

The Semantic Web in a philosophical perspective

Terje Aaberge, Sogndal, Norway

The semantic web is a proposal to make a more efficient web. By endowing the computer 'language' with a semantic structure defined by ontologies extracted from natural language, one hopes to facilitate the communication between human operators and computers and between computers. An ontology is a set of definitions that relate the terms and predicates of the vocabulary of the description language for a domain. It imposes a semantic structure that fix the meaning of terms and predicates that are polysemic in natural language and it serves as a basis for making inferences. Abstracted from the domain it limits the possible interpretations of the vocabulary. The extraction of ontologies from the semantic of a description language leans on Wittgenstein's metaphysics and picture theory from *Tractatus*, and language games from *Investigations*.

1. Introduction

A metaphor for the world wide web is that of a market place where each of the providers is represented by a web site by means of which he communicates with his customers. The web sites are linked or appear in the same result lists of queries in search engines, such that sites with similar offers are loosely grouped together. The most direct access to the different offers is given by the search engines that partially map the marketplace.

A customer looking for a particular item has to find the possible providers and then retrieve the item. Both of these tasks involves communication and might be arduous due to the lack of a common language shared by the providers, the computers and the customers. First of all, the maps established by the search engines give an incomplete account of the offers. Each offer is described by an index card that is established by agents on the basis of a purely syntactic analysis. In general, the relevance of results of a query that follows from looking through the index cards is thus lacking. Secondly, even if the customer finds a provider he might have difficulties of getting into agreement because the communication mediated by the computer is incomplete or unclear.

The semantic web is proposed as a solution to the problem of communication by defining computer 'languages' that may serve as interfaces between the human operators and the computers and between the computers (Daconta et al. 2003). It is not expected that it will be possible to create a unique language that will cover the content of the whole web, but that communities will create computer 'languages' for their domains of interest. To be understandable by the human operators they must be based on the informal description language that the community possesses and uses to describe the objects of the domain.

The means to do this is to extract the semantic of the informal description languages as ontologies. Ontologies endow computer 'languages' with semantic structures. Supplemented with logical rules they provide the computers with the ability to make inferences. A computer does not perceive the systems of a domain. It therefore has no semantic. However, the human operators can apply the semantic of the informal description language and thus conduct a meaningful communication with the computers

possessing a 'language' based on the ontologies extracted from the informal description language of the community. Moreover, the index cards produced by the agents employing the same formal 'language' will contain real information about the items associated with the domain. This information can be exploited by the search engine that will return more relevant answers to queries.

This paper presents an effort to put the semantic web into a philosophical setting and to show the relevance of some of Wittgenstein's ideas on language for its justification and the task of extracting ontologies from the semantic of the description language. First however, I will introduce some notions, define the framework and exemplify the tasks.

2. Description Languages and Theories

The notions considered in this paragraph are those of formal description language, theory, ontology, model, meta-model and computer 'language'.

The necessity to apply first order predicate logic as the syntax for the formal description language for a domain makes it appear as the juxtaposition of two languages an object language and a property language. Their vocabularies consist of the logical constants and three kinds of words, the names, variables and predicates, each kind having a particular syntactic role. A name refers to a unique object, a predicate to a property (predicate of the first kind) or a category of objects (predicate of the second kind) or a relation between objects. A variable refers to any of the objects in a category. As exemplified by the sentences "the water in bottle 3 is 5°C" and "5°C is a temperature", a predicate of the first kind in the object language is a name in the property language. The object language serves to describe the systems of the domain and the property language serves to describe the properties of the systems. This separation of the description language in two juxtaposed languages makes it possible to quantify over the properties also, not only the systems.

A theory is a formal description language endowed with ontologies defining semantic structures for the object and property languages. The ontologies are sets of implicit definitions of the predicates needed to describe the systems of the domain and their properties. They provide a formal representation of the semantic. However, they do not define a full semantic but limit the scope of possible interpretations.

A model of a system is a representation of the system in the property language. The model depicts the system such that literate interpreters knowing the system recognise its referent. A metamodel, on the other hand, is a set of rules of interpretation expressed in the metalanguage; these rules must be known to understand the ontology and the model. From the model we can extract a description of the system modeled. The degree of correspondence between the empirical description in the object language and the theoretical description in the property language determines the correctness of the model.

