Ontological View-driven Semantic Integration in Collaborative Networks

Yunjiao Xue[1], Hamada H. Ghenniwa[1], Weiming Shen[1, 2],

[1] Department of Electrical and Computer Engineering, The University of Western Ontario,
London, Ontario, Canada
[2] Centre for Computer-assisted Construction Technologies, National Research Council,
London, Ontario, Canada
yxue24@uwo.ca, hghenniwa@eng.uwo.ca, weiming.shen@nrc.gc.ca

Abstract. In collaborative enterprise networks, semantic heterogeneity is an important factor that hinders collaboration of various information systems. Ontology-driven semantic integration is an important category of solutions for the semantic integration problem. However, in many domains, there are no explicit and formal ontologies available. This paper proposes to adopt ontological views to address such challenges. It investigates the theoretical foundation of ontologies and ontological views. It presents a framework as a solution, based on the theoretical foundation, including the architecture of a semantic integration enabled environment, the modeling and representation of ontological views, and the semantic equivalence relationship discovered from the ontological views.

Keywords: Semantic Integration, Ontology, Ontological View, Collaborative Network, Semantic Equivalence Relationship.

1   Introduction

In a collaborative enterprise network, various information systems must collaborate to support information exchange and information processing and to meet other requirements. Due to the nature of being independently designed and built, the information systems, even for the same domain, are often heterogeneous in terms of (1) their supporting infrastructures (hardware, operating systems, communication facilities, etc); (2) syntactic representations of information; (3) schematic designs of information models, and (4) semantics of information, which will significantly hinder the collaboration between these systems. There have already been plenty of solutions for the first three areas of concern [9]. The final one, also known as the *semantic integration* problem [7], is attracting more and more attention from today's research communities.

As a category of solutions for the semantic integration problem, schema matching [8] aims at finding semantic relationships between schema elements such as database tables and table columns. Schema matches can be discovered by analyzing the

similarity of schema information, preservation of constraints, domain knowledge, and instance data. The limitation to this solution is the lack of a concept model.

Ontology-driven semantic integration is another category of solutions for the semantic integration problem [4]. The traditional solutions are based on available ontologies. The *ontology integration* can be applied to discover semantic correspondences among these ontologies [10]. However, in many domains, there are no pre-established explicit ontologies available.

This paper is dedicated to investigate the theoretical foundation of ontologies and ontological views, and analyze ontological view-driven solutions to address the semantic integration problem. The rest of the paper is organized as follows: Section 2 defines some fundamental concepts; Section 3 analyzes a hypothesis for this work; Section 4 discusses the framework for ontological view-driven semantic integration, and Section 5 concludes the paper.

2   Fundamental Concepts

2.1   Ontology and Ontological View

Research on *semantics* bases itself on computer-based information systems. Generally speaking, semantics refers to the intended meaning of a subject. In another words, the semantics of information in the information systems refers to the users' interpretation of the computer-based representation of the world [5]. It reflects the way that users relate computer-based representations to the real world. *Ontology* plays an important role in specifying the information semantics. An ontology is a formal and explicit specification of a conceptualization [3]. Simply, an ontology specifies the concepts and relationships between the concepts in a domain [1]. In the following we present a set of definitions based on [2] that are necessary to formally define ontology.

A *domain* D is a set of concepts, i.e., D = $\{C_1, C_2, \ldots, C_n\}$ where each C_i is a concept. A *domain space* is a structure <D, W>, where D is a domain and W is a set of maximal states of affairs of such a domain (also called *possible worlds*). Given a domain space <D, W>, a *conceptual relation* ρ^n of arity *n* is a function from a set W of possible worlds to the set of all *n*-ary relations on D, 2^{D^n}, ρ^n: $W \rightarrow 2^{D^n}$. A *conceptualization* of domain D is defined as an ordered triple **C** = <D, W, $\Re$>, where $\Re$ is a set of conceptual relations on the domain space <D, W>.

