Introduction to Knowledge Organization Systems (KOS)

Eero Hyvönen
Aalto University, Semantic Computing Research Group (SeCo) *http://seco.cs.aalto.fi*
University of Helsinki, HELDIG *http://heldig.fi*

eero.hyvonen@aalto.fi

Learning Objectives

Learn the idea of knowledge organizing systems in different disciplines:

- Philosophy
- Linguistics
- Terminology
- Library and Information Sciences
- Computer Science

Outline

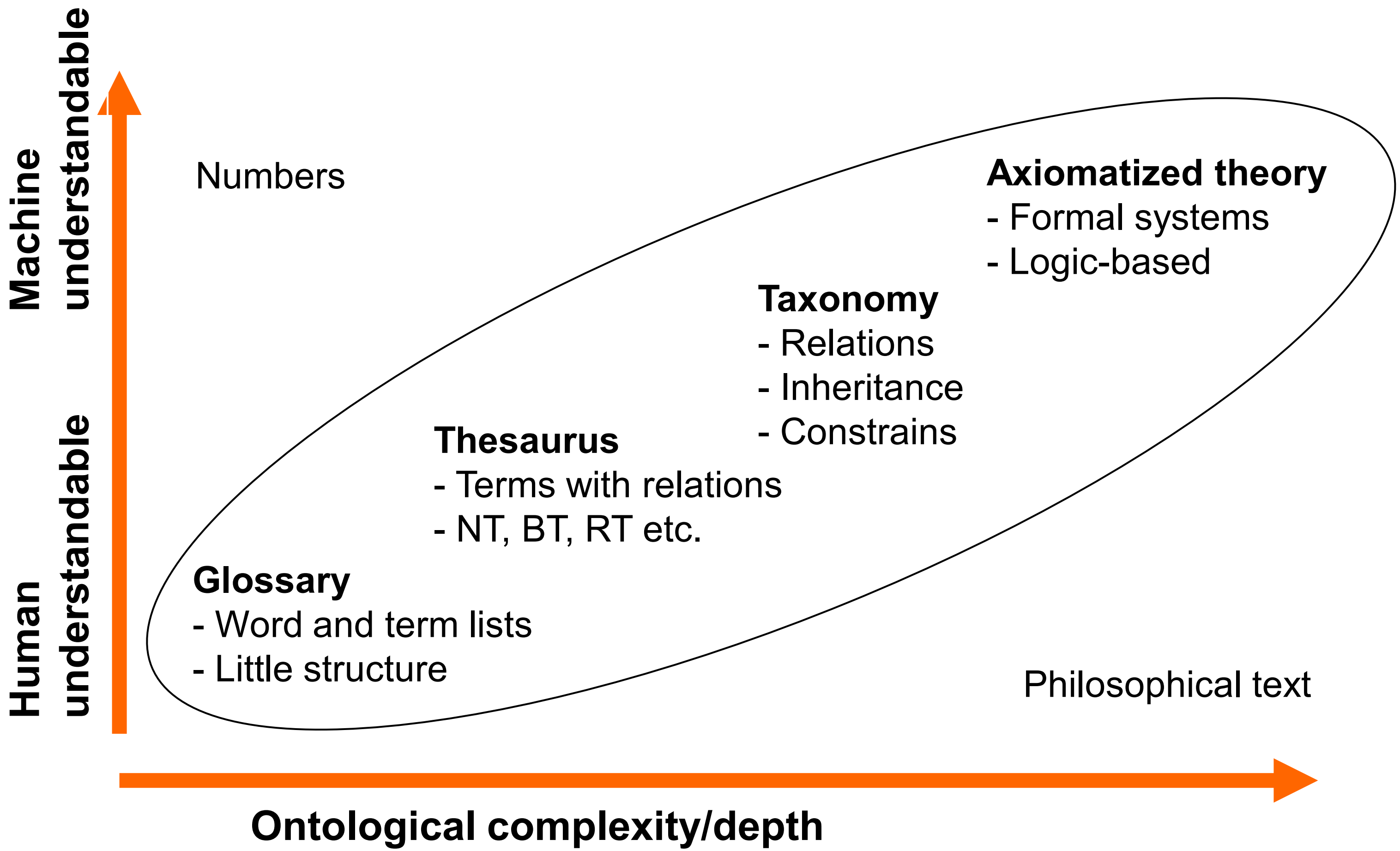
What are Knowledge Organization Systems?

