Visual Modeling of Ontology Views for e-Sciences Using XSemantic Nets

Rajugan, R.[1], Elizabeth Chang[2] and Tharam S. Dillon[3]
[1,2,3] *DEBII, Curtin University of Technology, Australia*
{rajugan,Elizabeth.Chang, Tharam.Dillon}@cbs.curtin.edu.au

Abstract

Ontologies are the foundation of the Semantic Web (SW) and one of the keys necessary to its success. Conversely, e-Science domains are usually characterized by heterogeneous semi-structured data models and formats. In recent years, building ontologies are gaining momentum in the e-Science domains to provide explanatory semantics in an agreed, yet shared encodings. Conversely, ontology views hold the promise of; (a) provide a manageable portion of a larger ontology for the localized applications and users, (b) enable precise extraction of sub-onlogies of a larger ontology that commits to the main ontology, (c) enable localized customization and usage of the portion of a larger ontology and (d) enable interoperability between large ontology bases and applications. Therefore, it is interesting to look at ontology views and their desired applications in the context of e-Sciences, as there exists no agreed upon standard, methodology or formalism to specify, define and materialize such ontology views. Thus, in this paper, we elaborate on our own research direction towards proposing a meaningful ontology view formalism and its associated semantics for the e-Science domain.

1. Introduction

The World Wide Web (WWW) is widespread and constantly growing with vast amounts of information being disseminated using a variety of technologies available to-date. But, unlike databases, it has created the problem of *standardization* vs. *customization*, where anyone can represent what s/he wants according to one's own representation and annotations. Thus, the traditional web is not semantically rich, and ad-hoc queries are usually done using search engines that query the web using strings rather than search by the meaning of data. Synonymous to this, a similar set of arguments holds true for the e-Science domains, where available data for a given domain does not conforms a given data structure and/or schema and are usually characterized by heterogeneous semi-structured data structure and formats. This has triggered a series of new research directions such as metadata description languages, search engines [1], knowledge representation [2] and Semantic Web (SW) [3] techniques to organize, describe, manage, disperse and query e-Science data in an orderly manner using the SW as the new medium. To overcome some of problems associated with these issues, researchers have recently proposed investigate ontologies to model and represent e-Science data and their associated semantics in a shared and agreed form. Examples of such works include the Protein Ontology (PO) in the biomap project [4-6], the Human Disease Classification (HDC) [7] and the work done at [8].

The Semantic Web, envisioned by Berners-Lee [9], enables users to define and represent information structures, using widely accepted metadata standards to annotate new vocabularies that are independent of any language or syntax used to implement it. These accepted metadata standards are formulated into an *ontology*, a notion which is the foundation of the SW and one of the keys to the success of SW.

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

1.1. Ontology

An ontology is the notion of *shared conceptualization* [10]. It can also be stated as a specification of a conceptualization of a problem domain [11, 12] or may be viewed as a *shared conceptualization* of a domain that is commonly agreed to by all parties [13]. Formally, an ontology may be stated as "a specification of a Conceptualization" [10]. Here, "conceptualization" refers to the understanding of the *concepts* and the *relationships* between the concepts that may exist or do exist in a specific domain or a community [13]. Also, a specification of the conceptualization refers to the notion or the representation of the commonly agreed, domain specific shared knowledge. There are many other ways of expressing ontologies, but such simplified definitions may incorrectly lead to oversimplification of the ontology concept, i.e. ontology is a collection of words, definitions and concepts, a similar notion of knowledge bases in the classical artificial intelligence communities. Therefore, it should be noted that an ontology is more than a knowledge base, and not merely a collection of distributed metadata repositories, or annotated traditional databases.

In the context of SW, theoretically, the entire world is modelled as one super-ontology [11, 12], providing great compatibility and consistency across all sub-domains. This illustrates the known characteristics of an ontology, where, in any domain, an ontology tends to grow larger than its original size, introducing problems such as: (i) being too large to be utilized in its original scale by potential applications; and (ii) capable of being understood by the community due to its size.

