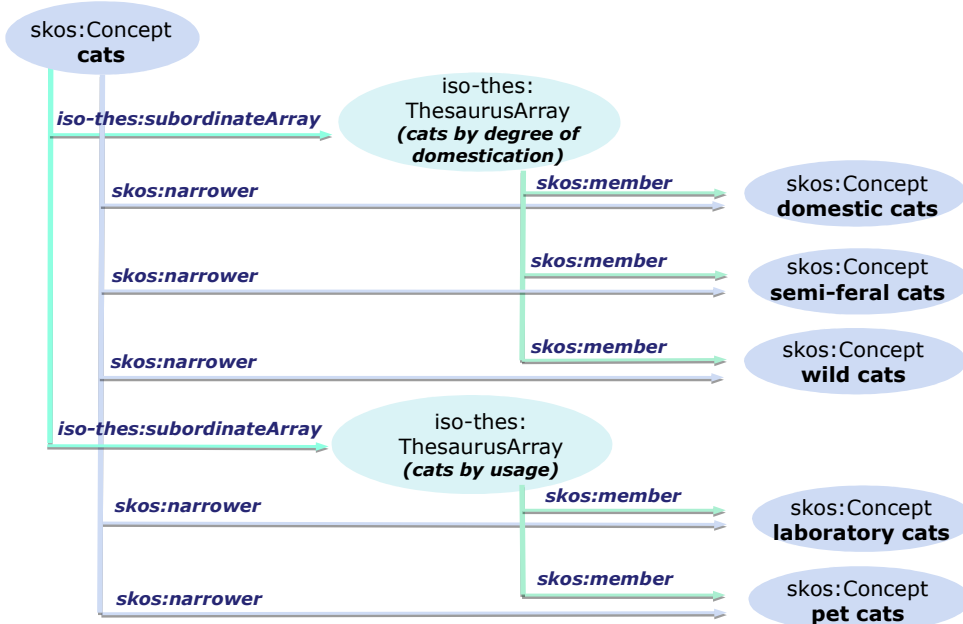
Although all thesauri have arrays of sibling terms, it is not always necessary to model them explicitly and so this class is optional. The ThesaurusArray and NodeLabel classes of the model are needed particularly to support the generation of displays incorporating node labels and/or non-alphabetic sequences of sibling. [...]

Some systems organize the display by treating node labels as though they were concepts, with relationships to broader and narrower concepts. This is done only for display purposes and not for the normal functions of broader and narrower relationships. Only a concept can have true broader and narrower relationships. (ISO 25964-1, 15.2.17 Arrays and node labels)

## Concept hierarchy and concept groups

Generic Hierarchy

.... Agents Facet
.... Living Organisms
.... Eukaryota (domain)
.... Animalia (kingdom)
.... Chordata (phylum)
.... Vertebrata (subphylum)
.... Mammalia (class)
.... Carnivora (order)
.... Felidae (family)
.... Felis (genus)
.... Felis domesticus (species)
....
....
....
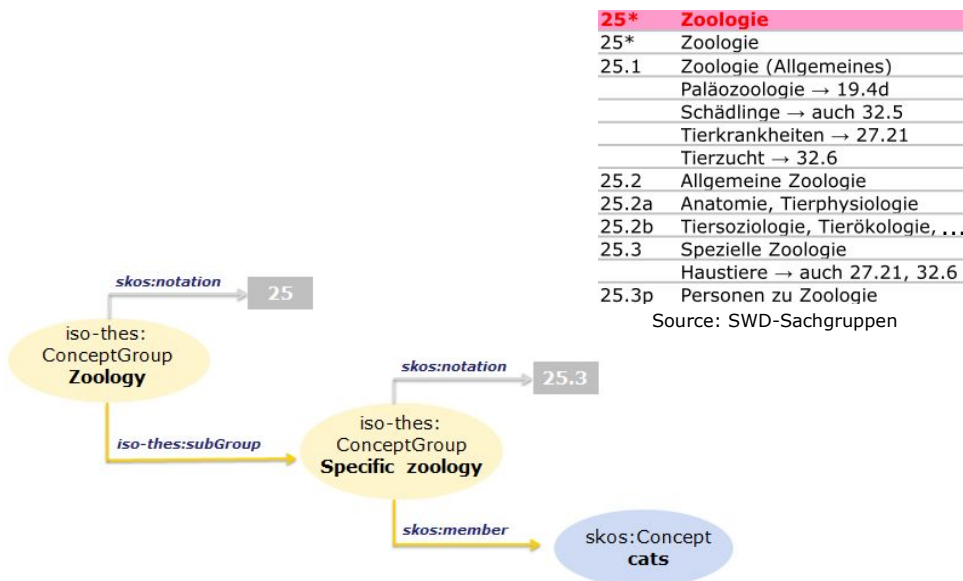
Source: Art & Architecture Thesaurus