Five perspectives to KOS

What are Knowledge Organizing Systems (KOS)

“KOS is mostly used to refer to functional items designed for organizing knowledge and information, and making their management and retrieval easier”
(Encyclopedia of Knowledge Organization, https://www.isko.org/cyclo/kos)

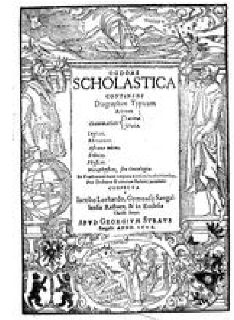
Different Kind of KOS

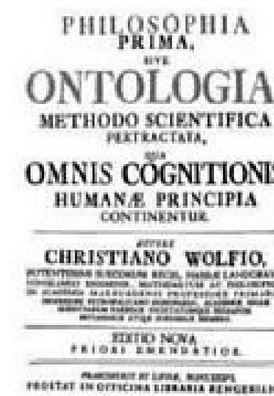


KOS in Philosophy: Ontology

Perspectives of ontology: Philosophy

- Ontology = Study of the essence of being
  - *Apart from the particular existing things*
- Examples of ontological studies
  - Plato's world of ideas in metaphysics
  - Aristotle's (384–322 B.C.) 10 Categories
- Medieval logicians: first semantic net
  - *Genus (supertype) vs. species (subtype)*
- "Ontology" as a discipline with a name
  - *R. Göckel, J. Lorhard, 1613*
  - *Kant (1787), Peirce, Husserl, Whitehead, Heidegger, …*
- Today often theoretical studies in formal logic
  - *Foundational categories & logic behind everything*



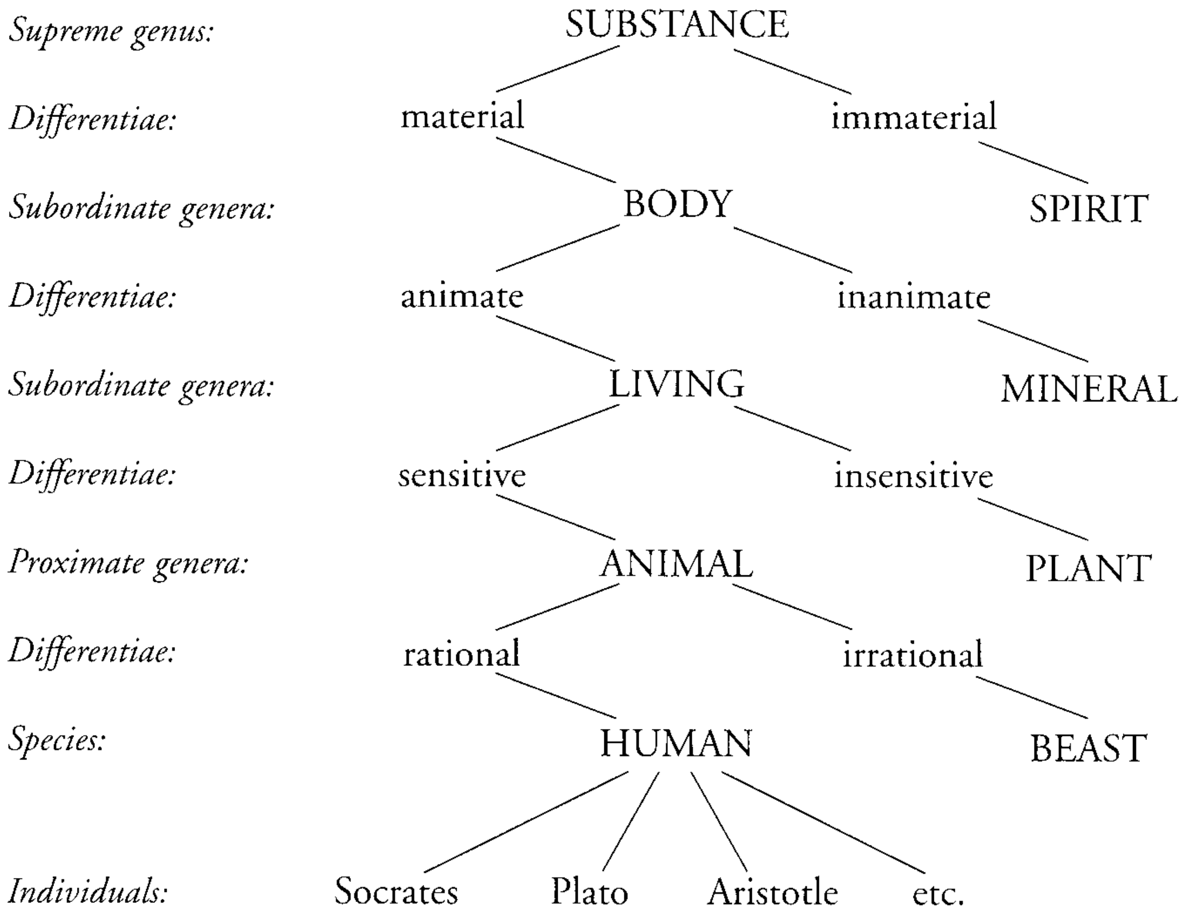


Aristotle's 10 Categories

Substance	A cat
Quantity	The cat is 50 cm high
Quality	The cat is black
Relation	The cat is half the size of …
Where	The cat is in the house
When	The cat came out yesterday
Position	The cat sat
Having	The cat has a rat
Action	The cat is running
Affection	The cat desires fish

(Sowa, 2004)

Hierarchical categories:
Tree of Porphyry of Aristotle's "Substance"



(Sowa, 2004)

FIGURE 1.1 Tree of Porphyry, translated from a version by Peter of Spain (1239)

Aristotle's Syllogisms -> Logic

- Four types of propositions:

Type	Name	Pattern
A	*Universal affirmative*	Every *A* is *B*.
I	*Particular affirmative*	Some *A* is *B*.
E	*Universal negative*	No *A* is *B*.
O	*Particular negative*	Some *A* is not *B*.