The different languages referred to above and the theory is languages in the sense that they possess a se-

mantic inherited from the informal description language. Abstracting the theory from the domain however produces a formal system that serves as a computer 'language'. It has no complete semantics but possesses a semantic structure defined by the ontologies.

3. Ontologies for the Object Languages

A domain consists of a set of (physical) systems that possess properties and relations. A system is uniquely identified and described by the properties it possesses. This is done by means of the atomic sentences that attach properties to the system, i.e. they are concatenations of the name of the system and the predicates that refer to the properties of the system. The basis for such a description is logical atomism. Each atomic sentence stands for an atomic fact. The conjunction of atomic sentences that applies to a system provides a *picture* of the system and serves to distinguish it from other systems.

Some properties are mutually exclusive in the sense that they cannot be possessed by a system at the same time; for example, a system cannot at the same time be red and green. This relation of exclusiveness of properties serves to categorise the predicates of the first kind. Each such category is then the range of a map from the set of systems of the domain to the predicates of the first kind. The map, called an observable, relates properties of the category to the system. Colour is thus an observable. Other examples of observables are form, temperature, position in space, mass, velocity etc.

It is necessary to distinguish between two kinds of observables. This is a result of the problem encountered when one wants to describe change and it is illustrated by the following statement:

change does not exist, because if something *changes* than it is no longer the *same* and we cannot say that anything has changed.

This semantic problem was a central theme in Greek philosophy. One of their solutions, which have become a basis for physics, is to distinguish between two kinds of properties, properties that do not change in time and thus serves to identify the system and properties that change. The latter are called state properties. The properties of the systems are thus categorised as identification and state properties and the corresponding observables as identification and state observables respectively. The state properties form a space called the state space of the systems.

The systems can be classified with respect to the identification observables. One starts with one of the observables and uses its values to distinguish between the systems and construct classes, one for each value. The procedure can be continued until the set of observables is exhausted. The result is a hierarchy of classes with respect to the set inclusion relation.

The classes are referred to by predicates of the second kind which thus are ordered naturally in a taxonomy that constitute a linguistic representation of the classification. The taxonomy together with the definitions of the classes is an ontology for the object language. The class definitions impose a semantic structure that mirrors the class inclusion relations and create semantic relations between the predicates.

In the object language the meaning of a name is the object it refer to, the meaning of a predicate is given either by an operational definition or the extension.

4. Ontologies for the Property Languages

The construction of an ontology for the property language can be illustrated by the development of Euclidean geometry. The domain is here the set of two-dimensional systems. The only interesting property of a system is its form. We assume that the observed forms are described by figures that can be constructed by ruler and compass and traced on a piece of paper by a pencil. These are the points, lines, and the figures that enclose a finite area, i.e. the circles, triangles and higher order polygons. Each of the corresponding categories are represented by a predicate (of the first kind), Point, Line etc. The corresponding property is denoted by the names point, line etc. (in the property language). They are associated with (operational) definitions leading to their construction by compass and ruler.

These categories can again be divided. Thus, the category Circle may be divided into categories of circles with given radius, the category of Triangle may be divided into categories of equilateral triangles and non-equilateral triangles etc. Each of the subdivisions introduces new predicates that are accompanied by a definition that serves to distinguish between the systems that are elements of the category and those that are not.

By studying the figures and the way they are constructed we may discover relations between them that can be expressed as sentences. These sentences are then 'categorised' as definitions and theorems; all the theorems can be proved from the definitions. The separation is partly based on convenience and tradition; the proofs should be as simple and direct as possible. The set of definitions constitute an ontology for the domain of plane geometry. *Abstracted* from the domain they define a semantic structure that limits the scope of possible interpretations.

An interpretation is determined by the relation of some of the names and predicates of the ontology to external 'objects'. The other terms and predicates are then given meaning by the definitions. Terms and predicates whose interpretation is a sufficient basis for the semantic of a theory are said to be primary. All the other terms and predicates are defined in terms of the primary terms and predicates by means of the definitions. The definitions that only contain primary terms and predicates are called axioms (Blanché 1999).

The axiom system constitutes a foundation for a mathematical theory. From this foundation the whole structure can be constructed. However, to do so we need to introduce additional concepts. Thus, considering for example the Euclid axiom system,

any two points lie on a straight line;

two lines meet in at most one point;

any finite line element can be produced as far as you wish;

it is possible describe a circle with any centre and any radius;

all right angles are equal;

given any line, and any point not on the line, then there exists *exactly one* line parallel to the first line passing through the given point;

we see that there is no mention of the concept of triangle. This secondary concept must be introduced by a separate and thus secondary definition. The introduction of new

concepts is not automatic but the result of conscious choices.

