An Ontology-based Approach to Semantic Integration in e-Bussiness

Rômulo Passos, Vivianne Medeiros, Roberto Barros
Centro de Informática – Universidade Federal de Pernambuco (UFPE)
e-mail: {ranop, vnm, roberto}@cin.ufpe.br

Abstract

Nowadays, solutions in large-scale electronic commerce have to deal with several challenges in information integration. Some issues in this task involve coping with the explosion of the number of mappings and the problem of standardization itself. Much work appeared trying to solve integration in e-Business, however most of them are aimed at a particular solution. The purpose of this work is to bring a method of using a main, top layer ontology intermediating the domains. This ontology must be the most generic possible within the underlying domains, which specialize and instantiate it according to individual necessities. The advantage is to associate certain ontology concepts with Semantic Web standards and technologies in order to preserve independency in data distribution, even though preserving the consistency of information. This architecture is validated by the implementation of a prototype using a Web interface and design, and the Ontology Web Language (OWL, a W3C recommendation).

1. Introduction

In the past years, companies have independently developed their own information systems which data are heterogeneous and inconsistent across enterprises. The emergence of Web built up the necessity of creating electronic marketplaces, virtual places in which buyers and sellers meet to negotiate and exchange goods and services, by sharing information that is often obtained as hypertext catalogs from different companies [1].

In particular, B2B parties usually adopt different ways to specify their products, and as there are an enormous number of standards, several integration problems appear that require the development of special integration architecture.

The problem increases when solutions have to deal with dimensions and data heterogeneity of the Web in order to accommodate the semantic differences between information sources. Solutions capable of capturing the semantics of Web pages are incipient, what has led to the demand and adoption of ontologies. This technology appears as a means to logically and consistently structure the information spread in the network.

In [4], ontologies are defined as “formal and consensual specifications of conceptualizations that provide a shared and common understanding of a domain, an understanding that can be communicated across people and application systems”.

Basically, two characteristics of Internet make more difficult the access of specific and relevant information: the number of sources and the absence of precise semantic definition in the available information to make them understandable by programs and systems. It means that it is necessary to aggregate value to the information so that it can be inferred by humans and also by intelligent agents. In this context, ontologies are applicable.

Recently, Web interoperability hinged on the promises of XML to provide a standard for a shared common language. Unfortunately, the lack of XML semantics proves to be an obstacle for the development of Web services that can autonomously act on the electronic market [14].

Although ontologies can be adopted in situations where the capability to represent semantics is important enough to overcome XML’s maturity advantages [7], most of current state-of-the-art business integration services are based on XML mediated frameworks. On the other hand, ontology based implementations up to now have developed information integration based in ontology mappings (in XSLT translation fashion).

Thus, the main purpose of this work is to provide a solution that applies theoretical solutions to some key problems of e-business content management, while adding some differentials, such as adopting Semantic Web paradigm for information integration approach. It was used Ontology Web Language (OWL), a W3C standard, to define the formal concepts, besides it was built a Web-based interface to submit queries to the sources.

This paper shows the application of a tool – called OntoShop – that makes use of OWL ontology layers, providing information distribution while imposing an approach of hierarchy over information. In this system, clients can look for information about products and services available in servers in a unique view fashion.

The remainder of the paper is organized as follows. In Section 2, related work and contributions are detailed. Section 3 provides a description about the technologies that compose the design of the architecture. Section 4 gives a detailed explanation of the architecture proposed. Section 5 presents a prototype built to validate the structure adopted. Finally, Section 6 provides a summary and perspectives for future work.

2. Related Work

The topic of information integration has extensive importance to Data Management research. Mediators have been used for information integration by recognized systems such as Information Manifold [8], TSIMMIS [6], OBSERVER [10], among others. Even though mediators and the benefits of ontologies are known to data integration systems, the work presented here aims to achieve new hints.

In particular, benefits of using ontologies are well known, but Semantic Web paradigm and technology standards are not yet well explored.

In the context of mapping complexity for information integration, which means that each conceptualization has to be mapped one another, one solution is for developers to write code that translates between terminologies of pairs of systems. However, this solution is not suitable for large-scale integration.

