Aalto University
School of Science

SeCo

Ontologies

CS-E4410 Semantic Web, 30.01.2019

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

Outline

Four perspectives of ontology

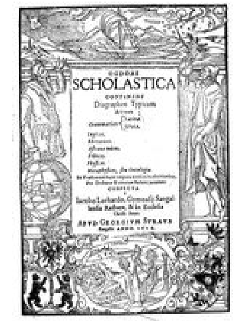
- Philosophy
- Linguistics
- Terminology
- Computer science

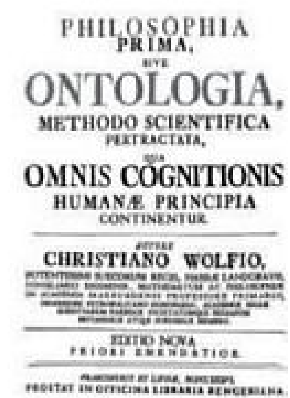
Two Semantic Web standards for representing ontologies (in addition to RDFS)

- Simple Knowledge Organization System SKOS (latter part of today's lecture)
  - *Standard vocabulary for representing simple ontologies*
- Web Ontology Language OWL (next lecture)
  - *Rich and complex knowledge system based on description logic*

Perspectives of ontology: Philosophy

- Study of the essence of Being
  - *Apart from the particular existing things*
- Plato’s world of ideas: metaphysics
- Aristotle’s (384–322 B.C.) 10 Categories
- Medieval logicians: first semantic net
  - *Genus (supertype) vs. species (subtype)*
- “Ontology” as a discipline
  - *R. Göckel, J. Lorhard, 1613*
  - *Kant (1787), Peirce, Husserl, Whitehead, Heidegger, …*
- Today: theoretical bias
  - *Foundational categories & logic behind everything*


SCHOLASTICA


PHILOSOPHIA
PRIMA,
SIVE
ONTOLOGIA,
METHODO SCIENTIFICA
PERTRACTATA,
QUA
OMNIS COGNITIONIS
HUMANÆ PRINCIPIA
CONTINENTUR.
AUCTORE
CHRISTIANO WOLFIO,
EDITIO NOVA
PRIORI EMENDATIOR.

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

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

(Sowa, 2004)