For each possible world $w \in W$, the *intended structure* of *w*, according to a conceptualization **C** = <D, W, $\Re$>, is the structure $\mathbf{S}_{w\mathbf{C}}$ = <D, $\mathbf{R}_{w\mathbf{C}}$>, where $\mathbf{R}_{w\mathbf{C}}$ = $\{\rho(w) \mid \rho \in \Re\}$ is the set of extensions (relative to *w*) of the elements of $\Re$. We use $\mathbf{S}_{\mathbf{C}}$ = $\{\mathbf{S}_{w\mathbf{C}} \mid w \in W\}$ to denote all the intended structures (or intended world structure) of **C**.

A *logical language* **L** is a composition of a vocabulary V and a set of models of the language. V contains constant symbols and predicate symbols. A *model* of **L** is a structure <**S**, I>, where **S** = <D, **R**> is a world structure and I: V→D∪**R** is an

interpretation function assigning elements of D to constant symbols of V, and elements of **R** to predicate symbols of V.

An *intensional interpretation* of a logical language **L** with a vocabulary V is a structure <**C**, $\mathfrak{I}$>, where **C** = <D, W, $\mathfrak{R}$> is a conceptualization and $\mathfrak{I}$: V→D∪$\mathfrak{R}$ is a function assigning elements of D to constant symbols of V, and elements of $\mathfrak{R}$ to predicate symbols of V. This intensional interpretation is called *ontological commitment* for **L**, denoted as **K** = <**C**, $\mathfrak{I}$>. We also say that **L** commits to **C** by means of **K**, where **C** is the underlying conceptualization of **K**. **K** constrains the intensional interpretation of **L**, i.e., the language is used in an intended way for a domain instead of an arbitrary way.

Given a language **L** with a vocabulary V and an ontological commitment **K** = <**C**, $\mathfrak{I}$> for **L**, a model <S, I> is *compatible* with **K** if: i) S∈S_C; ii) for each constant symbol c∈V, I(c) = $\mathfrak{I}$(c); iii) there exists a world w such that for each predicate symbol p∈V, I maps the predicate into an admittable extension of $\mathfrak{I}$(p), i.e. there exists a conceptual relation ρ, such that $\mathfrak{I}(p) = \rho \wedge \rho(w) = I(p)$.

Given a language **L** and an ontological commitment **K**, the set $\mathbf{I_K(L)}$ of all models of **L** that are compatible with **K** is called the *set of intended models* of **L** according to **K.** Given a language **L** with an ontological commitment **K**, an *ontology* for **L** is a set of axioms designed in a way such that the set of its models approximates as best as possible the set of intended models of **L** according to **K**.

According to the above definition, an "ontology" is a designed artifact that is committed to a conceptualization according to an ontological commitment. It reflects a view on the conceptualization. Since the conceptualization can be viewed in various ways, there actually is not just one unique "ontology" for it. Instead, different views of the conceptualization may exist. We define the formal and explicit specification of each view as an *ontological view*. Accordingly, its intensional interpretation is called an *ontological commitment of view*. There can be multiple ontological views for a single conceptualization.

Different languages can be employed to specify the ontological views. Further, if two languages are employed for ontological views with partially overlapping intended models, it is possible for the corresponding ontological views to be semantically integrated. Formally, given an ontological view O with intended models $\mathbf{I_K(L)}$ and another ontological view O' with intended models $\mathbf{I_{K'}(L')}$, O and O' are integrate-able (denoted by $\Diamond$) if and only if $\mathbf{I_K(L)}$ overlaps with $\mathbf{I_{K'}(L')}$. That is,

$$(\mathbf{I_K(L)} \neq \mathbf{I_{K'}(L')}) \wedge (\mathbf{I_K(L)} \cap \mathbf{I_{K'}(L')} \neq \varnothing) \leftrightarrow (O \Diamond O').$$

2.2 Semantic Integration

Information systems are built based on *information models* which are explicit models specifying information in the systems, such as a database schema. Given a set of information models IM_1, IM_2, …, IM_n, the semantic integration upon them includes two aspects:

(1) For any two elements e_i and e_j from IM_i and IM_j, if they refer to the same concept in terms of the domain of discourse, independent of the way they are represented, this fact can be discovered.

