What Is an Ontology?

Chapter · May 2009

DOI: 10.1007/978-3-540-92673-3_0

Steffen Staab
Universität Koblenz-Landau

What Is an *Ontology*?

Nicola Guarino[1], Daniel Oberle[2], and Steffen Staab[3]

[1] ITSC-CNR, Laboratory for Applied Ontology, 38100 Trento, Italy, nicola.guarino@cnr.it
[2] SAP Research, CEC Karlsruhe, 76131 Karlsruhe, Germany, d.oberle@sap.com
[3] University of Koblenz-Landau, ISWeb, 56016 Koblenz, Germany, staab@uni-koblenz.de

Summary. The word "ontology" is used with different senses in different communities. The most radical difference is perhaps between the philosophical sense, which has of course a well-established tradition, and the computational sense, which emerged in the recent years in the knowledge engineering community, starting from an early informal definition of (computational) ontologies as "explicit specifications of conceptualizations". In this paper we shall revisit the previous attempts to clarify and formalize such original definition, providing a detailed account of the notions of *conceptualization* and *explicit specification*, while discussing at the same time the importance of *shared* explicit specifications.

1 Introduction

The word "ontology" is used with different meanings in different communities. Following [9], we distinguish between the use as an uncountable noun ("Ontology," with uppercase initial) and the use as a countable noun ("an ontology," with lowercase initial) in the remainder of this chapter. In the first case, we refer to a philosophical discipline, namely the branch of philosophy which deals with the *nature* and *structure* of "reality." Aristotle dealt with this subject in his Metaphysics[1] and defined Ontology[2] as the science of "being *qua* being," i.e., the study of attributes that belong to things because of their very nature. Unlike the experimental sciences, which aim at discovering and modeling reality under a certain perspective, Ontology focuses on the

[1] The first books of Aristotle's treatises, known collectively as "Organon," deal with the nature of the world, i.e., physics. Metaphysics denotes the subjects dealt with in the rest of the books – among them Ontology. Philosophers sometimes equate Metaphysics and Ontology.

[2] Note, that the term "Ontology" itself was coined only in the early seventeenth century [13].

S. Staab and R. Studer (eds.), *Handbook on Ontologies,* International Handbooks on Information Systems, DOI 10.1007/978-3-540-92673-3,

nature and structure of things per se, independently of any further considerations, and even independently of their actual existence. For example, it makes perfect sense to study the Ontology of unicorns and other fictitious entities: although they do not have actual existence, their nature and structure can be described in terms of general categories and relations.

In the second case, which reflects the most prevalent use in Computer Science, we refer to *an* ontology as a special kind of information object or computational artifact. According to [7, 8], the account of existence in this case is a pragmatic one: "For AI systems, what 'exists' is that which can be represented."

Computational ontologies are a means to formally model the structure of a system, i.e., the relevant entities and relations that emerge from its observation, and which are useful to our purposes. An example of such a system can be a company with all its employees and their interrelationships. The ontology engineer analyzes relevant entities[3] and organizes them into *concepts* and *relations*, being represented, respectively, by unary and binary predicates.[4] The backbone of an ontology consists of a generalization/specialization hierarchy of concepts, i.e., a taxonomy. Supposing we are interested in aspects related to human resources, then *Person*, *Manager*, and *Researcher* might be relevant concepts, where the first is a superconcept of the latter two. *Cooperates-with* can be considered a relevant relation holding between persons. A concrete person working in a company would then be an instance of its corresponding concept.

In 1993, Gruber originally defined the notion of an ontology as an "explicit specification of a conceptualization" [7].[5] In 1997, Borst defined an ontology as a "formal specification of a shared conceptualization" [1]. This definition additionally required that the conceptualization should express a *shared* view between several parties, a consensus rather than an individual view. Also, such conceptualization should be expressed in a (formal) machine readable format. In 1998, Studer et al. [15] merged these two definitions stating that: "An ontology is a formal, explicit specification of a shared conceptualization."

---

[3] Entity denotes the most general being, and, thus, subsumes subjects, objects, processes, ideas, etc.

[4] Unfortunately, the terminology used in Computer Science is problematic here. What we call "concepts" in this chapter may be better called "properties" or "categories." Regrettably, "property" is used to denote a binary relation in RDF(S), so we shall avoid using it. Also, Smith made us aware that the notion of "concept" is quite ambiguous [14]. A way to solve the terminological conflict is to adopt the philosophical term "universal," which roughly denotes those entities that can have instances; particulars are entities that do not have instances. What we call "concepts" correspond to unary universals, while "relations" correspond to binary universals.

[5] Other definitions of an ontology have surfaced in the literature, e.g., [16] or [11], which are similar to Gruber's. However, the one from Gruber seems to be the most prevalent and most cited.

All these definitions were assuming an informal notion of "conceptualization," which was discussed in detail in [9]. In the following, we shall revisit such discussion, by focusing on the three major aspects of the definition by Studer et al.:

- What is a *conceptualization*?
- What is a proper *formal, explicit specification*?
- Why is '*shared*' of importance?

It is the task of this chapter to provide a concise view of these aspects in the following sections. It lies in the nature of such a chapter that we have tried to make it more precise and formal than many other useful definitions of ontologies that do exist – but that do not clarify terms to the degree of accuracy that we target here.

Accordingly, the reader new to the subject of ontologies may prefer to learn first about applications and examples of ontologies in the latter parts of this book and may decide to return to this opening chapter once he wants to see the common *raison d'être* behind the different approaches.