- Examples of syllogisms:

Barbara	**Celarent**
A: Every animal is material. A: Every human is an animal. A: ∴ Every human is material.	E: No spirit is a body. A: Every human is a body. E: ∴ No spirit is a human.
Darii	**Ferio**
A: Every beast is irrational. I: Some animal is a beast. I: ∴ Some animal is irrational.	E: No plant is rational. I: Some body is a plant. O: ∴ Some body is not rational.

- Key idea: thinking can be formalized!

(Sowa, 2004)

Formal Ontology

Branch of philosophy

- Well-defined mechanical models of human reasoning
- Using formal methods in the study of being
- Developing formal (logical) ontological theories
- Combination of philosophy and AI

Theories

- Theory of parts & wholes
- Theory of time
- Naïve physics vs. traditional physical models
- ...

Formal (domain independent) ontologies are used for creating domain specific ontological models

- Interoperability through shared principles

Linguistics: Perspectives to KOS

Language-based Perspectives to KOS

- Semantic glossaries
  - Reorganizing dictionaries based on meanings
- Semantic thesauri
  - Representing relational structures between meanings
- Terminologies
  - For defining terms
  - Based on Concept Analysis

Roget's Thesaurus: A Semantic Glossary

Idea: organizing words according to meaning, not alphabetically

Everything in 1000 categories

- Nouns, adjectives, verbs, …
  - *1852: 15,000 words*
  - *1975: over 100,000 words*
  - *1992: over 250,000 words*
- Neighboring categories semantically related
  - *E.g., 266="Journey"; 267="Navigation"*
- Not a formal model
  - *Targeted for human interpretation*
  - *E.g., for finding alternative expressions in writing*

Roget's Thesaurus: Example

Top level: 6 classes

- *1. Abstract relations*
- *2. Space*
- *3. Matter*
- *4. Intellect*
- *5. Volition*
- *6. Affections*

CLASS 2. Space
Space in general
I Abstract space
180 Indefinite space {Noun: space, extension, extent, expanse,…
Verb: reach, extend,…
Adj: spacious, roomy, …
Adv: extensively, …}
181 Definite region …
182 Limited space …
II Relative space
183 Situation …
…

Semantic Thesauri:
WordNet – A Lexical Database for English

- **Words (nouns, verbs, adj., adv.) are organized into synonym sets, i.e., 117 000 synsets/concepts**
- **Synsets are organized in conceptual hierarchies**
- **Key ideas:**
  - Link meanings (synsets) in addition to words in semantic relations : hyponymy, meronymy, (nouns) troponymy (verbs), antonynymy (adj)
- **Several language versions exist**

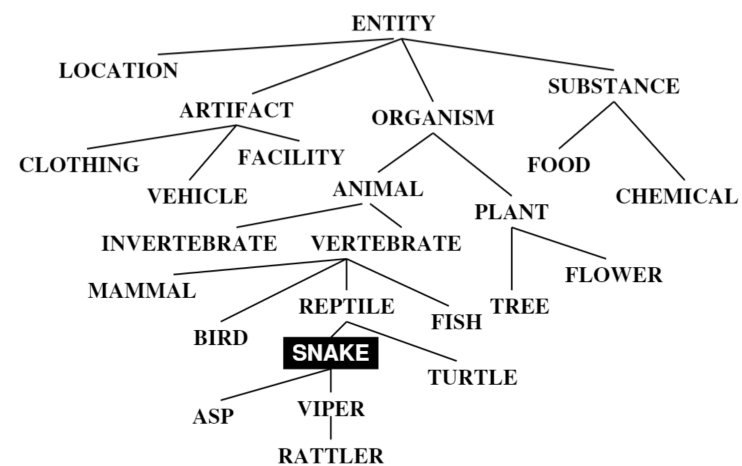


Figure 1. A simplified portion of the Wordnet taxonomy of nominal concepts above the noun "rattler".

