Laboratory for Applied Ontology

Institute of Cognitive Science and Technology
Italian National Research Council

Classifying (Medical) Ontologies

Stefano Borgo

Laboratory for Applied Ontology (LOA)

Institute for Cognitive Sciences and Technology (ISTC-CNR)

Trento-Roma, Italy

www.loa-cnr.it

Classification: a relative notion…

A triviality:

ontologies are complex artifacts

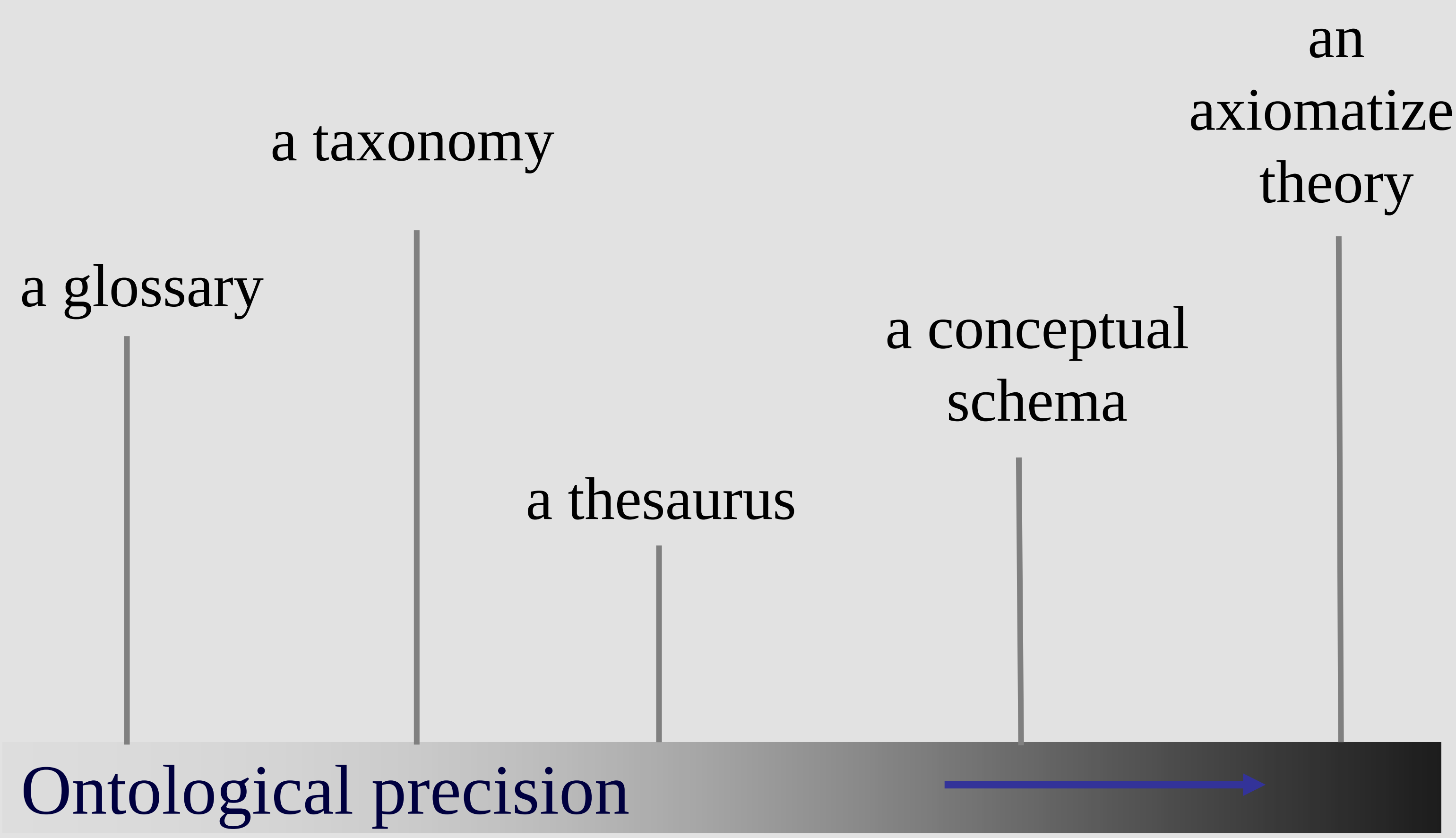
Consequence:

ontologies may differ in several aspects

- Formalism (taxonomy, frame, axioms, conceptual graphs…)
- Purpose (retrieval, NLP, sharing, modeling…)
- Domain (management, learning, medicine, foundations of…)
- Construction (top-down, bottom-up, middle, merging…)
- Complexity (tangledness, splitting, depth…)
- Coverage, Implementation, Size, Motivations,…

…**Precision**

The “PRECISION” axis



What do you mean by “ontology”?