2 What is a *Conceptualization*?

Gruber [7, 8] refers to the notion of a conceptualization according to Genesereth and Nilsson [5], who claim: "A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly."

Despite the complex mental nature of the notion of "conceptualization," Genesereth and Nilsson choose to explain it by using a very simple mathematical representation: an extensional relational structure.

Definition 2.1 (Extensional relational structure) An extensional relational structure*, (or a* conceptualization according to Genesereth and Nilsson*), is a tuple* $(D, \mathbf{R})$ *where*

- D *is a set called the universe of discourse*
- $\mathbf{R}$ *is a set of relations on* D

Note that, in the above definition, the members of the set $\mathbf{R}$ are ordinary mathematical relations on D, i.e., sets of ordered tuples of elements of D. So each element of $\mathbf{R}$ is an *extensional* relation, reflecting a *specific* world state involving the elements of D, such as the one depicted in Fig. 1, concerning the following example.

Example 2.1 *Let us consider human resources management in a large software company with 50,000 people, each one identified by a number (e.g., the social security number, or a similar code) preceded by the letter I. Let us assume that our universe of discourse D contains all these people, and that we are only interested in relations involving people. Our* $\mathbf{R}$ *will contain some unary relations, such as Person, Manager, and Researcher, as well as the binary relations reports-to and cooperates-with.*[6] *The corresponding extensional relation structure* $(D, \mathbf{R})$ *looks as follows:*

- $D = \{I000001, ..., I050000, ...\}$
- $\mathbf{R} = \{Person, Manager, Researcher, cooperates\text{-}with, reports\text{-}to\}$

Relation extensions reflect a specific world. Here, we assume that Person comprises the whole universe D and that Manager and Researcher are strict subsets of D. The binary relations reports-to and cooperates-with are sets of tuples that specify every hierarchical relationship and every collaboration in our company. Some managers and researchers are depicted in Fig. 1. Here, I046758, a researcher, reports to his manager I034820, and cooperates with another researcher, namely I044443.

- $Person = D$
- $Manager = \{..., I034820, ...\}$
- $Researcher = \{..., I044443, ..., I046758, ...\}$
- $reports\text{-}to = \{..., (I046758, I034820), (I044443, I034820), ...\}$
- $cooperates\text{-}with = \{..., (I046758, I044443), ...\}$

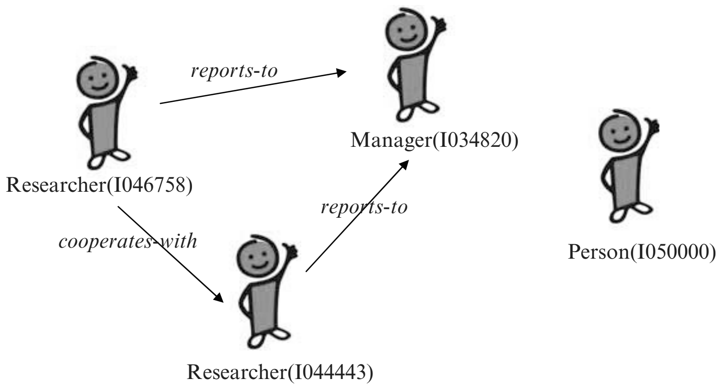

Fig. 1. A tiny part of a specific world with persons, managers, researchers, and their relationships in the running example of human resources in a large software company

[6] The name of a person could also be assigned via relations, e.g., *firstname(I046758,'Daniel')* and *lastname(I046758,'Oberle')*.

Despite its simplicity, this extensional notion of a conceptualization does not really fit our needs and our intuition, mainly because it depends too much on a specific state of the world. Arguably, a conceptualization is about concepts. Now, should our concept of *reports-to* change when the hierarchical structure of our company changes? Indeed, as discussed in [9], a conceptualization should not change when the world changes. Otherwise, according to the Genesereth and Nilsson's view given in Definition 2.1, every specific people interaction graph, such as the one depicted in Fig. 1, would correspond to a different conceptualization, as shown in Example 2.2.

Example 2.2 *Let us consider the following alteration of Example 2.1 with* $D' = D$ *and* $\mathbf{R}' = \{Person, Manager, Researcher, \text{reports-to'}, \text{cooperates-with}\}$ *where* $\text{reports-to'} = \text{reports-to} \cup \{(I034820, I050000)\}$.

Although we only added one new reporting relationship, it is obvious that $(D, \mathbf{R}) \neq (D', \mathbf{R}')$ and, thus, we have two different conceptualizations according to Genesereth and Nilsson.

The problem is that the extensional relations belonging to $\mathbf{R}$ reflect a specific world state. However, we need to focus on the meaning of the underlying concepts, which are independent of a single world state: for instance, the meaning of *cooperates-with* lies in the particular way two persons act in the company.

In practice, understanding such meaning implies having a rule to decide, observing different behavior patterns, whether or not two persons are cooperating. Suppose that, in our case, for two persons *I046758* and *I044443* to cooperate means that (1) both declare to have the same goal; (2) both do something to achieve this goal. Then, the meaning of "cooperating" can be defined as a *function* that, for each global behavioral context involving all our universe, gives us the list of couples who are actually cooperating in that context. The reverse of this function *grounds* the meaning of a concept in a specific world state. Generalizing this approach, and abstracting from time for the sake of simplicity, we shall say that an *intensional* relation[7] (as opposed to an *extensional* relation) is a function from a set of maximal world states (the global behavioral contexts in our case) into extensional relations. This is the common way of expressing intensions, which goes back to Carnap [2] and is adopted and extended in Montague's semantics [4].

