Guidelines for Constructing Reusable Domain Ontologies

Muthukkaruppan Annamalai
Department of Computer Science & Software Engineering
The University of Melbourne
Victoria 3010, Australia
mkppan@cs.mu.oz.au

Leon Sterling
Department of Computer Science & Software Engineering
The University of Melbourne
Victoria 3010, Australia
leon@cs.mu.oz.au

ABSTRACT

The growing interest in ontologies is concomitant with the increasing use of agent systems in user environment. Ontologies have established themselves as schemas for encoding knowledge about a particular domain, which can be interpreted by both humans and agents to accomplish a task in cooperation. However, construction of the domain ontologies is a bottleneck, and planning towards reuse of domain ontologies is essential. Current methodologies concerned with ontology development have not dealt with explicit reuse of domain ontologies. This paper presents guidelines for systematic construction of reusable domain ontologies. A purpose-driven approach has been adopted. The guidelines have been used for constructing ontologies in the Experimental High-Energy Physics domain.

1. INTRODUCTION

The World Wide Web has become the *de facto* medium for distributed user community to share digital information, posing a challenge for users to effectively utilise the accessed information. The next generation agentised web promises to dispense with some of the human effort through machine processable metadata linked to ontologies. The working definition of ontology is a specification of shared conceptualisation [4]. An ontology presents a shared understanding of how the world is organised in a particular aspect of the domain and specifies the meaning of terms that make up the vocabulary in the domain of discourse – a necessity for information access and interoperability.

Constructing ontologies from scratch to support domain applications requires a great deal of effort and time. Alternatively, reusable domain ontologies provide opportunities for developers to exploit and reuse existing domain knowledge to build their applications with much ease and reliability. A common belief is that reusable ontologies ought to be conceived and developed independent from application and context of its use. The consequence of adhering strictly to this notion is that the developed reusable domain ontologies are: **a)** usually over-generalised and omit useful knowledge; **b)** often are also sparse constructs because it is not easy to determine which part of the concrete domain knowledge can be reused, particularly when the capturing of the domain knowledge is attempted in a top-down fashion; and **c)** necessitates modification and considerable extension work before it can be utilised. A better alternative is to be able to develop reusable ontologies without over-compromising their usability in the domain.

The approach we propose is to first ask which kind of domain knowledge should be encoded; bring together the relevant pieces of knowledge; only then identify which of those pieces of knowledge can be reused and isolate them. This paper outlines a strategy to develop reusable domain ontologies in this manner. We will discuss this issue in the context of a case study to develop reusable Experimental High-Energy Physics (EHEP) ontologies for the Belle collaboration *(http://belle.kek.jp/belle/)*. The distributed scientific community collaboratively sets-up experiment, accumulate event data, generate simulation data, construct software tools to analyse the data, carry out data analysis, publish their findings, and progressively build on each other's work. We aim to show that suitable ontologies be developed and reused to facilitate this scientific community to produce and share information effectively on the agentised web [1, 2]. The reusable domain ontologies will serve as a basis for communication, integration and sharing of information pertaining to experimental analysis within the collaboration.

The paper is organised as follows. In Section 2, we sketch out our basic strategy for developing reusable domain ontologies. This strategy is further elaborated and illustrated in Sections 3 and 4. The development guidelines are also given in here. Finally, in the concluding section we summarise the contribution of this paper.

2. STRATEGISING THE DEVELOPMENT OF REUSABLE DOMAIN ONTOLOGIES

We have devised a strategy for developing reusable domain ontologies by putting together the key points in traditional and modern ontology development methodologies and principles, such that of [3, 4, 5, 8, 9, 10]. The notable features in these methodologies will be pointed out as we advance through this paper.

There are two types of ontologies in our environment, namely, *domain* and *purposive* ontologies. A domain ontology captures an area of the domain generically, and serves as a reusable knowledge entity that can be shared across this domain. The domain ontologies are loosely coupled to one another, reflecting the association between the different facets of the domain captured by the respective ontologies. On the other hand, a purposive ontology explicitly defines the terms for supporting specific purpose or use. The purposive ontology encodes specialised domain knowledge by composing various reusable domain ontologies and then affecting the necessary application-specific extensions.

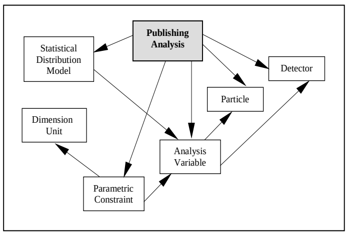

Figure 1: Interconnected Ontologies

The ontological commitment required to support a particular need is embodied in a set of reusable domain ontologies as illustrated in Figure 1. The filled box denotes the purposive ontology modelled for publishing EHEP analysis. Each empty box represents generic domain ontology. An arrow points at the ontology that holds the definition of the referred terms. A generic ontology is linked to another via concept relations. This links depicts the dependencies between the ontologies in this conceptualisation of the domain. This kind of loose coupling allows scalable modifications of the domain ontologies.