A way to achieve a solution to the problem of mapping complexity is to formally specify the meaning of the terminology of each system and to define a translation between each system terminologies and an intermediate terminology.

Some basics of knowledge engineering, sub area of Artificial Intelligence, say that no solution to the Internet should be general, and the biggest prove resides in the fact that the tests in Information Retrieval show that restricting domains of knowledge improve their performance [5]. Moreover, ontologies can improve the applications on the Web because information can be referred and reached more precisely, instead of using the potentially ambiguous keywords. Beyond that, the delimitation of scope of the domain is more precisely defined.

None of the mentioned systems is functionally designed for the Semantic Web. The paradigm of the Semantic Web intends that each site provide its own ontology, based on technology standards, such as OWL. Thus, in order to achieve Semantic Web intents, ontologies have to be built from each participating entity, similarly to today's construction of HTML sites.

As the system presented here does not deal directly with multidatabases, but with multiple ontologies, presuming their construction by each business entity, the approach discussed in this paper presumes the abstraction of each data management system. This approach has some particular advantages to e-business environment. Each enterprise maintains its own data structures and elements private, only providing a view of the logic that is desired to reveal, by using ontologies as an interface to the Web.

So, the presented approach does not have to deal directly with the problem of query transformations, as all the logic has the start point in a central ontology and is managed by a unique inference engine. Moreover, it does not have to cope with data transformations, because each ontology extend a main ontology. This is not faced as a lack of functionality, but on the contrary that means the project explores Semantic Web potentials and facilities of ontology automation in order to achieve equivalent or even better results, when compared to other information integration systems.

Ontology paradigm here is not only used in a centralized fashion, but also distributed, intended to achieve scalability and extendibility, in the patterns of Semantic Web. It is not intended to directly resolve data and concepts discrepancies, but to act as a broker to complementary concepts that extend a main ontology.

Other information integration systems are aimed to act as database source integrators. This approach is aimed to act as an ontology sources integrator, supposing each information source is in the same general knowledge domain, but each having its specific ontology definitions as the extension of the main one.

Differently from this work, others used more traditional languages, such as OIL or RDF Schema for ontology implementations, and there are a few projects in OWL considering its potentials, what is explained by the fact OWL is new to this area (adopted as W3C standard in early 2004). Other ontology languages lack of conceptualization instruments that OWL can provide, and none of the mentioned systems use the ontology language to both store logic and the data itself.

Unlike Information Manifold, this work uses the extensibility of OWL itself for selecting sources and retrieving information from them, in an algorithm based in searching for information in the hierarchized layers of the ontology, as described, while not being necessary to adopting any extra solution to achieve it. The only need is to have references in the system to the sites to be retrieved.

The project presented in this paper is focused in alignment of content standards. The solution shown here is not proposed to be a full e-Business solution for negotiations and enterprise schemes. It is intended for enterprises to provide an interface for consistently exchanging information (e.g., products and services) among sets of enterprises disperse in the World via Web. Heterogeneity of e-Commerce cannot be captured by one standard, in spite of the need of personalized views on them. Therefore, scalable mediation service between different standards is essential. Here is described how ontologies can contribute to a solution for this problem, focusing on the alignment of business concepts and classifications.

3. State of the Art

3.1. e-Business Integration Issues

The Web explosion has transformed information systems from single isolated node to entry points into a worldwide network of information exchange and business transactions. As a result, solving queries in this context may involve retrieving and integrating information from multiple heterogeneous sources.

The problems that have to be faced are mainly due to both structural and application heterogeneity, as well as semantic differences between information sources. The latter can cause different kinds of conflicts, ranging from simple contradictions in meaning of terms to different primitives used to represent the same information.

Accordingly, one of the main challenges for the designers of the e-commerce infra-structures is the information sharing, retrieving data located in different sources thus obtaining an integrated view to overcome any contradiction or redundancy [1].

3.2. Ontologies

In the study of databases, ontologies have been used mainly in heterogeneous databases and data warehouses as global conceptual models, resulting of an entity-relationship set of definitions, in order to interact information.

Constructing a shared domain ontology from the beginning is difficult as knowledge acquisition problems have to be handled. However, a number of specific parts of the business integration domain have been modelled in depth within standardization initiatives managed by large consortiums.