To formalize this notion of intensional relation, we first have to clarify what a "world" and a "world state" is. We shall define them with reference to the notion of "system," which will be given for granted: since we are dealing with computer representations of real phenomena, a system is simply the given piece of reality we want to model, which, at a given degree of granularity, is

[7] To underly their link with conceptualizations, Guarino has proposed to call such intensional relations "conceptual relations" in [10].

"perceived" by an observing agent (typically external to the system itself) by means of an array of "observed variables."[8]

In our case, this system will be an actual group of people interacting in certain ways. For the sake of simplicity, we shall assume to observe this system at a granularity where single persons can be considered as atoms, so we shall abstract, e.g., from body parts. Moreover, we shall assume that the only observed variables are those which tell us whether a person has a certain goal (belonging to a pre-determined list), and whether such person is actually acting to achieve such goal. Supposing there is just one goal, we have $50{,}000 + 50{,}000 = 100{,}000$ variables. Each combination of such variables is a world state. Two different agents (outside the observed system) will share the same meaning of "cooperating" if, in presence of the same world states, will pick up the same couples as instances of the *cooperates-with* relation. If not, they will have different conceptualizations, i.e., *different ways of interpreting their sensory data*. For instance, an agent may assume that sharing a goal is enough for cooperating, while the other may require in addition some actual work aimed at achieving the goal.

Definition 2.2 (World) *With respect to a specific system S we want to model, a* world state *for S is a maximal observable state of affairs, i.e., a unique assignment of values to all the observable variables that characterize the system. A* world *is a totally ordered set of world states, corresponding to the system's evolution in time. If we abstract from time for the sake of simplicity, a world state coincides with a world.*

At this point, we are ready to define the notion of an intensional relation in more formal terms, building on [9], as follows:

Definition 2.3 (Intensional relation, or *conceptual relation*) *Let S be an arbitrary system, D an arbitrary set of distinguished elements of S, and W the set of world states for S (also called worlds, or possible worlds). The tuple $<D, W>$ is called a* domain space for S*, as it intuitively fixes the space of variability of the universe of discourse D with respect to the possible states of S. An* intensional relation *(or conceptual relation) ρ^n of arity n on $<D, W>$ is a total function $\rho^n : W \rightarrow 2^{D^n}$ from the set W into the set of all n-ary (extensional) relations on D.*

Once we have clarified what a conceptual relation is, we give a representation of a conceptualization in Definition 2.4. Below, we also show how the conceptualization of our human resources system looks like in Example 2.3.

Definition 2.4 (Intensional relational structure, or *conceptualization*) *An* intensional relational structure *(or a* conceptualization *according to Guarino) is a triple $\mathbf{C} = (D, W, \Re)$ with*

[8] It is important to note that, if we want to provide a well-founded, grounded account of meaning, this system needs to be first of all a physical system, and not an abstract entity.

- D *a universe of discourse*
- W *a set of possible worlds*
- $\Re$ *a set of conceptual relations on the domain space* $<D, W>$

Example 2.3 *Coming back to the Examples 2.1 and 2.2, we can see them as describing two different worlds compatible with the following conceptualization* **C***:*

- $D = \{I000001, ..., I050000, ...\}$ *the universe of discourse*
- $W = \{w_1, w_2, ...\}$ *the set of possible worlds*
- $\Re = \{Person^1, Manager^1, Researcher^1, cooperates\text{-}with^2, reports\text{-}to^2\}$ *the set of conceptual relations*

For the sake of simplicity, we assume that the unary conceptual relations, viz., Person^1*, Manager*^{1}*, and Researcher*^{1}*, are rigid, and, thus, map to the same extensions in every possible world. We do not make this specific assumption here for the binary reports-to*^{2} *and cooperates-with*^{2}*:*

- *for all worlds* w *in* W*:* $Person^1(w) = D$
- *for all worlds* w *in* W*:* $Manager^1(w) = \{..., I034820, ...\}$
- *for all worlds* w *in* W*:* $Researcher^1(w) = \{..., I044443, ..., I046758, ...\}$
- $reports\text{-}to^2(w_1) = \{..., (I046758, I034820), (I044443, I034820), , ...\}$
- $reports\text{-}to^2(w_2) = \{..., (I046758, I034820), (I044443, I034820), (I034820, I050000), ...\}$
- $reports\text{-}to^2(w_3) = ...$
- $cooperates\text{-}with^2(w_1) = \{..., (I046758, I044443), ...\}$
- $cooperates\text{-}with^2(w_2) = ...$

3 What is a Proper *Formal, Explicit Specification*?

In practical applications, as well as in human communication, we need to use a *language* to refer to the elements of a conceptualization: for instance, to express the fact that *I046758* cooperates with *I044443*, we have to introduce a specific symbol (formally, a predicate symbol, say *cooperates-with*, which, in the user's intention, is intended to represent a certain conceptual relation. We say in this case that our language (let us call it **L**) *commits* to a conceptualization.[9] Suppose now that **L** is a first-order logical language, whose nonlogical symbols (i.e., its *signature*, or its *vocabulary*) are the elements of the set $\{I046758, I044443, cooperates\text{-}with, reports\text{-}to\}$. How can we make sure

[9] Of course, properly speaking, it is an *agent* who commits to a conceptualization while using a certain language: what we call the *language commitment* is an account of the competent use of the language by an agent who adopts a certain conceptualization.

