Adapted from: PRETORIUS, A.J., *"Lexon Visualisation: Visualising Binary Fact Types in Ontology Bases"*, Chapter 2, Unpublished MSc Thesis, Brussels, Vrije Universiteit Brussel, 2004.

Ontologies - Introduction and Overview

A. Johannes Pretorius
Semantic Technology and Applications Research Laboratory
Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
hannes.pretorius@vub.ac.be

1. Introduction

Tim Berners-Lee, creator of the World Wide Web, foresees a future when the Web will be more than just a collection of web pages (Berners-Lee et al., 2001). In this future, computers themselves will be able to consider the meaning, or *semantics*, of information sources on the Web. This will enable computer programs to perform complex tasks autonomously and to communicate amongst one another, and with humans, by being able to *meaningfully* interpret the wealth of knowledge that is available on the Web.

It is anticipated that humans will make use of, or even own, one or several such programs or *agents*. Agents might be expected to run errands for human users, like searching and making an appointment for a consultation with a medical doctor. In performing such a task, an agent is likely to take into account aspects such as the schedules of both parties, the practitioner's professional rating, potential health insurance issues and the locale of the consultation rooms. All of this will happen in the background, on the *Semantic Web*, with the patient only being informed by the agent of the date and time of the appointment. This he or she may then confirm or decline.

2. The Semantic Web

As already alluded to, Berners-Lee has coined this vision of the future the *Semantic Web*. For scenarios like the one sketched above to become a reality, however, the current Web needs to evolve into more than just a platform that provides humans with access to information and more than just an infrastructure that allows individuals and organisations to host websites. What is required is for it to become a mechanism that permits machines to meaningfully collect content from miscellaneous sources, process the information and exchange the results with other agents and human users. Only when the Web becomes *machine accessible*, it is argued, will it reach its full potential.

2.1. The World Wide Web

The Web in its current form is a remarkable success in its own right. The number of active users and information resources on the Web is increasing at an accelerating pace; any attempt at an estimate is futile and instantly outdated (Bray, 1996). To get some sense of its phenomenal growth, consider that the original work for the World Wide Web was done by Tim Berners-Lee at CERN (European

Laboratory for Particle Physics) in Geneva in the late 1980s. By early 1994, more than 50000 copies of Marc Andreessen's initial Mosaic web browser were being downloaded monthly from the NCSA's (National Center for Supercomputer Applications) public server at the University of Illinois (Mitchell, 1999). This was even before the Web had been embraced as a strategic resource by corporations world-wide.

The success of the World Wide Web can be attributed to its simplicity; little technical expertise is required to add content to it. Also, access to the Web is relatively uncomplicated. The decentralised nature of the Web makes it possible for service providers globally to offer connectivity to users. Linking to the Web is never further than a dialup connection away, with more users being connected permanently to high-bandwidth networks offering fast access to the Web.

Finding meaningful information among the millions of information resources on the Web is an entirely different matter though. The sheer volume of information currently available on the Web makes its limitations apparent; finding relevant information is a daunting task. Users are currently restricted to browsing (following hyperlinks from one web page to the next) and syntactic key-word searches (where the patterns of the characters that make up the key-words are matched to strings occurring in resources on the Web) for finding significant information. This often leads to a rather frustrating exercise based largely on guess-work and (often misguided) intuition. It suffices to say that there is a lot out there, but no way to determine what a particular information resource is about without actually viewing it yourself. That is if you can find it in the first place.

2.2. Semantics and the Web

The Semantic Web is intended to offer a solution to the inherent dilemma described above. More concretely, research aimed at realising the Semantic Web is an attempt to transform the current Web into an information space with a *semantic organisational foundation*; an information space that makes information semantically accessible to machines by considering its meaning (Maedche, 2003).

An information resource on the Semantic Web will not only contain data, but will also consist of *metadata* which describe what the data are about. This will allow agents and their human users to identify, collect and process suitable information sources by interpreting the semantic metadata based on the task at hand. They are also free to exchange results and to communicate by sharing such resources. The semantic foundation mentioned above will be provided by *ontologies* (Berners-Lee et al., 2001).

3. Ontologies

Although it is currently somewhat of a catchword, there exists some disagreement amongst researchers as to the definition of *ontology* (Noy and Hafner, 1997 and Pisanelli et al., 2002). For this reason and the fact that the term often takes on different meanings in different contexts, a short historical overview of the concept of ontology is in order.