Another interesting characteristic that is worth mentioning here is the notion of *generic* vs. *specific* ontologies. This is one of the key features that distinguish ontology bases from traditional data repositories. Given an e-Science domain, it can be represented using both generic (and upper) ontologies and specific (or lower) ontologies. Simply stated, a specific ontology commits to the use of all generic ontology concepts and specification [14]. The specific ontologies are also known as *ontology commitments* [11]. The specific ontology may involve additional constraints and additional elaborations [14]. However, in this research we do not elaborate on this feature of ontologies, as the detailed discussion on ontology and ontology classification is outside the scope of this paper. In the following sections, we look at some of the concepts that are relevant to our research and to this paper.

1.2. Sub-Ontology

As described before, an ontology base tends to grow larger with usage. Therefore, there exists an interest to extract a portion (i.e. sub-section or sub-ontology) of the main ontology that is of interest to the user for a given task. This would have benefited both users from over-utilizing valuable computing resources and time in carrying out their tasks, and would have provided other benefits such as privacy, security and efficient access. Another reason for localized sub-ontology is the processing time involved executing semantic queries over large-scale ontology bases (for example, HDC, the PO, etc.).

For example, if we take the HDC ontology base or the PO, at a given point in time, only a sub-section of the vast ontological structure is required by a medical analyst (or biochemist looking at protein chains/sequences) to complete his/her task. But due to its size, HDC (or PO) in its entirety is a drawback as it presents the user with a vast amount of un-related vocabularies for a given task, thus slowing or preventing the user from making time-critical decisions. Also, given the vast amount of classified data (such as medication, patient history etc.) in the HDC, providing privacy, security and access to all the information present in the HDC for number of users, is an unmanageable task. But, given a sub-ontology [15, 16] extracted from the base ontology for a given user profile, it is a manageable task. Similarly, if we consider the PO[1] developed at our research group, the complexity involved in dynamically querying, materializing and visualising such vast protein structures is a challenging task. A sub-ontology is not just a portion of the existing ontology [16]. Rather, it is a view which may involve keeping the required nodes and concatenations of the relationships though nodes have been removed. However, there exist many issues that need addressing in the case of ontology views (i.e. sub-ontologies); (i) First, unlike database or semi-structured views [15, 17], sub-ontologies are not just an extracted portion of the main ontology base, but a collection of concepts, relationships and concerns that itself is a new interpretation of the base ontology [16], (ii) Second is the meaningful representation of such sub-ontologies that are easily understood by humans as well as easily transformed to machine- (or user-application) readable notations that are at a level of abstraction that is capable of interpreting, querying and processing of concepts, relationships and constraints, (iii) Third, are the complementary issues associated with sub-ontologies [10, 18, 19], such as view maintenance, versioning and materialization, that are synonymous with database views, but deserve detailed studies of their own, in the context of ontologies and (iv) Fourth is the issue of meaningful, yet efficient extraction of sub-ontologies from distributed base ontologies [12, 20].

Thus, these issues creates the need to investigate successful database technologies, such as views, in the context of SW, where ontology views [12, 17] can be used for: (i) Ontology extraction; (ii) Ontology versioning; (iii) SW-enabling traditional data sources; (iv) Sub-ontology extraction, in an industrial setting; and (v) Applications of SW. Therefore, given the benefits and concerns of sub-ontologies, the motivation for this research consists of several related sub-problems and these are: (i) the development of an approach that allows the separation of *implementation* aspects and the *conceptual* aspects of the ontology views so that there is a clear separation of concerns allowing analysis and design of ontology views to be separated from their implementation (i.e. view extraction and elaboration); (ii) the development of semantics for ontology views as described in point (i) as analogous to data and XML views [15]; (iii) to achieve point (i) above, requires extensions and elaborations of the *conceptual operators* presented in [15], in the context of LVM for XML, to permit extraction and elaboration of construction of ontology views at a higher level of abstraction; (iv) to define representations to express these ontology views at the conceptual level analogous to the LVM. This requires specification of customized graphical notations and the development of automated schemata transformation rules for mapping conceptual models to OWL (or RDF-S) like representations; (v) to investigate feasible deployment strategies for the proposed ontology views. (vi) to define a view development methodology for *ontology views* that utilizes the conceptual models and carries out automated transformation to a *template* (such as in RDF-S) for the views as well as a representation of the view construction in an appropriate query language such as XQuery [21], RDQL, etc.; and (vii) to apply the concepts, methods and transformations developed in (i) to (iv) above to specific domains, particularly: (a) PO in bioinformatics, (b) ontology approach in HDC.

