Terminological Principles used for Ontologies

Bodil Nistrup Madsen, Hanne Erdman Thomsen

Abstract

Designers of IT systems, such as database applications, information retrieval systems and natural language querying systems, very often talk about using ontologies as a foundation for these systems. Topics such as development of ontologies and formats for ontologies are found in many presentations in conferences, e.g. in conferences that deal with the Semantic Web. However, very often it becomes clear that people have very different understandings of what an ontology is, and many different classifications and definitions of ontologies exist. The aim of our paper is to contribute to concept clarification, and to argue that this may be obtained by using principles of terminology work.

Introduction

Systematic terminology work has for many years included the elaboration of concept systems, which may be characterized as a kind of domain specific ontologies. The ‘father’ of terminology, Eugen Wüster, started describing concept systems at least 50 years ago (see e.g. Wüster 1979), and the German standardization organization, DIN, published a standard on concept systems, DIN 2331, already in 1980. The Technical Committee TC 37 *Terminology and other Language Resources* of the International Standardization Organization, ISO, has published guidelines on the establishment of concept systems (ISO 704:2000). These guidelines are currently under revision to reflect new developments within this field.

In 2007 ISO TC 37 set up an Ontology Task Force with the aim of proposing a strategy for the work on ontologies within TC 37[1]. 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

1. The authors of the present abstract are both engaged in the work of this Task Force.

those types, e.g. ontology, taxonomy and thesaurus. One of the results will be systematic overview in the form of an *ontology of ontologies* which will comprise proposals for definitions of the different types of ontology.

This paper is organized as follows. In section 1 we will give an introduction to the principles of terminological ontologies that we will use in the rest of the paper. Section 2 gives examples of classifications of ontologies, based on Gómez-Pérez et al. (2003) and various other authors. In section 3 we present a first draft of a part of an ontology of ontologies that is based on the principles described in section 1, and in section 4 we compare this proposal with the examples of ontologies in section 2. As already mentioned there are very different understandings of what an ontology is. In section 5 we will discuss the distinction between ontology, understood as "concept model" and ontology, understood as "conceptual data model".

1 Principles of terminological ontologies

Here we will briefly present some central principles of terminological ontologies, which form a useful basis for a method for concept clarification and systematisation. Some of these principles have been used for many years in terminology work, and some have been developed in the research and development project CAOS - Computer-Aided Ontology Structuring, whose aim is to develop a computer system designed to enable semi-automatic construction of concept systems, or ontologies. CAOS is a joint research project carried out by researchers at the Department of International Language Studies and Computational Linguistics, Copenhagen Business School. The system is intended to be interactive and presupposes an end-user with a terminological background. CAOS supports terminological concept modelling. The backbone of this concept modelling is constituted by characteristics modelled by formal *feature specifications*, i.e. attribute-value pairs[2]. Our use of feature specifications is subject to a number of principles and constraints. For a more detailed description of these principles, see for example: Madsen, Thomsen, Vikner (2004).

Figure 1 shows an extract of an ontology for types of prevention, which is based on an ontology, developed by Working Group 07, *Prevention, Health Promotion and Public Health,* the Danish Council for Health Terminology. The results of the work of this working group may be seen on the web-site of the Danish National Board of Health: http://begrebsbasen.sst.dk.

2. This approach to the modelling of characteristics was proposed in Thomsen (1998) and Madsen (1998). Cf. also Carpenter (1992).

In figure 1 the characteristics of the 6 co-ordinate concepts that are types of prevention are described by means of feature specifications. These feature specifications consist of a feature and a value, e.g. "TARGET GROUP: population". The boxes across the concept relations represent subdivision criteria. We use this representation of characteristics because it clarifies the relation between characteristics and subdivision criteria: For any given concept the feature of a non-inherited characteristic may be a subdivision criterion for the superordinate concept. The clarification makes it much easier to identify subdivision criteria and differentiating characteristics in practical terminology work.