that such symbols are *interpreted* according to the conceptualization we commit to? For instance, how can we make sure that, for somebody who does not understand English, *cooperates-with* is not interpreted as corresponding to our conceptualization of *reports-to*, and vice versa? Technically, the problem is that a logical signature can, of course, be interpreted in arbitrarily many different ways. Even if we fix *a priori* our interpretation domain (the *domain of discourse*) to be a subset of our *cognitive domain*, the possible interpretation functions mapping predicate symbols into proper subsets of the domain of discourse are still unconstrained. In other words, once we commit to a certain conceptualization, we have to make sure to only admit those *models* which are *intended* according to the conceptualization. For instance, the intended models of the *cooperates-with* predicate will be those such that the interpretation of the predicate returns one of the various possible extensions (one for each possible world) of the conceptual relation denoted by the predicate. The problem however is that, to specify what such possible extensions are, we need to *explicitly* specify our conceptualization, while conceptualizations are typically in the mind of people, i.e., *implicit*.

Here emerges the role of ontologies as "explicit specifications of conceptualizations." In principle, we can explicitly specify a conceptualization in two ways: *extensionally* and *intensionally*. In our example, an *extensional* specification of our conceptualization would require listing the extensions of every (conceptual) relation for all possible worlds. However, this is impossible in most cases (e.g., if the universe of discourse D or the set of possible worlds W are infinite) or at least very impractical. In our running example, we are dealing with thousands of employees and their possible cooperations can probably not be fully enumerated. Still, in some cases it makes sense to partially specify a conceptualization in an extensional way, by means of examples, listing the extensions of conceptual relations in correspondence of selected, stereotypical world states. In general, however, a more effective way to specify a conceptualization is to fix a language we want to use to talk of it, and to constrain the interpretations of such a language in an *intensional* way, by means of suitable axioms (called *meaning postulates* [2]). For example, we can write simple axioms stating that *reports-to* is asymmetric and intransitive, while *cooperates-with* is symmetric, irreflexive, and intransitive. In short, an ontology is just a set of such axioms, i.e., a logical theory designed in order to capture the intended models corresponding to a certain conceptualization and to exclude the unintended ones. The result will be an *approximate* specification of a conceptualization: the better intended models will be captured and non-intended models will be excluded (cf. Fig. 2).

The axioms for intensionally and explicitly specifying the conceptualization can be given in an *informal* or *formal* language **L**. As explained in the introduction, [15] requires that the explicit specification must be formal in addition to what proposed in [1, 7]. 'Formal' refers to the fact that the expressions must be machine readable, hence natural language is excluded. Let us now discuss all the notions above in a more formal way.

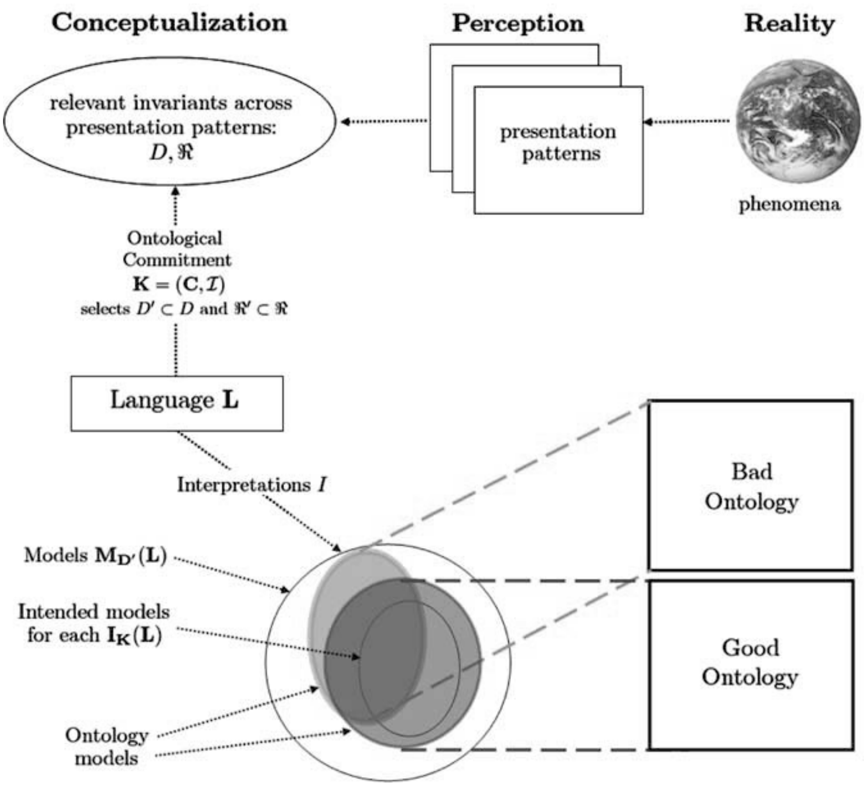

Fig. 2. The relationships between phenomena occurring in reality, their perception (at different times), their abstracted conceptualization, the language used to talk about such conceptualization, its intended models, and an ontology

3.1 Committing to a Conceptualization

Let us assume that our language $\mathbf{L}$ is (a variant of) a first-order logical language, with a vocabulary $\mathbf{V}$ consisting of a set of constant and predicate symbols (we shall not consider function symbols here). We shall introduce the notion of ontological commitment by extending the standard notion of a (extensional) first order structure to that of an *intensional* first order structure.

