Proceedings of the International Multiconference on Computer Science and Information Technology pp. 799–808
ISSN 1896-7094

Shared Ontologies to Increase Systems Interoperatibiliy in University Institutions

Richard Gil, Ana María Borges, and Leonardo Contreras

Department of Processes and Systems,
Simón Bolívar University, Caracas, Venezuela
{rgil,aborges,leocon}@usb.ve

Abstract. The increasing demand of quality information has forced that the methodological resources of system engineering come perfecting gradually. The practical materialization of diverse concepts and relations, are requiring that the semantic technology and ontological engineering obtain all their advance and potential. The University Institutions must be at the top of these findings. It is the purpose of this paper to show a first version of ontological integration of some subsystem models that interact inwards and outside of this type of institutions. These subsystems were selected intentionally of previous studies. The Protégé-PROMPT tool will be used as resource for the integration of some experimental ontology of such subsystems, using as methodological reference a framework that estimates an approach of shared terminology to be able to integrate in a compatible ontology. As result a method which allows obtaining a shared ontology is proposed.

Key words: Interoperability, Ontologies Integration, Shared Ontology, Protégé-PROMPT, University Institutions, Methodology.

1 Introduction

Obtain interoperability and systems integration is obvious an ambitious goal, an extended technical literature on this subject demonstrates the extent of it. Numerous papers have being published from the theory to the practice-experimental [1–3]. These papers give many current contributions in the scope of the ontologies. Most of them trend to smooth the way to implement more and better diverse information systems toward the desired integration. In this sense, and looking for methodological resources to approach viable solutions in our specific cases of study—University Institutions—it is required to emphasize the efforts of investigators to implement and to describe tools toward integration [4, 1, 5, 6]. Different authors use different technologies, among them, mathematical or algebraic formalization, algorithms, network-neuronal, rules of inference, heuristic, analysis of graphs, etc., that have enriched the existing alternatives to approach some of the problem linked with the heterogeneity of systems. Adopting these technologies appropriately, under a concept of "semantic protocol", it can make flexible the heterogeneous systems of such organizations, taking care of their own dynamic of its internal changes (structural or of processes) reacting in prospective as response to the external and continuous changes of a globalize world. In addition with the mentioned "desirable protocol" it would be allowed to conserve the necessary stability for the planned interchange between the constituent subsystems that always will be influenced by the continuous adjustments in the same trajectory (natural evolution) of the original dominion.

For this study, variants to a similar methodology to obtain interoperability proposed by Stuckenschmidth y van Harmelen [7] will be made. It is originally support on a frame of reference and methods to create shared ontologies. In this research we take three ontologies of subsystems of University Institutions that are being developing now, to ontologically integrated using a tool of most outstanding of technical Literature [8, 9], as Protégé-PROMPT. This methodological variant, cradle in this appropriate experimental area of the case of study, will allow to validate a practical way for the possibility of increasing interoperability between different dominions for diverse systems.

2 Systems Interoperability

Several previous studies as the one by Obrst [10] clarifies better the terms of systems integration and interoperability. System integration is more related to the source of data; systems interoperability with software applications and their levels of aggregation (Fig. 3, page 368 cited, shows the dimensions of interoperability and integration).

By the increasing necessity of semantic interoperability intra and inter systems and also within it and with other organizations, in this research it is considered pertinent to focus on recent tendencies of systems integration, known as Cooperative Information Systems suggested by Antonios and van Harmelen [11]:

"Cooperative Information Systems are servicing a diverse mix of demands characterized by "Content-Community-Commerce". These demands are originating in current trends for "off-the-shelf", software solutions, such as enterprise resource planning and e-commerce systems".