3.1. Origins

An important distinction that should be drawn is between the notions of *Ontology* and *ontology* (Guarino, 1998). The difference is subtle but important. The former, written with a capitalised 'O', is an uncountable noun with no plural. It refers to the philosophical discipline that studies the nature of being. It is an old discipline introduced by Aristotle, which attempts to address questions such as: *'What is being?'* and *'What characteristics do all beings have in common?'*. In this sense, it would be correct to say: *'I think that Ontology is an interesting discipline'*.

When written with a lowercase 'o', and still considered in a philosophical sense, *ontology* refers to a system of categories (or frames of reference) that account for a certain view of the world. Gruber (1993: 199) calls it a *"systematic account of existence"*. In this form it is a countable noun for which a plural form, ontologies, exists. An ontology is independent of natural language, but reliant on a particular philosophical view. An individual is in possession of some *concepts* in his or her mental model (for instance, the concept of a *'Mouse'*, a small rodent that could be perceived as cute or intimidating, depending on your point of view) which are independent of language (*'Mouse'* in English, *'Muis'* in Afrikaans or *'Maus'* in German all refer to the same concept). This distinction is often made explicit by referring to *lexical* knowledge (the term used to refer to the concept) and *non-lexical* knowledge (the concept itself). In this sense, it would be correct to say *'I find Aristotle's ontology interesting'*.

3.2. Designed artefacts

The concept of *ontology* (written with a lowercase 'o') was first borrowed from the realm of Philosophy by Artificial Intelligence researchers and has since become a matter of interest to computer and information scientists in general. In Computer Science literature, the term takes on a new meaning, not entirely unrelated to its philosophical counterpart. An ontology is generally regarded as a *designed artefact* consisting of a specific *shared vocabulary* used to describe entities in some *domain of interest*, as well as a *set of assumptions* about the intended meaning of the terms in the vocabulary (Guarino, 1998). According to Gruber (1995), such an approach makes it possible to design ontologies with a specific goal in mind and also allows for the resulting artefact (the ontology) to be evaluated against objective design criteria.

A frequently cited, and perhaps the most prevalent, definition of ontology is attributed to Tom Gruber who defines it as "*an explicit specification of a conceptualisation*" (Gruber, 1993: 199). This definition borrows from the Artificial Intelligence literature on *Declarative Knowledge*, which is concerned with the formal symbolic representation of knowledge (Genesereth and Nilsson, 1987). In this field, formal logical languages, such as first-order predicate calculus, are used to declaratively describe models of the world. This is necessary because natural languages are too ambiguous for machine interpretation.

Fundamental to the Declarative Knowledge approach is the notion of *conceptualisation*. That is, an abstract and simplified view of that world, or *domain of interest*, which is being represented. This domain could be a part of reality or an entirely fictitious environment. Such a conceptualisation consists of *objects* or *entities* that are assumed to exist in the domain of interest as well as the *relationships* that hold between them (Genesereth and Nilsson, 1987). These relationships can also be interpreted as *roles* that the entities play with respect to one another (Halpin, 1999). Collectively, the set of objects about which knowledge is being expressed is referred to as the *universe of discourse*.

As mentioned above, the universe of discourse and the relationships that hold in it are expressed in a declarative formal vocabulary that collectively constitutes the knowledge about a domain; that 'which is known' about it. In this light, Noy and

Hafner (1997) state that the task of intelligent systems, in general, is to formally represent concepts. Consequently, every system *commits* to some conceptualisation implicitly or explicitly. In the light of Gruber's definition, an explicit specification of such a conceptualisation is called an ontology.

3.3. Ontological commitment

According to Gruber (1993) the ontology of a shared domain can be described by defining a set of representational *terms*. These terms (lexical references) are associated with entities (non lexical referents) in the universe of discourse. Formal axioms are also introduced to constrain their interpretation and well-formed use. In this respect, an ontology is viewed as the explicit statement of a logical theory. Indeed, in the context of the Semantic Web, "*ontologies describe domain theories with the intent of the explicit representation of the semantics of the domain data"* (Maedche, 2003: 3). Although such ontologies often assume the form of a taxonomic class hierarchy, they are not restricted to hierarchies in any way. In fact, ontologies may take on the form of much more general and complex structures (Gruber, 1995).

In order for knowledge to be shared amongst agents, agreement must exist on the topics about which are being communicated. This raises the issue of *ontological commitment* which Gruber (1993: 201) describes as "*the agreements about the objects and relations being talked about among agents*". Therefore, when agents *commit* to an ontology, there is an *agreement* with respect to the semantics of the objects and relations represented. Furthermore, there is agreement to use the shared vocabulary in a *coherent* and *consistent* manner. In this way, the vocabulary is formalised with the aim of making the mapping from terms to things as exact as possible (Maedche, 2003).