- There are different strategies to provide knowledge structures (*engineering artifacts*) suitable to organize information
- Strategies depend on the application use and correspond to different meanings for the term “ontology”.

NOTE: we focus on structures for content, thus we avoid discussing languages, markup languages, indexing, content management, implementations and the like

Ontologies can be roughly divided in four groups:

- Non-ontologies (I’m not kidding)
- Linguistic (terminological) ontologies
- Implementation driven ontologies
- Formal ontologies

Linguistic ontologies (1/2)

- **Glossaries**: scattered lists of terms with glosses in natural language.

  Formally, a glossary is a ***(Labeled) Set***
  (elements are "defined" in natural language).

- **Controlled vocabulary**: collection of terms that have been enumerated explicitly by a *registration authority*. In theory, all terms in the list are unambiguously defined (not true in practice). Requirement: any ambiguous term has different instance-names to distinguish the different meanings it refers to. If several terms are used to mean the same concept, one is identified as preferred (the others are synonyms).

  Formally, a controlled vocabulary is a ***Set of 1-Trees***
  (set of trees of depth at most 1, only one edge-label, elements are defined in NL).

Linguistic ontologies (2/2)

- **Taxonomies**: a controlled vocabulary organized into a hierarchical structure. There might be more then one parent-child relationship in a taxonomy (es. whole-part, broader-narrower, genus-species, type-instance). In some cases, a term can have multiple parents so the term can occur in different places of the taxonomy (however, it must have the same children everywhere).

  Formally, a taxonomy is a complex (***Label-restricted***) ***Set of Dags***
  (set of fully labelled dags of unconstrained depth)

- **Thesauri**: these are taxonomies coupled with equivalence/association relations (generally: synonym of, related to, similar to, and so on). The number of relations may vary but it is anyway quite small (<20). It is the most complex type of controlled vocabulary.

  Formally, a thesaurus is a (***Label-restricted***) ***Multi-graph***
  (set of fully labeled graphs, each edge-label isolates a set of graphs, edge-labels are *more or less* fixed).

Non-ontologies *called* ontologies (1/1)

- **Catalogs:** a catalog is simply a set of terms, that is, it provides no constraint (formal or informal) to characterize their meaning.

    Formally, a catalog is a pure ***Labeled Set***.

    (it weakens glossaries by dropping the glosses)

- **Topic Maps:** An ISO standard for describing knowledge structures and associating them with information resources. The topics, associations, and occurrences that comprise topic maps allow them to describe informally complex structures. Topic Maps are centralized (all information is contained in the map). Note that anything (an object, a feature, a role, a concept) can be a topic.

    Formally, a topic map is a ***(nested) Hyper-graph***
    (both nodes and edges have zero or more labels; any string of characters, sound, icon,… can be a label)

    (it weakens thesauri by using unrestricted (edge-)labels and "undefined" n-ary relations)

Implementation driven ontologies (1/2)

- **Conceptual Schema**: Set of terms, attributes and relations with explicit descriptions (definitions), rules for their use, and perhaps cardinality constraints. Differently from linguistic ontologies, the set of attributes and relations is not fixed to a (more or less) given list, the choice depends on the modeler and the purpose of the ontology. Indeed, the main task is to guarantee data consistency and this drives the introduction of constraints.

  Formally, a conceptual schema is a full ***Hyper-graph*** (set of fully labeled graphs, all labels are “defined”).

Implementation driven ontologies (2/2)

- **Knowledge Bases:** Formal systems that captures the meaning of the adopted vocabulary via logical formulas. A KB is considerably richer than a conceptual schema since the underlying languages are more expressive. The purpose is not simply retrieval (for which frames suffice) but reasoning. However, the main task is still data consistency. The classical distinction between terminological part (T-box) and assertional part (A-box) can be taken as a distinction between the ontology adopted by the system and the data classified by the system.

Formally, a knowledge base is a ***Logic theory***

(it is not possible to characterize it within the graph terminology).

Formal ontologies: the notion

The usual intuition of an ontology as "*a specification of a conceptualization of a knowledge domain*" spans the systems we have seen from glossaries to KBs (and beyond).

Formal ontology deepens this intuition requiring a clear semantics for the language, clear motivations for the adopted distinctions as well as strict rules about how to specify terms and relationships.

This is obtained by relying on ontological analysis (in the philosophical sense) and by using formal logic (usually DL up to subsets of HOL) where the meaning of the terms is guaranteed by formal semantics.

The complexity of a representation system splits into two distinct aspects:

- the organization of knowledge structure and
- the specific information for an application domain.

Formal ontologies look at the first issue only.