The main challenges and outstanding potential of application in the cooperative information systems are:

1. Knowledge Management (organized validated, consulted knowledge, etc.).
2. Electronic commerce Business-to-Client type (B2C, based on "intelligent agents").
3. Electronic commerce Business-to-Business type (B2B, abstraction of dominions using "agents").
4. Customized agents (agreed to own interests and preferences, and memorized by "agents").
5. Technology of Web semantic:
    a. Explicit Meta-data: to capture data meaning.

b. Ontologies (classes, relations, restrictions, logic relations and subgroup relations).

c. Logic (to solve logics of predicate and rules to support validation and consistency).

d. Agents (they work independently and with prediction to support decisions).

All these trends will be addressed in the sub-dominions of the subsystems considered to give a prospective character to the proposals.

Many factors affect the application interoperability, since not only it tries itself to solve the increasing of the organization internal information requirements, but that is desired to be able to integrate it with external information of other organizations (B2B, Business-to-Government (B2G)), or with those of the different users and/or clients (B2C, Government-to-Client (G2C)). The integration must be dynamic, flexible and transparent (intelligent agents) and in addition it will be customized (personal agents), based on common or determined rules (semantic logics and restrictions).

Vernardat [12] emphasizes the following:

"The systems integration consists of arming the different parts of a system guaranteeing at the same time the compatibility of the assembly as well as the good operation of the complete system.

. . . the interoperability of the systems is an average one to obtain integration, based on a weak connection between the parts of the system, but seated in the capacity of the parts to communicate among them to accede and to resort to its functionalities."

". . . The central paper of the models (based on ontologies) in the development of the interoperability capacities influences remarkably in obtaining the quality, since the processes of engineering of the interoperability of applications, are the platform of flows of the models."

It is for that reason that this proposal try to approach ontological integration as a methodological strategy that allows to obtain system interoperability [13, 10], and to experiment with specific cases of dominions of subsystems oriented to satisfy information services at University Institutions [14–16].

3 Integration Using Ontologies

Review of the appropriate interpretation of the term ontological integration and its strategies or methods to obtain it, such as alignment, merging, articulation, mapping, fusion, integration, morphism and others, will be tried to adopt a definition that clarify the meaning of it.

Kalfoglou and Schrlemmer [5] affirm that usually S (of O) is a mathematical structure, that can usually consist of a set of concepts (classes) keeping certain order and symbols whose arguments are defined by the concepts and their hierarchies; in addition they define a total and partial ontological transformation of the following form:

"A total ontological mapping from $O_1 = (S_1, A_1)$ to $O_2 = (S_2, A_2)$ is a morphism $f : S_1 \rightarrow S_2$ of ontological signature, such that $A_2| = f(A_1)$, i.e. all interpretations that satisfy O_2's axioms also satisfy O_1's translated axioms."

"In order to accommodate a weaker notion of ontology mapping, we will say that there is a partial ontology mapping from $O_1 = (S_1, A_1)$ to $O_2 = (S_2, A_2)$ if there exists a sub-ontology $O'_1 = (S'_1, A'_1)$ ($S'_1 \subseteq S_1$ and $A'_1 \subseteq A_1$) such that there is a total mapping from O'_1 to O_2."

There are several ontologies integration "domains". They can be classified according to the traditional approach, as well as the emergent approach. From the traditional approach the ontologies evolutions is classified by the emergent approach as the share information between pairs (P2P), the schemes integration as the intelligent communication agents, the catalogue integration as the composition of Web services, and the data integration as the answers to questions on the Web (query).

There are other classifications which besides to include the techniques of ontologies match based on schemes, that allow to group the techniques arranged by the level of the match technique from syntactic to semantic through the structural. They considered the desegregation level, from the ontology-element-level to the structure-or morphology-level.

The Problem of the ontological mapping can be because in the phase of conceptualization, the semantics of the dominion will be encoded in different ways [17]. The potential problems include:

- Different ontologies may use different names to represent the same concept, or the same name to represent different concept.
- Different ontologies may use different values to represent the same concept.
- The same concept may be represented at the class level in one ontology and at the instance level in another one.
- Different ontologies may use different structural representations of concepts.