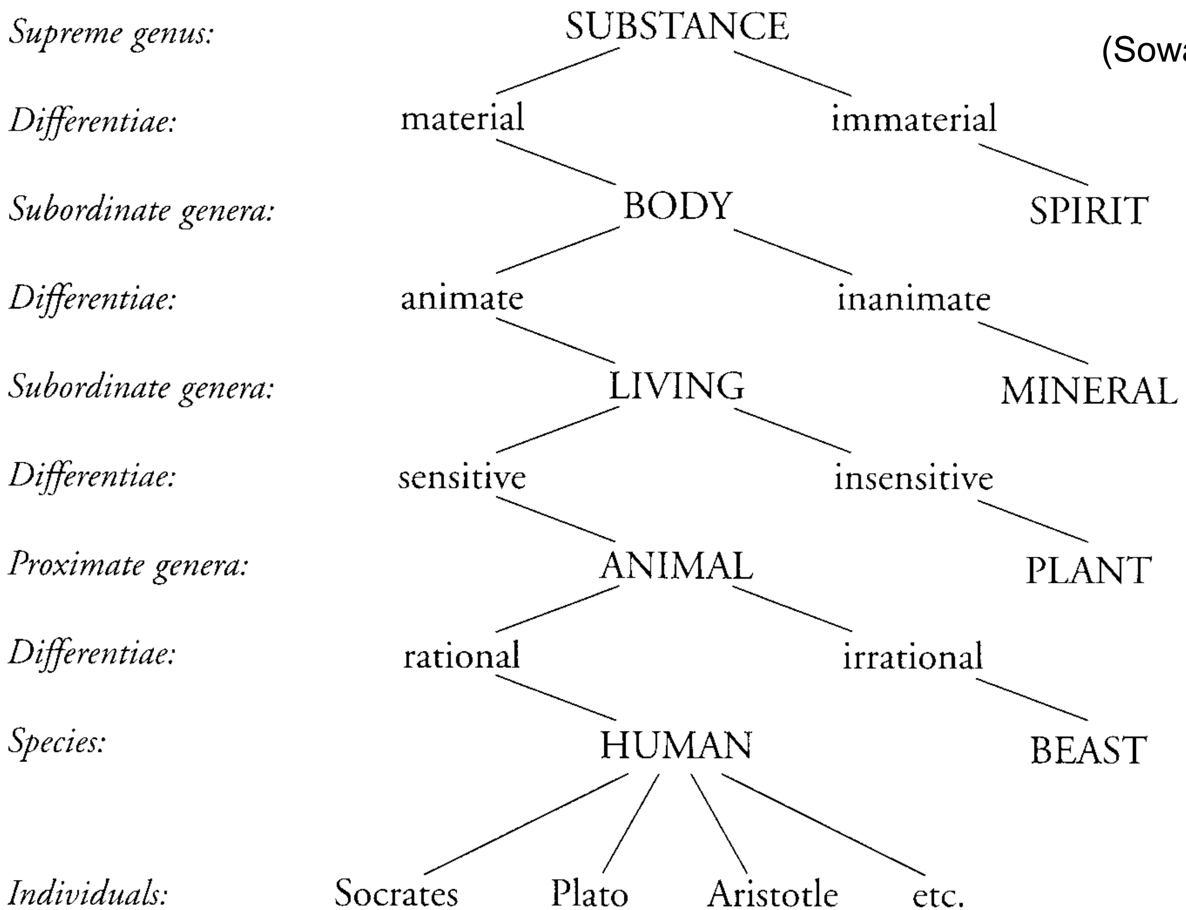


FIGURE 1.1 Tree of Porphyry, translated from a version by Peter of Spain (1239)

Aristotle's Syllogisms

- 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:

(Sowa, 2004)

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.

Formal ontology

Branch of philosophy

- 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
- ...

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

- Interoperability through shared principles

Perspectives of ontology: Linguistics

- Meanings vs. words
- Peter Mark Roget's Thesaurus 1852
  - *Tool for analysis and classification of ideas that helps human communication*
- Terminology: dictionaries & vocabularies
  - *Concept analysis*
- Thesauri
  - *Indexing/classifying/retrieving data*
  - *Language translation*
- Semantic thesauri
  - *WordNet, EDR (Electronic Dictionary Research), …*

Roget’s Thesaurus

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
  - *Only for human interpretation*

Roget's Thesaurus: example

Top level: 6 classes

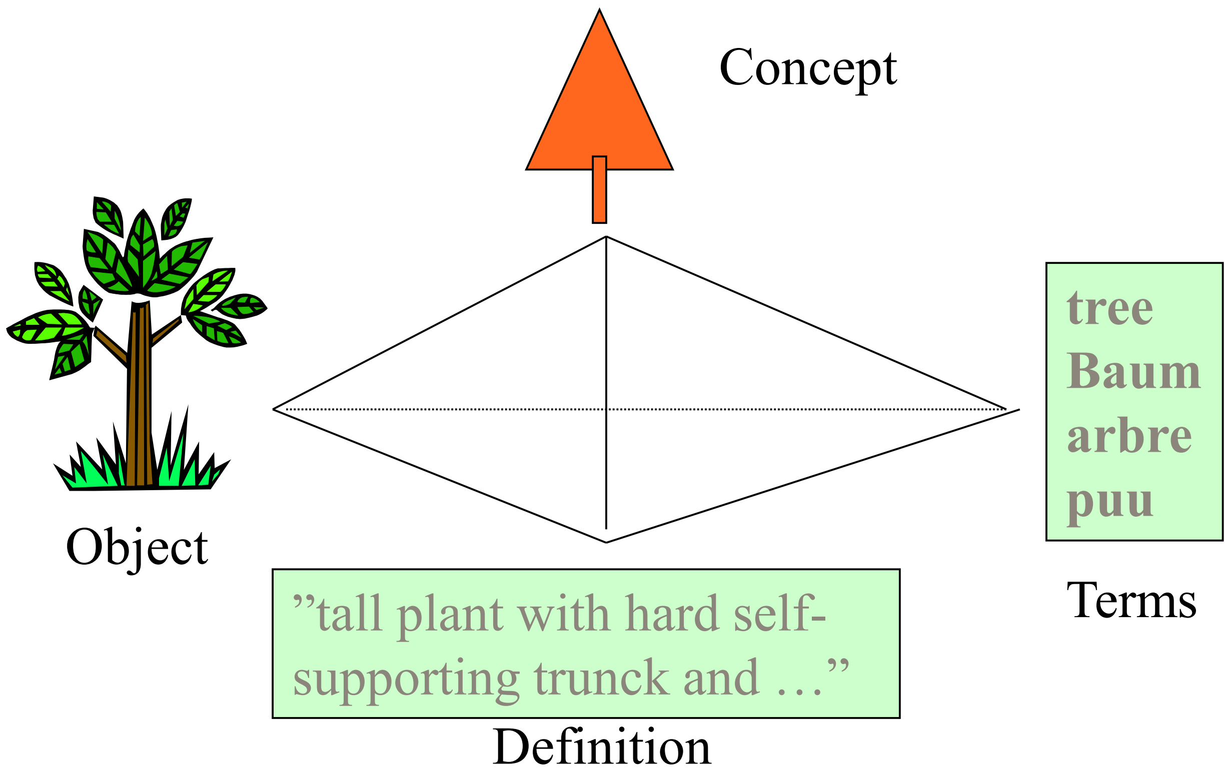
- *1. Abstract relations*
- *2. Space*
- *3. Matter*
- *4. Intellect*
- *5. Volition*
- *6. Affections*

```
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  …
      …
```