Ontological commitments allow for a number of agents to meaningfully communicate about a domain without necessarily operating on a globally shared theory individually. In the context of multiple agents, a *common ontology* serves as a *knowledge-level* specification of the ontological commitments of a set of participating agents. (This is opposed to an individual agent's internal and implementation specific *symbol-level* representation of knowledge). A common ontology defines the vocabulary with which queries and assertions are exchanged among agents, thereby providing the means to bridge the semantic gap that exists between the lexical representations of information and its non-lexical conceptualisation (Maedche, 2003).

As mentioned above, in order to facilitate meaningful communication, an agent must commit to the semantics of the terms and relationships in the common ontology. This includes axioms about properties of objects and how they are related. It is important to note that even though commitment to a common ontology is a guarantee of *consistency*, it does not guarantee *completeness* with respect to queries and assertions made using the vocabulary defined in the ontology (Gruber, 1993 and Gruber, 1995). No guarantee is made with respect to which queries an agent must be able to answer. In fact, the axiomatisation of an ontology is not a functional specification of an agent's *behaviour*. Instead, it typically only specifies some of the formal constraints of objects in the domain of discourse.

3.4. Example 1

Consider Figure 1 for a more concrete synthesis of the above discussion. In the figure there exists a shared vocabulary of terms with which two agents communicate. In this case, the string *'Mouse'* constitutes one such term. However, nothing is gained if no semantics or meaning is attributed to these terms. To facilitate this, an agent's internal model attributes meaning to a term by establishing a link between the lexical reference and that concept in the world that it represents or refers to.

Now, as far as an agent is concerned, the term *'Mouse'* might just as well refer to a furry rodent as to a computer input device. Conversely, if there is a *commitment* to an ontology (a computer hardware ontology, in this case) then there exists agreement that *'Mouse'* refers to an input device. In fact, it is likely to be formally specified that the term *'Mouse'* has a subsumption relationship with respect to the term *'Input_device'* and that an *'Input_device'* plays certain roles with respect to other concepts. In this example, the agreement that exists with respect to the meaning of *'Mouse'* allows for meaningful and unambiguous communication between agents. To facilitate such communication the ontology serves as 'background knowledge' that rules out incorrect references.

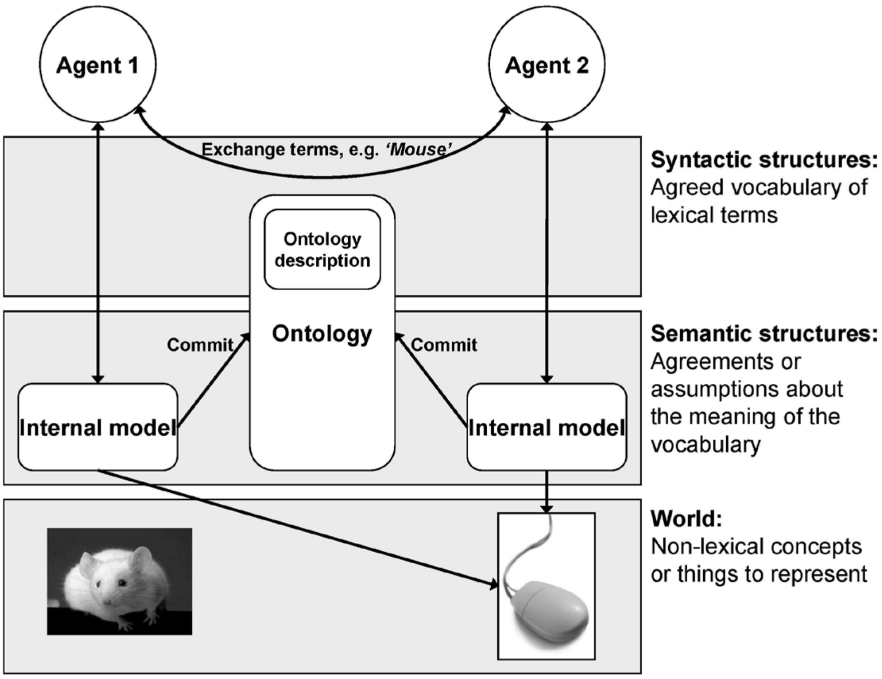

Figure 1. Graphical representation of an ontology (adapted from Maedche, 2003).