The problems involved in mapping ontologies have been tackled in three ways:

- Manual approach: similar concepts are identified by human experts and mappings by hand.
- Semi-automated approach: to automate some processes of the mapping to attend the user.
- Total-automated approach: to try to complete all the process without human intervention.

One of more commonly used is the semi-automated, and is in this category where Anchor-PROMPT is placed. This is the approach use in this research [17]:

"...this one take as input pairs of related terms ("anchors") defined by the user, and then performs the remainder of the mapping process automatically: from the set of previously defined anchors, new pairs of semantically close terms are produced by analyzing the paths between the anchors in the corresponding ontologies".

A model to represent efficiency and effectiveness phases [18] can facilitate the compression of the phases of the different match approaches.

Abels, Haak and Hahn identified based on the literature, that the methods used by ontology integration approaches are [19]: text similarity; keyword extraction; language based methods; identification of word relations; type similarity / domains and ranges; class inheritance analysis; structural analysis / taxonomic structure; data interpretation and analysis of key properties, and graph- mapping.

4 Proposal of Methodological Variant for Integration

Noy and Musen [20] expose that “given the diversity in the tasks that the tools supports, their modes of interaction, the input data they rely on, it is impossible to compare the tools to one another and to find one or even several measure to identify ‘the best’ tool”.

Therefore, the authors suggest that the evaluation must be user-oriented. “A user may ask either what is the best tool for his tasks or weather a particular tool is good enough for his task” [20]. Therefore they recommend for the evaluation pragmatic criteria such as:

- Input requirements: elements from the source ontologies do the tool use.
- Level of user interaction: batch or interactive mode, partial or final results.
- Types of output: what is the result of working with the tool?
- Content of output: which elements of the source ontologies are correlated in the output”.

In this sense and searching of an operative, practical and structured answer, a methodological framework was taken from Stuckenschmidth and van Harmelen [7]. They presuppose an approach of shared terminology to be able to integrate in a compatible ontology (Fig. 2.2, page 37 cited).

As far as the three alternative ways suggested by these authors of how to employ ontologies for the integration, in this research we use the approach of multiple or local ontologies. We use it because it is try “to provide a new ontology source that allows to relate the other technologies” [7] (Fig. 2.1, page 32 cited).

The authors of the framework make a formal definition of shared terminology, shared ontology and source ontology that facilitates the understanding of each one of the resources required to derive the desirable shared ontology, these are [7]:

“A shared terminology is a tuple $\langle$W,l$\rangle$, where W is a set of words and $l : WxW \rightarrow$ {syno,hyper,hypo} is a partial function from the set of all pairs of terms into a set of identifiers specifying whether the first term in a synonym, a hypernym or hyponym of the second.”

“A shared ontology is a tuple $\langle$ST,T,R,A$\rangle$, where $ST = \langle W_L, l_L \rangle$ is a shared terminology. T is a basic set of terms, R is a set of relations $R \subseteq TxT$ and A is a set of axioms of the form $T_i \sqsubseteq T_j$ if the following conditions hold:

$T \subseteq W_l$,

For each pair of words (W_i, W_j),
If $l((W_i, W_j))$ = hyper then $W_j \sqsubseteq W_i$ is in A
If $l((W_i, W_j))$ = hypo then $W_i \sqsubseteq W_j$ is in A
If $l((W_i, W_j))$ = syno then $W_i \sqsubseteq W_j$ and $W_j \sqsubseteq W_i$ are in A"

"A source ontology is a tuple ⟨S,C,d⟩, where $S = \langle ST_S, T_S, R_S, A_S \rangle$ is a shared ontology, C is a set of class names not from the set of terms in S, L is a terminological language and d is a function that assigns expressions δ_i to class names C_i in C such that:

δ_i only refers to relations in R_S,

L is defined over T_S"

At the methodology proposed by such authors, the process steps are enumerated (Fig. 4.1, page 69 cited) as fallow:

1. Finding common concepts: common or semantically translated concepts.
2. Definition of properties: attributes that describe the concepts.
3. Finding property values: ranges and level of resolution of variables (values possible).
4. Adapt ontology: use and "invent" variations when specific terms are required.
5. Refine definitions: refining the concepts, the "bridges terms" and the vocabularies.