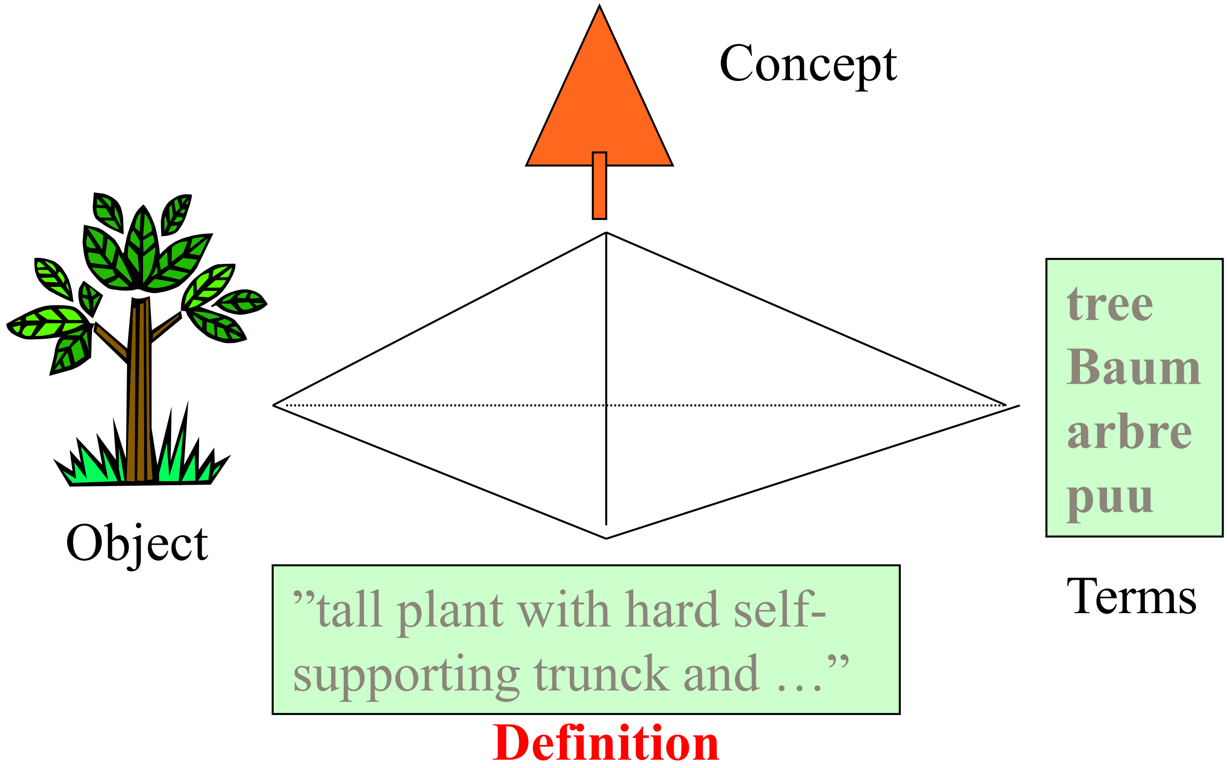
(M. Ciaramita et al., 2008)

Terminology

- Two meanings of “Terminology”:
  1. *Terminology = the group of specialized words or meanings relating to a particular field*
  2. *Terminology = the study of such terms and their use*
- Goal: Terminology defines terminologies human users
  - *In Finland: e.g., Finnish Terminology Centre TSK (http://www.tsk.fi/tsk/)*
- Based on **concept analysis** of word meanings
  - Standardized methodology (by ISO)
- Following presentation is based on:
  - *Heidi Suonuuti: “Guide to terminology”, 2001*

Concept Analysis:

Extended Odgen-Richards triangle to tetraed



Concept vs. Term Relationships

Monosemy

- One term - one concept

Polysemy

- One term - many *related* concepts
  - *E.g. "head" (of arrow) vs. "head" (of human)*

Homonymy

- One term - many *unrelated* concepts
  - *E.g., "bank" (institution) vs. "bank" (of a river)*

Synonymy

- One concept – many terms
  - *E.g. "apartment"="flat"*

Specifying Term Definitions

Based on identifying delimiting characteristics (properties)

- Delimiting characteristics differentiate concepts
  - *E.g., concept "tree":*
    - "have a root" — not delimiting from, e.g., pushes or flowers
    - "have a self-supporting trunk" — delimiting

Intensional and extensional definitions can be used

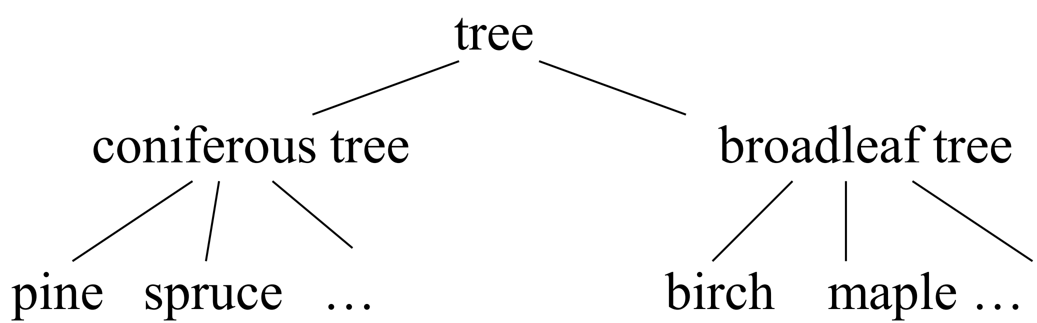
- Intension = sum of general characteristics
  - *Describe only delimiting characteristics*
  - *Other characteristics come from hierarchies of concepts*
    - Tree = "long-living **plant**, have a self-supporting trunk, …"
- Extensional = list of objects
  - *Tree = {pine, maple, spruce, …}*
  - *Weekday = {Sunday, Monday, Tuesday, Wednesday, …}*