ConceptGroup

25 Zoology
25.3 Spezific zoology

Source: Gemeinsame Normdatei
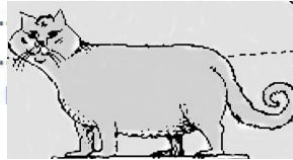
We can also leave the hierarchy untouched and create additional structural elements to aggregate concepts, possibly from different hierarchy paths, under a certain aspect or view. This often results in classification-like structures which are sometimes called fields, themes, micro-thesauri, subject areas, subject groups, or the like. These kinds of structures are represented by **concept groups** in the ISO standard. The example above is from the Gemeinsame Normdatei of the German National Library.

Concept groups according to ISO 25964-1

25*	Zoologie
25*	Zoologie
25.1	Zoologie (Allgemeines)
	Paläozoologie → 19.4d
	Schädlinge → auch 32.5
	Tierkrankheiten → 27.21
	Tierzucht → 32.6
25.2	Allgemeine Zoologie
25.2a	Anatomie, Tierphysiologie
25.2b	Tiersoziologie, Tierökologie, …
25.3	Spezielle Zoologie
	Haustiere → auch 27.21, 32.6
25.3p	Personen zu Zoologie

Source: SWD-Sachgruppen

The „SWD-Sachgruppen" (more recently called „GND-Systematik") of the German National Library is a concept scheme in parallel to the „Gemeinsame Normdatei" (GND). The subject headings of the GND are assigned to the „Sachgruppen" (subject groups) which are best represented as iso-thes:ConceptGroups.

