International Journal of Computer Science and Applications,
©Technomathematics Research Foundation
Vol. 9, No. 2, pp. 32 – 41, 2012

ONTOLOGY AND ONTOLOGY CONSTRUCTION: BACKGROUND AND PRACTICES

TERJE AABERGE, RAJENDRA AKERKAR

Vestlandsforsking
Sogndal, Norway
taa@vestforsk.no
rak@vestforsk.no

When developing semantic applications, constructing the underlying ontologies is a crucial part. This paper presents a view-point on the notion of 'ontology' and of ontology construction. Ontology construction cannot be seen in isolation, but must be considered in relation to its use. We also present two ways of conceiving ontology construction and ultimately, we survey common categories in the construction methods of ontology learning.

Keywords: Ontology, Semantic Web, Semantics, Pragmatics, ontology construction

1. Introduction

A language has words (signs), sentences (ordered finite sequences of words) and arguments (ordered finite sequences of sentences). The composition of the sentences satisfies rules of syntax and according to their use, determine a classification of the words into names (terms) and predicates of fixed arities; the arguments satisfy rules of inference that from some premises derive a conclusion through some intermediary steps. The derivation depends exclusively on the syntactic form of the sentences forming the argument. But the words also carry meaning that is inherited by sentences and arguments. Randomly chosen collections of words and sentences however are meaningless.

Meaning is acquired through interpretations which relate the linguistic entities to something outside the language. There are two ways of representing meaning formally, referred to as *extensional* and *intensional*[1] and corresponding to the complementary theories of meaning, semantics and pragmatics[2]; semantics focuses on the relation between words and what they stand for, their denotata, while pragmatics concern how context contributes to meaning. This is expressed by the following elaboration of the semiotic triangle:

[1] Intensional should not be confused with intentional. An *intensional* interpretation gives the sense of a predicate referring to a property of an individual by specifying the necessary and sufficient conditions that determine whether the individual possesses the property. On the other hand, *intentional* applies to purposeful actions.

[2] The two aspects of meaning, intension and extension correspond to Frege's 'Sinn' and 'Bedeutung' (Frege 1971).

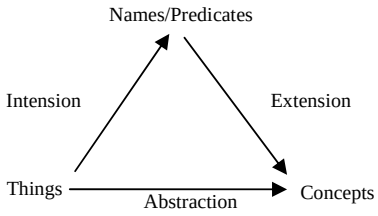

Figure 1

where the arrows symbolises that an object is associated a name and characterised by predicates, that to each name and predicate corresponds a concept which is the mental idea of a thing, formally represented as a (mathematical) set of things, i.e. an thing in itself. This is exemplified by the semiotic triangle for the Eiffel Tower:

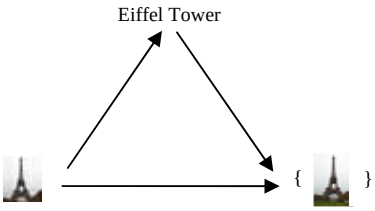

Figure 2

where the curly parenthesis {} symbolises the conceptualisation.

The semiotic triangle express manifestations of the subject/object separation, the separation between the observing (human) subject and its object of investigation that results in the distinction between a thing and its properties or between things and their relations. It is reflected in the subject/object form of atomic sentences, the basic syntactic structural element of the language. By connecting a thing and a property (or things and their relation) an atomic sentences span two levels of abstraction. As we move on the ladder of abstraction what has been an object will become a subject. The structure displayed by the semiotic triangle is however generic. It applies to the characterisation of ontological distinctions emerging for any two subsequent levels of abstraction.

This paper is structured as follows. The concept and creation of meaning and, definition of ontology are discussed in Section 2. Different approaches of ontology construction are presented in Section 3. Finally, Section 4 summarizes the conclusions of the paper.

2. Meaning

2.1 *Interpretations*

The *extension* of a name or predicate is given by a set of thing(s) (Figure 2). It is the formal representative of the concept carrying the meaning of the word. Thus, the

extension of a name is the singular set of the individual carrying the name[3], the extension of a one-place predicate the set of objects to which the predicate applies. The extensional meaning of a predicate that stands for a kind of relation is the set of ordered pairs that satisfy the relation etc. Extensional (model theoretic) semantics conceives the structure of the domain to be imposed by the structure of language. Accordingly, an extensional interpretation is represented by a map from the vocabulary to a conceptual model of the domain pictured as the set consisting of the singular sets of the individuals, sets of individuals, sets of ordered pairs of individuals etc. The interpretation map being a bijection (one-to-one), maps a name or predicate to its extension.