5 Applying the Approach

This study contemplates the integration of three ontological models (based on OWL) of a proposal for a University Institutions System [21]. First, two ontologies that were developed by others were integrate and have been taken from Internet. Both were developed in English language and deals with University dominions associate to the Administration—Management University Model. After that we have integrated the ontology shared resulting with the third ontology of the Company / Organization Model previously developed in Spanish language [22]. The proposed methods intend to gain some experience with integration using partial ontological models for futures universitary case or another practical applications.

As far as the methodological variant proposed in this paper, based on the original one developed by Stuckenschmidth and Harmelen [7], two important adjustments are made. One is related to information sources, obvious conditioned by the dominion of the application (University Institutions) and the other, somewhat more important, by the characteristics and scope of the model tool: Protégé-PROMPT.

5.1 Knowledge Sources:

In order to obtain the "more possible standardized" shared ontology, it will require to sharpen the vocabulary and the conceptualization. To reach those are used: Main Upper level ontologies, domain ontologies, dictionaries and thesauri with the typical terms of the dominion and additionally linguistic thesauri [23, 7].

a) Domain Ontologies:

a.1) UpperCyc Ontology. (http://www.cyc.com) This ontology was developed by CyCorp Corporation. The Cyc knowledge base was constructed in English language. The Cyc knowledge base (KB) is a formalized representation of a vast quantity of fundamental human knowledge: facts, rules of thumb, and heuristics for reasoning about the objects and events of everyday life. The Cyc KB is divided into many (currently thousands of) "microtheories", each of which is essentially a bundle of assertions that share a common set of assumptions; some "microtheories" are focused on a particular knowledge domain, a particular level of detail, a particular interval in time, etc. At the present time, the Cyc KB contains nearly two hundred thousand terms and several dozen hand-entered assertions about/involving each term. New assertions are continually added to the Cyc knowledge base by human knowledge enterers. Additionally, term-denoting functions allow for the automatic creation of millions of non-atomic terms. For this work the version 1.0.2.Win32 of the OpenCyc tool for Windows had been used.

a.2) Suggested Upper Merged Ontology (SUMO). (http://suo.ieee.org) This ontology is being created as part of the IEEE Standard Upper Ontology Working Group. The goal of this Working Group is to develop a standard upper ontology that will promote data interoperability, information search and retrieval, automated inference, and natural language processing. The SUMO has been translated into various representation formats, but the language of development is a variant of KIF. SUMO is a collection of approximately 1000 well-defined and well-documented concepts, interconnected into semantic network and accompanied by a number of axioms. The axioms mostly reflect common-sense notions that are generally recognized among the concepts. SUMO is intended as a domain-independent substrate for designing domain ontologies.For this work SUMO155 had been used.

a.3) SWAT Projects—The Lehigh University (LUBM). The Lehigh University Benchmark is developed to facilitate the evaluation of semantic Web repositories in a standard and systematic way (http://swat.cse.lehigh.edu/projects/lubm). The benchmark is intended to evaluate the performance of those repositories with respect to extensional queries over a large data set that commits to a single realistic ontology. It consists of university domain ontology, customizable and repeatable synthetic data, a set of test queries, and several performance metrics. They are currently conducting experiments to determine the tradeoff space for various implementations of OWL repositories.

b) Domains and Linguistic thesauri:

b.1) WordNet. Is a large lexical database of English, developed by The Cognitive Science Laboratory of Princeton University (http://wordnet.princeton.edu). Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. Version 2.1 for the English language, used in this work, contains more than 120 thousand words and near 100 thousand concepts. It is organized in sets interrelated of synonymous (synset) and antonyms (antset), hypernymy and hyponymy (subclass-of and superclass-of), meronymy and holonymy (part-of and has-a).

b.2) Multilingual Extended WordNet. The words networks are structured of form similar to WordNet (`http://gplsi.dlsi.ua.es/wsd/ext_index.html`). Use a multi languages index that allows the searches of words meaning, the comparison of synset and its interrelations through different languages.

