

Modelling Ontology Use for Information Systems

Roland Kaschek

Department of Information Systems, Massey University, New Zealand

Abstract. This paper characterizes ontologies as models. It looks then at ways of using ontologies for information systems. These ways are discussed in terms of ontology role, role justification, and relevance. Space limitations prevent from discussing IS ontology papers and elaborating on conclusions.

Introduction

Among the ontologies used in information systems (IS) are the ones of Bunge (see, e.g. [Web03]) and Chisholm (see, e.g. [MK01]). Regarding information systems the differences between these ontologies are significant. The most widespread use of an ontology in information systems appears to be the one that comes with the use of semantic models. Examples for semantic models are the Entity-Relationship model (ERM), see [Che76], State Charts, Dataflow Diagrams and Petri Nets. They have a built-in ontology that aids modelers in: identifying, classifying and relating to each other phenomena in a universe of discourse and by doing so obtain a specification of a consistent conceptualization.

For example, in ER modelling one is guided to (1) define the scope of the universe of discourse D ; (2) recognize phenomena within D that appear to be relevant for a task at hand; (3) classify these phenomena as entity, attribute, or relationship; and (4) relate classified phenomena to each other following the ERM's syntactic conventions. Defining the scope of U is more formalized in Function Point Analysis, see [GH96], than in ER modelling. The conceptualizations obtained with a semantic model intentionally often are shared by a number of involved individuals.

Models

This paper draws from Stachowiak's general model theory (GMT) (see [Sta92], [Sta83], [Sta73]). A similar theory was recently used in [M*03]. A recent discussion of the GMT for software engineering is in [Lud03]. Let a cultural unit be something to which one, in a given culture, can intelligibly refer to. This concept according to Eco, see [Eco94, p. 75], was introduced by Schneider. Let A be a group of agents and M, O specifications of cultural units U_M, U_O . Let A refer to O by means of M . Codify this with the predicate $\mu(O, M, A)$ and call it model relationship. Call the roles of O and M in it original and model respectively. Stachowiak has characterized models by: (1) a mapping property, i.e.,

each model is a model of something, i.e., its original; (2) a truncation property, i.e., the model in general lacks some of the specification parts of the original; and (3) a pragmatic property, i.e., the model is subdue to a purpose and its use for this purpose is only justified with respect to particular users, their objectives, applied techniques and tools, and period of time etc. This paper deviates from his views on models: originals do not have to predate the model, the cognitive dimension of models is explicitly recognized, as is a plenty property. The latter states that specification M can be richer than O . Consider a map M as a model of an intuitive specification W for a world C_W . M can easily be transported and folded, as the map is made of paper. This significantly impacts the map's usability.

The Reference mode, i.e., the kind of reference A makes with the model to the original is a dimension of the model relationship $\mu(O, M, A)$. It helps addressing questions like: Do all agent instances share the view of model and original or do they have different views? In which way does the agent relate the model to the original? is it a descriptive, prescriptive, prognostic, idealizing or constitutive way? In the descriptive, prescriptive, prognostic and idealizing reference mode the agent A uses M for specifying what O is like, should be like, expects it to be like under particular stated conditions, and what O will be like under ideal conditions respectively. In the constitutive reference mode A replaces O by M and in this sense constitutes U_O by means of M .

For an example of a model that occurs frequently in information systems consider a table in a relational database and its primary key. For each primary key there is a table (mapping property). Usually there is a number of table columns that do not belong to the key (truncation property). The primary key only is valid for distinguishing between any two tuples of the table if the table is maintained properly and only for those who are entitled for using it etc. (pragmatic property). In contrast to the table the primary key can be the target of a foreign key reference (plenty property).

Ontologies and their use

Following the frequently used definition of Gruber, see [Gru93, p. 199], an "... ontology is an explicit specification of a conceptualization". And a "... conceptualization is an abstract simplified view of the world that we wish to represent for some purpose." In [UG96, p. 96] and [SMJ02] ontologies more specifically are considered as shared conceptualization.

An ontology model A shared conceptualization is a cultural unit. Let an ontology X be used by an agent A to specify a cultural unit U_W . Let a specification W exist that allows for justification of the use of X to specify U_W . Let furthermore U_X be the cultural unit associated with X prior to agreeing that X actually should be used as a specification of U_W . Note that U_X may be void. Conceptualizing the consideration of an ontology X for U_W by an agent A thus leads to the model relation $\mu(W, X, A)$ in which a constitutive mode of reference occurs, i.e., a replacement of W by X .

Modelling ontology use Let the predicate $u(A, M, D, L, T)$ denote an ontology use, i.e., let it mean that a group A of agents uses ontology M with respect to a domain D , for a task T in which a language L is involved. Dimensions of an ontology use can be identified as (1) role, (i.e., the way the ontology is used); (2) justification (addressing whether A is entitled for using M in role R because M is an ontology); and (3) relevance (addressing whether it is relevant for using M in role R that M is an ontology). To each of these dimensions will be associated a scale. For both of the latter dimensions the scale values "++", and "--" will be used in the obvious meaning. The scale for 'role' comprises providers of: correctness criterium (ccp), communication medium (cmp), knowledge (kp), meaning (mp), quality concept (qcp), and world view (wvp). Let T be a task. The ontology m in the role ccp, cmp, kp, mp, qcp, and wvp is used for distinguishing the correct from the wrong sentences of L ; for providing agents in A the media L for talking about D ; for providing knowledge regarding D to members of A ; for associating items in D as a meaning to the sentences of L ; for establishing the concept of quality regarding items in D ; and specifying a conceptualization of D respectively. For an ontology use u the triplet $\alpha_u = (\rho, \iota, \omega)$, with ρ, ι, ω scale values of role, justification and relevance respectively is called use assessment. An ontology use u should be considered as problematic if its assessment α_u scores "--" for 'justification' or 'relevance'.