(2) For any element e_i from IM_i, if it is required to be communicated to IM_j (if applicable), it can be converted into another element (referring to the same concept) that is correct in both representation and semantics in IM_j such that IM_j can handle it in a semantically reasonable manner.

3   Hypothesis for Semantic Integration

Based on how information systems are built in a collaborative enterprise network, we can safely assume that there are explicit information models available and the models are committed to the intended models that overlap. The information models are not restricted to a particular language or paradigm, such as relational, XML, or Objected-Oriented (OO). Further, the modeling languages of the information models adopt symbols based on a natural language foundation such as English. The constant symbols such as English words refer to concepts under an ontological commitment.

An ontological view, as an explicitly represented model, can be created from an information model. The ontological views provide a common base that eliminates syntactical and schematic heterogeneities among information models, therefore the semantic integration can be conducted at the ontological view level.

Before we present the hypothesis of this research, we formally define the *semantically equivalent mapping* (or *equivalence mapping*) between languages:

Given a source language $\mathbf{L}_S$ with ontological commitment of view $\mathbf{K}_S = <\mathbf{C}, \mathfrak{I}_S>$ and vocabulary V_S, a target language $\mathbf{L}_T$ with ontological commitment of view $\mathbf{K}_T = <\mathbf{C}, \mathfrak{I}_T>$ and vocabulary V_T, the two languages share the same conceptualization $\mathbf{C} = <\mathbf{D}, \mathbf{W}, \mathfrak{R}>$, a semantically equivalent mapping is a function from V_S to V_T, m: $V_S \rightarrow V_T$, assigning symbols in V_T to the ones in V_S which share the same intensional interpretation, i.e., i) for constant symbols $c_S \in V_S$ and $c_T \in V_T$, $m(c_S) = c_T$ if and only if $\exists d \in \mathbf{D}$, such that $\mathfrak{I}_S(c_S) = \mathfrak{I}_T(c_T) = d$ and ii) for predicate symbols $p_S \in V_S$ and $p_T \in V_T$, $m(p_S) = p_T$ if and only if $\exists \rho \in \mathfrak{R}$ such that $\mathfrak{I}_S(p_S) = \mathfrak{I}_T(p_T) = \rho$.

In this context, we base our research on the following hypothesis:

If the semantically equivalent relationships between concepts (specified by symbols in languages) in multiple ontological views can be discovered, then these ontological views, as well as the information models from which the ontological views are developed , can be semantically integrated.

To support this hypothesis, we introduce the following two propositions.

(1) A concept in a conceptualization can be externalized by a constant symbol in a language under an ontological commitment.

Prove: According to the definition of the intended model, given a language $\mathbf{L}$ with an ontological commitment $\mathbf{K}$, the set $\mathbf{I}_K(\mathbf{L})$ of all models of $\mathbf{L}$ that are compatible with $\mathbf{K}$ is defined as the set of intended models of $\mathbf{L}$ according to $\mathbf{K}$. Therefore, for any two models m_1 and m_2 in $\mathbf{I}_K(\mathbf{L})$, m_1 and m_2 are compatible with $\mathbf{K}$. That is, for each constant symbol c in the vocabulary of $\mathbf{L}$, there is $\mathbf{I}_1(c) = \mathfrak{I}(c)$ for m_1 where $\mathbf{I}_1$ is the interpretation function of m_1, and $\mathbf{I}_2(c) = \mathfrak{I}(c)$ for m_2, where $\mathbf{I}_2$ is the interpretation function of m_2, and $\mathfrak{I}$ is the interpretation function in $\mathbf{K}$. That is, under the given ontological commitment $\mathbf{K}$, a constant symbol c is always interpreted as a concept in the domain of discourse.