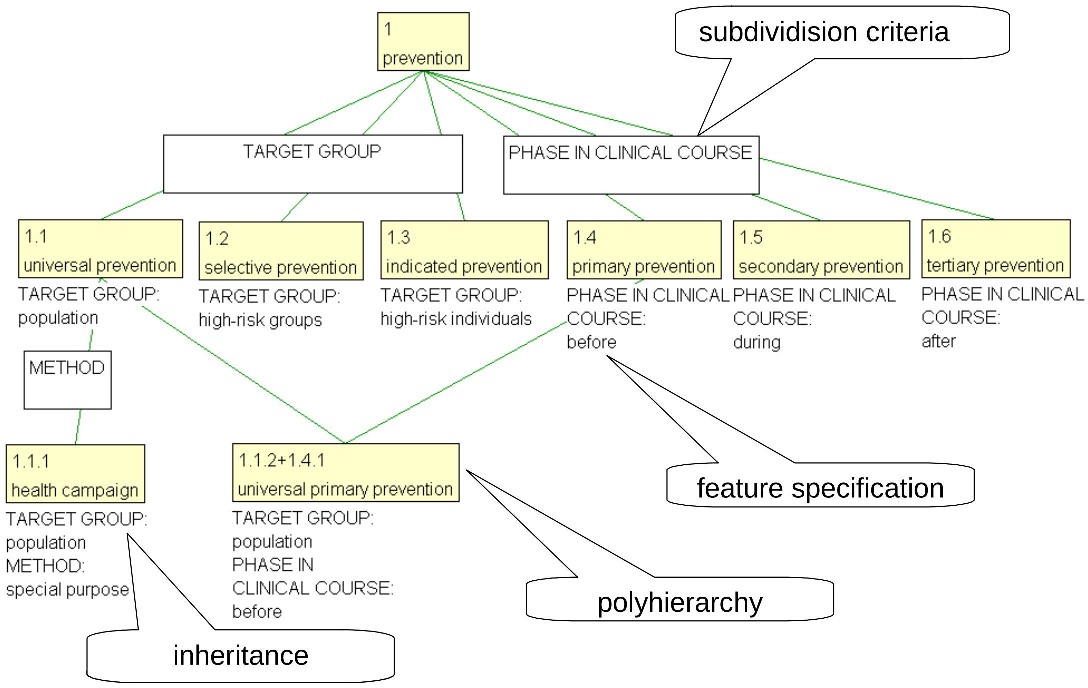

Figure 1: Extract of an ontology for types of prevention[3]

Figure 1 shows an example of inheritance of characteristics. The concept *health campaign* is a subordinate concept to *universal prevention*, and thus it inherits the feature specification "TARGET GROUP: population". Figure 1 also shows an example of polyhierarchy. The concept *universal primary prevention* has as superordinate concepts both the concept *universal prevention* and *primary prevention*. This also gives an example of inheritance, since the concept *universal primary prevention* inherits the feature specifi-

3. The diagram in Figure 1 is made with i-Model, the terminological concept modelling tool integrated with the terminology management system i-Term developed by DANTERM-centret.

cation of both superordinate concepts. Polyhierarchy is only allowed, if the two superordinate concepts of the subordinate concept belong to two different groups, i.e. if they fall within two different subdividing dimensions.

Figure 2 shows an example of dimension specifications for the concept *prevention* from Figure 1. This is a facility that we use in the CAOS prototype. Here the values associated with the corresponding attributes on the subconcepts of the concept *prevention* are specified, and the dimension specifications written in bold face are *subdividing* dimensions.

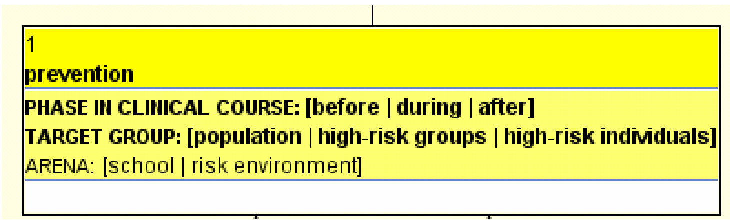

Figure 2: Dimension specifications for the concept "prevention"[4]

Subdividing dimensions must be chosen in such a way that each subordinate concept has one and only one feature specification containing as an attribute a subdividing dimension of the superordinate concept. That is, there can be no overlapping subdividing dimensions. A primary feature specification containing a subdividing dimension as an attribute is called a delimiting feature specification. Thus, a concept may contain one and only one delimiting feature specification. The concept *selective prevention* in Figure 1 could have been characterized by two feature specifications: "TARGET GROUP: population" and "ARENA: risk environment", but only one of these, "TARGET GROUP: population", is delimiting. In terminology work this is the delimiting characteristic to be used in definition writing.

In the following sections we will use the principles of terminological ontologies for describing concepts in an ontology.

2 Classification of ontologies – examples

Many authors have proposed classifications or categorizations of ontologies. One example is Gómez-Pérez et al. (2003:25 ff), who collect the classifications of various other authors, i.e. Mizoguchi et al. (1995), van Heijst et al. (1997), Guarino (1998) and Lassila & McGuiness (2001). In the following we will first show the ontologies of the four authors as terminological ontologies, and then discuss how they can be merged according to terminological

[4] The diagram in Figure 2 is made with the CAOS prototype.

principles, and compare this with the presentation in Gómez-Pérez et al. (2003).

Mizoguchi et al. (1995) have four main types of ontologies, which are distinguished according to the purpose of the ontology. They actually give several "dimensions which contribute to making the characteristics clear" (p. 50), but these are described only for task and domain ontologies, and therefore we have left them out in Figure 3.