Some pragmatic concerns regarding ontology roles are: (1) truth of propositions is not always sufficient for a task T . Assume a postman knows a letter's recipient name and street of residence but not the house number. If the street has many houses he is likely not to deliver the letter. What he knows is correct but not precise. (2) proving a proposition can let it appear as unrefutable. Quantum Theory is used as an ontology enabling such proofs. At 2 April 2004 the Web edition of the "Neue Zürcher Zeitung" (Science Column) reported in the article (in German:) "The Quantum Lab in the Matchbox" on an encryption device cipherings of which provably cannot be deciphered without having access to the respective key. (3) wrong models are used in favor of correct ones if their differences are insignificant with respect to a task T . In primary and secondary education for example the wrong ball-model of molecules (Chemistry) and orbs (Physics) are used rather than (more) correct ones. Note, finally, that it is a well established approach in software engineering to relate quality to use and usability, see, e.g. [ISO91].

Let the predicate $\mu(O, M, A)$ be true and A use M in the constitutive reference mode. Consider an ontology use $u = u(A, M, D, L, T)$. Using M as wvp is justified if the overlap of D and C_O is counted as significant. What counts as significant depends on T . It is relevant for using M as wvp that M actually is an ontology because that makes M capable of specifying C_O . Using M as ccp regarding L turns discriminating true from false propositions in L into drawing implications from M . It is justified to use M in this role. However, it is not relevant for this use that M is an ontology. For a theory to be used in this role it would be relevant that the important propositions can be implied efficiently. What counts as 'important' or 'efficient' depends on T . Using M as kp is jus-

tified if $C_O = D$. It is relevant for this use that M is an ontology. Using M as cmp can be justified if $D = C_O$. Using M in this role would not be justified if enough members of A are not fluent in L . What counts as enough depends on T . It is not relevant for this use of M that M is an ontology. It is justified to use M as mp, if L enables A to talk about C_O . It is not relevant for using M in this role that M is an ontology. Note that justification of using M as mp may involve a fluency issue like the one discussed with respect to the role cmp. The characteristics of a theory suitable for being used as mp depend on T . Using M as a qcp can be justified if for T an understanding of quality as ontology compliance is adequate. An ontology does not qualify per-se for the qcp. It is less important that a quality concept means ontology compliance than it is that this concept helps doing T .

References

- [Che76] Peter P. Chen. The Entity-Relationship Model: Toward a Unified View of Data. *ACM Transactions on Database Systems*, 1(1):9–37, 1976.
- [Eco94] Umberto Eco. *Einführung in die Semiotik*. Wilhelm Fink Verlag GmbH & Co. KG, München, 1994.
- [GH96] David Garmus and David Herron. *Managing the software process: a practical guide to functional measurements*. Yourdon Press Computing Series. Prentice Hall PTR, Upper Saddle River, New Jersey, 1996.
- [Gru93] Thomas Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199 – 220, 1993.
- [ISO91] Information technology—Software product evaluation—Quality characteristics and guidelines for their use, 1991. International Standard 9126:1991(E).
- [Lud03] Jochen Ludewig. Models in software engineering - an introduction. *Software and Systems Modeling*, 2(1):5 – 14, March 2003. Reworked version of Ludewig’s presentation at ”Modellierung 2002”.
- [MK01] Simon Milton and Ed. Kazmierczak. An ontological study of data modelling languages using Chisholm’s ontology. In *European-Japanese conference on information modelling and knowledge bases*, Amsterdam et al., 2001. IOS Press.
- [SMJ02] Peter Spyns, Robert Meersman, and Mustafa Jarrar. Data modelling versus Ontology engineering. *ACM SIGMOD Record*, 31(4):12 – 17, December 2002.
- [Sta73] Herbert Stachowiak. *Allgemeine Modelltheorie*. Springer Verlag, Wien, New York, 1973.
- [Sta83] Herbert Stachowiak. Erkenntnisstufen zum Systematischen Neopragmatismus und zur Allgemeinen Modelltheorie. In Herbert Stachowiak, editor, *Modelle-Konstruktionen der Wirklichkeit*, pages 87–146. Wilhelm Fink Verlag, München, 1983.
- [Sta92] Herbert Stachowiak. Modell. In Helmut Seiffert and Gerard Radnitzky, editors, *Handlexikon Zur Wissenschaftstheorie*, pages 219–222. Deutscher Taschenbuch Verlag GmbH & Co. KG, München, 1992.
- [UG96] Mike Uschold and Michael Gruninger. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 11(2):93 – 136, 1996.
- [Web03] Ron Weber. Conceptual modelling and ontology: possibilities and pitfalls. *Journal of Database Management*, 14(3):1 – 20, July - September 2003.