3.5. Design criteria

When ontologies are viewed as designed artefacts, the question of when and how to represent something in such an ontology becomes a *design decision*. As mentioned in Section 3.2, this approach allows for such an artefact to be evaluated against objective criteria, rather than basing it on some notion of *truth*. To this end, Gruber (1995) suggests five design criteria for ontologies which are considered in more detail below:

1. **Clarity:** definitions should be formal, complete, objective and independent of social or computational context. This results in restricting the number of possible interpretations of a concept, thereby contributing to the effectiveness of communication between agents.
2. **Coherence:** only inferences consistent with existing definitions should be allowed.
3. **Extendibility:** the ontology should be designed to serve as a conceptual foundation for a range of anticipated tasks. It should be possible to extend the ontology without altering the existing definitions. If this is the case, it is not

necessary to include a vocabulary sufficient to express the knowledge related to all anticipated extensions since the mechanisms to define the required specialisations are already in place.

4. **Minimal encoding bias:** conceptualisations should be specified at a knowledge-level. Representation choices should not be based on convenience of notation or implementation issues at a symbol-level.
5. **Minimal ontological commitment:** the minimum ontological commitment sufficient to support the intended knowledge sharing activities should be allowed. Accordingly, the weakest theory necessary to facilitate communication consistent with the conceptualisation should be specified. This allows agents to more easily extend the ontology where needed for their individual purposes.

In the light of these criteria it is worth noting that it is only necessary for a shared ontology to describe a vocabulary for *communicating about* a domain. In contrast, a *knowledge-base* contains the knowledge needed to solve problems or answer queries about such a domain by committing to an ontology. The difference between an ontology and a knowledge base can also be described in terms of their different objectives (Maedche, 2003); an ontology aims to capture the conceptual structures of a domain while a knowledge-base aims to specify a concrete state of the domain. Put differently, an ontology consists of *intensional* logical definitions (characteristics that distinguish concepts), while a knowledge-base comprises of *extensional* parts (entity instances). Guarino (1998) makes the distinction between *state-independent* and *state-dependent* information.

3.6. Formal definition

The following formal definition is adapted from Maedche (2003). An ontology structure O is defined as

$$O = \{C, R, A^O\},$$

Where:

1. C is a set whose elements are called *concepts*.
2. $R \subseteq C \times C$ is a set whose elements are called *relations*. For $r = (c_1, c_2) \in R$, one may write $r(c_1) = c_2$.
3. A^O is a set of axioms on O.

To cope with the lexical level, the notion of a *lexicon* is introduced. For an ontology structure $O = \{C, R, A^O\}$ a lexicon L is defined as

$$L = \{L^C, L^R, F, G\},$$

Where:

1. L^C is a set whose elements are called *lexical entries for concepts*.
2. L^R is a set whose elements are called *lexical entries for relations*.
3. $F \subseteq L^C \times C$ is a *reference for concepts* such that
$$F(l_C) = \{c \in C : (l_C, c) \in F\} \text{ for all } l_C \in L^C,$$
$$F^{-1}(c) = \{l_C \in L^C : (l_C, c) \in F\} \text{ for all } c \in C.$$
4. $G \subseteq L^R \times R$ is a *reference for relations* such that
$$G(l_R) = \{r \in R : (l_R, r) \in G\} \text{ for all } l_R \in L^R,$$
$$G^{-1}(r) = \{l_R \in L^R : (l_R, r) \in G\} \text{ for all } r \in R.$$

It is noted that this definition allows for a lexical entry to refer to several concepts or relations *(homonymy)* and for one concept or relation to be referred to by several lexical entries *(synonymy)*. Furthermore, Maedche (2003) includes a concept hierarchy $H^C \subseteq C \times C$ to express the inherent hierarchical structure of concepts. It is felt that this is redundant, especially in light of the present discussion, since a hierarchical structure can be explicitly defined in terms of *R*. Finally, the mapping from elements in the lexicon to elements in the ontology structure corresponds to the notion of an *interpretation* in declarative semantics (Genesereth and Nilsson, 1987). Such a (semantic) interpretation associates elements of a language to elements of a conceptualisation.

Based on the above definitions, an ontology can be formally defined to be a structure $\langle O, L \rangle$ where *O* is an ontology structure and *L* is a corresponding lexicon. In the light of the intuitive discussion of ontologies in Section 3.2, the ontology structure *O* plays the role of an e*xplicit specification of a conceptualisation* of some domain while the lexicon *L* provides the *agreed vocabulary* to communicate about this conceptualisation.

3.7. Example 2

As an illustration of the above definitions, consider the following example. Let $O = \{C, R, A^O\}$ be an ontology structure such that $C = \{c_1, c_2\}$ and $R = \{r\}$ where $r(c_1) = c_2$. Suppose that $A^O = \varnothing$. Also, let $L = \{L^C, L^R, F, G\}$ be a corresponding lexicon such that $L^C = \{'Mouse', 'Input\_device'\}$, $L^R = \{'is\_a'\}$, $F('Mouse') = c_1$, $F('Input\_device') = c_2$ and $G('is\_a') = r$.