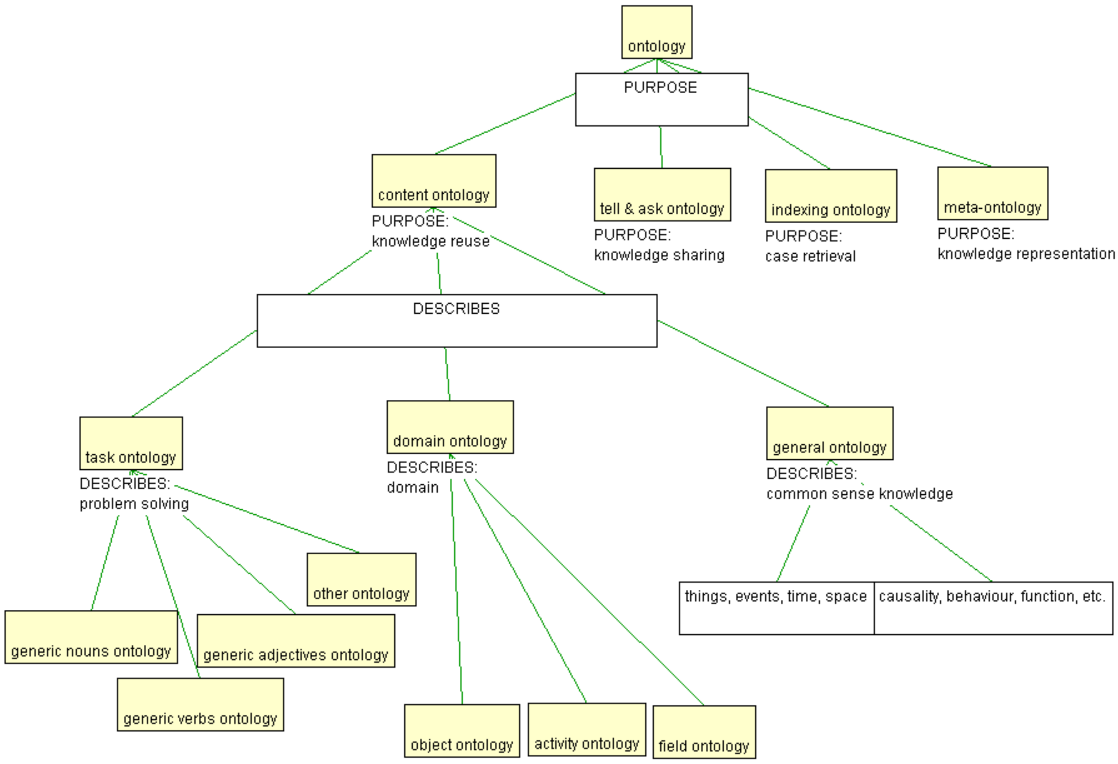

Figure 3: Ontologies according to Mizoguchi et al. (1995)

Content ontologies are further subdivided into task, domain and general ontologies depending on what they describe, and these are further subdivided, but no criterion is given explicitly.

Van Heijst et al. (1997) use two subdivision criteria at the top level, i.e. AMOUNT AND TYPE OF STRUCTURE and ISSUE OF THE CONCEPTUALIZATION. In Figure 4 these criteria are abbreviated. The use of two criteria will make it possible to locate for example a *terminological domain ontology* in the diagram.

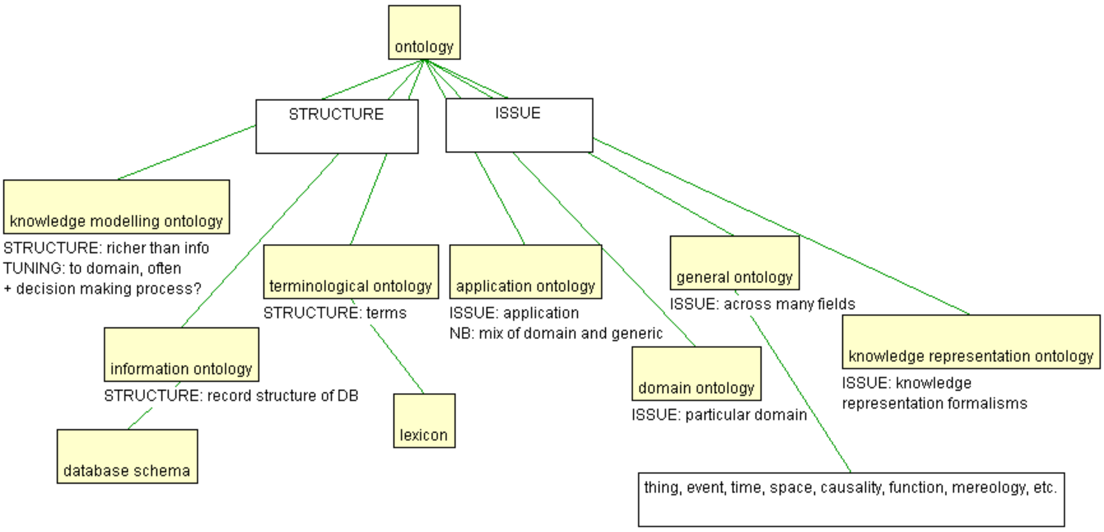

Figure 4: Ontologies according to van Heijst et al. (1997)

From the characteristics in Figures 3 and 4 combined with examples given in the above mentioned two papers we can deduce that *meta ontology* in Figure 3 is equivalent to *knowledge representation ontology* in Figure 4.

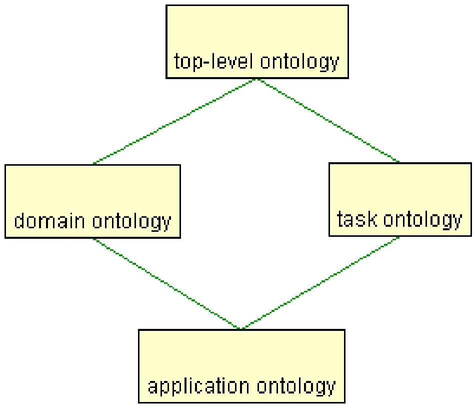

Figure 5: Ontologies according to Guarino (1998)

Guarino (1998:9) has the small diagram shown in Figure 5, which is often cited in the ontology literature. It should be noted, however, that the relations in Guarino's illustration do not represent generalizations among the types of ontologies, but rather generalizations between the concepts in these ontologies when they are linked together, cf. Guarino (1998:10): "Domain ontologies … describe the vocabulary related to a generic domain … by specializing the terms introduced in a top-level ontology". According to the explanations of the concepts (types of ontologies) given by Guarino, the ontology should rather be structured as in Figure 6. Interestingly, this is very similar to part of the ontology by Mizoguchi et al. (1995) in Figure 3. A problem with the ontology in Figure 6 is the polyhierarchy since, according to the above mentioned terminological principles, *application ontology* should not be able to inherit two characteristics with the same attribute. We will get back to this problem later in the paper.