Ontology classifications [5] are very interlaced with the genericity intent. It is important to observe classification in order to correctly define the purpose of the ontology. General ontologies are the most referenced ontologies in literature, because many applications take them as their basis.

So, it is interesting to adopt general ontologies in order for them to be reused and to serve as intermediaries of integration schemes.

3.3. The Semantic Web

The Semantic Web was first envisioned by Tim Berners-Lee [2]. The philosophy is the same as that of World-Wide Web – anyone can publicize information or consume anyone else's information.

For many, the Semantic Web symbolizes the next generation in the evolution of the WWW. Lee first envisioned a Web where not only humans, but also machines – to be more precise, softwares called intelligent agents – would be able to crawl and acquire information from the Web. Although the Web today is in spread use, its current structure, based in standards only for human interfaces and having no way of defining precisely information – information is only in natural language –, makes it very difficult to apply automation to the Web.

So far, most Semantic Web research has focused on defining standards for communicating facts, rules, ontologies, etc. XML, RDF, RDF-schema, OWL and others form a necessary basis for the construction of the Semantic Web.

This so-called "new Web" aims to provide services and automation in an environment where machines can help humans. However, if not enough people represent information at all, or not richly enough, or not in sufficient number, to make these services viable, the Semantic Web will not come into reality.

By applying concepts and standards of the Semantic Web, the approach here described goes in the direction of making the solution available worldwide and contributing for spreading information. Moreover, these concepts give the basis to build a more enriched information system.

3.4. The problem of mapping complexity

Current solutions for interoperation in e-Business based in ontologies tend to use a mapping approach to integrate information. In this context, three Business Enterprise Ontologies are involved in B2B integration across two enterprises: the one at the sender side, the one at the receiver side and the one that is used as an exchange format, i. e. the messages types and values of elements within messages. The two organizations map two ontologies each: from the local to the interchange ontology and the interchange ontology to the local one. If an organization has several trading partners using different B2B protocols, several mappings from the local to several interchange ontologies have to take place. In this case each organization has to deal with many ontologies at the same time [3]. This leads to a n.m ontology mapping complexity.

Despites there is some work that facilitate or automate the mapping task, the paired mapping scheme limits the possibilities, as each company has to be involved in B2B integration by creating a (limited) number of mappings.

Furthermore, building mappings between the required business enterprise ontologies is time consuming and requires a lot of knowledge specific to the different standard applications as well as B2B protocols. Only large corporations can afford to maintain all the knowledge to build and maintain their own mappings. However, by far not all organizations can afford to run major standard applications and make use of the translation services provided by third party companies [3]. As a solution to this issue, [11] proposes the use of a unified ontology, mediating the business elements, in order to decrease the complexity to n+m.

4. Architecture

The architecture here proposed intends to provide an effective and extensible mechanism for integrating diverse sets of data sources. It is expressed in a top-down fashion, since it starts with domain analysis to elucidate key concepts by consulting corporate data standards, and information models in order to reach a commitment of communities to common ontologies. It distinguishes two levels of ontologies:

- Top-level ontology, which is shared by different communities and model the basic entities;

- Specific ontology, which is structured according to each data provider. It is composed by hierarchies derived from top-level ontology to state subclass-of relationships and exposes its information as a mean of the main ontology through definition of correspondences between them. It also stores instances of the available data.

In a nutshell, this approach is based on the idea of groups reaching an agreement on what are the primordial entities of their world. Thus, this information is modeled as a top-level ontology which gathers all the general concepts relevant to the area. After the top-level ontology is specified, more specific ontologies can be created to model the data sources they represent by using terms defined in the main ontology. These new concepts are specified in more detail using inheritance, so they will have all the basic properties defined in the higher-level ontology plus the additional rules that are important to their definition, besides they can be reached through queries expressed in terms of the main ontology. It means that the subclasses inherit all properties of the parent class and add some more of their own to refine general descriptions of the level from which they inherit, as they represent information according to the concepts defined in the top-level ontology, it supports in a transparent way the retrieval of information to queries expressed in terms of the main ontology.