The intensional meaning is based on a directed graph model that conceives the domain as consisting of things with properties and relations. The things are represented by nodes and the relations by arrows (edges). An intensional interpretation is represented by maps from the domain to the vocabulary of the language: a bijection that maps the individuals (or relations) to names and *observables* that maps individuals (or relations) to predicates thus providing pictures (Wittgenstein 1961). The instances of these maps express the *semantic relations* of the interpretation. In this case it is thus the structure of the domain that determines the structure of language (Aaberge 2010).

Observables are identified by mutual exclusion of properties. Two properties that cannot simultaneously be possessed by an individual are represented by predicates belonging to the range of the same observable; an individual cannot at the same time be red and green, colour is therefore an observable. It maps an individual to the predicate representing its colour. Other observables are weight, position in space, temperature etc. An observable represents a kind of measurements and is associated with an operational definition exhibiting a standard of measure which provides a *translation* between properties and predicates, laws on which the measuring device is based and rules of application of the measuring device. For example, the measurement of the colour of an individual consists in holding a colour chart representing the standard of measure for the colours against the individual. If the mental pictures that the observer gets of the colour of the individual and the colour marked red on the colour chart coincide, then red is taken to denote the result of the measurement. The intensional meaning of the predicates is provided by the operational definitions.

An extensional interpretation can be derived from an intensional one. The inverse image of a name is its extension and the inverse image of the value (predicate) of an observable, i.e. the set of all individuals that is mapped to the same predicate by the observable, is the extension of the predicate. The opposite is however not the case. Intensional interpretations cannot be constructed from extensional ones. The reason for this asymmetry is that intensional interpretations contain more information than extensional ones.

---

[3] The meaning of a name is often referred to the bearer of the name, its referent, however, as Wittgenstein has pointed out, “When Mr N. N. dies, one says that the bearer of the name dies, not that the meaning dies” (PI 40, Wittgenstein 1968).

The forms of interpretation are connected to the complementary theories of meaning, semantics and pragmatics. While an interpretation so to speak is the answer, semantics and pragmatics also address the problem of how is created.

2.2 *Creation of Meaning*

Semantics for formal languages is extensional; meaning is embodied as a cognitive entity and represented by a set. Semantics is thus essentially captured by model theory.

Pragmatics is rooted in the language-action perspective. It concerns how context contributes to the meaning, i.e. how the context of a dialogue adds meaning to the messages exchanged between the interlocutors. The point is illustrated by the following story about Victor Hugo and the publication of 'Les miserables': after having sent the manuscript to his Belgian editor, Hugo, who lived in exile on Guernsey, was getting anxious about the public reception of his work and therefore sent the editor a telegram with just one sign '?'. Promptly, he got the answer '!'.

But pragmatics also concerns how meaning is arrived at. This is done either through a continued dialogue between interlocutors or through scientific investigations which also are dialogues but with 'nature'. 'Nature' is serving as one of the interlocutors responding to suggestions and hypotheses put forward by an investigator. An operational definition describes the context for an investigation and provides the intensional meaning of the predicates expressing results of the investigation.

To possess a language means to understand and correctly apply words, sentences and arguments, i.e. it is to master a technique, a skill that can be tested. The mastery is to know the rules of the *language game* and apply them correctly (Wittgenstein 1968). These are the rules of syntax which are common to all the language games constituting a language and the *ontologies*. An ontology is a linguistic representation of the structural properties of a conceptual model for a domain of discourse. Each language game is thus associated an ontology whose choice is determined by the context of the dialogue. The ontology carries information that is not part of the messages exchanged, but if known to both parties it is used to determine extract the information content.

The introduction of linguistic signs and the creation of meaning is a dynamic process. Language is a social activity, "a form of life" that evolves organically by modifying and incorporating language games, and they are learned iteratively through such games with respect to a winning strategy that aims at common understanding. The dynamical evolution is subjected to 'laws' laid down as rules of methodology determining how to arrive at the rules for correct use of vocabulary.