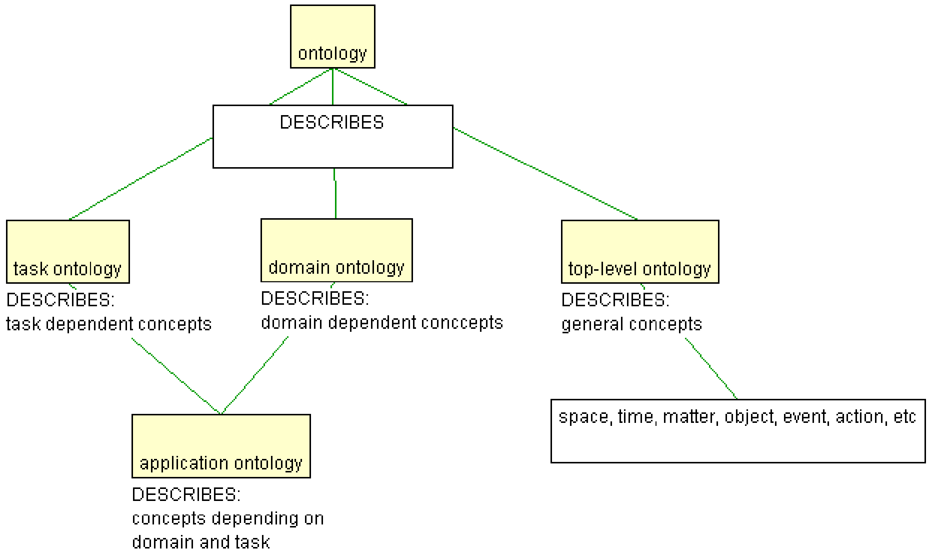

Figure 6: Ontologies according to Guarino (1998) - revised

Lassila & McGuiness (2001) present a range of ontologies categorized according to their richness of structure, shown in a terminological ontology in Figure 7.

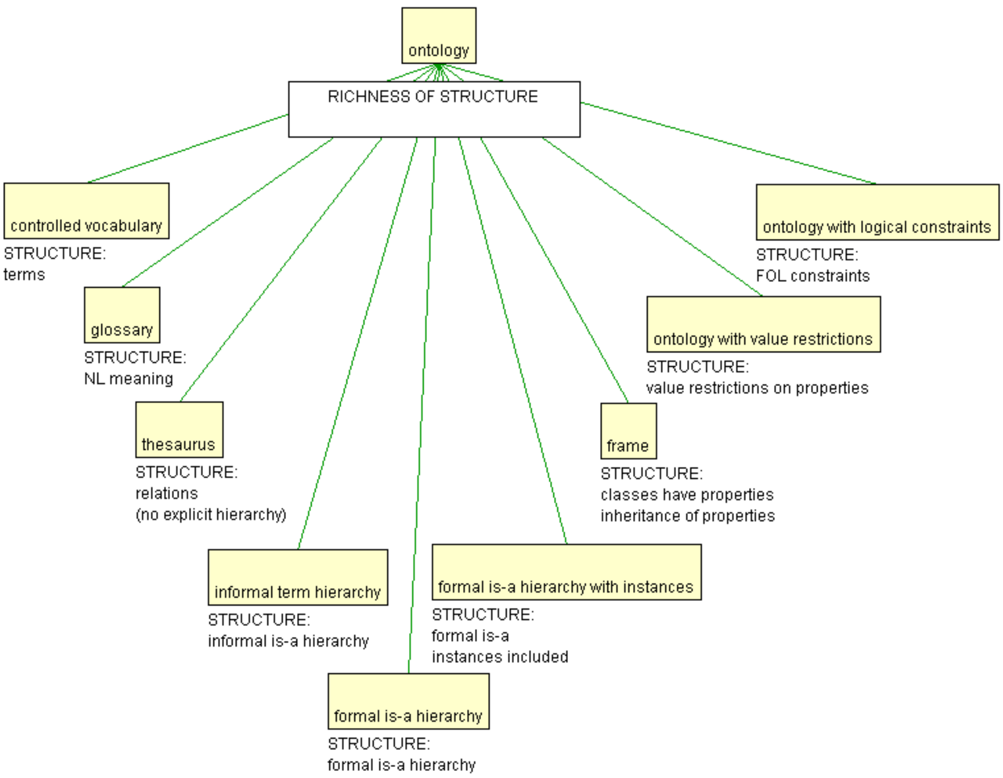

Figure 7: Ontologies according to Lassila & McGuiness (2001)

This subdivision according to structure is more fine-grained than the one by van Heijst et al. (1997) (Figure 4). *Terminological ontology* (in Figure 4) is probably equivalent to *ontology with informal is-a hierarchy* (in Figure 7), and *knowledge modeling ontology* (in Figure 4) could be equivalent to either *glossary* or *ontology with logical constraints* (in Figure 7). The exact equivalences are difficult to clarify from the sources.