Concept Systems

- Concepts are not independent but are related
- Concept systems are used for making definitions
- Relation types between concepts:
  1. *Generic relations (hyponymy)*
  2. *Partitive relations (meronymy)*
  3. *Associative relations*

1. Generic Relation (Hyponymy)

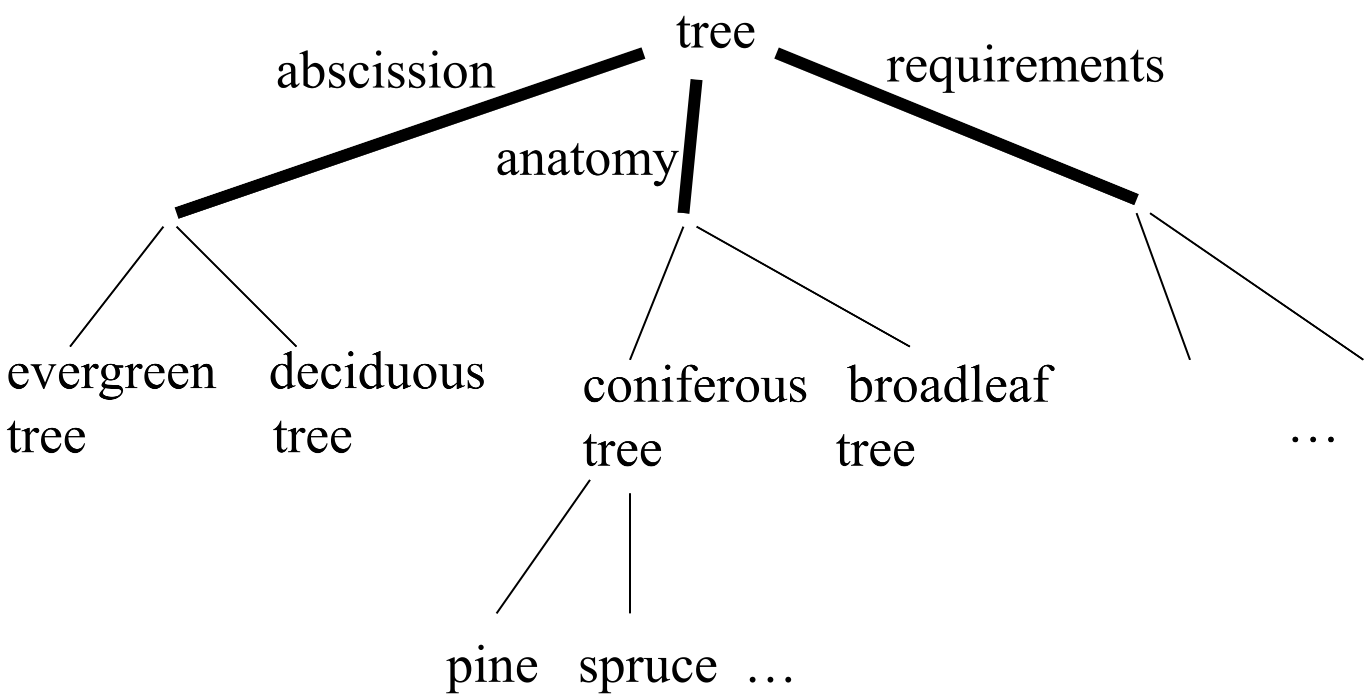
- Concepts share general characteristics but one
- Concept hierarchy: super/subordinate



- Problem: several branching possibilities
  - *Anatomy: coniferous vs. broad leaf*
  - *Requirements: light-demanding vs. tolerant*
  - *Abscission: evergreen vs. deciduous*

Representing Parallel Independent Subdivisions

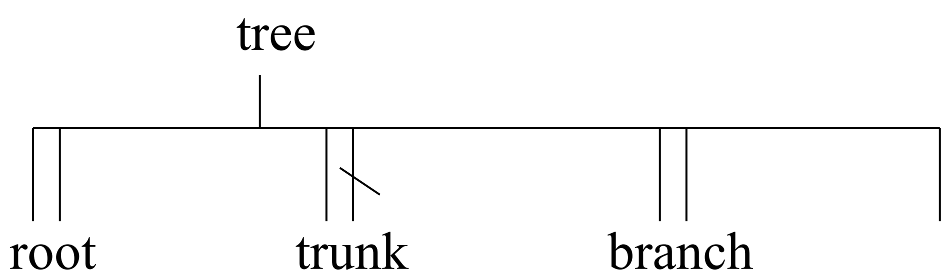
Using three subdivision dimensions:



2. Partitive Relation (Meronymy)

Part-whole relation

- Examples:
  - *Atoms in a molecule*
  - *Legs of a chair*
- Optional, single, and multiple parts



- Also subcategorization using different criteria is possible
  - *Tree -> permanent vs. non-permanent organs*

2. Partitive Relations (Meronymy)

part / whole	branch / tree
member / set	tree / forest
piece / whole	piece-of-cake / cake
material / object	aluminum / airplane
phase / process	childhood / growing-up
place / region	Helsinki / Finland

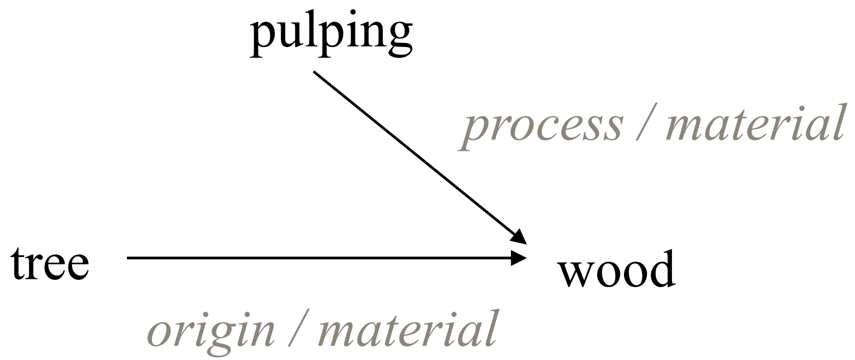
(C. Fellbaum, 1998)

3. Associative Relations

cause / effect	spring / leafs in trees
producer / product	bird / nest
activity / actor	nesting / bird
activity / location	nesting / tree
object / location	nest / tree
object / activity	apple tree / fruit gathering
tool / function	paper machine / paper making
material / product	wood / paper
etc…	etc…

3. Associative Relation

Concept analysis notation example



Why Terminology?

- Provides useful methodology for defining concepts definition
- *Normative* goal
  - *Analyze, select, harmonize, and define a concise set of terms to* ***be used in human communications***
- Does not provide formal enough *descriptive* representations for machine semantics
  - *But concept analysis methodology is useful there, too*

KOS in Library and Information Sciences

Major Approaches to KOS in Library and Information Sciences

- Thesauri
  - Indexing and information retrieval
- Classifications
- Taxonomies

”Classical” Thesauri

Semantically arranged networks of terms/keywords

- Keywords are used for cataloging/indexing the meaning of contents in a standard way
- So that contents can found in later in information retrieval using the same thesaurus

Widely used in libraries, archives, museums etc.

- *Library of Congress Subject Headings (LCSH)*
- *General Finnish thesaurus YSA, MASA, MUSA, Allärs, …*

Based often on the following relations

• BT	Broader term
NT	Narrower term
RT	Related term
USE	“See” for a recommended term
UF	Used for; opposite of USE
SN	Scope note for definitions etc.

Notice: these relations correspond to those used in terminology

Thesauri are based on ISO standards

Thesaurus Example

Banks
NT Deposits
NT Investments
NT Loans

Business [loans]
BT Loans

Deposits
BT Banks
RT Investments
RT Loans

Home equity [mortgage]
BT Mortgage

Investments
BT Banks
RT Deposits
RT Loans

Loans
BT Banks
RT Deposits
RT Investments
NT Business
NT Personal
NT Mortgage

Mortgage [loans]
BT Loans
NT Home equity
NT Purchase
NT Vacation residence

Purchase [mortgage]
BT Mortgage

Vacation residence [mortgage]
BT Mortgage

Semantic Limitations of Thesauri

Meaning of relations is often unclear

- BT/NT is used for super/subordinate, but also for part-of
- RT has lots of different interpretations
  - *Cause/effect, tool/product, ...*
  - *Similarly as associative relations in terminology*

More explicit semantics are needed for computers

- E.g., delimiting characteristics are implicit
- Semantics vague

However, traditional thesauri are still useful resources!

Example: Thesaurus Limitations for Term Expansion

- Furniture
  NT Mirrors
  Mirrors
  NT Makeup mirrors
- OK, but the results of query ”Find all furniture” would contain also makeup mirrors in term expansion!