On the other hand, since **I** is a function in any model, it is guaranteed that c is interpreted as only one concept, say C, under **K**. In other words, it is an explicitness of concept C. Therefore, even C is implicit, c can be taken as its representative. Since c is explicit, it can be used for processing the concept that it represents. □

(2) The semantically equivalent relationship between symbols under an ontological commitment implies the same concept reference.

Prove: Given symbols v_1 and v_2 from two ontological views such that v_1 maps to a concept C_1 in a conceptualization and v_2 also maps to a concept C_2 in the same conceptualization (Proposition 1), if v_1 and v_2 have a semantically equivalent relationship between them, then they have the same semantics, i.e., the same concept reference. Therefore, it can be concluded that C_1 and C_2 are actually the same concept in the conceptualization. Consequently, information models corresponding to v_1 and v_2 are semantically equivalent. □

The first proposition indicates that each ontological view has a specific *representation* based on a language since the ontological view is an explicit model. The second proposition shows that the *semantic similarity* between representations of models can be used to approximate the semantically equivalent relationships between the models themselves. Semantic similarity is a metric upon explicitly represented models computed from either a linguistic or structural perspective. Such a metric implies that two models may have the same semantics because their representations are linguistically or structurally similar to each other.

4   A Framework for Ontological-View Driven Semantic Integration

We propose a framework to achieve ontological view-driven semantic integration in open environments. This framework includes three main aspects: the architecture of a semantic integration enabled environment, ontological view modeling and representation, and the semantic equivalence relationship discovery.

(1) Architecture of Semantic Integration Enabled Environment

We propose a novel architecture that extends the traditional data/information architecture into three layers (see Figure 1), including:

(a) A data management and integration layer. This layer provides abstraction for the binary digits and organizes the digits into various types of elemental data such as numbers, characters, and strings.

(b) An information management and integration layer. This layer associates data to information models, providing specifications to data and converting data into information.

(c) A semantic management and integration layer. This layer deals with the semantics of information, resolves semantic heterogeneities, and ensures that information with the same semantics is handled in semantically consistent ways. The management and integration of this layer are addressed by a set of semantic integration services.
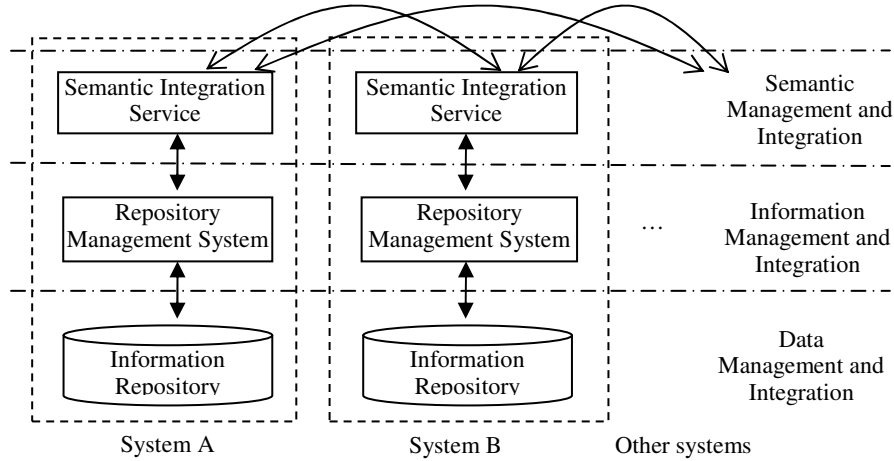
Fig. 1. Architecture of the semantic integration enabled environment.