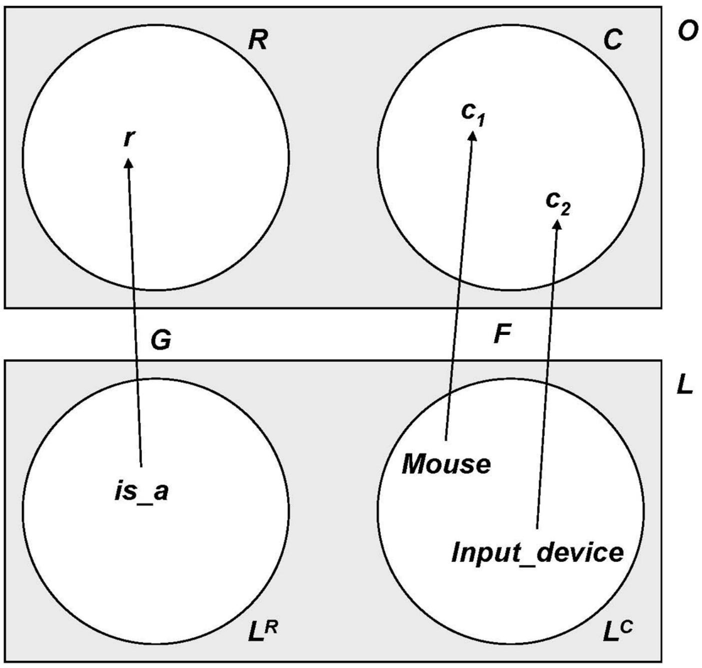

Figure 2. Graphical depiction of an instantiated ontology.

Figure 2 depicts the elementary ontology described above graphically. The top of the figure depicts the ontology structure *O*. It consists of two concepts c_1, c_2 and a relation *r* that relates them. This corresponds to a conceptualisation of a domain of interest without any lexical references. The latter is provided by the lexicon *L* depicted at the bottom of the figure. L provides the lexical references for O by means of F and G. F and G respectively map lexical reference strings to the concepts and relations defined in O. For instance, for $r \in R$ one may consider $G^{-1}(r)$ to get the lexical reference, '*is_a'*, corresponding to r and vice versa.

4. Further applications of ontologies

Gruber (1993) defines one of the long term goals of ontology research as that of specifying content-specific agreements to enable libraries of reusable knowledge components and knowledge based services that can be invoked over networks. A contemporary interpretation of this may be to provide *web services* which can be invoked over the Semantic Web. However, it is not only the Semantic Web that serves as a thrust and justification for research on ontologies.

According to Noy and Hafner (1997) ontologies are designed and built for various reasons including *natural language application*s, *theoretical investigation* and *simulation and modelling*. Natural language applications include knowledge acquisition from text and semantic information retrieval (these applications can easily be related back to the Semantic Web, of course). An example of a natural language application is WORDNET, an online lexical reference system (as cited in Noy and Hafner, 1997). The goal of theoretical investigations is to investigate the philosophical foundations of knowledge models (Sowa's ontology, for instance, as cited in Noy and Hafner, 1997). Finally, simulation and modelling implies constructing models of scientific knowledge for further investigation. One application of this approach is the modelling of objects and processes in molecular biology and biochemistry in GEN-SIM (as cited in Noy and Hafner, 1997).

Above all, ontologies are being put to use to facilitate *knowledge sharing and reuse* (Noy and Hafner, 1997). Pisanelli et al. (2002) argue that *information systems* in general, no longer operate as 'islands', but that they are often interdependent. Sharing and reusing knowledge across such systems is not easy since different systems typically use different terms to refer to the same concepts and vice versa. In this respect it is useful to define a common vocabulary in which shared knowledge may be represented. There is an increased demand for data sharing in such *open systems* which must rely on concrete conceptual frameworks to give precise meaning to the data.

Noy and Hafner (1997) also state that there exists widespread agreement in the ontology engineering community that it would be of great benefit to integrate ontologies so that they can share and reuse each other's knowledge. Currently a large number of distributed computerised resources such as lexicons, thesauri and glossaries are available to information systems (Meersman, 2001). By implementing ontologies, these resources can be integrated with information systems to greatly ease knowledge sharing and hence increase the individual and collective systems' overall value.