Definition 3.1 (Extensional first-order structure) *Let* $\mathbf{L}$ *be a first-order logical language with vocabulary* $\mathbf{V}$ *and* $S = (D, \mathbf{R})$ *an extensional relational structure. An* extensional first order structure *(also called* model *for* $\mathbf{L}$*) is a tuple* $M = (S, I)$*, where* I *(called* extensional interpretation function*) is a total function* $I : \mathbf{V} \to D \cup \mathbf{R}$ *that maps each vocabulary symbol of* $\mathbf{V}$ *to either an element of* D *or an extensional relation belonging to the set* $\mathbf{R}$.

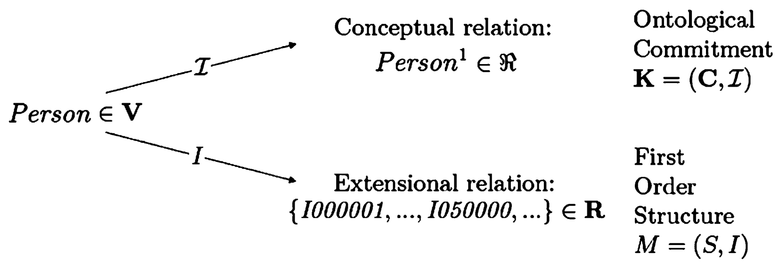

Fig. 3. The predicate symbol *Person* has both an extensional interpretation (through the usual notion of model, or extensional first-order structure) and an intensional interpretation (through the notion of ontological commitment, or intensional first order structure)

Definition 3.2 (Intensional first-order structure) (also called: ontological commitment) *Let* $\mathbf{L}$ *be a first-order logical language with vocabulary* $\mathbf{V}$ *and* $\mathbf{C} = (D, W, \Re)$ *an intensional relational structure (i.e., a conceptualization). An* intensional first order structure *(also called* ontological commitment*) for* $\mathbf{L}$ *is a tuple* $\mathbf{K} = (\mathbf{C}, \mathcal{I})$*, where* $\mathcal{I}$ *(called* intensional interpretation function*) is a total function* $\mathcal{I} : \mathbf{V} \to D \cup \Re$ *that maps each vocabulary symbol of* $\mathbf{V}$ *to either an element of* D *or an intensional relation belonging to the set* $\Re$*.*

It should be clear now that the definition of ontological commitment extends the usual (extensional) definition of "meaning" for vocabulary symbols to the intensional case, substituting the notion of model with the notion of conceptualization. Figure 3 captures this idea.

Example 3.1 *Coming back to our Example 2.1, the vocabulary* $\mathbf{V}$ *coincides with the relation symbols, i.e.,* $\mathbf{V} = \{$*Person, Manager, Researcher, reports-to, cooperates-with*$\}$*. Our ontological commitment consists of mapping the relation symbol Person to the conceptual relation* $Person^1$ *and proceeding alike with Manager, Researcher, reports-to, and cooperates-with.*

3.2 Specifying a Conceptualization

As we have seen, the notion of ontological commitment is an extension of the standard notion of model. The latter is an extensional account of meaning, the former is an intensional account of meaning. But what is the relationship between the two? Of course, once we specify the intensional meaning of a vocabulary through its ontological commitment, somehow we also constrain its models. Let us introduce the notion of *intended model* with respect to a certain ontological commitment for this purpose.

Definition 3.3 (Intended models) *Let* $\mathbf{C} = (D, W, \Re)$ *be a conceptualization,* $\mathbf{L}$ *a first-order logical language with vocabulary* $\mathbf{V}$ *and ontological* *commitment* $\mathbf{K} = (\mathbf{C}, \mathcal{I})$. *A model* $M = (S, I)$, *with* $S = (D, \mathbf{R})$, *is called an* intended model *of* $\mathbf{L}$ *according to* $\mathbf{K}$ *iff*

1. *For all constant symbols* $c \in \mathbf{V}$ *we have* $I(c) = \mathcal{I}(c)$
2. *There exists a world* $w \in W$ *such that, for each predicate symbol* $v \in \mathbf{V}$ *there exists an intensional relation* $\rho \in \Re$ *such that* $\mathcal{I}(v) = \rho$ *and* $I(v) = \rho(w)$

The set $\mathbf{I_K}(\mathbf{L})$ *of all models of* $\mathbf{L}$ *that are compatible with* $\mathbf{K}$ *is called the set of* intended models *of* $\mathbf{L}$ *according to* $\mathbf{K}$.

Condition 1 above just requires that the mapping of constant symbols to elements of the universe of discourse is identical. Example 2.1 does not introduce any constant symbols. Condition 2 states that there must exist a world such that every predicate symbol is mapped into an intensional relation whose value, for that world, coincides with the extensional interpretation of such symbol. This means that our intended model will be – so to speak – a description of that world. In Example 2.1, for instance, we have that, for w_1, $I(Person) = \{I000001, ..., I050000, ...\} = Person^1(w_1)$ and $I(reports\text{-}to) = \{..., (I046758,\ I034820), (I044443,\ I034820), (I034820,\ I050000)\ , ...\} = reports\text{-}to^2(w_1)$.

With the notion of intended models at hand, we can now clarify the role of an ontology, considered as a logical theory designed to account for the intended meaning of the vocabulary used by a logical language. In the following, we also provide an ontology for our running example.