The value of reusable ontologies is long recognised by ontology researchers. The research on the technology for supporting knowledge sharing and reuse originated in Stanford's Knowledge Systems Laboratory (KSL) [7]. It spurred the development of reusable, general ontologies (such as Ontology of Time, Money, Measure, etceteras) for the KSL's Ontolingua server. The design principles espoused in [4] were commonly used to develop these ontologies that are meant to be shared across different knowledge domains.[1] While the past attempt was concerned with the development of domain-independent single ontologies, we focus on constructing inter-depending ontologies that are small and easier to reuse in a particular domain.

There appears to be no clear methodology for building domain ontologies that can be reused with minimal extensions. Moreover, a disagreeable prospect arises from the conventional idea that reusable domain ontologies ought to be conceived and developed independent from their application, due to the following reasons: **a)** Unlike general knowledge, which ground out in primitives with assumptions that are ordinarily understood, domain knowledge deals with domain-specific jargon; **b)** It is not feasible to think of knowledge needs of all foreseeable domain applications; **c)** The intended meaning of some terms can be different according to context of application; and **d)** It is also impractical for knowledge engineers to describe all knowledge they know about the domain. Hence, it is difficult to see how it would be possible for knowledge engineers to rely on their intuition alone to build ontologies that adequately capture the reusable knowledge in their domain.

Given this situation, we maintain that it is only sensible if we allow our current needs to dictate the creation of usable domain ontologies that are also reusable.[2] Modelling according to the purpose and use helps to determine what features of the domain knowledge should be encoded and provide a focus for knowledge acquisition. Consequently, we advocate development of purposive ontologies, and simultaneously pursue the creation of reusable domain ontologies.

Our basic strategy is spelt out as follows: **a)** Adopt a bottom-up view of the domain to conceptualise the knowledge required to support a specific need, and build the conceptual model; **b)** Identify potentially reusable chunk in the conceptual model and generalise it; and **c)** Formalise the generalised domain model into reusable domain ontology. In what follows, we will elaborate this strategy and illustrate it using simple examples.

3. PURPOSE-DRIVEN CREATION OF REUSABLE DOMAIN ONTOLOGIES

We begin by modelling the purposive ontology, which defines the vocabulary for the purpose of describing the experimental analysis in collaboration documents, such as research notes and research papers. The ontology can be applied for document annotation, query-answering and information retrieval.

3.1 Modelling the Purposive Ontology

We set out to accomplish this task by creating the concept model, which will serve as the foundation for the purposive ontology. Individuals in this conceptualisation are typically defined as concepts, and are constrained by properties, relations and axioms. The taxonomic and cross relationships among the concepts are explicitly specified. We used Protege-2000 *(http://www-protege.stanford.edu)*, a frame-based modelling tool to construct the concept model. Figure 2 describes a partial hierarchy of top-level concepts in the model that was developed to conceive this purposive ontology.

The model is elaborated from scientific collaboration documents, mainly books, research papers, existing standard HEP terminology and consulting the EHEP physicists. Our initial discussion with the EHEP physicists and related literature review enabled us to recognise some of the distinct concepts required for describing a typical EHEP experimental analysis. They are Signal event, Background event, Kinematic variables, Topological variables, Particle, etceteras. We called them the hook-concepts, as they serve as hooks (or links) for structuring additional concepts into the concept model. Initially, our competency questions [5], that are the questions we want the ontology to answer, revolved around these hook-concepts.[3] Subsequently, new concepts affiliated with the existing concepts emerge, which are organised in the hierarchical model by bottoming-up and middling-out processes [10]. The internal structures of these concepts are defined to limit their possible interpretation and relations in

[1] Clarity, Coherence, Extendibility, Minimal encoding bias and Minimal ontological commitment are prescribed as suitable design principles.

[2] As a matter of fact, the modern ontology development methodologies [8, 9] follow the tradition of knowledge engineering and back the development of application-oriented ontologies.

[3] In essence, the competency questions identify the kind of domain knowledge that should be encoded. Examples of competency questions are: What are the kinematic cuts performed in $B \to \rho\pi$ event analysis? What are the suppressed background events?