(2) Ontological View Modeling and Representation

Many of the paradigms used to build information models, such as relational and OO, follow the *concept-property* construct, where a concept is modeled as a set of properties. In this work we believe it will be normal to adopt the *concept-property* construct for modeling ontological views. Therefore, we shall adopt a paradigm to support modeling concepts, properties, and relationships between concepts such as *isA* and *partOf*.

FRAME [6] is adopted for modeling the ontological views. It is a knowledge modeling approach that provides a clear and explicit structure adequate for modeling ontological views, particularly in describing the properties of concepts. A FRAME has a four-level structure, including FRAME (representing a concept), SLOT (capturing properties and relationships), FACET (capturing details of each SLOT), and DATA (providing specific information for instances of concepts). When modeling concepts, usually the DATA level is not used.

To specify the ontological views based on FRAME, we propose the FOSL (FRAME-based Ontological view Specification Language). FOSL is a logical language created from the following vocabulary:

(1) Constant symbols: the set of $FR \cup S \cup F \cup V$, where FR is a set of symbols referring to frames (concepts), S is a set of symbols referring to slots (properties), F is a set of symbols referring to facets, and V is a set of values that the facets can take.

(2) Predicate symbols: the following predicate symbols are considered:

a) A binary predicate *hasProperty* applied on $FR \times S$. *hasProperty(fr, s)* indicates that frame *fr* has a slot *s*.

b) A binary predicate *hasFacet* applied on $S \times F$. *hasFacet(s, f)* indicates that slot *s* has a facet *f*.

c) A binary predicate *hasValue* applied on $F \times V$. *hasValue(f, v)* indicates that facet *f* has a value *v*.

d) A binary predicate *isA* applied on $FR \times FR$. *isA(fr_1, fr_2)* indicates that frame fr_1 is a type of frame fr_2, i.e., the concept modeled by fr_1 is a specialization of the concept modeled by fr_2.

e) A binary predicate *partOf* applied on FR × FR. *partOf*(fr_1, fr_2) indicates that frame fr_1 is a part of frame fr_2, i.e., the concept modeled by fr_1 is a part of the concept modeled by fr_2.

We adopt XML-based representation for FOSL. An ontological view can be modeled as a set of FRAMEs and represented in an XML document. The document is supported with multiple <concept> tags for concepts (FRAMEs) respectively. Under a <concept> tag the SLOTs are tagged as <relationship> or <property>. The FACETs of each SLOT are tagged as <facet> which is described by two attributes: *name* and *value*.

To create an ontological view from an information model, the concepts, properties, and relationships need to be extracted. Given that an information model *M* is specified by language $\mathbf{L}_M = <\mathbf{S}_M, \mathrm{I}_M>$ with vocabulary V_M and the ontological view model is specified by language $\mathbf{L}_O = <\mathbf{S}_O, \mathrm{I}_O>$ with vocabulary V_O, the creation of an ontological view is to find a mapping *m* between $\mathbf{L}_M$ and $\mathbf{L}_O$ such that $m(\mathrm{I}_O) \subseteq \mathrm{I}_M$. The mapping adopts a set of rules for each modeling paradigm to identify what constructs in the information model can be mapped to concepts, properties, facets of properties, and relationships, and then maps them into corresponding constructs in the FRAME model.

(3) Semantic Equivalence Relationship Discovery from Ontological Views

The semantic equivalence relationship is deduced from the *semantic similarity metric* between symbols. A semantic similarity metric is a combination *<A, t>* where *A* is an approach to compute the similarity between symbols and *t* is a threshold. The approach *A* can be viewed as a function $A: S \times S \to R$ where *S* is the set of symbols and *R* is the set of real numbers. If $A(s_1, s_2) > t$, s_1, $s_2 \in S$, then it can be confidently believed that two symbols are semantically equivalent, i.e., s_1 and s_2 have a semantic equivalence relationship.

