Ontologies vs. classification systems

Bodil Nistrup Madsen
Copenhagen Business School
Copenhagen, Denmark
`bnm.isv@cbs.dk`

Hanne Erdman Thomsen
Copenhagen Business School
Copenhagen, Denmark
`het.isv@cbs.dk`

Abstract

What is an ontology compared to a classification system? Is a taxonomy a kind of classification system or a kind of ontology? These are questions that we meet when working with people from industry and public authorities, who need methods and tools for concept clarification, for developing meta data sets or for obtaining advanced search facilities. In this paper we will present an attempt at answering these questions. We will give a presentation of various types of ontologies and briefly introduce terminological ontologies. Furthermore we will argue that classification systems, e.g. product classification systems and meta data taxonomies, should be based on ontologies.

1 Introduction

In recent years many authors have discussed the nature of ontologies and proposed various definitions and subtypes of ontologies for various purposes, among them Gruber (2007), Guarino (1998), Gómez-Pérez et al. (2004). According to

Figure 1: Some concepts related to knowledge structuring.

CEN (2004) ontologies and taxonomies are types of knowledge structuring, as shown in Figure 1.

The ontology in Figure 1 comprises concepts (boxes with systematic notations) and subdivision criteria (boxes with text in capital letters). The concepts are related by means of type relations (lines between the concept boxes) and further described by means of feature specifications each consisting of an attribute-value pair (e.g. PURPOSE: knowledge representation).

According to the ontology in Figure 1 one may distinguish *models* and *classification systems* as follows: The purpose of a model is to *give a simplified representation of knowledge about phenomena*, whereas the purpose of a classification system is *the subdivision of phenomena into classes that form the basis for ordering 'things'*.

Very often a conceptual data model, represented by means of an ER diagram or an UML diagram, is referred to as ontology. Our recommendation is to use the term ontology only as defined here.

2 Various types of ontologies

In 2007, ISO Technical Committee 37, *Terminology and Other Language Resources* (ISO TC 37), set up an Ontology Task Force with the aim of proposing a strategy for the work on ontologies within TC 37. As a basis for this strategy, the Task Force will develop an overview of related ongoing projects, existing standards and proposals for future projects within TC 37 as well as an overview of examples of ontologies and projects 'outside' TC37. The first step in the work of the Ontology Task Force is to describe different types of knowledge representation resources, and to clarify the differences between these. One of the results is a systematic overview in the form of an ontology of ontologies which comprises proposals for definitions of the different types of ontology.
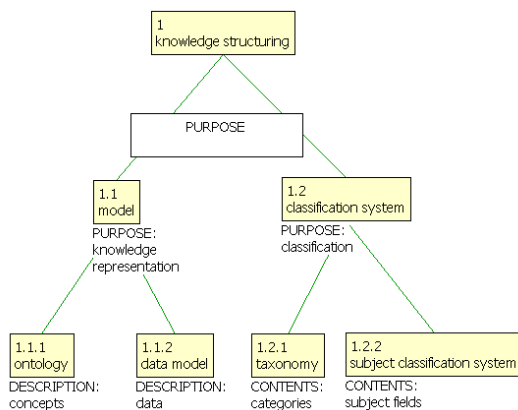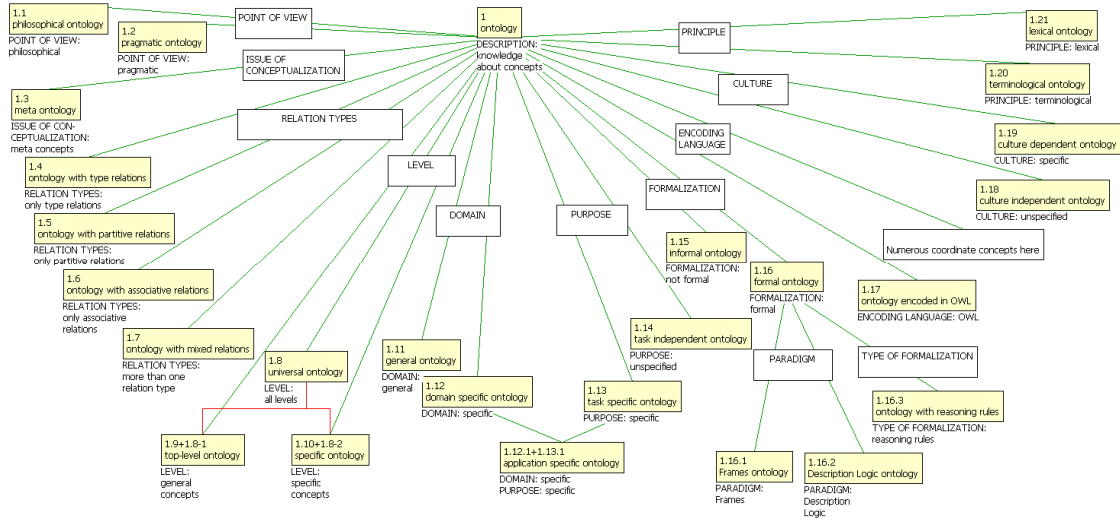
Figure 2: Ontology of ontologies.