Perspectives of ontology: Terminology

- Based on **concept analysis** of word meanings
- Standardized methodology
  - *Various ISO standards*
- Well-defined vocabularies for human users, not for machines
  - *In Finland: e.g., The Finnish Terminology Centre TSK*
- Following presentation is based on:
  - *Heidi Suonuuti: "Guide to terminology", 2001*

Concept analysis:
Extended Odgen-Richards triangle to tetraed



Concept vs. term

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"*

Term definitions

Characteristics (properties)

- Delimiting characteristics differentiate concepts
  - *E.g., concept "tree":*
    - "have a root" not delimiting
    - "have a self-supporting trunk" delimiting

Intensional and extensional definitions

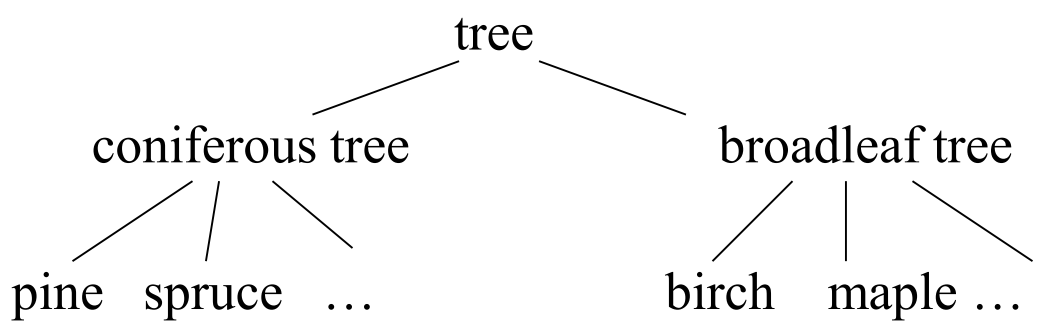
- Intension = sum of general characteristics
  - *Describe only delimiting characteristics*
  - *Other characteristics come from the hierarchy*
    - Tree = "long-living **plant**, have a self-supporting trunk, …"
- Extensional = list of objects
  - *Tree = {pine, maple, spruce, …}*

Concept systems

- Concepts are not independent
- Relations between concepts
  - *Generic (hyponymy)*
  - *Partitive (meronymy)*
  - *Associative*

Generic relation (hyponymy)