Many thesauri group concepts using a classification structure that exists in parallel to the hierarchies of thesaurus concepts based on BT/NT relationships. Groups created by the classification are often based on disciplines, subject areas or areas of business activity. [...] In general, there is not a BT/NT relationship between a ConceptGroup and the concepts that it contains. Concepts may be gathered into ConceptGroups from many different facets or hierarchies of the thesaurus, and the notation used for the classification into groups may be quite distinct from any notation that might be used for the concepts themselves. Groups may have subgroups, being nested to any level. (ISO 25964-1, 15.2.18 Concept groups)
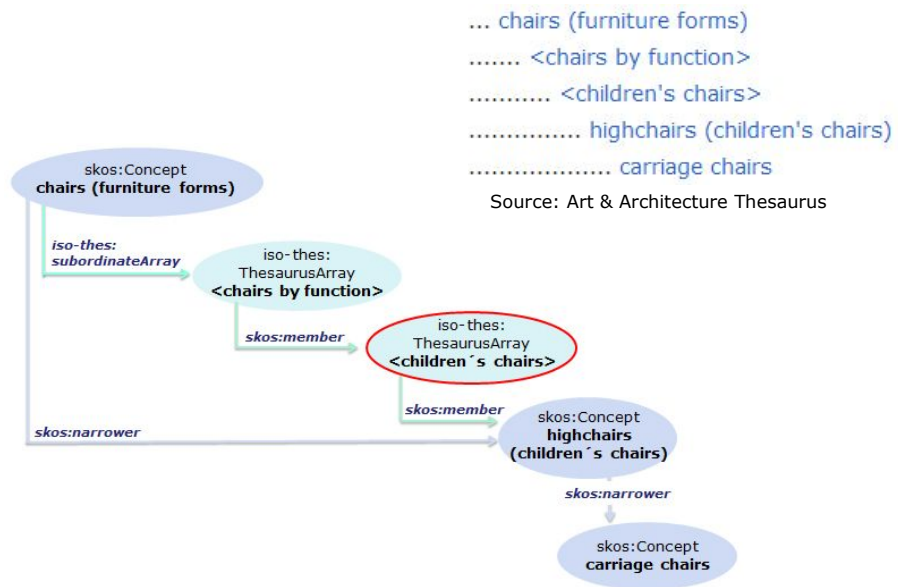
Node Labels and Top Terms
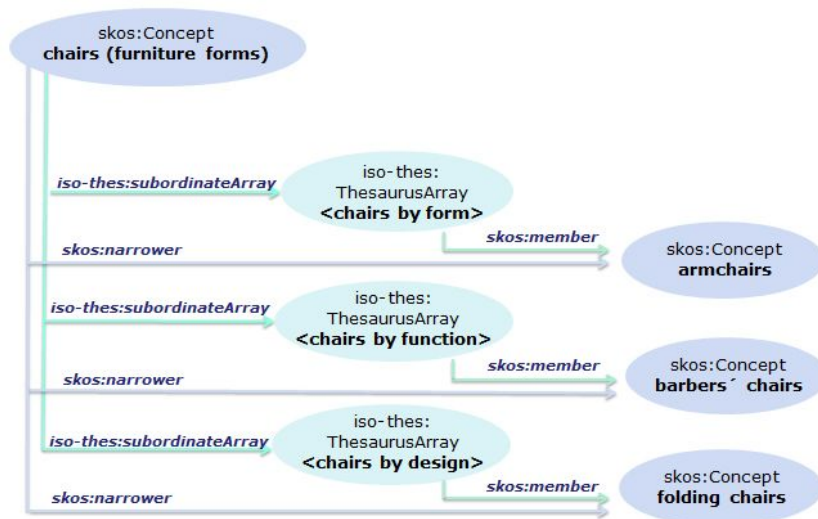
Art& ArchitectureThesaurus

The hierarchies of the Art & Architecture thesaurus are structured by so-called „guide terms" which all serve the function of **grouping sibling concepts**. However, not all of them comply with the definition of node labels given by the ISO 25964 standard. „Chairs by form", „chairs by function" and „chairs by design" are node labels according to the standard, showing characteristics of division.

... chairs (furniture forms)
.... <chairs by function>
.... <children's chairs>
.... highchairs (children's chairs)
.... carriage chairs

Source: Art & Architecture Thesaurus

skos:Concept
chairs (furniture forms)

iso-thes: subordinateArray

iso-thes:
ThesaurusArray
<chairs by function>

skos:member

iso-thes:
ThesaurusArray
<children´s chairs>

skos:narrower

skos:member

skos:Concept
**highchairs
(children´s chairs)**

skos:narrower

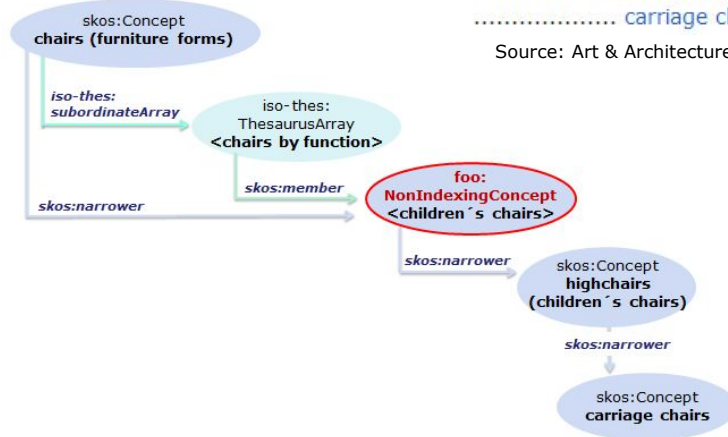skos:Concept
carriage chairs

Art& ArchitectureThesaurus

In this example, the guide term „childrens's chairs" would not comply with the ISO definition of a thesaurus array. It bundles sibling concepts like node labels do, but does not designate a characteristic of division. It is rather a concept which is not used for indexing.