Figure 2 presents this ontology of ontologies. The description of the concepts is to a great extent based on Guarino (1998). In this ontology characteristics and subdivision criteria are introduced that clearly distinguish the types of ontologies, e.g. LEVEL, DOMAIN and PURPOSE. The broken lines between concepts represent part-whole relations.

The ontology in Figure 2 may be characterized as a terminological ontology, i.e. an ontology that is based on the terminological method, making use of characteristics and subdivision criteria, cf. ISO 704 (2000).

A terminological ontology is a domain specific ontology. We use the term *terminological ontology* as a synonym of the term *concept system*, which is normally used in terminology work, cf. for example ISO 704 (2000). Gruber (2007) describes an ontology in the following way: *An ontology specifies a vocabulary with which to make assertions, which may be inputs or outputs of knowledge agents (such as a software program). … an ontology must be formulated in some representation language …* In our view, the demand for a representation language narrows the concept, i.e. Gruber's definition describes the concept *formal ontology* in Figure 2.

3 Ontologies as the basis for classification systems

As already mentioned, we distinguish *ontology* and *classification system* with respect to purpose. However, we strongly recommend that a classification system is built on the basis of a terminological ontology or by using the principles of terminological ontologies.

In the extract of the product classification system eCl@ss in Figure 3, it is evident that by using principles of terminological ontologies, this system could be structured in a more logical way, and thus could be intuitively easier to use: *automobile*, *aircraft*, *railborne vehicle* and *water vehicle* are distinguished with respect to "channel of transportation". For example *automobiles* are meant for traveling on streets or roads while *aircrafts* are designed to travel through the air. *Farming vehicles* and *hoisting, lifting vehicles* are characterized with respect to purpose. The order of the classes does not make this clear.
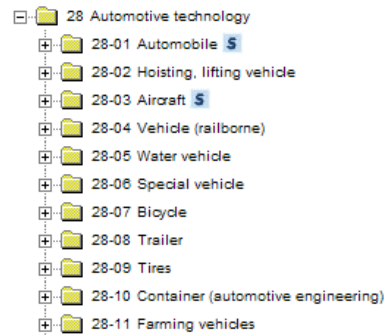

Figure 3: Extract of a product classification system.

Figure 4 presents an ontology with concepts corresponding to the classes in Figure 3. Since some of the classes in Figure 3 do not refer to automobiles, the top concept chosen is *vehicle*.
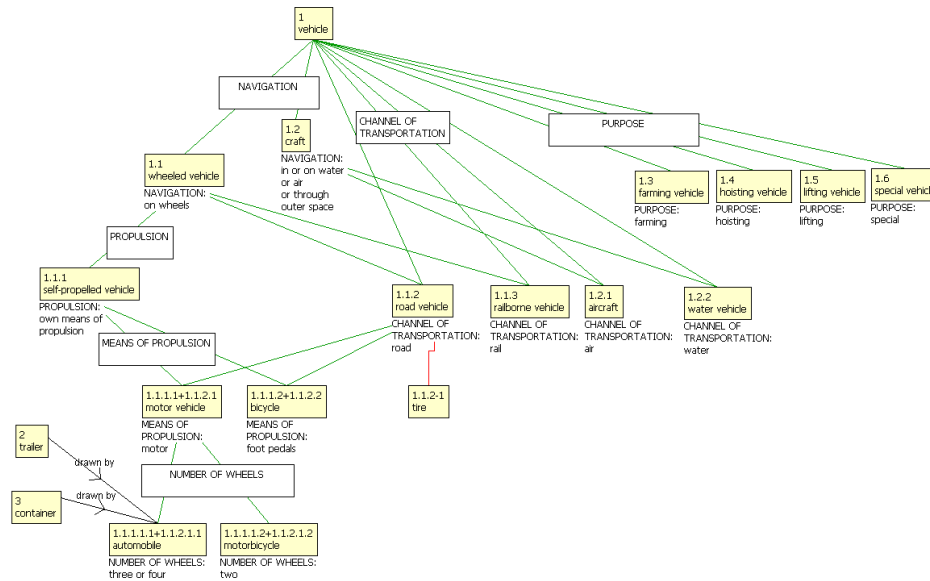
28

Figure 4: Ontology of vehicles.