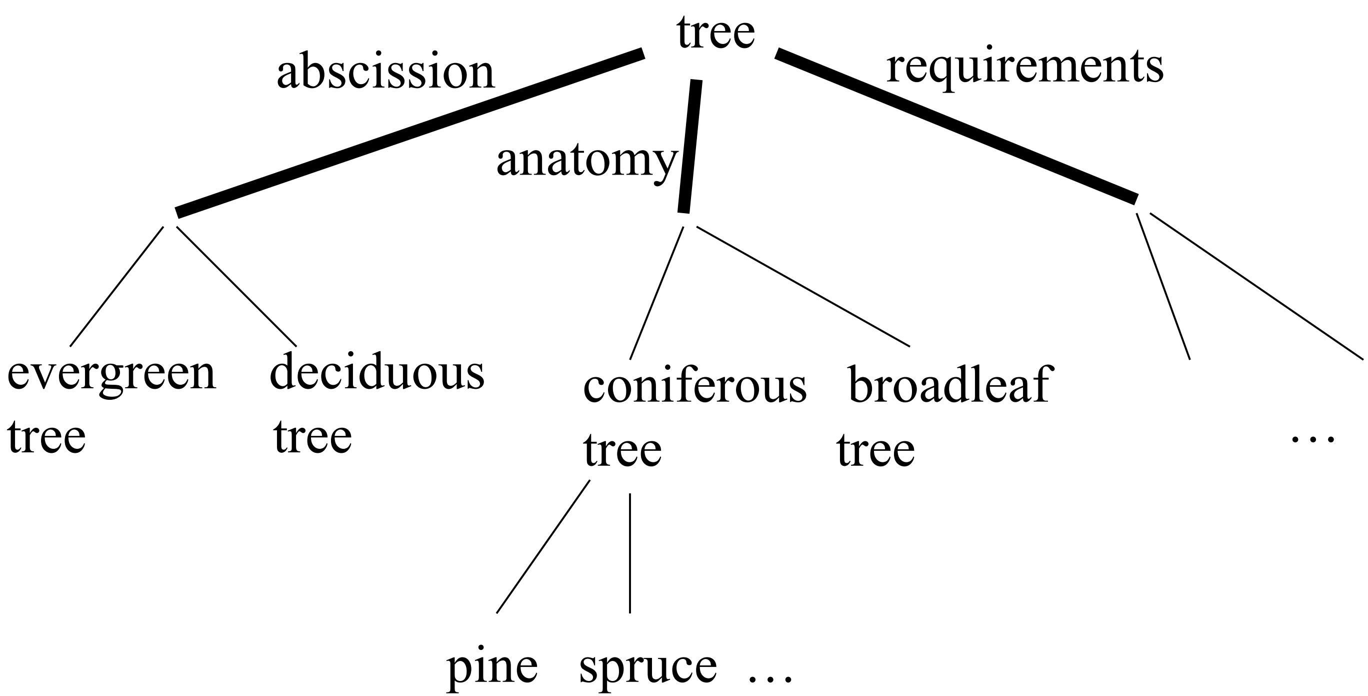
- Concepts share general characteristics but one
- Concept hierarchy: super/subordinate



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

Representing parallel independent subdivisions

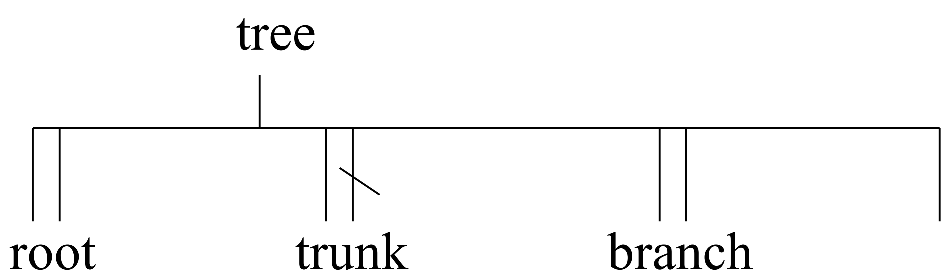
Using three subdivision dimensions:



Partitive relation (meronymy)

Part-whole relation

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



- Also along different dimensions
  - *Tree -> permanent vs. non-permanent organs*

Different meronymy relations

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

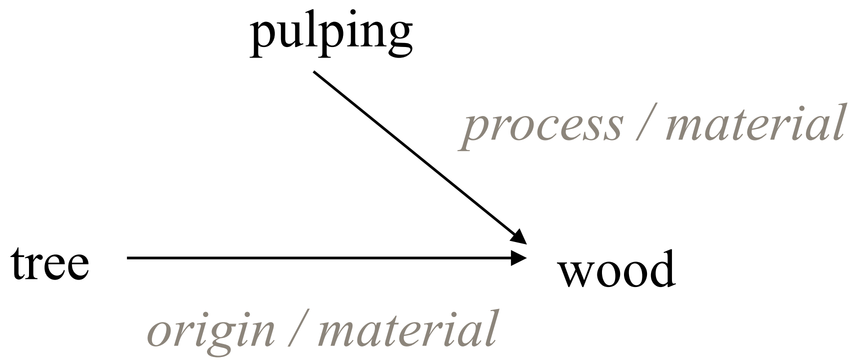
(C. Fellbaum, 1998)

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…

Associative relations

Arrow notation



Why terminology?