Finally, it is generally accepted that *knowledge management*, the corporate control of an organisation's business data and metadata, is of vital strategic importance. In this respect, it has been shown how ontologies can be effectively put to use to model *business semantics* (Zhao et al., 2004) and *collaborative business processes* (Zhao and Meersman, 2003). However, ontologies do not only provide the means to address issues such as complexity, scalability and size in business data. In addition, research results with regards to methodologies and tools also promise to assist with the *introduction* and *management* of ontologies in such complex settings (Meersman, 2001). An example of one such methodology is AKEM (Application

Knowledge Engineering Methodology), discussed by Zhao and Meersman (2003) and Zhao et al. (2004).

5. Ontology bases – the DOGMA approach

As mentioned above, there exists disagreement with regards to the definition of ontology. According to Noy and Hafner (1997) and Pisanelli et al. (2002) the main point of disagreement is whether or not axioms should form part of an ontology. Proponents seem to fall mainly into two camps. There are those that see an ontology simply as a description with the aim of organising concepts, while a second group regards it as a *complete theory* consisting of both a formal vocabulary and defined axioms that allow further deductions or inferences to be made. Based on an ontology description, the latter group regards axioms as a means to represent more information about concepts and their relations to each other as well as an instrument with which to place constraints on their properties and roles.

One way to address the above issue is to adopt an approach that neatly separates the specification of concepts from their axiomatisation. The DOGMA framework developed at the VUB STARLab *(Vrije Universiteit Brussel Semantic Technology and Applications Research Laboratory)* adopts an orthodox database model-theoretic approach to ontology representation and development (Meersman, 2000). In this approach concepts and relations are regarded as separate from constraints and derivation rules. Consequently, the DOGMA framework consists of two layers: an ontology base and a commitment layer. Before considering these layers in turn, further arguments for this approach are considered below.

5.1. Semantic independence and agreement

In previous sections, the case has already been made for the need of *agreement* in order to achieve meaningful cooperation, interoperation and communication between agents, information systems and ultimately humans. As outlined in Section 3, ontologies provide the mechanism to realise these objectives. According to Meersman (2000) an ontology can be viewed as a mathematical object which serves as a *domain* for a *semantic interpretation function*. In order to facilitate agreement, this semantic interpretation function together with derivation rules and constraints are deliberately left outside of the ontology description. Furthermore, the elements that constitute this description are made to be as simple as possible.

The point of departure for this view is the observation that ontologies have the potential to lead to *semantic independence* of information and knowledge based systems (and therefore the Semantic Web as a whole), similar to the way that databases afford the ability to specify and manage data structures outside of applications (Meersman, 2000). Gruber (1993) also suggests that the role of ontologies can be considered similar to that of the *conceptual schemas* of database systems.

A Database Management System (DBMS) aims to provide data schemata for some conceptualisation that are independent of application instances. Such a conceptual schema provides logical descriptions of data, allowing application programs to interoperate without having to share data structures. Ontologies could take this notion to its logical extreme by providing the structure and semantics for entire domains. From this point of view, ontologies allow for the specification and management of domain semantics *external* to applications. Data play the role of atomic facts and ontologies define the vocabulary used to compose complex expressions from these facts. From such a well-defined vocabulary, applications are free to compose a large number of coherent sentences.

A number of observations from the methodological treatment of database and information systems development need to be considered, however (Meersman, 2001):

1. There is no such thing as absolute meaning or semantics; it can only result from *agreement*.
2. Domain axioms, constraints, derivation rules and procedures are essential to achieving an understanding of a domain's semantics. However, it is extremely difficult to get *agreement* about them. This is often only possible for a specific context or application instance.
3. Meaning or semantics make communication about a domain possible. Although concepts in such a domain are independent of language, they are necessarily rooted and described in a language. Unambiguous lexical references to concepts are hard to *agree* on.

Consequently, to facilitate *agreement* the elementariness of the ontology specification is essential (Meersman, 2000). This specification is handled by the first layer of the DOGMA framework, the ontology base. An *ontology base* consists of a set of *lexons* which model abstract concept and relationship types in a domain of interest. These generic constructs are then used to specify application specific logic in the second layer of the framework, the commitment layer.

5.2. Lexons and the ontology base

A lexon is a conceptual construct that depicts a generic semantic relationship in some domain of interest (Meersman, 2000 and Zhao and Meersman, 2003). An ontology base is defined as a set W of lexons where a lexon assumes the form of a 5-tuple

$$l = \langle \gamma, t_i, r_{i-j}, t_j, r_{j-i} \rangle \text{ for } l \in W,$$

Where:

1. $\gamma \in \Gamma$ is the *context identifier*,
2. $t_i, t_j \in T$ are *terms* referring to entities in the semantic relationship,
3. $r_{i-j}, r_{j-i} \in R$ are *roles* in the semantic relationship,
4. Γ, T and R are strings over an alphabet A^+.