6 The Subsystems: Ontological models

Solution domain (University Institution) attempt a general model previously proposed [21]. The subsystems that conforms the semantic model are five, but those emphasized down (3 and 4) underlying are the worked ones for integration during this experimental approach, the original were: 1) Academic—Educational, 2) Production—Extension. 3) Company / Organization 4) Administration—Management University 5) Research and Development.

6.1 Specific Semantic models selected:

- 3) Company / Organization Model: Management model for company of service technologies (TICs).

- Sale and distribution of technology services (Paradigm ITIL).
- Construction Strategy: incremental using by referring upper ontology.
- Application scenario: human communication and access to data via transforming ontologies.

- 4) University Administration—Management Model: Management model e-learning and Government-e from existing ontologies and reference models taken from technical Literature.

- Architecture and Government-e supported in Web semantic.
- construction strategy: ontologies from others authors [24–26].
- Application scenario: access to data via shared ontologies and transforming ontologies.

7 Methodological Proposal Results

A method proposal is developed to obtain a shared ontology the study cases using Protégé-PROMPT, and the diverse sources of knowledge mentioned in 5.1. The shared ontology is derivate in four recursive phases trying to improve the definition of concepts and terminology (towards a Lightweight ontology):

1. Administration—Management University two-generic ontology integration (using PROMPT).
2. Upper ontologies comparison and aggregation (using UpperCyc, SUMO)
3. Terms and linguistic structure validation (using Extended Wordnet).
4. Company / Organization ontology merging (using PROMPT).
5. Terms and linguistic structure validation (using Extended Wordnet, Spanish).
6. Terms comparison and aggregation from a specific domain ontology (using LUBM).

8 Conclusions and recommendations

1. Integrate subsystems using ontologies is facilitated if versions of shared ontologies are achieved. In our case of study this strategy was feasible because the sources and terminology are "convergent enough", they can be specified in the same language (OWL) and the referring models are somewhat generic.

2. The integration strategy seems to be generalized to the other sub dominions of this study case, by the similarity in the strategy of development of the specific ontologies, which would be verified in later researches.

3. The methodological framework that was adapted as far as tools and knowledge sources, seems to be appropriate to the domain nature for University Institutions in addition to consider too, the diversities of languages sources (Spanish and English), where the ontologies could be conceptualize. This strategy must be a valid option for terminological integration of similar domains specified with sources multi-languages, characteristic of a globalize world.

4. Later research could be allowing to assuring the adaptability and changes possibility of the shared ontologies derived, will be the challenge of next studies and experimentations.

References

1. Wache H., Vögele T., Visser U., Stuckenschmidt H., Schuster G., Neumann H., Hübner S.: *Ontology-based Integration of Information—A Survey of Existing Approaches,* International Joint Conference of Artificial Intelligence (IJCAI 2001), Workshop: Ontologies and Information Sharing (2001).
2. Noy N.: *Semantic Integration: A Survey of Ontology-Based Approaches,* SIGMOD Record, Vol. 33, No. 4, 2004.
3. Buccella A., Cechich A., Brisaboa N.: *Ontology-Based Data Integration Methods: a Framework for Comparison,* Revista Colombiana de Computación, Vol. 6, No. 1 (2005).
4. Shvaiko P., Euzenat J.: *A Survey of Schema-based Matching Approaches,* Journal on Data Semantics (2005).
5. Kalfoglou Y., Schorlemmer M.: *Ontology Mapping: the State of the Art,* The Knowledge Engineering Review, Vol. 18:1, Cambridge University Press, UK (2003) 1–31.
6. Noy N., Stuckenschmidt H.: *Ontology Alignment: an Annotated Bibliography,* Dagstuhl Seminar Proceedings (2003).