Based on the 4 papers mentioned above, Gómez-Pérez et al. (2003:28-34) arrives at the typology presented in Figure 8. The left-hand side of this ontology, subdivided according to structure, is identical to Lassila & McGuiness (2001), and therefore we have not repeated the concepts in Figure 8. Gómez-Pérez et al. (2003) do not discuss to what extent the concepts in this range correspond to the three concepts given by van Heijst et al. (1997). They mention the concepts *lightweight* and *heavyweight ontology*, as the two ends of the range introduced by the subdivision criterion RICHNESS OF THE STRUCTURE in Figure 8. We have not included these, since it is not clear exactly which ontologies in Figure 7 are light- or heavyweight respectively.

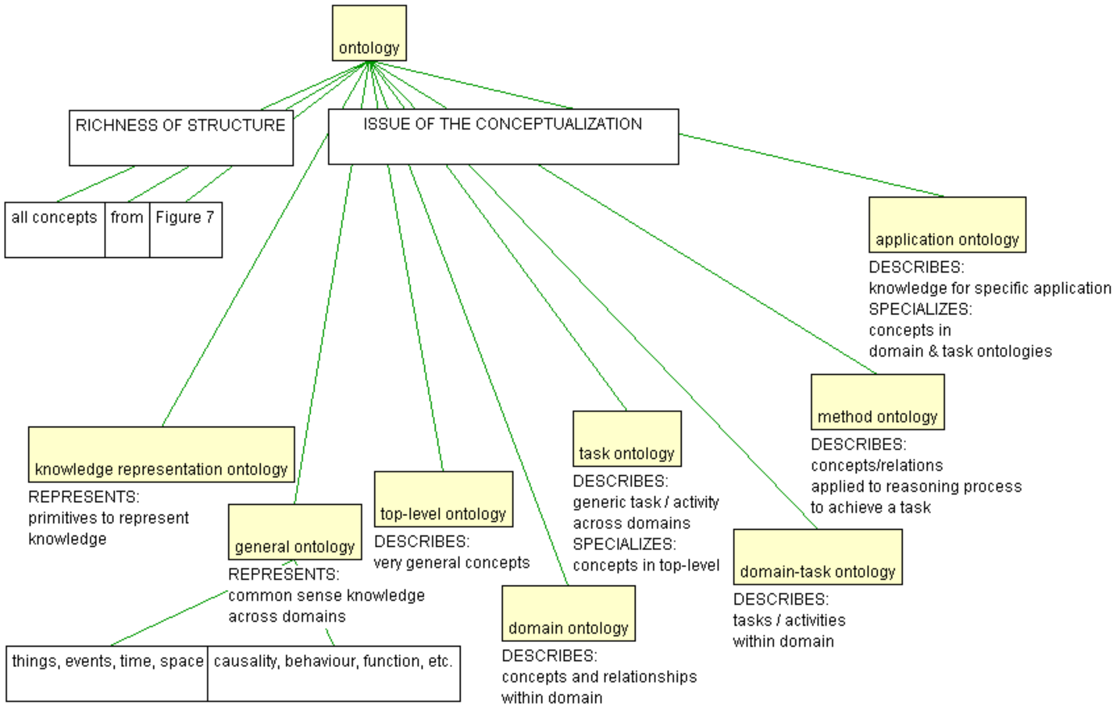

Figure 8: Ontologies according to Gómez-Pérez et al. (2003)

The characteristics given under the subdivision criterion 'issue of the conceptualization' are based on the explanations found in Gómez-Pérez et al. (2003:28-34). Note that here the authors include two concepts not introduced earlier: *domain-task ontology* and *method ontology*. These concepts are difficult to distinguish from *domain ontology* and *task ontology*, and it may be the case that *domain-task ontology* is better categorized as a subtype of *task ontology*.

We find that many of the distinctions in Figure 8 are not clear enough, consider for example the difference between *top-level ontology* and *general ontology*. It also needs to be clarified whether *top-level ontology* in Guarino (1998) is synonymous to *general ontology* in the other diagrams presented here, since the examples of what they contain (illustrated in white boxes in Figures 3, 4, 6 and 8) seem to be identical.

These were a few examples of the difficulties we find in the typology of Gómez-Pérez et al. (2003) and other authors. In section 3 we will introduce another way of structuring the domain of ontologies.

3 Proposal for an ontology for ontologies adhering to the principles of terminology

Figure 9 contains a first draft of a part of the ontology of ontologies that we are going to propose to the Ontology Task Force of ISO TC 37. The description of the concepts has to a great extent been based on Guarino (1998).

In this ontology we have introduced characteristics and subdivision criteria that clearly distinguish the types of ontologies, e.g. LEVEL, DOMAIN and PURPOSE.