In the above definition, $\gamma \in \Gamma$ refers to the *ideational context* in which the terms $t_i, t_j \in T$ and roles $r_{i-j}, r_{j-i} \in R$ are meaningful. According to Zhao and Meersman (2003: 2) the ideational context *"instils the lexon with particular semantic contents"*. In practise the ideational context consists of one or more knowledge resources such as documents and graphs which have been established, communicated, documented and agreed upon by a community of ontology engineers. Intuitively, a lexon $l = \langle \gamma, t_i, r_{i-j}, t_j, r_{j-i} \rangle$ may be interpreted as follows: in the ideational context γ, the entity t_i plays the role of r_{i-j} with respect to the term t_j. Conversely, the entity t_j plays the co-role r_{j-i} with respect to the entity t_i.

This definition leads to collections of lexons which individually depict binary relationships, also referred to as *binary fact types*. The binary nature of such relationships serves as a conceptualisation constraint to model the most basic and atomic fact types within a domain. The view is held that explicitly allowing for and including n-ary relationships would lead to the introduction of larger semantic units which are not elementary. Such non-elementary units have less flexibility and consequently lead to a lower degree of reusability. However, to allow for the introduction of higher order semantic relationships, it is possible to treat a lexon as a term in another lexon, a practise referred to as *reification* (Halpin, 1999).

It is fundamental to realise that a lexon in an ontology base represents a *fact type* of some abstract category or description and not a fact instance itself. Lexons are elementary semantic conceptions referring to semantic types and not tokens or instances of an application domain (Zhao and Meersman, 2003). It is the commitment layer which is concerned with the composition, constraining and instantiation of lexons to represent the semantics of a particular fact. Based on this exposition, an ontology base can be viewed as a set of *plausible elementary facts* that *may hold* in the domain under consideration (Meersman, 2000). This implies that an agent or application that operates in the domain should be *consistent* with this set of fact types but is able to further extend or specialise the ontology base by instantiating tokens at the level of the commitment layer. In this respect, Zhao and Meersman (2003: 2) define an ontology as *"an approximate semiotic representation of an agreed conceptualisation about a subject domain"*. In terms of the formal definition of ontology provided in Section 3.6 the ontology structure and lexicon are integrated into one representation with axioms excluded and left to the commitment layer.

5.3. The commitment layer – commitment revisited

As already mentioned, the second layer of the DOGMA framework is the *commitment layer*. In the DOGMA framework, a *commitment* constitutes a network of lexons which are logically connected, instantiated and constrained. Thus a commitment provides a partial but instantiated view of an underlying ontology base (Zhao and Meersman, 2003). However, there is an important difference between the commitment layer and the underlying ontology base; a commitment is *unambiguous* and *semantically consistent*. It defines tokens of application logic and semantics in terms of existing lexons and describes how instantiated lexons are *grounded* and *connected*. In this way the DOGMA framework allows application specific instantiation, integrity constraints and logical connections to be applied to lexons. This achieves consistency without necessary being complete, which corresponds to the initial notion of commitment introduced in Section 3.3.

Commitments may be achieved by selecting *perspectives*, imposing *constraints* or providing *instantiations* of lexons in an ontology base (Zhao and Meersman, 2003). Firstly, selecting a perspective involves adapting the 5-tuple lexon into a ternary structure. More concretely, given a lexon $\langle \gamma, t_i, r_{i-j}, t_j, r_{j-i} \rangle$, its commitment can take the form of two perspectives $\langle t_i, r_{i-j}, t_j \rangle$ or $\langle t_j, r_{j-i}, t_i \rangle$ where the first element is referred to as the *theme*, the second as the *transition* and the third the *rheme*. This corresponds to a linguistic structure consisting of subject, predicate and object parts with names borrowed from the school of *functional linguistics* (Zhao et al., 2004). The specific perspective taken depends on the choice of semantic focus and is called a *commitment statement*. In such a commitment statement, constraints can now be imposed on the three constituent parts. Finally, the theme, transition and rheme may also be instantiated with particular values or *instances*.

5.4. Advantages of the DOGMA approach

It is granted that the DOGMA approach simplifies the notion of an ontology but also offers significant advantages (Meersman, 2001). Firstly, a clear distinction is drawn between concepts as providers of meaningful relationships and application specific knowledge in the form of constraints, rules and procedures. As mentioned in Section 5.1, agreement on the sets of ideas in the form of an ontology base is also much easier when the explicit and formal semantic interpretation of the fact types is left to the application. This is further justified by the view that that there does not necessarily exist a demand for one unique and global conceptualisation. Rather, the need is for "*unambiguous communication about concepts that leaves every application free to make its own conceptualisation explicit*" (Pisanelli et al., 2002: 1).