Definition 3.4 (Ontology) *Let* $\mathbf{C}$ *be a conceptualization, and* $\mathbf{L}$ *a logical language with vocabulary* $\mathbf{V}$ *and ontological commitment* $\mathbf{K}$. *An* ontology $\mathbf{O_K}$ *for* $\mathbf{C}$ *with vocabulary* $\mathbf{V}$ *and ontological commitment* $\mathbf{K}$ *is a logical theory consisting of a set of formulas of* $\mathbf{L}$, *designed so that the set of its models approximates as well as possible the set of intended models of* $\mathbf{L}$ *according to* $\mathbf{K}$ *(cf. also Fig. 2).*

Example 3.2 *In the following we build an ontology O consisting of a set of logical formulae. Through* O_1 *to* O_6 *we specify our human resources domain with increasing precision.*

Taxonomic Information. We start our formalization by specifying that Researcher and Manager are sub-concepts of Person:
$O_1 = \{Researcher(x) \rightarrow Person(x), Manager(x) \rightarrow Person(x)\}$

Domains and Ranges. We continue by adding formulae to O_1 *which specify the domains and ranges of the binary relations:*
$O_2 = O_1 \cup \{cooperates\text{-}with(x, y) \rightarrow Person(x) \wedge Person(y), reports\text{-}to(x, y) \rightarrow Person(x) \wedge Person(y)\}$

Symmetry. cooperates-with can be considered a symmetric relation:
$O_3 = O_2 \cup \{cooperates\text{-}with(x, y) \leftrightarrow cooperates\text{-}with(y, x)\}$

Transitivity. Although arguable, we specify reports-to as a transitive relation:
$O_4 = O_3 \cup \{reports\text{-}to(x, z) \leftarrow reports\text{-}to(x, y) \wedge reports\text{-}to(y, z)\}$

Disjointness There is no Person who is both a Researcher and a Manager:
$O_5 = O_4 \cup \{Manager(x) \rightarrow \neg Researcher(x)\}$

3.3 Choosing the Right Domain and Vocabulary

On the basis of the discussion above, we might conclude that an ideal ontology is one whose models exactly coincide (modulo isomorphisms) with the intended ones. Things are not so simple, however: even a "perfect" ontology like that may fail to exactly specify its target conceptualization, if its *vocabulary* and its *domain of discourse* are not suitably chosen. The reason for that lies in the distinction between the *logical* notion of *model* and the *ontological* notion of *possible world*. The former is basically a combination of assignments of abstract relational structures (built over the domain of discourse) to vocabulary elements; the latter is a combination of actual (observed) states of affairs of a certain system. Of course, the number of possible models depends both on the size of the vocabulary and the extension of the domain of discourse, which are chosen more or less arbitrarily, on the basis of what appears to be relevant to talk of. On the contrary, the number of world states depends on the observed variables, even those which – at a first sight – are considered as irrelevant to talk of. With reference to our example, consider the two models where the predicates of our language (whose signature is reported above) are interpreted in such a way that their extensions are those described respectively in Examples 2.1 and 2.2. Each model corresponds to a different pattern of relationships among the people in our company, but, looking at the model itself, nothing tells us what are the world states where a certain pattern of relationships holds. So, for example, it is impossible to discriminate between a conceptualization where *cooperates-with* means that two persons cooperate when they are just sharing a goal, and another where they need also do something to achieve that goal. In other words, each model, in this example, will "collapse" many different world states. The reason of this is in the very simple vocabulary we have adopted: with just two predicates, we have not enough expressiveness to discriminate between different world states. So, to really capture our conceptualization, we need to extend the vocabulary in order to be able to talk of sharing a goal or achieving a goal, and we have to introduce goals (besides persons) in our domain of discourse. In conclusion, the degree to which an ontology specifies a conceptualization depends (1) on the richness of the domain of discourse; (2) on the richness of the vocabulary chosen; (3) on the axiomatization. In turn, the axiomatization depends on language expressiveness issues as discussed in Sect. 3.4.

3.4 Language Expressiveness Issues

At one extreme, we have rather informal approaches for the language **L** that may allow the definitions of terms only, with little or no specification of the

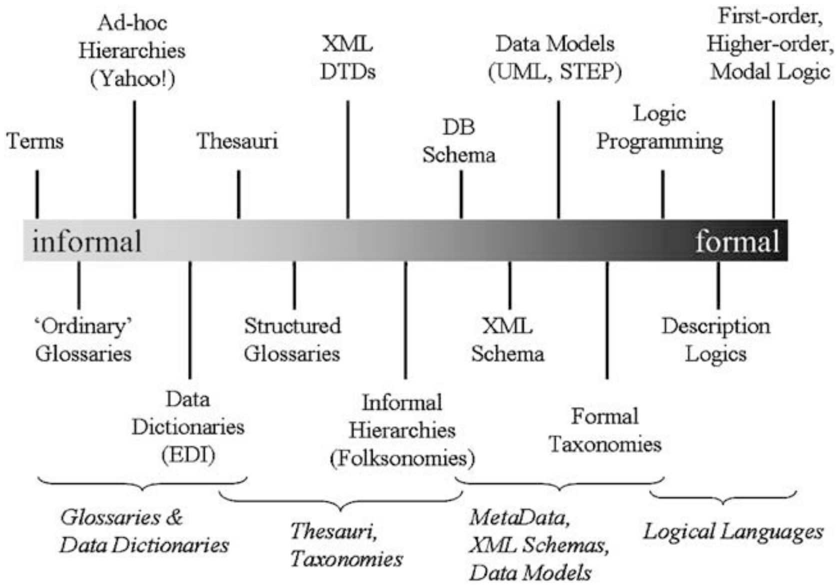

Fig. 4. Different approaches to the language **L** according to [17]. Typically, logical languages are eligible for the formal, explicit specification, and, thus, ontologies

meaning of the term. At the other end of the spectrum, we have formal approaches, i.e., logical languages that allow specifying rigorously formalized logical theories. This gives rise to the continuum introduced by [17] and depicted in Fig. 4. As we move along the continuum, the amount of meaning specified and the degree of formality increases (thus reducing ambiguity); there is also increasing support for automated reasoning (cf. Chapters "Tableau-Based Reasoning" and "Resolution-Based Reasoning for Ontologies").