- Provides useful tools and ideas for concept analysis and 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
  - *Methodology is useful there, too*

”Classical” thesauri

Semantically arranged terminologies/dictionaries

- Terms may be a mixture of words and concepts
- Studied in library and information science

Based traditionally on the following relations

BT	Broader term
NT	Narrower term
RT	Related term
USE	“See”
UF	Used for; opposite of USE
SN	Scope note

E.g., in Finland:

- General Finnish thesaurus YSA, MASA, MUSA, Allärs, …

Based on ISO standards

Use of thesauri

Used especially for

- Indexing information content (keywords)
- Information retrieval
  - *Keyword search*
  - *Term expansion: "tree" -> "pine", "birch",…*

Widely used, lots on indexed data

- E.g., libraries, museums, archives, …

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

Limitations

Meaning of relations

- BT/NT for sub/superordinate, part-of etc.
- RT has lots of different interpretations
  - *Cause/effect, tool/product, …*
  - *Similarly as associative relation in terminology*

Not formal enough for computers

- E.g., delimiting characteristics implicit
- Semantics vague
- Are still useful

Example: Thesaurus vs. ontology

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

Ontology

– Computer Science Perspective

Aalto University
School of Science

Department of
Computer Science

SeCo

History of ontologies in CS

Information Systems

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

Artificial Intelligence

- Since 60's
- Natural language understanding
- Knowledge representation
  - *= Logic + Ontologies + Computations*

WWW and the Semantic Web

- Since late 90's

What is an ontology?

“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

Ontology describes:

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

Components

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

A requirement for humans and machines to understand each other

Vocabularies vs. ontologies

”Vocabulary” often refers to a (small) metadata model, by which a particular domain can be represented

- RDF Schema
- Dublin Core
- Friend of a Friend (FOAF)

”(Domain) ontology” often refers to a (large) knowledge structure representing an application domain

- Keyword thesauri
- Place gazetteer
- Authority file system (for persons, organizations)
- ”Domain ontologies” are created using ”vocabularies”