Guide Terms in the AAT Thesaurus (3)

... chairs (furniture forms)

.... <chairs by function>

.... <children's chairs>

.... highchairs (children's chairs)

.... carriage chairs

Source: Art & Architecture Thesaurus

skos:Concept
chairs (furniture forms)

*iso-thes:
subordinateArray*

iso-thes:
ThesaurusArray
<chairs by function>

skos:member

foo:
**NonIndexingConcept
<children´s chairs>**

skos:narrower

skos:narrower

skos:Concept
**highchairs
(children´s chairs)**

skos:narrower

skos:Concept
carriage chairs

Art& ArchitectureThesaurus

To represent this type of node, a class „NonIndexingConcept" could be introduced as a subclass of skos:Concept.
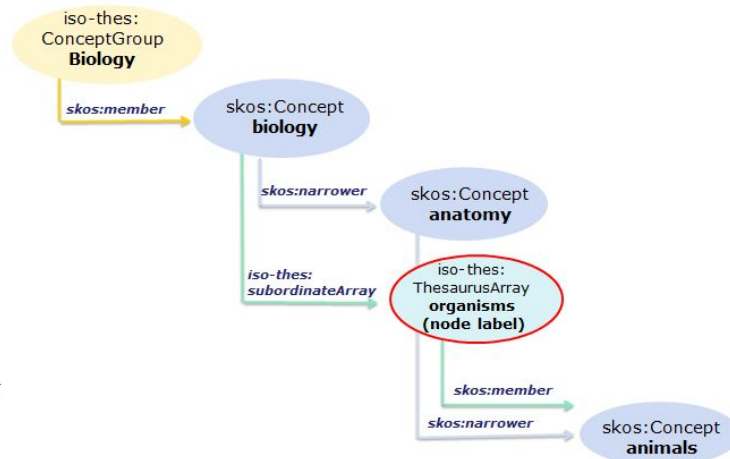
"Node label" in the Bioethic Thesaurus (1)

Subject Area IV:
Biology

biology
    anatomy
    anthropology [biology]
    behavioural biology +
    biochemistry
    biometry
    biophysics
    biosystematics +
    botany +
    cell biology +
    chronobiology
    developmental biology
    ecology +
    evolutionary biology +
    histology
    immunobiology +
    microbiology +
    molecular biology +
    morphology
    organisms (node label)
        animals +
        clones
        fungi +
        genetically modified organisms +
        harmful organisms
        introduced species +
        microorganisms +
        pathogenic organisms
        plants +
        renewable resources +
        soil organisms
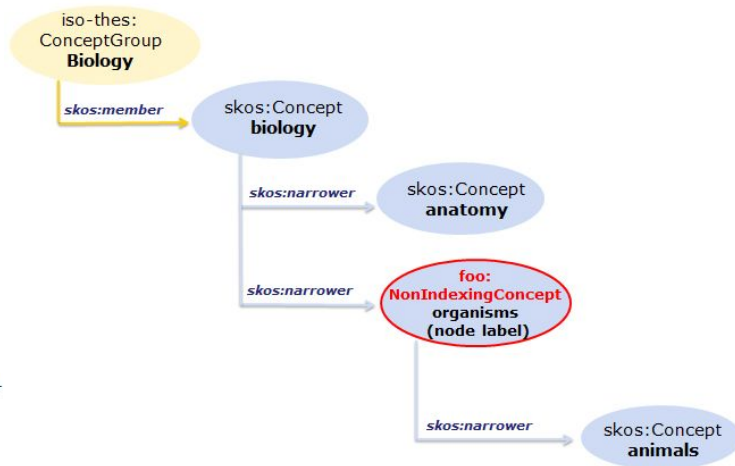   viruses +

Thesaurus Ethics in the Life Sciences

The thesaurus „Ethics in the Life Sciences" makes use of so-called node labels for clustering concepts. These are not used for indexing. The example of a „node label" here differs from the AAT example in that „organisms" is not part of a generic hierarchy, since „organisms" are not a kind of biology. If „organisms (node label)" would be represented as the node label for an iso-thes:ThesaurusArray, then the concept „animals" would become a narrower concept of the concept „biology", and a sibling of „anatomy". These are not correct statements.