In the ontology in Figure 4 the concepts are clearly delimited from each other by means of subdivision criteria: NAVIGATION, CHANNEL OF TRANSPORTATION, etc. It may be useful to introduce subdivision criteria also in a classification system in order to make this clear.

```
1              vehicle
1.1              wheeled vehicle
1.1.1              road vehicle
1.1.1-1              tire
1.1.1.1              motor vehicle
1.1.1.1.1              automobile
1.1.1.1.2              motorbicycle
1.1.1.2              bicycle
1.1.2              railborne vehicle
1.2              craft
1.2.1              aircraft
1.2.2              water vehicle
1.3            farming vehicle
1.4            hoisting vehicle
1.5            lifting vehicle
1.6            special vehicle
2            trailer
3            container
```

Figure 5: Extract of a classification system.

It is not intuitively understandable why the class *Bicycle* belongs to *Automotive technology* in Figure 3, but it may be because this class comprises *motor driven bicycles*. However, a closer look into the class *Bicycle*, reveals that the class also comprises the class *Bike*.

During the concept clarification process it turned out that there was a need for introducing the two concepts *wheeled vehicle* and *craft*

which were not in the classification in Figure 3. Based on the ontology in Figure 4, a classification list like the one in Figure 5 can be developed.

When building a classification system on the basis of an ontology, some simplifications will typically be made. In Figure 5 the concept *self-propelled vehicle,* which is a superordinate concept to *motor vehicle* and *bicycle,* is not found as a class. One may also consider to leave out the class *bicycle* for the above mentioned reasons.

As already mentioned, it may be useful to introduce subdivision criteria in order to make explicit the differences between the classes.

4 Classification systems compared to concept systems

A characteristic of a classification system is that the nodes are not always concepts, but often groups of concepts. This is true in the Semantic Types of UMLS (Unified Medical Language System), cf. Figure 6.

*The Semantic Network consists of (1) a set of broad subject categories, or Semantic Types, that provide a consistent categorization of all concepts represented in the UMLS Metathesaurus®, and (2) a set of useful and important relationships, or Semantic Relations, that exist between Semantic Type*s, cf. (Bodenreider, 2005) and the Semantic Network Fact Sheet (http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html).
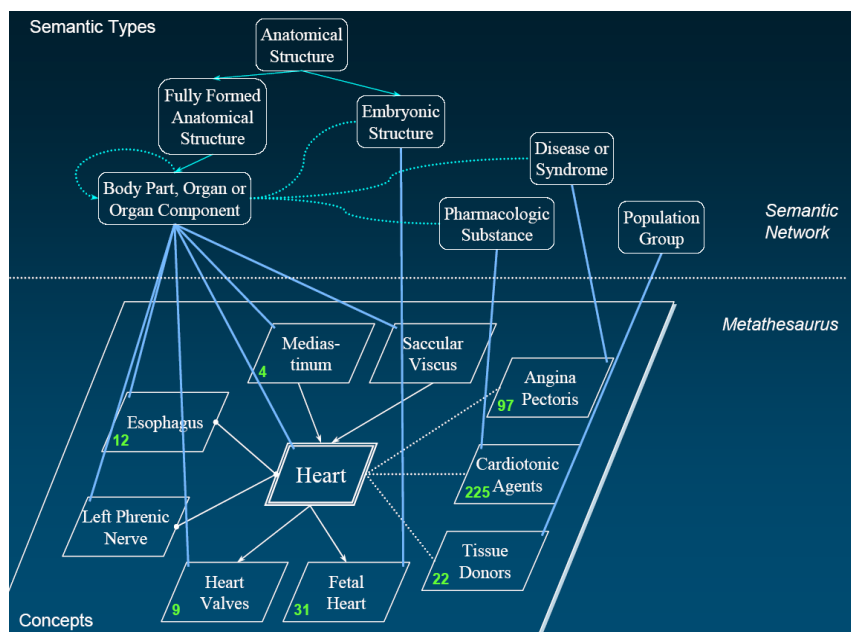
Figure 6: Example from UMLS.

An example of a semantic type is '*Body Part, Organ or Organ Component*', which conflates three concepts: *body part*, *organ* and *organ component*. In an ontology these three would be separate concepts (nodes).