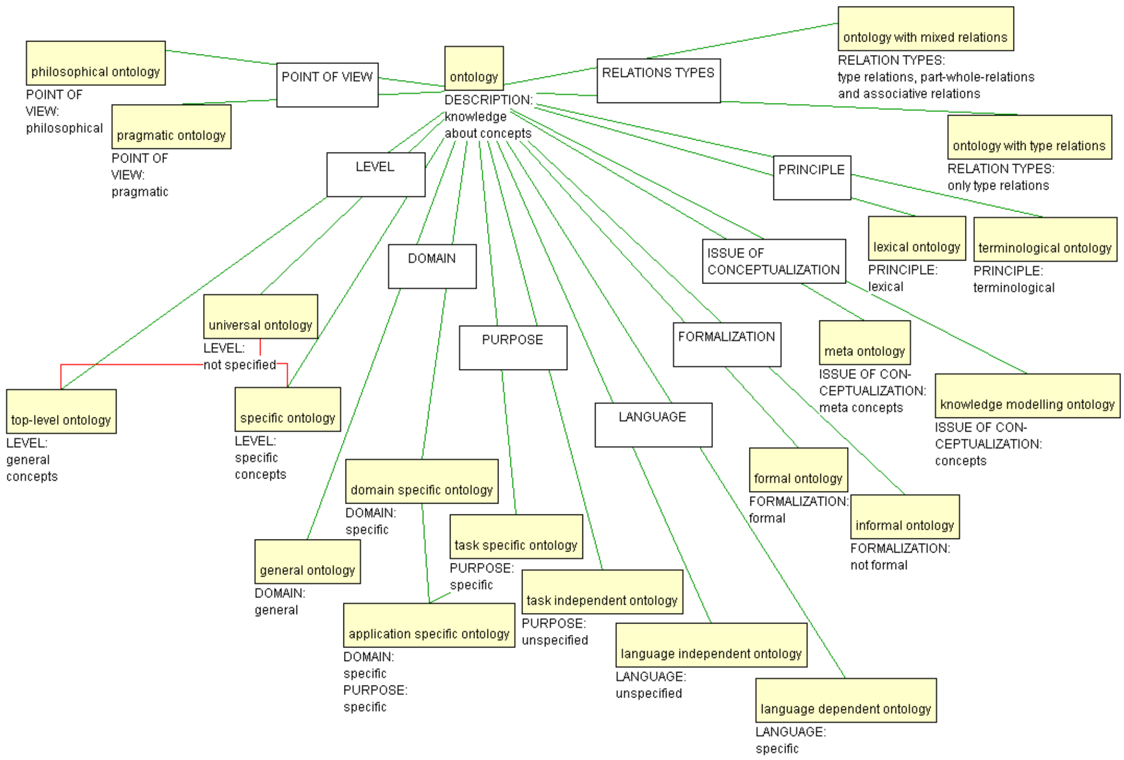

Figure 9: Part of an ontology for ontologies

With one exception all types of ontologies are found at the same level, having the top concept *ontology* as immediately superordinate concept, but belonging to different groups, each characterized by their own subdividing criteria. This means that it would be possible to describe one ontology as a subtype of several ontologies, e.g. an ontology could at the same time be a *formal ontology* and an *ontology with type relations*.

There is one explicit case of such a polyhierarchy in Figure 9: the concept *application specific ontology* is a subconcept of both *domain specific ontology* and *task specific ontology*, and inherits both characteristics “DOMAIN: specific” and “PURPOSE: specific”. In our view this is a better structure than the one found in Figure 3, where *task ontology*, *domain ontology* and *general ontology* are coordinate concepts at the same level, belonging to the same subdivision criterion.

According to the ontology in Figure 9, an ontology may at the same time be a *top-level ontology* and a *general ontology*, since these two concepts are subordinate concepts of two superordinate concepts, that belong to two different subdivision criteria. In our understanding a *top-level ontology* comprises general concepts with a view to level, while a *general ontology* comprises concepts that belong to general language, and not domain specific language.

The concept *universal ontology* constitutes a special case, since it may comprise both a *top-level ontology* and a *specific ontology*.

4 Comparison with the ontologies in section 2

The concepts at the lowest level in Figure 3, i.e. the subconcepts of *task ontology* and *domain ontology*, remain subconcepts of these concepts in Figure 9.

The concept *meta ontology* in Figure 3, based on Mizoguchi et al. (1995), corresponds to the concept *knowledge representation ontology* in Figures 4 and 8, based on van Heijst at al. (1997) and Gómez-Pérez et al. (2003) respectively. In our opinion the term *meta ontology* is more descriptive than *knowledge representation ontology*, and the characteristic in Figure 9 better describes the concept (“ISSUE OF CONCEPTUALIZATION: meta concepts”).

In our understanding the concept *content ontology* in Figure 3 corresponds to the concept *knowledge modelling ontology* in van Heijst et al. (1997), cf. Figure 4 and our ontology in Figure 9. However, it is somewhat difficult to tell from the descriptions of the concept in Mizoguchi et al. (1995) and van Heijst et al. (1997).

The concept *terminological ontology* in Figure 4 does not correspond to the concept that we have introduced in Figure 9. By *terminological ontology* we understand an ontology which adheres to the principles of terminology described in section 1. According to these principles an ontology is based on an analysis and specification of characteristics and use of subdivision criteria.

On the basis of some of the concepts in Figure 7, some more concepts can be introduced into our proposal for an ontology in Figure 9. For example, in Figure 7, we find the concept *formal is-a hierarchy*. Lassila & McGuiness (2001) characterizes this as a “strict subclass hierarchy”, which we understand as a hierarchy comprising only type relations (is-a relations), as opposed to “informal term hierarchy” which may comprise associative relations. Figure 9 comprises the concepts *formal ontology* and *informal ontology*. However, in this case ‘formal’ and ‘informal’ means something else than in Lassila & McGuiness (2001). In Figure 9 ‘formal’ and ‘informal’ refers to whether the concepts are described in a formal way or not. In Figure 9 we have introduced two concepts: *ontology with type relations*, corresponding to the concept *formal is-a hierarchy in* Lassila & McGuiness (2001), and *ontology with mixed relations*, i.e. ontology comprising type relations, part-whole-relations and associative relations. Hence the subdivision criterion would be RELATION TYPES. In Figure 9 it would be possible to have for example a concept *informal ontology with type relations*, which is a subconcept of the two concepts: *formal ontology* and *ontology with type relations*.