---

[1] http://proteinontology.info/

In this paper, we look at developing visual ontology views [17] using XSemantic net [15] notation for e-Science domains such as HDC and PO. Our main focus is to develop and extend the XSemantic net notation to model ontology views.

3. Related Work

The existing ontology view research ranges from low-level, implementation specific interpretations to high-level semantic aware view formalisms, such as in [16]. These view researches are constrained by: (a) available ontology representation language or notation; (b) the lack of standardized storage model (and structure) for storing and querying ontology bases; and (c) the lack of standardized query languages. However, for ease of understanding, here, we broadly grouped the ontology view research into two categories based on the type of storage structure and/or data model adapted to store and represent the "stored" (or base) ontology. They are: (i) ontology views that defined over ontology bases in purpose built repositories for storing and querying ontologies. For example, the works that fall in this category includes, the MOVE view system [16, 22], User Oriented Hybrid Ontology Development Environments [8], etc. and (ii) ontology views that are defined over ontology bases in OO (or relational) databases or extended OO/relational environments. Here, the works include, the CLOVE project [23], KAON project [24], etc. A detailed discussion on these view formalism can be found in our work in [25].

4. Our Work

In this paper, we propose a modeling notation to model LVM views for ontologies using XSemantic nets. In work with XML, we provided clear distinction between conceptual, logical and document levels views, as in the case of data engineering, there exists a need to clearly distinguish these levels of abstractions. But in the case of SW (e.g. ontologies), though there exists a clear distinction between conceptual and logical models/schemas, the distinction between the logical (or schema) level and document (or instance) level trends to overlap due to the nature of ontology bases, where concepts, relationships and values may present mixed sorts such as schemas and values [18].

Therefore, here, we provide a clear distinction between conceptual and logical views, but depending on the application, we allow an overlap between logical and document views. This is one of the main differences between the XML views and the SW-views. To our knowledge, other than our work, there exist no research directions that explore the conceptual and logical view formalism for the Semantic Web (SW) paradigm.

4.1. Conceptual Views

In the layered view model, the conceptual views are views that are defined at the conceptual level with conceptual level semantics using a higher-level modeling languages such as UML [26] or XSemantic nets [15]. Here, we use XSemantic nets. To understand the SW-view and its application in constructing ontology views, it is imperative to understand its concept and its properties. A formal definition of LVM views for ontologies ca be found in our works [15, 17].

The term ***context*** refers to the domain that interests an organization as a whole. It implies a meaningful collection of objects (or concepts), relationships (both structural and semantic) among these objects, as well as some constraints associated with the objects and their relationships, which are relevant to its applications. In this paper, to model conceptual views, we use XSemantic nets. Other modeling notations used to

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

model conceptual views can be found in [17]. XSemantic net provides a well defined, rich semantics to visually model a given domain into needed level of abstraction. In the case of ontology engineering, XSemantic nets provide rich collection of OO concepts and elements, namely; (i) classes (similar to concepts in ontology), (ii) attributes (iii) relationships (and cardinality constraints) between classes, (iv) relationships (and cardinality constraints) between class and its attributes and (v) a rich set of constraints. Some of the XSemantic net notations are given Fig. 1. Base on the V^c definition 1 above, in XSemantic nets, V^c_{obj} are shown using (simple/complex) nodes, V^c_{rel} using edges and $V^c_{constraint}$ using constraints defined over (a set of) node/(s) and (a set of) edges. The section below briefly address some of the unique characteristics of conceptual views, namely conceptual operators.