„Node label" in the Bioethic Thesaurus (2)

Thesaurus Ethics in the Life Sciences

The "node label" in this thesaurus could be represented as a NonIndexingConcept.
Note that the subordinate concepts are not all narrower concepts in the strict sense.

18

Facets, Hierarchy Names and Guide Terms in the AAT Thesaurus

	Record Type
Top of the AAT hierarchies	
.... Styles and Periods Facet ←	facet
.... Styles and Periods ←	hierarchy name
.... <styles and periods by general era> ←	guide term
.... ancient	concept
.... prehistoric	concept
.... protohistoric	concept
.... <three-age system> ←	guide term
.... Stone Age	concept
.... <transitional periods Stone Age to Bronze Age> ←	guide term
.... Bronze Age	
.... Iron Age	
.... Dark Ages (general style and period)	
.... Contemporary (style of art)	
.... Modern (styles and periods)	
.... pre-ceramic (general era)	
.... space age [N]	

Art & Architecture Thesaurus

Throughout the Art & Architecture Thesaurus, the concept hierarchy is built entirely from generic hierarchical relationships as specified in the ISO 25964 standard.

The generic hierarchy is not only supplemented with guide terms, but also with facets and hierarchy names, none of which are used for indexing. Neither of the three AAT node types (facet, hierarchy name, and guide term) as used in this example has a direct counterpart in the ISO standard. Using various kinds of non-indexing nodes for clustering concepts according to certain criteria can also be found in many other thesauri.

The standard leaves it undefined how facets and sub-facets are to be handled in a given application. Further investigation is needed on how to formalise facets.

Facets are provided for in the data model although not modelled explicitly, since different thesauri apply facet analysis in different ways. If facet names are included as top terms, they should be treated as normal preferred terms applying to top level concepts. Alternatively, they may be the names of concept groups, and/or they may appear in node labels. (ISO 25964-1, 15.2.19 Treatment of facets)

Top Terms in the ELSST Thesaurus

Preferred Term
EDUCATIONAL LEVELS

Scope Note
THE DEFINITIONS FOR THE LEVELS OF EDUCATION ARE BASED ON IS

Synonyms/UFs

Broader Terms	Narrower Terms
EDUCATIONAL SYSTEMS	PRE-PRIMARY EDUCATION PRIMARY EDUCATION SECONDARY EDUCATION TERTIARY EDUCATION

Top Terms(Click for hier Display)
EDUCATION (TT)

- EDUCATION
  - CURRICULUM
    - EXTRACURRICULAR ACTIVITIES
      - SCHOOL CLUBS
  - EDUCATIONAL BACKGROUND
  - EDUCATIONAL COURSES
    - DISTANCE LEARNING
    - PART-TIME COURSES
    - TRAINING COURSES
    - UNIVERSITY COURSES
  - EDUCATIONAL GROUPING
    - ABILITY GROUPING
    - EDUCATIONAL INTEGRATION
    - SCHOOL CLASSES
  - EDUCATIONAL OPPORTUNITIES
    - ACCESS TO EDUCATION
    - EDUCATIONAL CHOICE
    - EDUCATIONAL DISCRIMINATION
  - EDUCATIONAL PROVISION
  - EDUCATIONAL SYSTEMS AND INSTITUTIONS
  - FIELDS OF STUDY
  - HEALTH PROMOTION AND EDUCATION
  - SPECIAL EDUCATION
  - TEACHING
  - TECHNICAL EDUCATION
  - VOCATIONAL EDUCATION

ELSSTEuropean Language Social Science Thesaurus

Some thesauri use „Top Terms" (TT) to create a mere grouping of hierarchies, like in this example of the ELSST thesaurus. This application of top terms is not conformant with the understanding of „top concept" given in the ISO 25964 standard, where such concepts are inferred bottom-up from the hierarchical relationships between thesaurus concepts. In the ELSST, the concepts do not form consistent hierarchies amenable to the all-and-some-test. „Extracurricular activities" are not a kind of „curriculum" which in turn is not a kind of „education". „School classes" for instance are not a kind of education, since they belong to different fundamental categories (activities and objects, respectively). Re-declaring the ELSST „top terms" as concept groups would avoid these conflicts with the standard.