Formal ontologies (1/1)

- **Domain ontologies**: these are formal ontologies that focus on an application area (i.e., enterprise modeling, anatomy, astrophysics, etc.) The *purpose* is to provide a basic, stable and unambiguous description of concepts, entities and relations used in such a domain.

- **Core (reference) ontologies**: these are formal ontologies that furnish the organization of top-level (general) concepts used in (or across) some communities and application areas.
The *purpose* is to facilitate reliable exchange of information within those groups.

- **Foundational ontologies**: these are the most general formal ontologies. They deal with very general and basic terms like entity, event, process, spatial and temporal location, part-of, quality-of, participation and the like. The *purpose* of these ontologies is to characterize entities and relations that are common in all domains and to provide a consistent and unifying view.

"ontology" is used referring to…

Linguistic ontology

- Glossary
- Controlled vocabulary
- Taxonomy
- Thesaurus

But don't get surprised if you find someone calling ontology a catalog or a topic map.

Implementation driven ontology

- Conceptual Schema
- Knowledge Base

Formal ontology

- Domain ontology
- Core (reference) ontology
- Foundational ontology

Examples of medical ontologies (1/4)

MeSH: the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. Descriptors are arranged in both an alphabetic and a hierarchical structure. At the top level there are broad headings such as "Anatomy", "Organisms", "Diseases" and "Mental Disorders." The hierarchy is a forest with 15 heads and depth 11, at the bottom descriptors like "Ankle" and "Conduct Disorder" for a total of 22,568 descriptors. In addition, there are more about 200,000 headings called Supplementary Concept Records within a separate thesaurus.

There are also thousands of cross-references.

It is organized in a branching structure (tree).

Each descriptor may appear in several places.

Examples of medical ontologies (2/4)

UMLS: the Metathesaurus contains over 1 million biomedical concepts (definitions) and 2.8 million concept names from more than 100 controlled vocabularies used in patient records, administrative data,full-text databases and expert systems.

a) preserves the information (names, meanings, hierarchical contexts, attributes, and inter-term relationships present in its source vocabularies);

b) adds certain basic information to each concept; and

c) organized by concept or meaning. Alternative names for the same concept (synonyms, lexical variants, and translations) are linked together. It defines preferred terms.

d) the Is_A relation defines the main hierarchy. There is also a set of non-hierarchical relationships, which are grouped into five major categories: `physically related to,' `spatially related to,' `temporally related to,' `functionally related to,' and `conceptually related to.'

e) no automatic way to check inconsistences.

Nota: UMLS M. might contain cycles, undetected sibling concepts and polysemes plus other similar problems. We look at the general picture assuming UMLS has been (or could be) cleaned up.

Examples of medical ontologies (3/4)

Galen: (well, we have heard a lot about it already….) it provides language, terminology, and coding services for clinical applications (the aim is to store detailed clinical information about patients). The Common Reference Model of clinical terminology is an ontology in the formal sense and provides an application independent view of clinical terminology based on a description logic (GRAIL).
The GALEN model provides taxonomies which contain thousands of categories in a complex hierarchy.