By adopting the DOGMA approach it is also much easier to satisfy the design criteria outlined in Section 3.5. In fact, the two-layered DOGMA approach successfully addresses the design criteria of *extendibility*, *minimal encoding bias* and *minimal ontological commitment* (Zhao and Meersman, 2003). Furthermore, the naming of terms and roles in a natural language not only leads to minimal encoding bias, but also to an intuitive representation, easy evolution of the ontology base and aids human cognition by adding mnemonic value.

According to Zhao and Meersman (2003), similar to natural languages, an ontology base forms a semiotic system (a system of symbols or lexical references) consisting of semiotic resources (lexons) for communication and reasoning about a domain of interest. As a semiotic system, an ontology base constitutes a conceptual scheme which underlies application knowledge. Natural language systems exhibit three important characteristics which serve as further justification for the DOGMA approach. These features may be interpreted in terms of the DOGMA framework as outlined below:

1. There is a *multiplicity* of application specific perspectives and representations possible for an ontology base in the form of commitments to lexons.
2. The commitment layer allows for *unambiguous* and *consistent* semiotic expressions to be constructed from an ambiguous, inconsistent and redundant set of lexons.
3. Semantic *under-specification* allows for the maximal reuse of an ontology base.

Finally, from a methodological point of view, The DOGMA framework introduces a modular approach to "an otherwise monolithic representation of the generic and specific, declarative and operational semantics of an application domain" (Zhao et al., 2004: 2). This is achieved by having the commitment layer handle all changes, application requirements, and perspectives on the semantics of the application domain independent of the ontology base.

6. Conclusion

The notion of ontology originates from the discipline of Philosophy. It has evolved to its current meaning in the context of Computer and Information Science where it refers to a designed artefact which formally represents agreed semantics in a computer resource. This enables the sharing and reuse of information and allows for the interoperation of information resources. Ontology bases, which consist of elementary binary fact types called lexons, provide a simplification of this notion to facilitate agreement to a greater extent.

7. References

BERNERS-LEE, T., HENDLER, J., and LASSILA, O., "The Semantic Web", *Scientific American,* 284(5), May 2001, 34-43, 2001.

BRAY, T., "Measuring the Web", *Computer Networks and ISDN Systems,* 28(7-11-May), 992-1005, 1996.

GENESERETH, M. R. and NILSSON, N. J., *Logical Foundations of Artificial Intelligence,* San Mateo, California, Morgan Kaufman Publishers, 1987.

GRUBER, T., "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition,* 5(2), 199-220*,* 1993.

GRUBER, T., "Towards Principles for the Design of Ontologies Used for Knowledge Sharing", *International Journal of Human and Computer Studies,* 43(5/6), 907-928, 1995.

GUARINO, N., "Formal Ontology in Information Systems", *Proceedings of FOIS'98, Formal Ontology in Information Systems,* Trento, 3-15, 1998.

HALPIN, T.A., *Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design,* San Francisco, California, Morgan Kaufman Publishers, 1999.

MAEDCHE, A.D., *Ontology Learning for the Semantic Web,* Norwell, Massachusetts, Kluwer Academic Publishers, 2003.

MEERSMAN, R., "Can Ontology Theory Learn from Database Semantics?", *Proceedings of the Dagstuhl Seminar 0121, Semantics for the Web,* Dagstuhl, 2000.

MEERSMAN, R., "Ontologies and Databases: More than a Fleeting Resemblance?" D'Atri, A. and Missikoff, M. (eds.), *OES/SEO 2001 Workshop,* Rome, 2001.

MITCHELL, W.J., *City of Bits: Space, Place and the Infobahn,* Cambridge, Massachusetts, MIT Press, 1999.

NOY, N.F. and HAFNER, C.D., "The State of the Art in Ontology Design – A Survey and Comparative Review", *AI Magazine,* 36(3), 53-74, 1997.

PISANELLI, D.M., GANGEMI, A. and STEVE, G., "Ontologies and Information Systems: The Marriage of the Century?", *Proceedings of the LYEE Workshop,* Paris, 2002.

ZHAO, G. and MEERSMAN, R. (2003) Ontology Modelling for Collaborative Business Processes. *STARLab Technical Report.* Brussels: Vrije Universiteit Brussel.

ZHAO, G., GAO, Y. and MEERSMAN, R. (2004) Ontology-Based Approach to Business Modelling. *Proceedings of the International Conference of Knowledge Engineering and Decision Support (ICKEDS2004).* In press.