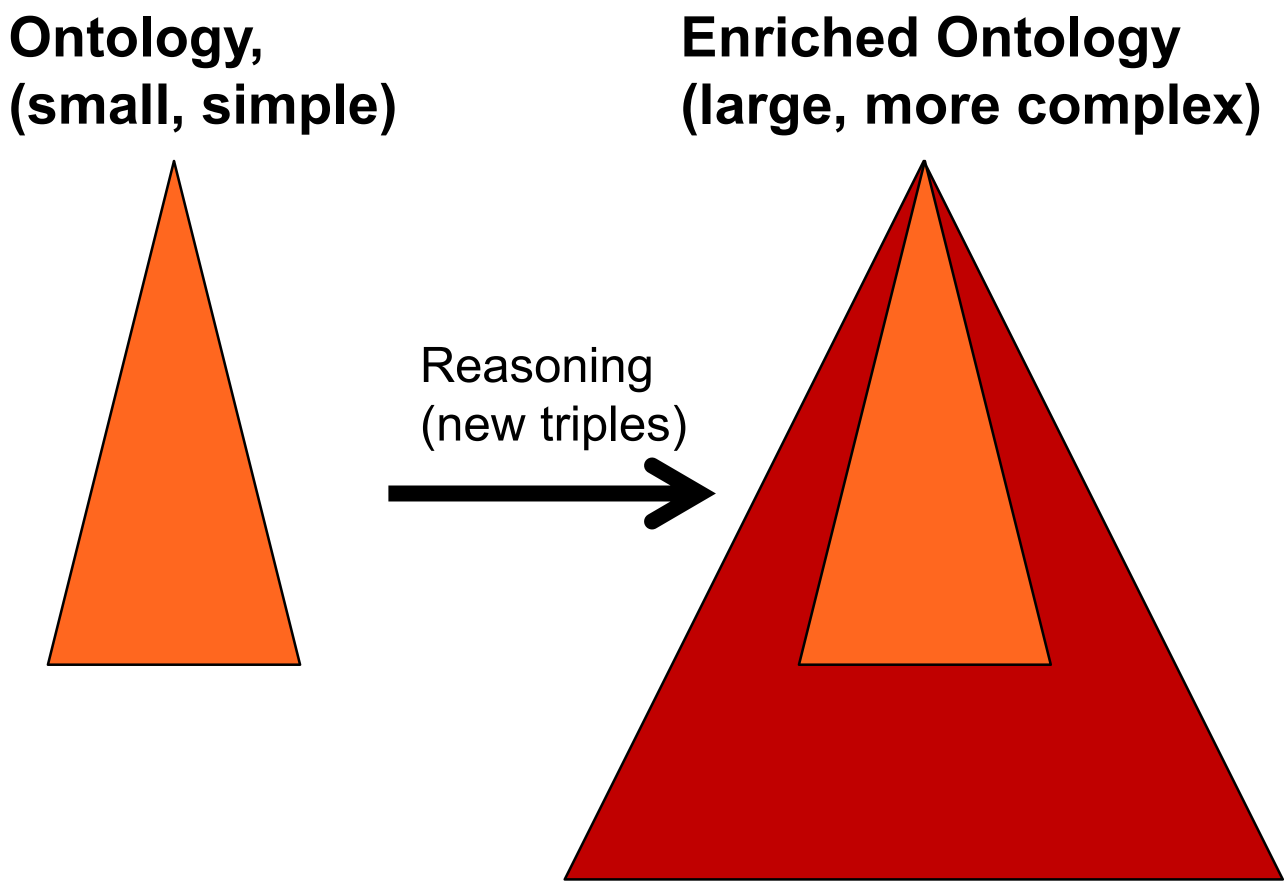
However, the terms may be mixed …

Components of an ontology

Core components

- Classes (types, kinds, sorts, categories, ...)
- Individuals (instances, objects)
- Attributes (properties, slots)
- Relations (properties, slots)
- Rules and axioms

Enriching ontology by reasoning



Why use ontologies?

- Semantic interoperability based on sharing concepts
- Benefits of reasoning
  - *Ontology development*
    - Deriving implicit subsumption hierarchies
    - Consistency checking possible
    - Ontology matching and merging
  - *Information retrieval*
    - Query expansion
    - Intelligence: finding implicit results based on reasoning
      - *Semantic search methods and recommending*
    - Using ontology structures to aid query formation and presenting results
    - Semantic analysis

Major domain ontology types

General concepts

- Used, e.g., as subject keywords, representing object types
- E.g., table, studying, philosophy, …

Actors

- Persons, groups, organizations
- Individuals with rich metadata
- E.g., Napoleon I, impressionists, Nokia Corp.

Places

- Points, paths, areas with geolocation
- Individuals with rich metadata
- Senate Square, Ring II, Finland

Events

- Glues Linked Data in chains and hierarchies
- Birth of Jesus, World War II

Classifications vs. Ontologies

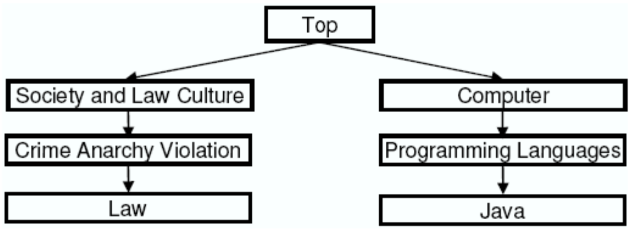


Fig. 1. An example of a classification (part of the Yahoo web directory).

(Giunchiglia et al., 2008)

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/

Summary

- Ontological systems have been discussed and created by
  - *Philosophers*
  - *Linguists*
  - *Terminologists*
  - *Computer scientists*
- Models and methodologies developed in CS are created for the usage of computers, too
- The semantic web is a rapidly developing new application domain where ontologies are used
- Further reading: Noy, McGuinness: Ontology Development 101: A Guide to Developing Your First Ontology, Stanford University, 2001.