Classifications

Hierarchical systems for categorizing things:

- So that they can be found using the classification index

EXAMPLES:

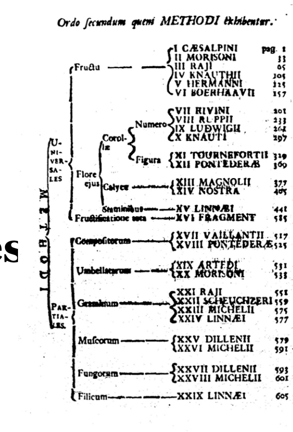
Library systems for organizing publications in shelves

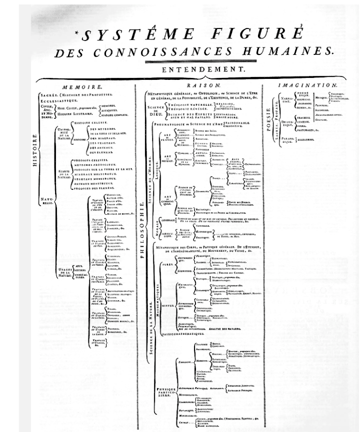
- E.g., Dewey Decimal Classification, UDK (in Finland)

Encyclopedias for organizing everything

- E.g., Diderot’s Encyclopedia

Your file system on computer





Classifications organize things but do not define their meaning

Classifications vs. Ontologies

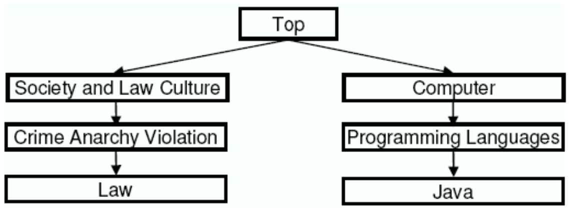


Fig. 1. An example of a classification (part of the Yahoo web directory).

Table 1. Comparison between classification schemes and ontologies

Category	Classification Schemes	Ontologies
Purpose	Organization of (large) document collections	Modeling of a domain
Language	Natural language, e.g. English	Formal language, e.g. OWL
Nodes	Usually represent complex concepts or individuals	Usually represent atomic concepts
Edges	Do not have well defined semantics	Have well defined semantics
Instances	Are not necessarily instances of the class to which they belong	Are instances of the class to which they belong
Users	Humans	Machines
Examples	DDC, LCC, UDC, etc.	MeSH ontology, Gene ontology[a], OpenCyc ontology[b], etc.

[a] http://www.geneontology.org/
[b] http://www.opencyc.org/

(Giunchiglia et al., 2008)

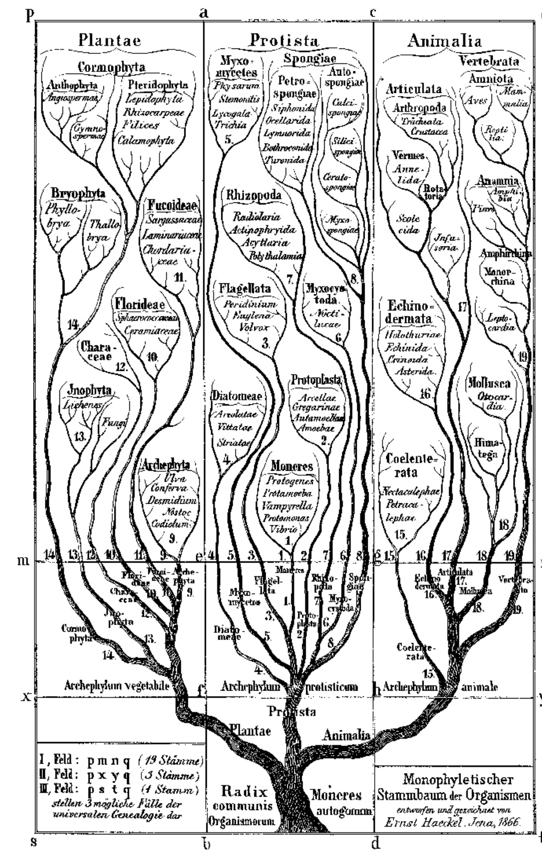
Taxonomies

Same idea as in classifications

- Systematic categorization of things

Term often used for classifying living organism

- Animals, plants, micro organisms
- Taxonomy = scientific discipline since 18th century
  - Initiated by Carl Linnaeus (1707-1778) in Sweden