In the Web setting pragmatics concerns the dialog between humans and virtual agents and the creators of web sites via the sites. Though Wittgenstein only discussed the pragmatic aspects of language with respect to humans' use of natural language, his analysis also applies to virtual agents, their use of formal languages and the communication between human users and agents. The pragmatic aspect of the Web thus consists of tools (like virtual agents), practices and theories describing why and how

people put, retrieve, and use information on the Web; it is broadly speaking about social interaction via the Web.

As the linguistic metaphor proposes, arriving from semantics to pragmatics requires that the practices of use are taken into account. Rather than being a repository of meaningfully represented data, the pragmatic web needs to adapt to the user interaction. One possibility might be that the system could adapt to the ways in which the represented data is used and viewed. By adapting to these interactions it should make "contextually" sensitive changes into the ways of representing and linking the data. Examples of this could be the weighting or highlighting some links over others.

2.3 *Ontologies*

The vocabulary of an object language for a given domain consists of names representing the individuals of the domain, predicates standing for properties and relations, and of logical constants.

The meaning of a predicate is in general not independent of the meaning of other predicates. This interdependence is expressed by *axioms* and *intensional* and *extensional definitions*. An extensional definition of a predicate is simply the list of the names of the individuals that constitute its extension. When the names are denoting identifiable individuals of the domain, the extension of the predicate representing its meaning is given. An intensional definition of a predicate (*definiendum*) is the conjunction of atomic sentences (*definientia*) stating the properties an individual must possess for the predicate to apply. When the meaning of the definientia is given the definition explains the meaning of definiendum. From an intensional definition of a predicate an extensional one can be derived; the extension of the predicate is the class of individuals that satisfies definientia in the intensional definition. The contrary is not possible. Intensional definitions cannot be derived from extensional definitions. The reason is that intensional definitions contain more information than extensional definitions. Finally, we can express restrictions on the possible meaning of predicates by means of *axioms*. An axiom is an implicit definition that relates the *primary terms* of the vocabulary. It follows that the semantics of a formal language is completely given by the interpretation of the primary terms.

An *ontology* for an object language is a non-logical vocabulary supplemented by a set of

- axioms
- intensional definitions
- extensional definitions

The axioms picture structural properties of the domain and limit the possible interpretation of the primary terms. The intensional and extensional definitions are terminological. They define new predicates from the primary terms[4] that serve to

[4] Or rather, all terminological definitions can be expressed by means of the primary term, but one will normally define new terms by terms already defined.

facilitate the discourse, e.g. instead of having to repeat the properties that an individual must possess to be of a certain kind, an intensional definition will introduce a predicate to denote the kind. Accordingly, the interpretation of the vocabulary and thus the language is fully determined by the interpretation of the primary vocabulary.

On the other hand, the axioms of a formal system determine the set of possible interpretations of the primary terms that turns the formal system into a language. An ontology thus contains information about the meaning of the vocabulary that can be exploited and which is exploited in computer systems. For example, a database storing information about companies will register each item by an Id and otherwise describe a company by giving its name, address, telephone number etc. The information is organised in a table whose ordinate lists the Ids and whose abscise lists the fields: `Name`, `Address`, `TelephoneNumber` etc. A row in the table is then the graph of the naming map or of an "observable" that to each Id associates a value. The Ids serve as unique representatives of the companies and the maps, the naming map and observables, supply a formal realisation of an intensional interpretation, i.e. without the pragmatic meaning provided by the context supplied by an operational definition.

The scheme thus provides a formal meaning of the description language: the domain level is represented by the Ids, the object language level by the set of given names, and predicates, i.e. addresses, telephone numbers, and the metalanguage level by the field names that are predicates in the metalanguage. An extensional interpretation is arrived at by taking the inverse images of the values og the naming map and observables. Thus, if Skjolden Hotel is the unique name of a company represented by the Id xy, the extension of Skjolden Hotel is the singleton set informally denoted `{xy}`, i.e. the set that only contains Skjolden Hotel. The extension of the predicate "Name" on the other hand is the set of all given names and thus a set of sets of individuals.

The method described is applied to create a semantic web based on the paradigm of Internet of Things (IoT) and Linked Data. The motivation behind is the expectation that adding formal meaning to the web resources will make it feasible for search agents to supply more relevant answers to queries than the present ones.

3. Principles of ontology construction

There are two ways of conceiving ontology construction, the bottom up approach that is predominant in the methodology of mathematics and a top down approach that is predominant in disciplines where the domain consists of objects of the world *a priori* given as in science.