In the research of schema matching and ontology mapping, multiple approaches have been developed to discover the semantic relationships between elements of the schemas or ontologies [7, 8]. These approaches can be applied to ontological views.

(a) Linguistic (Syntactical) Matching

Linguistic matching utilizes the vocabulary of the modeling language to discover the semantic equivalence relationships. Linguistic matching works on symbols that are mapped to concepts under an ontological commitment.

In linguistic matching, the principle is that the more syntactically similar two symbols are, the more likely they map to the same concept, the same property, or the same facet. To increase the precision of the comparison, the symbols will often be normalized and compared, sometimes with the help of natural language dictionaries to determine the synonym when the symbols are syntactically different.

(b) Structural (Semantic) Matching

The structural matching utilizes the semantic structures captured by the FRAME model to discover the semantic equivalence relationships if syntactically matching cannot provide sufficient clues. The principle is: even two symbols are syntactically different, they can be semantically similar if the structures around them are similar. FRAME model's tree-like structure is utilized to support the inference that two symbols are semantically equivalent if their properties are very similar, even though they are syntactically different.

5   Conclusion

Understanding and integrating heterogeneous information have become more important and challenging in collaborative enterprise networks. Semantic integration, as an important factor for successful information integration, has become one of the most active research areas. It has received extensive interest and attention from both the academic and industrial communities.

Our work on semantic integration fits into the evolvement by extending the traditional ontology-driven approaches to an ontological view-driven approach to overcome the grand challenges that were not thoroughly addressed by the traditional approaches. The most significant advancement is the removal of the assumption about the availability of explicit ontologies. Besides this, our study embeds the support for semantic integration into existing systems and makes them semantic-sensitive. This extension will provide a new level of capabilities for the information systems to interoperate at the semantic level.

References

1. Crubzy, M., Pincus, Z., Musen, M. A.: Mediating Knowledge between Application Components. In: Proceedings of the Semantic Integration Workshop of the Second International Semantic Web Conference (ISWC-03), Sanibel Island, Florida (2003)
2. Guarino, N.: Formal Ontology and Information Systems. In: Proceedings of FOIS'98, pp. 3--15. IOS Press, Trento, Italy (1998)
3. Gruber, T. R.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies, Vol. 43, Issues 5-6, pp. 907--928 (1995)
4. Hakimpour F., Timpf, S.: Using Ontologies for Resolution of Semantic Heterogeneity in GIS. In: Proceedings 4th AGILE Conference on Geographic Information Science, pp.385--395. Brno, Czech Republic (2001)
5. Meersman, R.: An Essay on the Role and Evolution of Data (base) Semantics. In: Proceedings of IFIP WG 2.6 Working Conference on Database Application Semantics, pp. 1--7. Chapman & Hall, Ltd., London, UK (1995)
6. Minsky, M.: A Framework for Representing Knowledge, In Winston, P. H. (ed.), The Psychology of Computer Vision, pp. 211--277. McGraw-Hill, New York (1975)
7. Noy, F. N.: Semantic Integration: A Survey of Ontology-based Approaches. SIGMOD Record, Special Issue on Semantic Integration, 33(4), pp. 65--70 (2004)
8. Rahm, E., Bernstein, P.A.: A Survey of Approaches to Automatic Schema Matching. The International Journal on Very Large Databases (VLDB), 10(4), pp. 334--350 (2001)
9. Sheth, A. P.: Changing Focus on Interoperability in Information Systems: from System, Syntax, Structure to Semantics. In: Goodchild, M.F., Egenhofer, M., Fegeas, R., Kottman, C. A. (eds), Interoperating Geographic Information System, vol. 47, pp. 5--29. Kluwer Academic Publishers, Norwell, MA, (1999)
10.Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hubner, S.: Ontology-based Integration of Information -- A Survey of Existing Approaches. In: Proceedings of the IJCAI'01 Workshop: Ontologies and Information Sharing, pp. 108--117. Seattle, Washington, USA (2001)