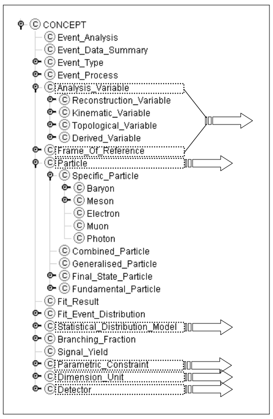

Figure 2: Externalising Reusable Domain Models

the ontology. This cycle is continued until the model is satisfactorily developed, that is when the set of compiled competency questions and their related answers can be clearly represented using the terms defined in the ontology. Each cycle evolves the model closer to the desired form.

3.2 Abstraction of Reusable Chunks

The concepts in the model are distinctly organised according to their role in elucidating particular aspects of the domain. As a rule, homogeneous concepts tend to cluster under a common parent concept and are potentially reusable in another situation (in whole or part). The reuse potential of these clustered concepts is reflected by their coherent nature.[4] The task of the knowledge engineer is to examine each cluster of concepts to identify the chunk that has reuse potential. The reusable chunk is isolated from the model and generalised into an independent reusable domain model. For example, the definition about *Particle* and its sub-classes in the concept model (shown in Figure 2) can be externalised from this model and componentised as reusable knowledge entity. Using this simple technique, we generated all the domain ontologies supporting the purposive ontology for describing experimental analysis as depicted in Figure 1.

Ontology Development Guideline I. We summarise the operational guidelines for developing purposive and domain ontologies (that does not reuse existing ontologies) as follows: **a)** Specify purpose and uses of ontology; **b)** Identify hook-concepts; **c)** Formulate competency questions; **d)** Identify new terms required to precisely formulate the competency questions and their generated answers; **e)** Define the new terms (concepts, properties, relations and axioms). Structure the concept into the concept model; **f)** Evaluate concept model against the set of competency questions and make the necessary changes; **g)** If still not satisfied with the level of details in the model, return to *step [c]*; **h)** Analyse concept clusters in the model that particularises specific areas of the domain. If a group of related concepts has reuse potential, generalise and shape them as a separate reusable domain model (sub-model); **i)** Link back the sub-models to the main model. Make the necessary application-specific extensions to the incorporated generalised domain knowledge; and **j)** Formalise the main model and the sub-models into purposive ontology and reusable domain ontologies, respectively. The ontological terms and definitions are expressed formally in a web-ontology specification language, such as DAML+OIL *(http://www.daml.org)*.

[4]The coherency among a set of concepts can be appropriately assessed by performing an ontological analysis [6] involving the concepts and relations in the model.

Future domain applications can exploit these reusable domain ontologies and may even churn out new reusable domain ontologies. This matter is discussed in the next section.

4. THE REUSE OF DOMAIN ONTOLOGIES

Using an existing domain ontology for supporting another application or to serve as a basis for building another ontology is cost-effective, provided the reusable ontology does not require much customisation effort. We describe two such scenarios that illustrate the reuse of existing domain ontologies and the creation of new versions of existing ontologies.

Purpose 1: Supporting Analysis Specification. We envision applications that allow physicists to partially automate experimental analysis such as skimming, tracking, vertexing and particle reconstruction. As a result, the vocabulary to describe the rudiments of these low-level analyses will need to feature in the purposive ontology built to support these applications.

We begin by sketching the concept model of this purposive ontology. The preliminary study shows that the purposive ontology will provide the vocabulary to describe the EHEP analysis, as in the previous case (refer to Section 3), but in greater detail. In particular, there will be additional analysis variables to be considered, and requires a much ‘finer’ representation.

All the domain ontologies created earlier are candidates for reuse. However, the *Analysis Variable* ontology will be revised to cater for this new requirement. Low-level analyses also involve tracks and clusters associated with event particles. Like *Particle* ontology, the knowledge about tracks and clusters can be held as separate entities. As we do not anticipate changes to the other existing ontologies, the purposive ontology being developed can make use of those reusable domain ontologies directly.

This developmental activity has generated two new domain ontologies, namely *Track* and *Cluster*; and a new version of *Analysis Variable* ontology, extended from its earlier version (see Figure 3). The distinction between the two versions of the *Analysis Variable* ontology can be made on the basis of additional definitions used to characterise the conceptualisation in the later version. Since no alteration to the existing definitions in this ontology was made, the new version is indeed backward compatible. Our rationale for constructing a new version of this ontology is to ensure that the prospective users who have adopted the original version can continue to rely upon their ontology, and not be overwhelmed with abstraction unrelated to their need.