7. Stuckenschmidt H., van Harmelen, F.: *Information Sharing on the Semantic Web,* Springer-Verlag, Berlin Heidelberg New York (2005).
8. Noy N., Musen M.: *PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment,* Proceedings of the 17th National Conference on Artificial Intelligence (AAAI 2000) (2000).
9. Noy N., Musen M.: *Using PROMPT Ontology-comparison Tools in the EON Ontology Alignment Contest,* Workshop on Evaluation of Ontology-based Tools (EON2004) at the Third International Conference on the Semantic Web (ISWC-2004), Hiroshima, Japan (2004).
10. Obrst L.: *Ontologies for Semantically Interoperable Systems,* Conference on Information and Knowledge Management (CIKM 2003), New Orleans, Louisiana, EE. UU. (2003).
11. Antonious G., van Harmelen F.: *A Semantic Web Premier,* The MIT Press (2004).
12. Vernardat F.: *Modelos Meta y Modelos para la Integración y la Interoperatividad de las Aplicaciones de Empresas de Producción,* PHD-Tesis, French (2006).
13. Kalfoglou Y., Schorlemmer M., Uschold M., Sheth A., Staab S.: *Semantic Interoperability and Integration,* Dagstuhl Seminar 04391 (2004).
14. López de Vergara J., Villagrá V., Berrocal J.: *An Ontology-based Method to Merge and Map Management Information Models,* Proceedings of the HP Openview University Association Tenth Plenary Workshop, Geneva, Switzerland (2003).
15. Park J., Ram S.: *Information Systems Interoperability: What Lies Beneath?* ACM Transactions on Information Systems, Vol. 22:4 (2004) 595–632.
16. Ding Y., Fensel D.: *Ontology Library Systems: the Key to Successful Ontology Reuse,* Proceedings of the First Semantic Web Working Symposium, Vrije Universiteit Amsterdan, Netherland (2001).
17. Villa R., Wilson R., Crestani F.: *Ontology Mapping by Concept Similarity,* Proceedings of the International Conference on Digital (2004).
18. Ehrig M, Staab S.: *Efficiency of Ontology Mapping Approaches,* International Workshop on Semantic Intelligent Middleware (2004).
19. Abels S., Haak L., Hahn A.: *Identification of Common Methods Used for Ontology Integration Tasks,* Proceedings of the First International Workshop on Information System, Bremen, Germany (2005).
20. Noy N., Musen M.: *Evaluating Ontology-mapping Tools: Requirements and Experience,* Workshop on Evaluation of Ontology Tools at EKAW, (2002).
21. Gil R., Contreras L., Ferrer J., Rivas C.: *Estrategia Ontológica para Abordar Heterogeneidad de Dominios en Instituciones Universitarias,* Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI 2007), Orlando, Florida, EE.UU. (2007).
22. Cumare L., Gil R.: *Bases para Desarrollar un Modelo Ontológico que Optimice Niveles de Servicio de Tecnología de Información Basados en las Mejores Prácticas de ITIL,* Caso de Estudio. Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI 2007), Orlando, Florida, EE.UU. (2007).
23. Gómez-Pérez A., Fernando-López, M. Corcho, O.: *Ontology Engineering,* Springer Springer-Verlag, London (2004).
24. Alani H.: *Ontology Construction from Online Ontologies,* Proceedings of the 15th International Conference on World, Edinburgh, Scotland (2006).
25. Taboada M., Martinez D., Mira J.: *Experiences in Reusing Knowledge Sources using Protégé and PROMPT,* International Journal of Human-Computer Studies 62 (2005) 597–618.
26. Noy N., Musen M.: *The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping,* International Journal of Human-Computer Studies, Elsevier (2003).