Here we have not discussed in detail all ontologies found in the examples in section 2, and we have not completed the ontology in Figure 9. This will be future work. However, some of the ontologies mentioned in for example Figure 7 should, in our view, not be categorized as ontologies, either because they are concerned only with the expressions (e.g. *controlled vocabulary*, *glossary*) or because their purpose is classification rather than description of knowledge (e.g. *thesaurus*).

In the next section we will discuss the concept *information ontology* which is found in Figure 4 (van Heijst et al. (1997).

5 Ontology vs. conceptual data model

Very often the term *ontology* is used for *data model*, either *conceptual data model* or *information model* (also referred to as *information ontology*, c.f. Figure 4). However, the concept *information model* is defined as a “data model that represents the organization of information in a manner that reflects the structure of an information system”, cf. for example CEN CWA 15045 (2004), and thus, in our opinion, it is not a type of ontology. In data modelling one distinguishes three concepts *conceptual data model*, *informa-*

tion model and *physical data model* that represent three levels in data modelling, where the first step is development of a conceptual data model.

An important issue to be discussed here is the distinction between ontology, understood as “concept model”, i.e. a model that may be defined as “model for the description of knowledge about concepts”, and ontology, understood as “conceptual data model”, i.e. a “data model that represents an abstract view of the real world”. This distinction is for example found in CEN CWA 15045 (2004), but in many cases the two concepts are not clearly distinguished, and confusion may arise because of the word “conceptual” in the term “conceptual data model”. There are some similarities between ontologies (concept models) and conceptual data models, but they also differ considerably.

Ontologies give information about concepts in the form of feature specifications and concept relations (information about meaning). Conceptual data models give information about the classes in the form of attributes and associations between the classes. The attributes of the data model give no information about the meaning of the classes, only a specification of what kind of information will be given about the entities represented by the classes in question.

Figure 10 shows an example of an extract of an ontology for terminological information categories, i.e. some concepts that may be relevant when one wants to establish a terminology database (source reference, e.g. for definitions, terms etc.).

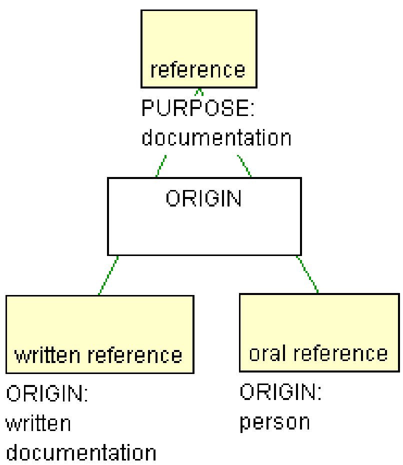

Figure 10: Extract of an ontology for terminological information

On the basis of the feature specifications it is possible to write definitions of the concepts.

Figure 11 shows the corresponding part of a conceptual data model for a terminology database (source reference, e.g. for definitions, terms etc.).

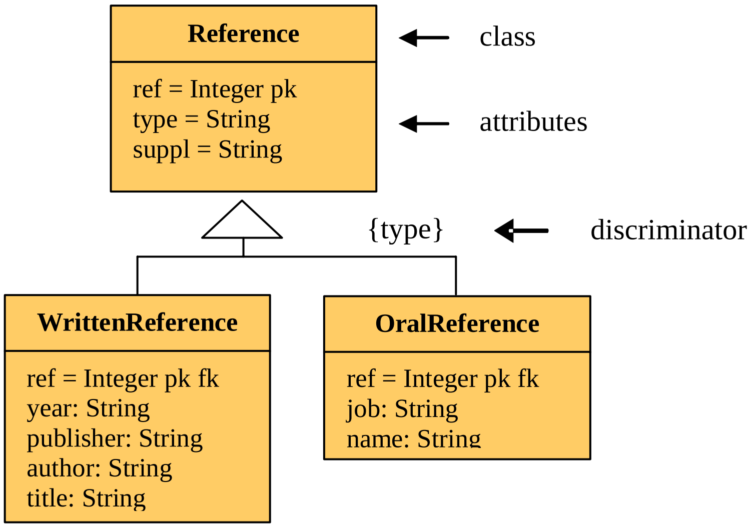

Figure 11: Conceptual data model for a part of a terminological database

When using UML for the elaboration of a conceptual data model for an IT system, the upper compartment of the class comprises the name of the class. The middle compartment contains attributes. Information on operations may be added in the lower compartment. Operations may for example specify input or search procedures that may be carried out on the class. This is not shown in Figure 11.

It is possible to add to the attributes information about data types as well as information about the primary key (pk) and foreign keys (fk); cf. Figure 11. Often this kind of information will not be found in the conceptual data model, but rather in the logical data model, which represents the next step in the data modeling process, i.e. a data model that represents the organization of information in a manner that reflects the structure of an information system. This step is closer to the implementation into tables of a relational database.