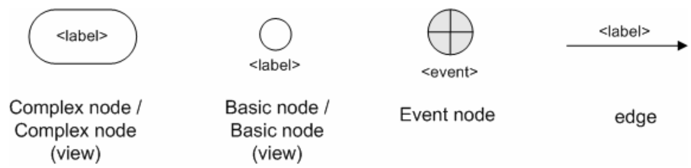

Figure 1

4.2. Conceptual Operators

A *Context* is presented in XSemantic nets using modeling primitives like *objects* (node), *attributes* (simple node), *relationships* (directed edges) and *constraints* in this study. To enable the construction of a valid conceptual view from a context, we introduced the notion of *conceptual operator* (λ). These operators are grouped into set operators, namely union, difference, intersection, Cartesian product and unary operators namely projection, rename, restructure, selection and joins, and can facilitate systematic construction of conceptual views from context. These conceptual operators can be easily transformed into query segments, user-defined functions and/or procedures for implementation. By doing so, they help the modeler to capture view construct at the abstract level without knowing or worrying about query/language syntax. The set of binary and unary operators provided here is a complete or basic set; i.e. other operators, such as division operator and compression operator [18] can be derived from these basic set of operators.

It should be noted here that, due to page limitations, we did not present a detailed discussion on logical and document level ontology view semantics, as it largely dependent on the underlying metadata description and/or representation language such as RDF, RDF-S, OWL, etc. However, for other language specific transformations (i.e. RDF, OWL, etc.), we intend to address this issue in the near future and at present we use other semantics and transformation methodologies proposed to carry out this task of mapping conceptual ontology views to logical and/or document views.

5. Conclusion and Future Work

In this paper, we presented an intuitive research direction to extend the LVM concepts to the Semantic Web based e-Science models using XSemantic nets. For future work, some further issues deserve investigation. First, the investigation of specifying and automated transformation of view specification, constraints and definitions of ontology views in the LVM to various SW language specific (e.g. OWL, RDF) schemata constructs. Second is the automation of the transformation process between the LVM conceptual operators to various SW (high-level) query language expressions (e.g. XQuery, RDQL etc.) with emphasis on quality and performance.

6. References

[1] R. W. P. Luk, H. V. Leong, T. S. Dillon, A. T. S. Chan, B. W. Croft, and J. Allan, "A survey in indexing and searching XML documents," *Journal of the American Society for Information Science and Technology*, vol. 53(6), pp. 415-437, 2002.

[2] K. Aberer, et-al., "Emergent Semantics Systems," First Int. IFIP Conf. on Semantics of a Networked World (ICSNW 2004) - Revised Selected Papers, Paris, France, 2004.

[3] W3C-SW, "The Semantic Web (http://www.w3.org/2001/sw/)," W3C, 2005.

[4] A. S. Sidhu, T. S. Dillon, and e. al., "Protein Ontology: Data Integration using Protein Ontology," in *Database Modeling in Biology: Practices and Challenges*, Z. Ma and J. Y. Chen, Eds. New York, NY: Springer, 2006, pp. to appear.

[5] A. S. Sidhu, T. S. Dillon, and e. al., "Ontological Foundation for Protein Data Models," First IFIP WG 2.12 & WG 12.4 Int. Workshop on Web Semantics (SWWS '05), Agia Napa, Cyprus, 2005.

[6] A. S. Sidhu, T. S. Dillon, and e. al., "Protein Ontology: Vocabulary for Protein Data," 3rd IEEE Int. Conf. on Information Technology and Applications (IEEE ICITA 2005), Sydney, Australia, 2005.

[7] M. Hadzic and E. Chang, "Role of the Ontologies in the Context of Grid Computing and Application for the Human Disease Studies," Semantics for Grid Databases, First Int. IFIP Conf. on ICSNW '04, France, 2004.

[8] HyOntUse, "User Oriented Hybrid Ontology Development Environments, (http://www.cs.man.ac.uk/mig/projects/current/hyontuse/)," 2003.