It is difficult to draw a strict line of where the criterion of formal starts on this continuum. In practice, the rightmost category of logical languages is usually considered as formal. Within this rightmost category one typically encounters the trade-off between *expressiveness* and *efficiency* when choosing the language **L**. On the one end, we find higher-order logic, full first-order logic, or modal logic. They are very expressive, but do often not allow for sound and complete reasoning and if they do, reasoning sometimes remains untractable. At the other end, we find less stringent subsets of first-order logic, which typically feature decidable and more efficient reasoners. They can be split in two major paradigms. First, languages from the family of *description logics (DL)* (cf. chapter "Description Logics"), e.g., OWL-DL (cf. chapter "Web Ontology Language: OWL"), are strict subsets of first-order logic. The second major paradigm comes from the tradition of *logic programming (LP)* [3] with one prominent representor being F-Logic (cf. chapter "Ontologies in F-Logic"). Though logic programming often uses a syntax comparable to first-order logics, it assumes a different interpretation of formulae. Unlike the Tarski-style model theory [18] of first-order and description logic, logic programming selects only a subset of models to judge semantic entailment of formulae. There are different ways to select subsets of models resulting in different semantics – all of them geared to deal more efficiently with larger sets of data than approaches based on first-order logic. One of the most prominent differences resulting from this different style of logical models is that expressive logic programming theories become non-monotonic.

4 Why is *Shared* of Importance?

A formal specification of a conceptualization does not need to be a specification of a *shared* conceptualization. As outlined above, the first definitions of "ontologies" did not consider the aspect of sharing [6, 8] and only later it was introduced by Borst [1]. Indeed, one may correctly argue that it is not possible to share whole conceptualizations, which are private to the mind of the individual. What can be shared, are approximations of conceptualizations based on a limited set of examples and showing the actual circumstances where a certain conceptual relation holds (for instance, actual situations showing cases where the *cooperates-with* relationship occurs). Beyond mere examples it is also possible to share meaning postulates, i.e., explicit formal constraints (e.g., the relationship *cooperates-with* is symmetric). Such definitions, however, presuppose a mutual agreement on the primitive terms used in these definitions. Since however meaning postulates cannot fully characterize the ontological commitment of primitive terms, one may recognize that sharing of conceptualizations is at best partial.

For practical usage of ontologies, it turned out very quickly that without at least such minimal shared ontological commitment from ontology stakeholders, the benefits of having an ontology are limited. The reason is that an ontology formally specifies a domain structure *under the limitation* that its stakeholder understand the primitive terms in the appropriate way. In other words, the ontology may turn out useless if it is used in a way that runs counter to the shared ontological commitment. In conclusion, any ontology will always be less complete and less formal than it would be desirable in theory. This is why it is important, for those ontologies intended to support large-scale interoperability, to be *well-founded*, in the sense that the basic primitives they are built on are sufficiently well-chosen and axiomatized to be generally understood.

4.1 Reference and Meaning

For appropriate usage, ontologies need to fulfill a further function, namely facilitating the communication between the human and the machine – referring to terminology specified in the ontology – or even for facilitating inter-machine and inter-human communication. The communication situation can

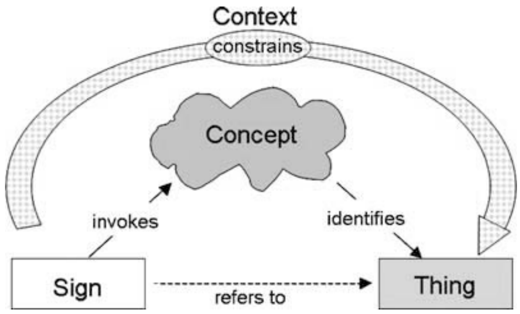

Fig. 5. Semiotic triangle

be illustrated using the semiotic triangle of Ogden and Richard [12], following thoughts by Peirce, Saussure, and Frege (cf. Fig. 5).

All agents, whatever their commitment to an ontology is, find themselves in a communication situation illustrated using the semiotic triangle: The sender of a message may use a word or – more generally – a sign like the string "Person" to stand for a concept the sender has in his own "mind." He uses the sign in order to refer to abstract or concrete things in the world, which may, but need not be, physical objects. The sender also invokes a concept in the mind of an actor receiving this sign. The receiver uses the concept in order to point out the individual or the class of individuals the sign was intended to refer to. Thereby, the interpretation of the sign as a concept as well as its use in a given situation depends heavily on the receiver as well as the overall communication context. Therefore, the meaning triangle is sometimes supplemented with further nodes in order to represent the receiver or the context of communication. We have illustrated the context by an instable arrow from sign to thing that constrains possible acts of reference. Note that the act of reference remains indirect, as it is mediated by the mental concept. Once the concept is invoked, it behaves (so to speak) as a function that, given a particular context (i.e., the world state mentioned in previous sections), returns the things we want to refer to. Moreover, the correspondences between sign, concept, and thing are weak and ambiguous. In many communication circumstances, the usage of signs can erroneously invoke the wrong concepts and represent different entities than intended to.