5 Ontologies as the basis for meta data taxonomies

In order to facilitate data exchange and interoperability, it is important to be able to describe elements of data collections systematically and unambiguously. This is the reason why metadata registries comprising sets of metadata categories with negotiated definitions and examples, exist in many fields.

When defining a set of metadata categories it is very useful to base it on a kind of systematization, e.g. a taxonomy, specifying main categories, categories and subcategories. Otherwise one may end up with an incomplete and inconsistent set of categories that is very difficult to use and to extend.

In order to obtain a well structured taxonomy we will argue that it should be based on the elaboration of a terminological ontology. In this way the concepts of the domain and their interrelations are clarified. In some cases it is even possible to generate a taxonomy on the basis of an ontology, i.e. some concepts of the ontology may more or less automatically be transformed into categories of the taxonomy. In other cases, the ontology renders the knowledge which forms the basis for the construction of the taxonomy.

6 Data categories for linguistic resources

ISO 12620:1999, *Computer assisted terminology management — Data Categories* specifies data categories used in terminological resources. These data categories are classified in three major groups and ten sub-groups:

Term and term-related data categories:
    A.1 term
    A.2 term-related information
    A.3 equivalence
Descriptive data categories:
    A.4 subject field
    A.5 concept-related description
    A.6 concept relation
    A.7 conceptual structures
    A.8 note
Administrative data categories:
    A.9 documentary language
    A.10 administrative information

This structure is not homogenous, i.e. it reflects various subdividing criteria (dimensions), and it does not give a very clear overview of the data categories.

One dimension is for example term-related information vs. concept-related description. Here it is not clear why e.g. *subject field* and *concept relation* do not fall within the group: *concept-related description*.

In 2003, it was proposed to set up a Data Category Registry (DCR) in TC 37 for all kinds

of lexical data. Since this DCR also includes data categories of dictionaries, the above structure was not very appropriate. Consequently it was decided to give up a classification of the categories. In our opinion it will, however, be difficult to ensure completeness, consistency, user-friendliness and extensibility of the above mentioned DCR, if there is no structure at all of the data categories.

7 Ontologies as the basis for meta data taxonomies

Figure 7 presents an extract of a terminological ontology for concepts pertaining to semantic information that may be registered in lexical data collections, such as e.g. termbases and electronic dictionaries. The three main types of semantic information are *subject classification*, *content specification* and *semantic relation*.

This ontology uses type relations, part whole relations and associative relations (lines with the designation of the relation type and an arrow indicating the direction of the relation).

The group of concepts on the right hand side, which are related by means of associative and part-whole relations, contribute to a better understanding of the concepts that are central for semantic information. For example, it is illustrated that a *content specification* describes the *inten-*

sion of a concept, and that the *intension* consists of *characteristic features*.

8 The Danish standard of lexical resources

The Danish Standard DS 2394-1:1998 comprises a taxonomy for the classification of lexical data, the STANLEX taxonomy. In STANLEX the main groups of information types are structured according to the linguistic disciplines: etymological information, grammatical information, graphical information, phonetic information, semantic information and usage. Examples of categories and sub categories are shown in Table 1.

9 From ontology to taxonomy

The 'backbone' of the ontology in Figure 7 consists of the top concept *semantic information* and the subordinate concepts which are related to this concept by means of type relations: *lexical paraphrase*, *analytic definition* etc. These concepts will typically form the background for categories to be included in a taxonomy. As already mentioned, the concepts that are related by means of part-whole relations or associative relations typically give a better understanding of the central concepts, but it will often not be relevant to introduce corresponding categories in a taxonomy.
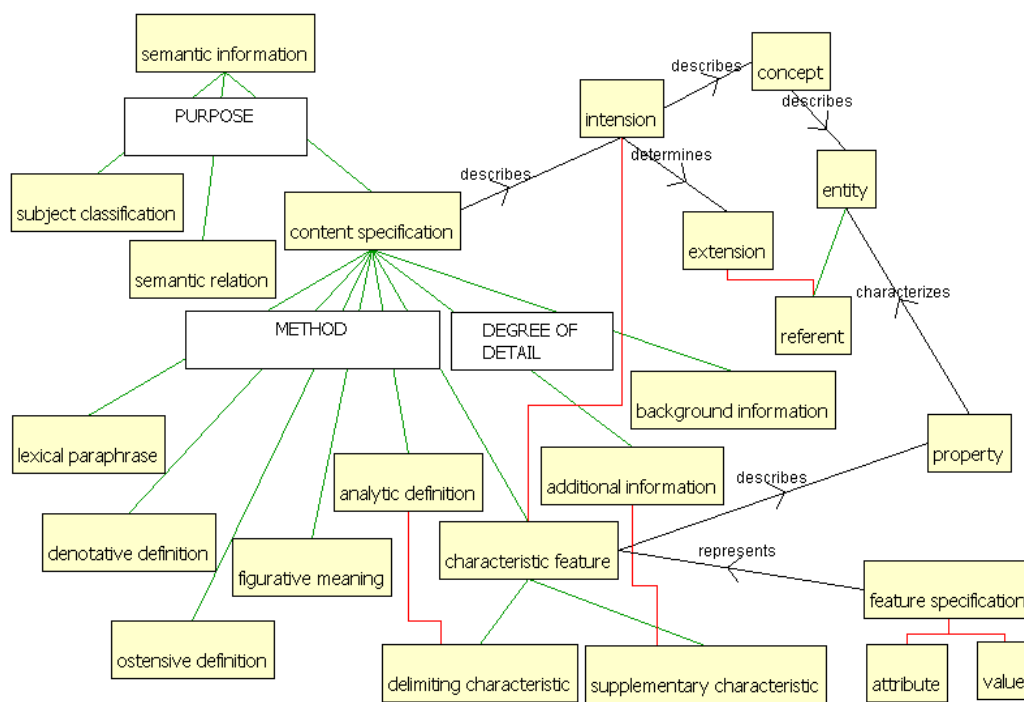