[9] T. Berners-Lee, "Weaving the web," Harper, San Francisco, 1999.

[10] T. R. Gruber, "A Translation Approach to Portable Ontologies," *Knowledge Acq.*, vol. 5, pp. 199-220, 1993.

[11] P. Spyns, R. Meersman, and J. Mustafa, "Data Modeling Versus Ontology Engineering," SIGMOD, 2002.

[12] C. Wouters, T. S. Dillon, J. W. Rahayu, E. Chang, and R. Meersman, "A Practical Approach to the Derivation of a Materialized Ontology View," in *Web Information Systems*, D. Taniar and W. Rahayu, Eds. USA: Idea Group Publishing, 2004.

[13] E. Chang, et al., *Trust and Reputation for Service Oriented Environments: Technologies for Building Business Intelligence and Consumer Confidence*. UK.: John Wiley and Sons, 2006.

[14] E. Chang, et al., *Trust and Reputation for Service Oriented Environments: Technologies for Building Business Intelligence and Consumer Confidence*. UK.: John Wiley and Sons, 2006.

[15] R.Rajugan, "A Layered View Model for XML with Conceptual and Logical Extension, and its Applications," in *Faculty of IT*. Sydney: University of Technology, Sydney (UTS), Australia, 2006, pp. 460.

[16] C. Wouters, "A Formalization and Application of Ontology Extraction," in *School of Engineering and Mathematical Sciences Faculty of Sciences, Technology and Engineering*. Melbourne: La Trobe University, Melbourne, Australia, 2006, pp. 299.

[17] R.Rajugan, E. Chang, and T. S. Dillon, "Sub-Ontologies and Ontology Views: A Theoretical Perspective," *Int. Journal of Metadata, Semantics and Ontologies (IJMSO), Inderscience Publishers*, vol. - to appear, 2006.

[18] C. Wouters, R.Rajugan, T. S. Dillon, and J. W. Rahayu, "Ontology Extraction Using Views for Semantic Web," in *Web Semantics and Ontology*, D. Taniar and W. Rahayu, Eds. USA: Idea Group Publishing, 2006, pp. 01 - 40.

[19] N. Noy and M. Musen, "Promptdiff: a fixed-point algorithm for comparing ontology versions," Proc. of the Eighteenth National Conf. on Artificial Intelligence, 2002.

[20] M. Bhatt, A. Flahive, C. Wouters, W. Rahayu, D. Taniar, and T. S. Dillon, "A Distributed Approach to Sub-Ontology Extraction," 18th Int. Conf. on Advanced Information Networking and Applications (AINA'04) - Volume 1, Fukuoka, Japan, 2004.

[21] W3C-XQuery, "XQuery 1.0: An XML Query Language (http://www.w3.org/TR/xquery)," in *XML Query Language (XQuery)*: The World Wide Web Consortium (W3C), 2004.

[22] C. Wouters, T. S. Dillon, J. W. Rahayu, E. Chang, and R. Meersman, "Ontologies on the MOVE," 9th Int. Conf. on Database Systems for Advanced Applications (DASFAA '04), Jeju Island, Korea, 2004.

[23] R. Uceda-Sosa, C. X. Chen, and K. T. Claypool, "CLOVE: A Framework to Design Ontology Views," 23th Int. Conf. on Conceptual Modeling (ER '04), Shanghai, China, 2004.

[24] KAON, "KAON Project (http://kaon.semanticweb.org/Members/rvo/Folder.2002-08-22.1409/Module.2002-08-22.1426/view)," vol. 2005, 2004.

[25] R.Rajugan, E. Chang, and T. S. Dillon, "Sub-Ontologies and Ontology Views: A Theoretical Perspective," *Int. Journal of Metadata, Semantics and Ontologies (IJMSO), Inderscience Publishers*, vol. - to appear, 2007.

[26] OMG-UML™, "UML 2.0 Final Adopted Specification (http://www.uml.org/#UML2.0)," vol. 2005: OMG, 2003.