The integration of the data is logically and physically shifted toward the providers of the data and their source-specific implementations. Shifting the integration to the provider will improve data quality because the providers have specific knowledge of which elements of their data model correspond to which elements of the top-level ontology. Another benefit is that each data provider has to map only once instead of writing code which translates between the terminologies of pairs of systems, thus decreasing the overall complexity of the system.

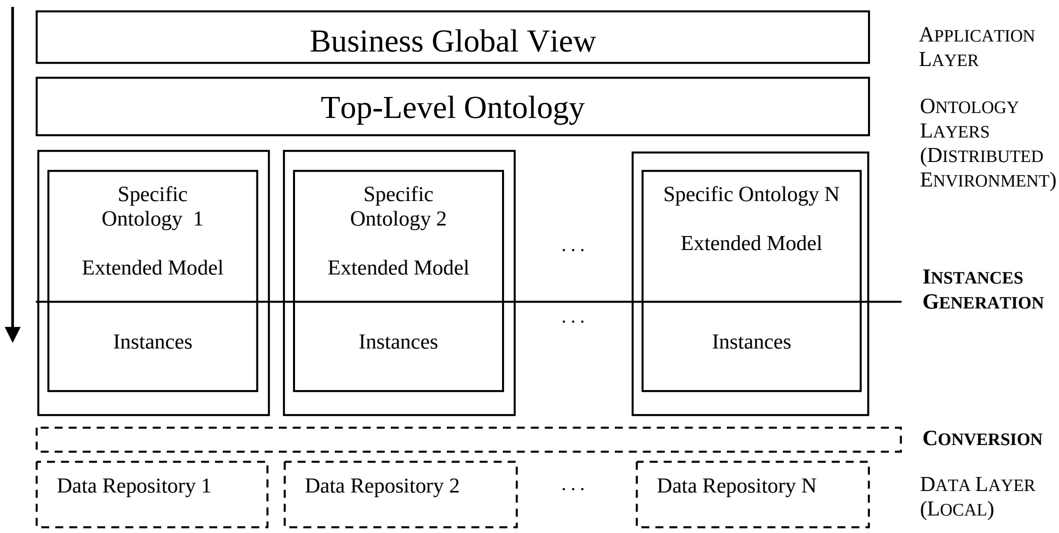

Figure 1: A view of the architecture in layers

Figure 1 shows the layered intention of this approach. The application layer gives a virtual integrated view of the underlying layers. It is virtual in the sense that the consistency is guaranteed by the time the user requests the modeled data.

The main purpose of the architecture here presented is to give the framework a unified interface to the semantic queries over heterogeneous data sets (structured and semi-structured), using the ontology approach.

This architecture is conceptually, in the database study perspective, a mediated data integration system, in virtual view mode, using an ontology as the common schema to overlap the heterogeneity of the data sets. Such common schema is described in OWL language and is supporting e-Business domain. It is a solution that offers queries with semantics, dealing with Semantic Web requisites.

In order to show these ideas in practice, it will be presented an example based on a case study in the next section. However, this structure is part of an effort to improve interoperability of distributed information systems and the approach itself is entirely general and can be applied to any domain facing similar design issues.

5. The Application

The pilot project “OntoShop” was built to validate the architecture and assure the concept, its usefulness and its extensibility. In this way, it was created ontologies to represent a supposed shopping center where the core ontology contained the basic building blocks and each store had to map their information according to it.

The top-level ontology was created following guidelines stated in [12], where it is proposed a generic e-Business Model Ontology that highlights the relevant e-business issues and elements that firms have to think of, in order to operate successfully in the Internet era. It is founded on an extensive literature and composed of four main pillars:

- Product innovation: What business the company is in, the product innovation and the value proposition offered on the market.
- Customer relationship: Who the company's target customers are, how it delivers them the products, and how it builds a strong relationship with them.
- Infrastructure management: How the company efficiently performs infrastructure or logistics issues, with whom, and as which kind of virtual enterprise.
- Financials: What is the revenue model (transaction, subscription/membership, advertising, commission, licensing) and the cost model (cost of goods sold, operating expenses for R&D, sales and marketing, general and administrative).

In order to extend the main ontology specializing and instantiating it, there were created some specific ontologies to represent the information of the data providers and maintaining the data distributed.