This problem is further aggravated when a multitude of agents exchanges messages in which terms do not have a prescribed meaning. Unavoidably, different agents will arrive at different conclusions about the semantics and the intention of the message.

When agents commit to a common ontology they can limit the conclusions possibly associated with the communications of specific signs, because not all relations between existing signs may hold and logical consequences from the usage of signs are implied by the logical theory specifying the ontology. Therefore the set of possible correspondences between signs, concepts and

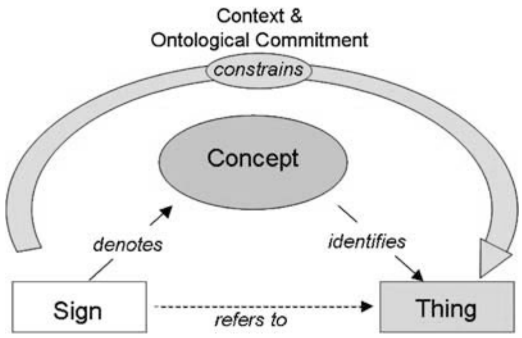

Fig. 6. Semiotic triangle revisited

real-world entities is strongly reduced – ideally up to a situation where the message becomes completely unambiguous (cf. Fig. 6). Thereby, not only the act of reference becomes clearer, but also the connection between sign and concept changes from a weakly defined relationship of "invokes" into a logically precise meaning of "denotes." Likewise, the meaning of a concept is now determined by a precise logical theory (contrast Figs. 5 and 6).

5 Discussion

In this chapter we have introduced three core aspects of computational ontologies: conceptualizations, specifications of conceptualizations, and shared ontological commitments. These are very broad categories suitable to investigate many different formalisms and fields of applications.

In fact, they are not even the only aspects of ontologies, which may be classified into different types, depending on the way they are used. For instance, the primary purpose of *top-level ontologies* lies in providing a broad view of the world suitable for many different target domains. *Reference ontologies* target the structuring of ontologies that are derived from them. The primary purpose of *core ontologies* derives from the definition of a super domain. *Application ontologies* are suitable for direct use in reasoning engines or software packages – and this list is not yet complete and will require many more experiences yet to be made.

Acknowledgments

We would like to thank our colleagues Aldo Gangemi, Susan Marie Thomas, Marta Sabou, as well as Pascal Hitzler for their fruitful reviews and discussions that helped to shape this contribution.

References

1. W. Borst. *Construction of Engineering Ontologies.* PhD thesis, Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1997.
2. R. Carnap. *Meaning and Necessity – A Study in Semantics and Modal Logic.* The University of Chicago Press, second edition, 1956.
3. S. K. Das. *Deductive Databases and Logic Programming.* Addison Wesley, 1992.
4. D. R. Dowty, R. Wall, and S. Peters. *Introduction to Montague Semantics*, volume 11 of *Studies in Linguistics and Philosophy.* Springer, Heidelberg, 1980.
5. M. R. Genesereth and N. J. Nilsson. *Logical Foundations of Artificial Intelligence.* Morgan Kaufmann, Los Altos, CA, 1987.
6. T. R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation.* Kluwer Academic Publishers, Deventer, The Netherlands, 1993.
7. T. R. Gruber. A Translation Approach to Portable Ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.
8. T. R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human Computer Studies*, 43(5–6): 907–928, 1995.
9. N. Guarino and P. Giaretta. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars, editor, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, pages 25–32. IOS Press, Amsterdam, 1995.
10. N. Guarino. Formal Ontology in Information Systems. In N. Guarino, editor, *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, June 6-8, 1998*, pages 3–15. IOS Press, Amsterdam, 1998.
11. J. Hendler. Agents and the Semantic Web. *IEEE Intelligent Systems*, 16(2): 30–37, 2001.
12. C. K. Ogden and I. A. Richards. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism.* Routledge & Kegan Paul Ltd., London, tenth edition, 1923.
13. P. Øhrstrøm, J. Andersen, and H. Schärfe. What has happened to ontology. In F. Dau, M.-L. Mugnier, and G. Stumme, editors, *Conceptual Structures: Common Semantics for Sharing Knowledge, 13th International Conference on Conceptual Structures, ICCS 2005, Kassel, Germany, July 17–22, 2005, Proceedings*, volume 3596 of *Lecture Notes in Computer Science*, pages 425–438. Springer, Heidelberg, 2005.
14. B. Smith. Beyond Concepts: Ontology as Reality Representation. In A. C. Varzi and L. Vieu, editors, *Formal Ontology in Information Systems – Proceedings of the Third International Conference (FOIS 2004)*, pages 73–85. IOS Press, Amsterdam, 2004.
15. R. Studer, R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1–2):161–198, 1998.
16. W. Swartout and A. Tate. Ontologies. *IEEE Intelligent Systems*, 14(1):18–19, 1999.
17. M. Uschold. Ontologies and Semantics for Seamless Connectivity. *SIGMOD Record*, 33(4):58–64, 2004.
18. R. L. Vaught. Alfred Tarski's Work in Model Theory. *The Journal of Symbolic Logic*, 51(4):869–882, 1986.