Terms representing top concepts, as an optional feature of alphabetical displays and single term displays, linking a concept to the top concept of the hierarchy in which it occurs. The class TopLevelRelationship enables this. A browse capability that starts by presenting editors with a list of top terms is also recommended. The attribute topConcept addresses this, specifying that a concept is at the top of a hierarchy, i.e. it has no broader concepts. This makes it easier, on importing a thesaurus, to pick out all the top level concepts from which to build a navigational tree. (ISO 25964-1, 15.2.13 Top level concepts)

Groups

Systematics of the STW Thesaurus for Economics (1)

V Economics
- V.00 Economy
- ▸ V.01 Economic Theory
- ▸ V.02 Microeconomics
- ▸ V.03 Macroeconomics
- ▸ V.04 Business Cycle, Growth, and Economic Structure
- ▾ V.05 Money and Financial Markets
  - V.05.01 Monetary Theory
  - V.05.02 Monetary Policy
  - V.05.03 Inflation
  - V.05.04 Interest Rate
  - V.05.05 Credit
  - ▾ V.05.06 Financial Markets
    - V.05.06.01 Money Market
    - ▸ V.05.06.02 Securities Markets
    - V.05.06.03 Forward Market
    - V.05.06.04 Foreign Exchange Market
  - V.05.07 Payments

V.05.06 Financial Markets

V.05.06 Finanzmärkte (german)

broader

- ▪ V.05 Money and Financial Markets ▾

narrower

- ▪ V.05.06.01 Money Market ▾
- ▪ V.05.06.02 Securities Markets ▾
- ▪ V.05.06.03 Forward Market ▾
- ▪ V.05.06.04 Foreign Exchange Market ▾

Descriptors

- ▪ Arbitrage
- ▪ Arbitrage pricing
  ...
- ▪ Yield

Persistent Identifier (for bookmarking and linking)

- ▪ http://zbw.eu/stw/thsys/70187

Standard-Thesaurus

This is an example for grouping concepts in a classification-like structure (called taxonomy or thesaurus systematics in this case). It is taken from the STW Thesaurus for Economics with kind permission by Joachim Neubert.
See for further information: Joachim Neubert: Bringing the „Thesaurus for Economics" on to the Web of Linked Data.
http://events.linkeddata.org/ldow2009/papers/ldow2009_paper7.pdf

Turtle representation of a systematics group of the STW

```
@base <http://zbw.eu/stw/> .

<thsys/71107>
    a skos:Concept, zbwext:Thsys ;
    rdfs:isDefinedBy <thsys/71107/about> ;
    rdfs:label "Securities Markets"@en, "Wertpapiermärkte"@de ;
    skos:notation "V.05.06.02"^^xsd:string ;
    skos:prefLabel "V.05.06.02  Securities Markets"@en,
                   "V.05.06.02  Wertpapiermärkte"@de .
    skos:broader   <thsys/70187> ; [Financial Markets]
    skos:narrower  <descriptor/...>,
                   <thsys/71105>,  [Bond Market]
                   <thsys/71106> ; [Stock Market]
    skos:inScheme <../stw> .
```