Technologies used to implementation of a prototype system are discussed in more details in the next sections.

5.1. OWL

In order to define data ontology, OntoShop employs the Web Ontology Language (OWL) for data modeling and storage.

OWL was chosen as the ontology language due to the facts that it was projected for Web and distributed applications, one of the key points of the solution, and secondly because OWL has more facilities for expressing meaning and semantics than XML, RDF, and RDF-S. Thus, OWL goes beyond these languages in its ability to represent machine interpretable content on the Web.

OWL is a revision of the DAML+OIL web ontology language. OWL adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes [13].

5.2. JODBC

Ontologies are a crucial role in giving intelligence to the Internet. It would permit pages to be annotated using logic formalisms that can be instantiated.

The knowledge-based systems (KBS) inference systems, based in logic programming, are composed basically by:

- A formalism, based in mathematic logic;
- A method or an strategy for solving the formalism (the inference engine);
- Strategies for controlling and scalability of inference or the conflict resolution methods, which might be guided and optimized.

In order to interact to user interface and answer user queries, it was necessary to give the prototype the ability to inference over ontologies with a Java inference engine extension.

An ontology management system which is being developed at IBM T. J. Watson Research Center is used to store ontologies and to make inferences. This system provides a Java API referred to as Java Ontology Base Connector (JOBC), which is the ontological equivalent of the Java Data Base Connector (JDBC). JOBC provides a simple-to-use but powerful mechanism for application programmers to use ontological information for creating applications without dealing with the details of ontologies [9].

5.3. Prototype system

The prototype developed consists of a number of interacting components: an ontology management system which provides a mechanism for dealing with ontological information at an appropriate level of abstraction, specific ontologies which contain information of data providers, and a user interface that allows the browsing of the top-level ontology to retrieve information expressed in the specific ones in a transparent way. To achieve this, the system stores the top-level ontology and needs to know the namespaces of the specific ontologies that are distributed over the Internet. Thus, the inference engine allows that queries expressed in terms of the core ontology can be applied to the specific ones since they are modeled with subclass-of relationships.

When a query is handed over to the system, the application program translated it to a query in a top-level ontology expressing it conform JOBC API. Thus, the JOBC driver works providing the infrastructure needed to support ontology management and the connection with the information sources; it consists of Java classes that will provide an implementation of the JOBC API, and contains a number of components such as an inference engine and a database. After getting a list of relevant resources, the JOBC driver will access the ontologies from each data provider utilizing the inference engine to provide a mechanism for interpreting the semantics represented as a set of rules. This enables the user queries to be answered, because when the requested fact is not immediately available, it can be inferred from available facts. Then, the result set is showed in a web browser at the client side. Figure 2 illustrates this process.