3.1 *The bottom up methodology*

The domain of a mathematical theory consists of conceptual entities. The task is to discover relations between basic entities, like numbers in arithmetic or points and lines in geometry, that are expressed by axioms. New entities are introduced by intensional

definitions, e.g. prime number and triangle, and theorems expressing possession of properties are inferred by means of logical deduction.

3.2 *The top down methodology*

With respect to the top down approach the object language for a domain is loosely given as a fragment of natural language. The task is then to specify the relations between the predicates on the basis of their pragmatic use. This consists in accomplishing the following tasks,

1. specify the domain of discourse
2. identify a primary vocabulary
3. establish the axioms
4. introduce secondary terms by intensional definitions
5. introduce further secondary terms by extensional definitions

Task number one is preliminary but important because if we do not delimit the domain properly we cannot establish a language of description with a well defined vocabulary. It is the nature of the individuals of the domain that determines the predicates needed for their descriptions. The primary vocabulary consists of the names of the individuals and predicates that represent properties and relations needed to describe the individuals. The structural properties of the domain are described by axioms in terms of a primary vocabulary. The formulation of axioms will in general not use the name, but a variable that is representing an individual of a certain kind as in the following example,

`“if x is the SonOf y and y is the BrotherOf z then z is the UncleOf x”`

which is an axiom relating the predicates “`SonOf`”, “`BrotherOf`” and “`UncleOf`”. `x`, `y` and `z` are here understood to represent unidentified persons. Notice that this axiom is picturing a structural property of family relations for a domain of persons. At the same time it expresses implicit dependencies of the possible meaning of the predicates. The secondary vocabulary is constructed from the primary terms.

The numbering of the tasks does not refer to the ordering of their execution. One better keeps in mind the different tasks to be accomplished and work iteratively.

3.3 *Ontology Construction Approaches*

The bottom up approach for ontology construction is based on linguistic analysis. This is primarily a human endeavour, however, methods exists that aids the work. In this section, we will discuss four types in ontology construction approaches of ontology learning.

3.3.1 *Dictionary-based ontology construction*

Based on the compilation concept of a dictionary, the hierarchy of concepts is mechanically formed. Conventional dictionaries contain items together with their synonyms, root words, word origins, etc. The definitions and relationships presented in the dictionary are used to determine the hierarchy relationships of concepts (Khan & Luo,

2002; Kietz, Maedche, & Volz, 2000; Tan, Han, & Elmasri, 2000). The dictionary-based construction method is the basis for other construction methods. The other methods are in some way related to the dictionary-based construction method either in the preliminary construction phase or in the final pruning and verification stage.

The dictionary-based approach is by and large limited to the volume size of the dictionary and can thus form domains having different scopes. It cannot offer considerable ontological framework without being combined with other modes. Using this method alone will prove incapable of adapting to the ceaselessly changing environment. On the whole, the ontology constructed using the dictionary-based method has a general description and is not domain specific. Once it is combined with another method it gives a essential ontological framework.

3.3.2 *Text-clustering-based ontology construction*

The text-clustering-based method is utilized to mechanize the creation of a conceptual hierarchy. It is based on related terms grouped together according to their synonyms. Every cluster is represented by a particular word or term that is believed to be more often used. Hence, repeating the exercise can derive the hierarchy of the terminologies. There are several problems found in using this method which restrict its usability (Hotho, Maedche, & Staab, 2001). This method is commonly regarded as an objective type of method that generates well-defined results considered optimum in certain aspects. Though, this is the contrary of what goes on in reality. Different users have different requirements for clustering because they have varying opinions and perspectives in viewing a particular document. Thus, we ought to adopt individual standards so as to accommodate differing perspectives in accomplishing the task of text clustering.

Furthermore, text clustering is normally required in high dimensional spaces to perform clustering computations since; every word or term is seen as an attribute of the entire document. However, experiments and mathematical analysis confirmed that clustering calculations in high dimensional spaces are inefficient because every data point possesses tendency of similar distances with other existing data points.

3.3.3 *Association rule based ontology construction*

Association rules can be used to generate combinations of different conceptual relationships. Association rules with stronger support, confidence and more extensive conceptual relationships can be placed on the upper level of ontology (Maedche & Staab, 2000). Nevertheless, if we use them as two discrete concepts, it will be hard to get the required support. As the document is made up of transactions required by the computation of association rules, separate strategies can result in distinct results.