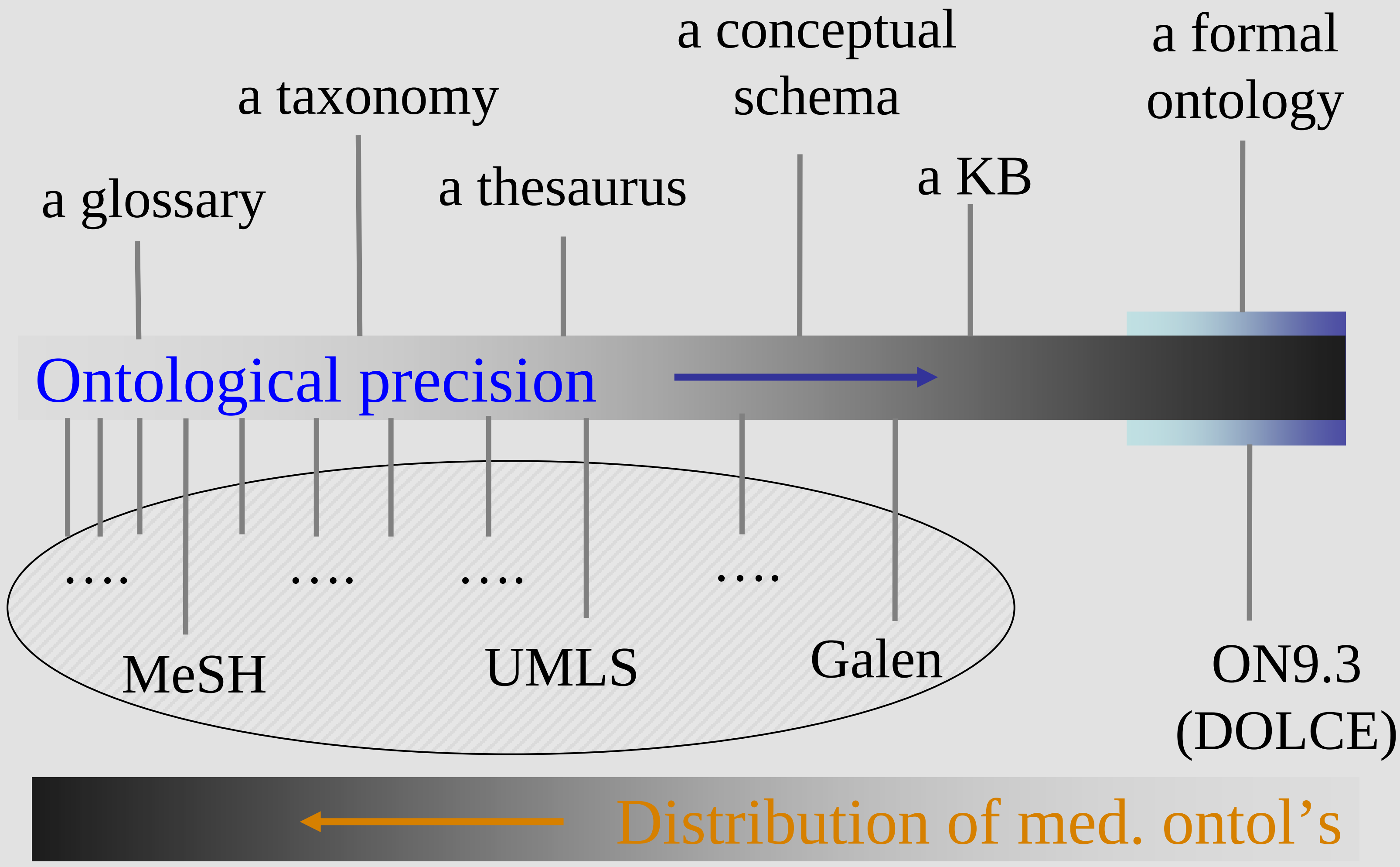
Examples of medical ontologies (4/4)

On9.3: it provides

1. a library of generic ontologies,
2. an integrated medical ontology (IMO) that integrates five medical top-levels (ICD10, UMLS, GALEN, SNOMED, GMN) providing relative mappings among the systems,
3. a formalized representation of some medical repositories and their classification within IMO

ON9.3 is attached to the DOLCE foundational ontology and thus it inherits its structure with the formal characterization of the basic notions and relations.

The resulting decorated PRECISION axis



DOLCE

a Descriptive Ontology for Linguistic and Cognitive Engineering

- **Strong cognitive bias*: descriptive* (as opposite to *prescriptive*) attitude**
- **Emphasis on *cognitive invariants***
- **Categories as *conceptual containers*: no "deep" metaphysical implications wrt "true" reality**
- **Clear *branching points* to allow easy comparison with different ontological options**
- *Rich axiomatization*
  - ***37 basic categories***
  - ***7 basic relations***
  - ***80 axioms, 100 definitions, 20 theorems***

Formal Ontological Analysis

- Theory of Parts
- Theory of Wholes
- Theory of Essence and Identity
- Theory of Dependence
- Theory of Qualities
- Theory of Composition and Constitution
- Theory of Participation
- Theory of Description

DOLCE's basic taxonomy

Endurant

- ***Physical***
  - **Amount of matter**
  - **Physical object**
  - **Feature**
- ***Non-Physical***
  - **Mental object**
  - **Social object**
- …

Perdurant

- ***Static***
  - **State**
  - **Process**
- ***Dynamic***
  - **Achievement**
  - **Accomplishment**

Quality

- ***Physical Qs***
  - **Spatial location**
  - …
- ***Temporal Qs***
  - **Temporal location**
  - …
- ***Abstract Qs***
  - …

Abstract

- ***Quality region***
  - **Time region**
  - **Space region**
  - **Color region**
  - …
- …

Core Ontologies
(applications of DOLCE using D&S, and OntoWordNet)

- Core ontology of biomedical terminologies (cf. UMLS)
- Core ontology of plans, task, and guidelines
- Core ontology of (Web) services
- Core ontology of service-level agreements
- Core ontology of transactions (bank, anti-money-laundering)
- Core ontology for the Italian legal lexicon
- Core ontology of regulatory compliance
- Core ontology of fishery (FAO's Agriculture Ontology Service)

FOIS-2004

International Conference on Formal Ontology in Information Systems

http://www.fois.org

Picture courtesy of Fototeca Web del Comune di Torino

November 4-6, 2004, Torino (Italy)

Thank you