Faceted Classification

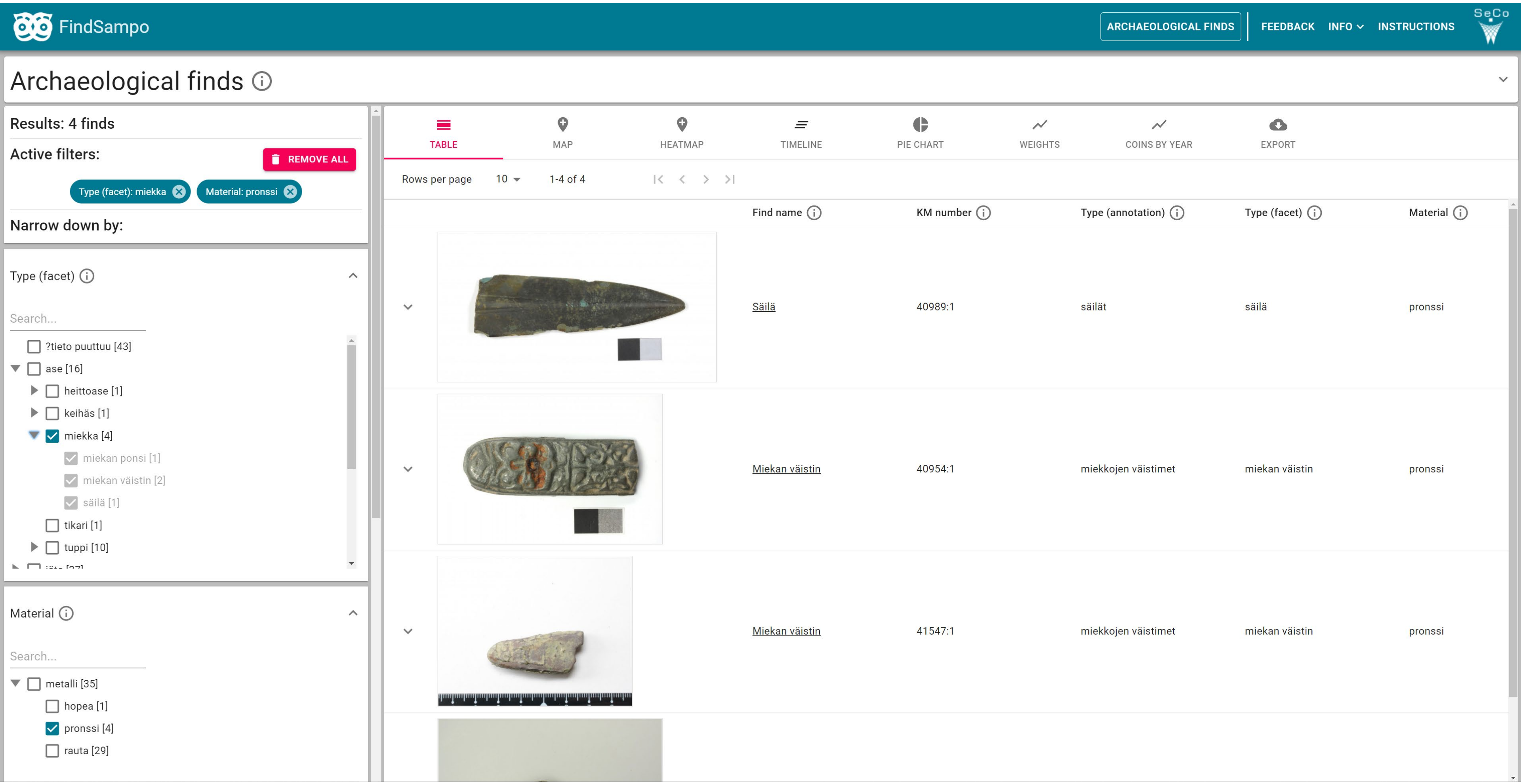
Things cannot often be classified along only one classification

- There would be too many categories
- E.g., a book can be at the same time about history, geography etc., is published at some time, is written in a language, targeted to children etc.

In faceted classification things are classified along several orthogonal classifications

- Idea developed by S. R. Ranganathan in the 1930's
- Faceted Search can be used for information retrieval
  - Making category selections on different facets in free order
  - Counting the number of hits for next selections to avoid dead ends

Example: Faceted Search in FindSampo



Computer Science Perspective to KOS: Ontologies

What is an ontology in Computer Science?

Ontology describes:

- The concepts/objects of the application domain
- The vocabulary used for referring to them

“An ontology is a formal, explicit specification of a shared conceptualization.” (Studer et al., 1998; based on Borst, 1997 and Gruber, 1993)

- Formal: well-defined syntax and semantics
- Explicit: can be represented and processed algorithmically, machine-understandability
- Shared: agreed upon in a community, facilitates communication
- Conceptualization: presents a model of the real world

Components

- Concept definitions and relations: for machines to understand
- Terminology: for humans to understand

A requirement for humans and machines to understand each other

History of Ontologies in CS

Information Systems

- 1967 G. H. Mealy
  - *Relating data with the real world*
- Object-oriented programming
  - *The main paradigm in programming practice since the 90's*

Artificial Intelligence

- Since 60's
- Natural language understanding research
- Knowledge representation research

  = *Logic* + *Ontologies* + *Computations*

WWW and the Semantic Web

- Since late 90's
- Ontologies are a key ingredient of the Semantic Web

Summary

- KOSs have been discussed and created by
  - *Philosophers*
  - *Linguists*
  - *Terminologists*
  - *Library and information scientists*
  - *Computer scientists*
- The semantic web is a rapidly developing application domain where ontologies are studied and used

Questions
More Information

Ontology as an Area of Philosophy

- https://en.wikipedia.org/wiki/Ontology

Ontologies in Computer and Information Sciences

- https://en.wikipedia.org/wiki/Ontology_(information_science)