3.3.4 *Knowledge base driven ontology construction*

Another approach is based on knowledge base. Obviously it requires the construction of knowledge bases beforehand in related domains. The knowledge base must include key

rules and examples (Alani et al., 2003). Once a user enters keywords to search for information, the rules in the knowledge base are used in order to filter data, while analogous examples are presented to make a likely comparison. When the proper result is picked out, rules in the knowledge base again are used to establish related ontology as well as giving the summary and results. The rules in the knowledge base can be considered as a sort of ontological expression and are used to build related ontology.

4. Concluding remarks

In this paper we have presented what an ontology is, its purpose and how to carry out the construction of an ontology for the object language of a domain. Ontology construction cannot be seen in isolation, but must be considered in relation to its use. In the course of discussion, two ways of conceiving ontology construction are elaborated and finally, we have reviewed four common categories in the construction methods of ontology learning.

References

Aaberge, T. (2010). Picturing Semantic Relations, In: Heinrich, R., Nemeth, E. and Pichler, W. (eds.) The 33nd International Wittgenstein Symposium, Kirchberg: LWS

Aaberge, T., Akerkar, R., Boley, H. (2011) "An Intensional Perspective on the Semantic and Pragmatic Web, In: *International Journal of Metadata, Semantics and Ontologies* (IJMSO), Vol. 6, No. 1, pp 74-80.

Akerkar, R. (2009). *Foundations of the Semantic Web*. London: Alpha Science International

Alani, H., Kim, S., Millard, D. E., Weal, M. J., Hall, W., Lewis, P. H. (2003). Automatic ontology-based knowledge extraction from web documents. IEEE Intelligent Systems, 18(1), 14–21.

Baeza-Yates, R., Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: Addison-Wesley.

Berners-Lee, T., Fischetti, M. (1999). Weaving the web: The original design and ultimate destiny of the world wide web by its inventor. San Francisco, CA: HarperAudio.

Buchli, F. (2003). Detecting software patterns using formal concept analysis. Bern, Switzerland: University of Bern.

Frege, G. (1971). *Écrits logiques et philosophiques*, Éditions Seuil (Paris)

Han, J., Kamber, M. (2001). *Data mining: Concepts and Techniques*. New York: Morgan Kaufmann.

Hotho, A., Maedche, A., Staab, S. (2001). Ontology-based text clustering, Proceedings of the IJCAI-2001 workshop text learning: Beyond supervision, Seattle.

Kang, S. H., Huh, W., Lee, S., Kim, Y. (2000). Automatic classification of WWW documents using a neural network, Proceedings of international conference on production research, Bangkok.

Khan, L., Luo, F. (2002). Ontology construction for information selection, Proceedings of the 14th IEEE international conference on tools with artificial intelligence, Washington, DC pp. 122–127.

Kietz, J. U., Maedche, A., Volz, R. (2000). A method for semi-automatic ontology acquisition from a corporate intranet, Proceedings of the EKAW'2000 workshop on ontologies and texts. France: Juan-les-Pins.

Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2), 72–79.

Maedche, A., Staab, S. (2000). Discovering conceptual relations from text, Proceedings of the 14th European conference on artificial intelligence, Berlin pp. 321–325.

Salton, G., McGill, M. (1983). *Introduction to modern information retrieval*. New York, NY: McGraw Hill.

Tan, K. W., Han, H., Elmasri, R. (2000). Web data cleansing and preparation for ontology extraction using WordNet, Proceedings of the first international conference on web information systems engineering, Hong Kong, Vol. 2 pp. 11–18.

Wei, J., Bressan, S., Ooi, B. C. (2000). Mining term association rules for automatic global query expansion: Methodology and preliminary results, Proceedings of the first international conference on web information systems engineering, Hong Kong, Vol. 1 pp. 366–373.

Wu, S. H., Day, M. Y., Hsu, W. L. (2001). FAQ—centered organizational memory, Proceedings of the IJCAI'2001 workshop on knowledge management and organizational memories, Seattle.

Wittgenstein, L. (1961). *Tractatus logico-philosophicus*, Routledge and Kegan Paul (London).

Wittgenstein, L. (1968). *Philosophical Investigations,* Blackwell Publishers Ltd. (Oxford).