Source: http://zbw.eu/stw/versions/8.10/thsys/71107/about.ttl

Source: STW Thesaurus for Economics

Here we have an excerpt from a Turtle representation of the dataset „Securities Markets" provided by STW.
When "skosifying" the "STW Thesaurus for Economics", a subclass of skos:Concept, zbwext:Thsys" was introduced to represent the systematics groups. "Using skos:Collection for the taxonomy was not an option, because this class cannot be nested to express hierarchies. Therefore we choose to subclass skos:Concept."
Joachim Neubert: Bringing the „Thesaurus for Economics" on to the Web of Linked Data. http://events.linkeddata.org/ldow2009/papers/ldow2009_paper7.pdf
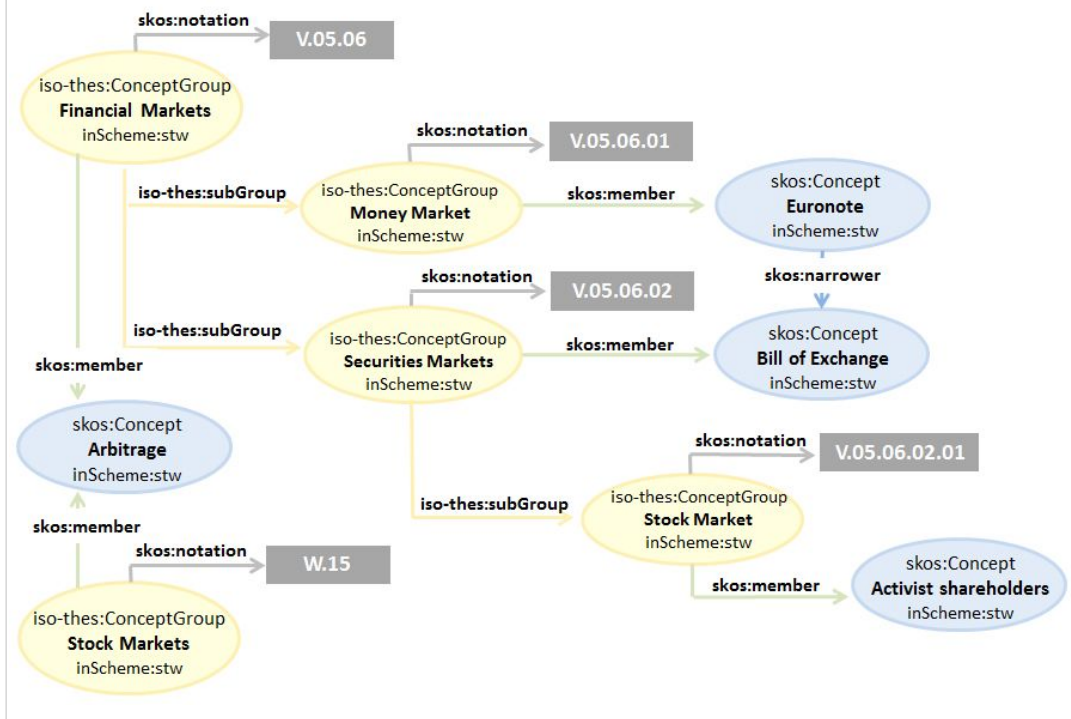
Turtle representation using iso-thes extensions

```
<thsys/71107>

    a iso-thes:ConceptGroup ;
    rdfs:label  "Securities Markets"@en, "Wertpapiermärkte"@de ;
    skos:notation "V.05.06.02"^^xsd:string  ;
    skos:prefLabel "Securities Markets"@en,
                    "Wertpapiermärkte"@de .
    iso-thes:subGroup <thsys/71105> ;
    iso-thes:subGroup <thsys/71106> ;
    skos:member <descriptor/...> ,
    skos:member <descriptor/...> ;
    skos:inScheme <stw> ;
    iso-thes:classificationOf <../stw>  .
```

This is an alternative representation of a STW systematics group using iso-thes:ConceptGroups. This representation takes into account the fact that the relationships among skos:Collections (or iso-thes:ConceptGroup, a subclass thereof) as well as between skos:Collections and skos:Concepts differ from the hierarchical semantic relationships between skos:Concepts, such as „skos:broader" or „skos:narrower".

Graphical representation of systematics groups of the „STW Thesaurus for Economics" using iso-thes elements.

Displays and user in interfaces

Source: Alibaba.com

Products ▾ | fertilizer

▼ Category
Organic Fertilizer (24364)
Nitrogen Fertilizer (17431)
Compound Fertilizer (9084)
Urea (6986)
Other Fertilizers (6688)
More..

▼ Type
Compost (2905)
Humic Acid (12938)
Manure (1247)
More..