The construction of ontologies for more complicated domains is based on this kind of analysis. The vocabulary established through such a construction is taken from natural language and the interpretation thus obtained will be the intended interpretation of the theory. The ontology will then fix the meaning of the words that in a natural language context are polysemic.

5. Method

The semantic of natural language represent the mental pictures humans possess of the external reality. To establish human understandable ontologies for a domain these pictures must be specified. There are several complementary methods to do this. The most important methods are dialogs, group tests, user tests and thought experiments (Speel et al.). They are all examples of ways of analysing language games.

It is the linguistic representations of the mental pictures that are investigated by these methods. The task of the analyst is to design language games that will uncover discrepancies with the mental pictures by means of dialogs, group tests and user tests which help us to see how words are used and thus apprehend their meaning from the context created. The thought experiments test the semantic coherency between the empirical descriptions in the object language and the theoretical descriptions in the property language. Prominent examples are the Zeno paradoxes.

6. Relevance of Wittgenstein's Ideas

Consider the case of a community possessing an informal language for the description of a (restricted) domain of interest. It serves as a medium for the recording of information about elements of the domain and as a vehicle for the communication of this information.

In *Investigations* Wittgenstein considers the application of a language as a set of games. As any game, each of them is associated with a set of rules that can be divided into syntactic and logical rules, and rules of application of words. His idea is that the meaning of words follows from their use in language games. To apply words correctly, the speaker must thus master the rules.

To be admitted to the community any potential member must learn the language, i.e. he must learn the rules of the language games. For a computer to be admitted as a member it must be endowed with the corresponding formal system (computer 'language'). This must be based on syntactic and logical rules that are a subset of those of first order predicate logic. Assuming this to be the case the problem left is to endow the computer with a semantic structure satisfying the rules of application of the words. This problem is "solved, not by giving new information but by arranging what we have always known" (PI, 109). One has to look at how words are used to determine their relative meaning in order to establish the definitions that constitute the ontology which thus represents the semantic structure of the informal description language. However, meaning is not given by definitions alone. It must be *grounded*. Such grounding is the reference to external objects provided by Wittgenstein's logical atomism and picture theory from *Tractatus*: a sentence is *true* if it pictures an existing state of affairs. It provides the ontology of the object language with a semantic (by correspondence).

The complete semantic of the description language is given by the relation between the object language and the property language. By this construction the semantic of the theory mirrors that of the informal language. It provides the semantic human operators apply in their communication with the computers.

7. Concluding remark

Humans use natural language to describe record and communicate. And we mostly manage to overcome the problems due to imprecise syntax and semantic by our knowledge of the possible meanings of the terms and the contexts in which they are used. The construction of a theory for a domain introduces ontologies that fix the meaning of the polysemic terms of natural language and make possible precise statements and inferences. Abstracted from the domain a theory becomes a formal system with a semantic structure defined by the ontologies.

A formal system can serve as a computer 'language' by means of which human operators communicate with a computer. A computer does not perceive the systems of a domain. It thus has no semantic. However, the human operators can apply the semantic of the theory and conduct a meaningful communication with the computer. Moreover, computers possessing the same ontologies can communicate among each others in a way that is meaningful for the operators.

Literature

Blanché, Robert 1999: *L'axiomatique*. Paris: Presses Universitaires de France.

Daconta, Michael C., Obrst, Leo J. and Smith, Kevin B. 2003: *The Semantic Web: A Guide to the Future of XML, Web Services and Knowledge Management*, Boston: Wiley Publishing Company.

Speel, P-H, Schreiber, A. Th., van Joolingen, W., van Heijst, G., Beijer, G.J.: *Conceptual Modelling for Knowledge-Based Systems*, <http://www.cs.vu.nl/~guus/papers/Speel01a.pdf>.

Wittgenstein, Ludwig 1961: *Tractatus logico-philosophicus*. London: Routledge and Kegan Paul.

Wittgenstein, Ludwig 1968: *Philosophical Investigations*. Oxford: Basil Blackwell.

Email: terje.aaberge@vestforsk.no