Purpose 2: Supporting Detector Description. Another planned application aims to provide information about the Belle detectors. The purposive ontology developed for

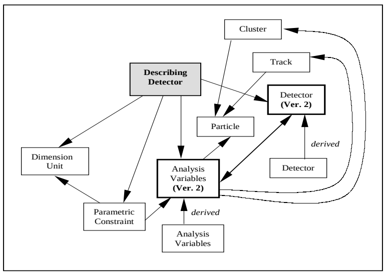

Figure 3: Reusing Existing Domain Ontologies

this application will be used to build a knowledge base about detectors' particle identification capabilities.[5]

The existing *Detector* ontology merely describes the detector and regions where the presence of tracks and clusters are sensed and is referred to by definitions in *Analysis Variable* ontology. The new requirement necessitates definitions in Detector to refer to definitions in *Analysis Variable* instead (see Figure 3). The new version of *Detector* ontology will have to provide the vocabulary to describe the Belle detectors in greater details, including *inverse* relationships between definitions in *Detector* and *Analysis Variable*.

Ontology Development Guideline II. We now recapitulate the operational guidelines for developing purposive and domain ontologies by reusing existing domain ontologies: **a)** Specify purpose and uses of ontology; **b)** Sketch the model of the purposive ontology; **c)** Identify existing reusable domain ontologies that can be used to support the modelling process; **d)** Construct the unsupported portion of the model based on *Guideline I steps [b] - [g]*; **e)** Identify reusable region in the model. If found, convert the related concepts into an independent reusable domain model. It corresponds to *Guideline I step [h]*; **f)** Select reusable ontologies (identified earlier) and make necessary changes to accommodate application needs. Sometimes, the modified ontologies may be redeveloped as new versions of the existing ontologies; and **g)** Rebuild the model of the purposive ontology by linking the sub-models, and formalise the ontologies. It is similar to *Guideline I steps [i] and [j]*.

5. CONCLUDING REMARKS

We have presented guidelines for creating reusable domain ontologies for a scientific user-community. In our environment, the different domain ontologies organise and structure the knowledge of diverse parts of the domain. The domain ontologies are constructed as small reusable knowledge components that can be easily shared across applications. These ontologies are loosely coupled to one another reflecting their association in the real world.

The conceptualisation is guided by a common set of competency questions generated when application-dependent purposive models are conceived. This provides the reason for believing that the same kind of reusable chunks will emerge from the different purposive models, even if the order in which the applications are considered is varied. The concepts in the domain ontologies are captured in some generality to make reuse possible. We are consistent with making specific knowledge more generic.

Sometimes it would be necessary to allow co-existence of different versions of a domain ontology to accommodate different needs in the domain. For example, in Section 4 we exemplified the creation of new versions of existing ontologies. These versions are seen as distinct domain ontologies with dissimilar reuse potential. Herein lays the larger issue of the management of the domain ontologies. A mechanism to control the different versions of domain ontologies is essentially required.

Acknowledgments: We are grateful to our associates in the Physics department, particularly Glenn Moloney and Lyle Winton for providing insights into EHEP experimental analysis.

6. REFERENCES

[1] M. Annamalai, L. Sterling, and G. Moloney. A collaborative framework for distributed scientific groups. In *Proceedings of AAMAS'02 Workshop on Ontologies in Agent Systems*, 2002.

[2] L. Cruz, M. Annamalai, and L. Sterling. Analysing high-energy physics experiments. In *Proceedings of AAMAS'02 Workshop on AgentCities*, 2002.

[3] M. Fernandez, A. Gomez-Perez, and N. Juristo. Methontology: From ontological art towards ontological engineering. In *Proceedings of AAAI'97 Spring Symposium on Ontological Engineering*, 1997.

[4] T. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.

[5] M. Gruninger and M. S. Fox. Methodology for the design and evaluation of ontologies. In *Proceedings of IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995.

[6] N. Guarino and C. Welty. Ontological analysis of taxonomic relationships. In *Proceedings of ER'00 Conference on Conceptual Modelling*, 2000.

[7] R. Neches, R. Fikes, T. Finin, T. Gruber, T. Patil, R. Senator, and W. R. Swartout. Enabling technology for knowledge sharing. *AI Magazine*, 12(3):16–36, 1991.

[8] A. T. Schreiber, J. M. Akkermans, A. A. Anjewierden, R. Dehoog, N. R. Rhadbold, W. V. D. Velde, and B. J. Wielinga. *Knowledge Engineering and Management - The Common*KADS *Methodology*. University of Amsterdam, 1998.

[9] S. Staab, R. Studer, H. P. Schnurr, and Y. Sure. Knowledge process and ontologies. *IEEE Intelligent Systems*, pages 26–34, Jan/Feb 2001.

[10] M. Uschold. Building ontologies: Towards a unified methodology. In *Proceedings of the British Computer Society Specialist Group Conference on Expert Systems*, 1996.

[5] An example piece of knowledge may look like this: Aerogel Cerenkov Counter discriminates Kaon over Pion (event variable) with 93% efficiency for momentum greater than $700 MeV/C$.