▼ State
Granular (26852)
Powder (21831)
Liquid (4224)

▼ Release Type
Quick (25255)
Controlled (7120)
Slow (16014)

fertilizers
(by type of plant nutrient)
— Calcium fertilizers
— Magnesium fertilizers
(by number of plant nutrients)
— Compound fertilizers
— Fertilizer combinations
(by substance group)
— Inorganic fertilizers
— Organic fertilizers
— Organomineral fertilizers
(by aggregate state)
— Liquid fertilizers
— Liquid gas fertilizers
(by nutrient release time)
— Slow release fertilizers

Source: AGROVOC

fertilizers (2867) +
• NT Calcium fertilizers
• NT Compound fertilizers
  • • NT NPK fertilizers
  • • NT Nitrogen phosphorus fertilizers
    • • • NT Nitrophosphates
  • • NT Nitrogen potassium fertilizers
  • • NT Phosphorus potassium fertilizers
• NT Fertilizer combinations
  • • NT fertilizer pesticide combinations
• NT Inorganic fertilizers
• NT Liquid fertilizers
• NT Liquid gas fertilizers
• NT Magnesium fertilizers
• NT Micronutrient fertilizers
• NT Organic fertilizers
  • • NT Biofertilizers
  • • NT Composts
    • • • NT Potting composts
  • • NT Fish manure
  • • NT Green manures
  • • NT Guano
• NT Organomineral fertilizers
• NT Phosphate fertilizers
  • • NT Rock phosphate
  • • NT Superphosphate
• NT Potash fertilizers
• NT Slow release fertilizers
• NT Sulphur fertilizers
• NT nitrogen fertilizers
  • • NT ammonium fertilizers
  • • NT nitrate fertilizers

Making use of thesaurus arrays and node labels can support user interfaces based on **faceted browsing**. The example on the left is from a trade catalog listing commercial sources for fertilisers. Similar views can be generated for document sets indexed with the Agrovoc thesaurus.

Descriptor display of German Parliament Thesaurus PARTHES

Schlagwort:	☑	**Haustier**
Thesaurus:		Sachbegriffe
Definition:		NUR VERWENDEN für Heimtiere (z. B. Hauskatzen, Hunde)
Benutzt für:		
		Heimtier
Oberbegriffe:		
	☐	➜ Nutztier
Unterbegriffe:		
	☐	➜ Hund
	☐	➜ Katze
Verwandte Begriffe:		
	☐	➜ Tierhaltung
	☐	➜ Tierhandel
	☐	➜ Tierheim
Zuordnung:		
		➜ NU.01.01 - Tiere
		➜ LE.05 - Tierhaltung, Tierzucht, Tierschutz

*Thesaurus*für Parlamentsmaterialien PARTHES

In this descriptor display of the PARTHES Thesaurus, users can expand their selection of search vocabulary along different relationships.

Systematic display of German Parliament Thesaurus PARTHES

Auswahl

- ⊞ RE Religion, Ethik und Philosophie
- ⊞ MK Medien, Information und Kommunikation
- ⊟ NU Natur und Umwelt
  - ⊟ NU.01 Lebewesen
    - **NU.01.01 Tiere**
    - NU.01.02 Pflanzen
  - NU.02 Landschaften, Ökosysteme, Biotope
  - NU.03 Klima, Wetter, Luft
  - NU.04 Boden
  - NU.05 Wasser und Gewässer
  - NU.06 Umweltbelastungen,

Details

- ☐ → Hai
- ☐ → Hase
- ☐ → Haustier
- ☐ → Huhn
- ☐ → Hund
- ☐ → Insekt
- ☐ → Kampfhund
- ☐ → Kaninchen
- ☐ → Katze

*Thesaurus*für Parlamentsmaterialien PARTHES

Nested concept groups are used here to generate a systematic display for top-down browsing. Assigned descriptors are shown in the „Details" column, where they can be selected for a search query.

Hierarchical display of German Parliament Thesaurus PARTHES

*Thesaurus*für Parlamentsmaterialien PARTHES

In this view of PARTHES, the tree display is generated from the concept hierarchy. The „Details" column contains the relationship context of the selected concept for additional selection or navigation.

Make things as simple as possible, but not simpler.