Figure 7: Ontology of semantic information.

Main group	Category	Subcategory
Semantic information	Subject classification	• Classification system • Normative subject classification • Nonnormative subject classification
	Semantic relations	• Concept system • Position of concept in concept system • Generic relation • Partitive relation • Successive relation • Causal relation • Associative relation • Antonymy • Metonymy • Equivalence within one language • Equivalence between two or more languages • Equivalence constraint
	Content specification	• Lexical paraphrase • Analytic definition • Denotative definition • Ostensive definition • Additional information • Background information • Characteristic feature • Figurative meaning

Table 1: Categories and subcategories of Semantic Information.

The nodes in a taxonomy represent categories, not concepts, and a taxonomy category may sometimes correspond to more concepts. This may be more user friendly, since the user of the taxonomy will then not have to worry about subtle distinctions. For example, in Figure 7, the concept *additional information* refers to information in the form of supplementary characteristics, while *background information* gives further information about historical, technical, legal or other aspects of the semantics of the lexical entry. In a taxonomy, one might decide to 'merge' the two concepts *additional information* and *background information* into one category, since it may be difficult for the user of the taxonomy to choose between them.

Sometimes the taxonomy will not comprise the 'lowest' levels of a hierarchy in the corresponding ontology. For example there may not be a need for distinguishing between *delimiting characteristics* and *supplementary characteristics* in the taxonomy. This is the case in the Danish Standard of lexical data categories.

In some cases it may be relevant to convert concepts of an ontology participating solely in associative or part-whole relations into categories in a taxonomy. For example it may be relevant to include the categories *feature specifica-*

tion, *attribute* and *value* from Figure 5 as taxonomy categories.

10   Conclusion

In this paper we have argued that by applying principles of terminological ontologies when constructing a product classification system or a metadata taxonomy, it is possible to obtain a clear and intuitively understandable structure and in this way to obtain completeness, consistency, user-friendliness and extensibility. In some cases an ontology may be mapped directly into a classification system, but in other cases it will be necessary and useful to introduce adjustments into the classification system compared to the ontology. The principles that we introduce here are relevant for the development of all kinds of classification systems.

References

Bodenreider, Olivier. 2005. Consistency between Metathesaurus and Semantic Network Workshop on *The Future of the UMLS Semantic Network.* NLM.

DS 2394-1. 1998. *Lexical data collections – Description of data categories and data structure – Part 1: Taxonomy for the classification of information types*, Danish Standards.

CWA 15045. 2005. *CEN Workshop Agreement: Multilingual Catalogue Strategies for eCommerce and eBusiness*.

eCl@ss: http://www.eclass-online.com/

Gruber, Tom. 2008. *Ontology. Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag. Looked up on February 23 2009: http://tomgruber.org/writing/ontology-definition-2007.htm

Gómez-Pérez, Asunción; Mariano Fernández-López & Oscar Corcho. 2004. *Ontological Engineering – with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. London: Springer Verlag.

Guarino, Nicola. 1998. Formal Ontology and Information Systems. *Formal Ontology in Information Systems, Proceedings of the First International Conference (FOIS'98)*, Amsterdam: IOS Press.

ISO 704. 2000. *Terminology work — Principles and methods*. Genève: ISO.