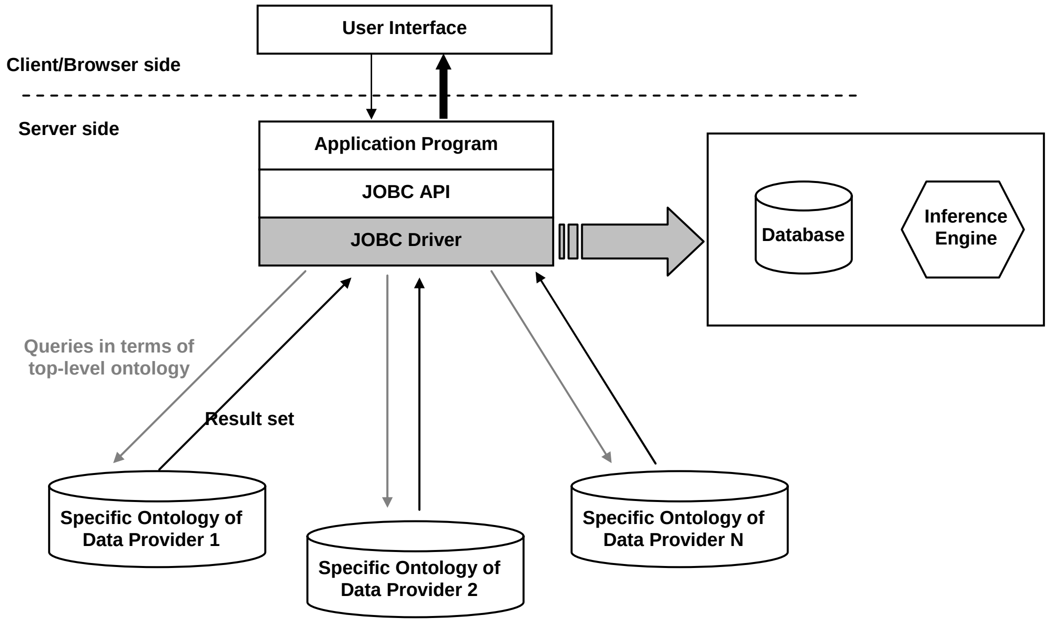

Figure 2: The flow of a query submitted to the system

6. Conclusions and Further Work

As seen, the demand for semantics in information integration among enterprises has been a major issue to the realization of e-Business. Semantic Web technologies and ontology techniques come to fill the holes that make e-Business applications tied and limited to specific standards.

In the same time a mediation ontology decreases costs, by diminishing complexity in enterprises' interoperation, it tends to increase companies profitability, as the space of trade is open worldwide to autonomous agents that can capture and process information.

The prototype shown here tries to give a solution to some of these issues, focusing in the data integration problem. For the development of the project, some tools and technologies had to be used, as OWL language, which promises to be a standard for information definition and distribution.

Some workarounds had to be made to overcome limitations in the inference engine. As a future work some other inference engines should be tested, or even an implementation extension of JODBC could be done to fit this application.

References

1. Benetti, I. et al.: SI-Designer: an Information Integration Framework for e-Commerce.: In Proceedings of Workshop on E-Business and the Intelligent Web, pp. 97–104, 2001.
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web.: Scientific American, 284(5):34–43, 2001.
3. Bussler, Christoph.: Business-to-Business Integration needs a Meta Business Enterprise Ontology.: Oracle Corporation, Redwood Shores, USA.
4. Ding,Y., Fensel, D., Klein, M., Omelayenko, B., and Schulten, R.: The role of ontologies in eCommerce. S.Stab & R. Studer (eds.) Handbook on Ontologies, Springer, 2003.
5. Freitas, F.L.G.: Ontologias e a Web Semântica. Universidade Católica de Santos.: Brazilian Computer Society Artificial Intelligence National Meeting, São Paulo, Brazil, 2003.
6. Garcia-Molina, H. et al.: Integrating and Accessing Heterogeneous Information Sources in TSIMMIS.: In Proceedings of the AAAI Symposium on Information Gathering, Stanford, California, 1995.
7. Kim, H. M.: Predicting How Ontologies for the Semantic Web Will Evolve.: Schulich School of Business, York University, Toronto, Canada.
8. Kirk, T., Levy, A. Y., Sagiv, Y. and Srivastava, D. The Information Manifold. C. Knoblock and A. Levy, Stanford, California, 1995.
9. Lee, J., Goodwin, R., Ye, Y., Akkiraju, R.: Towards Enterprise-Scale Ontology Management. IBM T. J. Watson Research Center, NY.
10. Mena, E., Illarramendi, A., Kashyap, V., and Sheth, A. OBSERVER: an approach for query processing in global information systems based on interoperation across pre-existing ontologies. Distributed and Parallel Databases, 8(2):223–71, 2000.
11. Omelayenko, B. and Fensel, D.: A Two-Layered Integration Approach for Product Information in B2B E-commerce.: In Proceedings of the 2nd International Conference on Electronic Commerce and Web Technologies, Munich, Germany, September 2001.
12. Osterwalter, A., Pigneur, I.: An e-Business Model Ontology for Modeling e-Business. 15th Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy. Bled, Slovenia, 2002.
13. OWL Web Ontology Language Overview.: Deborah L. McGuinness and Frank van Harmelen, Editors. W3C Recommendation, February 2004, http://www.w3.org/TR/2004/REC-owl-features-20040210/.
14. Paolucci, M., Sycara, K., Nishimura, T., Srinivasan, N.: Toward a Semantic e-Commerce. Recent Trends and Developments.: In Proceedings of 6th International Conference on Business Information Systems, Colorado Springs, Colorado, USA, 2003.