The attributes that are used in standard UML for data modelling serve a different purpose from the feature specifications in an ontology. They specify what kinds of information may be related to each class, and consequently to each instance represented in the IT system. The values of the attributes will exist only in an IT system (e.g. in a database), and they will give information about instances; cf. also Madsen, Thomsen, Vikner (2002, 2005).

The discriminator of UML corresponds to the subdivision criterion (subdividing dimension) in CAOS. However, in standard UML it is not possible to represent several dimensions from which one may be chosen as the subdividing dimension, and there is no notation for the specification of dimension values, as is the case in CAOS. Furthermore there is no (formal) connection between the discriminator and the attributes of subordinate concepts.

In the literature on data bases, data modelling is also referred to as conceptual modelling or semantic modelling, which may contribute to the confusion concerning the two concepts *concept modelling* and *conceptual data modelling*. Furthermore, the concept ontology is often used for a conceptual data model in the literature on ontologies; cf. Gómez-Pérez et al. (2004: 22) and Madsen, Thomsen, Vikner (2005).

In the above example all three concepts of the ontology in Figure 10 give rise to classes in the data model in Figure 11. However, this is not always the case. An ontology for terminological information categories may comprise the concept “extension”, but it will not be relevant to include e.g. a class “Extension” in a data model for a terminological database. Other concepts of the ontology will not give rise to classes in the data model, but rather attributes or even values of attributes. For example, the concept “delimiting characteristic” will not give rise to a class in a data model, but rather correspond to an attribute, which belongs to a class “Concept”, or maybe it will be a value of an attribute “type of characteristic”.

In our opinion the distinction between *concept model* and *conceptual data model* is very important, and our recommendation is that the development of a conceptual data model should always be preceded by the elaboration of a concept model (ontology). This is the most systematic method for clarifying the meaning of the concepts underlying the data model, and to obtain a common understanding of the types of information, that will be handled by the IT system.

Conclusion

In our paper we have argued that when clarifying concepts with a view to setting up an ontology it is an advantage to use terminological principles, which makes use of feature specifications and subdivision criteria, and which adheres to the principles concerning the rules for polyhierarchy.

We have used these principles to describe examples of the categorization of ontologies made by various authors, and have demonstrated that this may help in getting a better overview of different author’s understanding of ontologies.

We have given a first draft of a part of an ontology of ontologies that makes use of the terminological principles, and we have compared this draft ontology with the description of ontologies presented by other authors.

From our description it may be concluded that it is a good solution to describe each concept by means of one characteristic, which means that the concepts may be grouped according to one subdivision criterion. This allows for introducing polyhierarchy, where one concept may be a subconcept of several superordinate concepts, and therefore inherit characteristics from all superordinate concepts.

Furthermore we have briefly discussed the distinction between ontology, understood as “concept model”, i.e. a model that may be defined as model for the description of knowledge about concepts, and ontology, understood as “conceptual data model”, i.e. a formal model for the description of data in an IT system. In our opinion this is a very important distinction, and our recommendation is that the development of a conceptual data model should always be preceded by the elaboration of a concept model.

References

Carpenter, Bob; *The Logic of Typed Feature Structures*. Cambridge, Mass.: Cambridge University Press, 1992.

CEN CWA 15045 (2004). Multilingual Catalogue Strategies for eCommerce and eBusiness.

DIN 2331 (1980) Begriffssysteme und ihre Darstellung, Berlin.

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

Guarino, Nicola; Formal Ontology and Information Systems. In Nicola Guarino (ed.): *Formal Ontology in Information Systems, Proceedings of the First International Conference (FOIS'98)*, June 6-8, Trento, Italy, Amsterdam: IOS Press, 1998.

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

Lassila, Ora & Deborah McGuiness; *The Role of Frame-based Representation on the Semantic Web*. Technical Report KSL-01-02. Knowledge Systems Laboratory. Stanford University, 2001.

Madsen, Bodil Nistrup; Typed Feature Structures for Terminology Work - Part I. In: *LSP - Identity and Interface - Research, Knowledge and Society. Proceedings of the 11th European Symposium on Language for Special Purposes*, Copenhagen, august 1997, Copenhagen Business School, 1998.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner; Data Modelling and Conceptual Modelling in the Domain of Terminology. In: Melby, Alan (ed.): P*roceedings of TKE '02 - Terminology and Knowledge Engineering*, INRIA, Nancy, 2002.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner; Principles of a system for terminological concept modelling. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation, Vol. I.* Lisbon, 2004.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner; "Multidimensionality in terminological concept modelling". In: Bodil Nistrup Madsen, Hanne Erdman Thomsen (eds.): *Terminology and Content Development, TKE 2005, 7th International Conference on Terminology and Knowledge Engineering*, Copenhagen, 2005.

Thomsen, Hanne Erdman; Typed Feature Structures for Terminology Work - Part II. In: *LSP - Identity and Interface - Research, Knowledge and Society*. *Proceedings of the 11th European Symposium on Language for Special Purposes*, Copenhagen, august 1997, Copenhagen Business School, 1998.

Van Heijst, G., A. Th. Schreiber, B.J. Wielinga; Using explicit ontologies in KBS development. In: *International Journal of Human-Computer Studies*. 45: 183-292.

Wüster, Eugen; *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Copenhagen: LSP Centre, Copenhagen School of Economics. (1st edition: 1979, Wien: Springer Verlag), 1985.