

Integrating Information, Applications and Services on the Web

Juan C. Lavariega, Lorena G. Gómez-Martínez

ITESM-CII, Campus Monterrey
Instituto Tecnológico y de Estudios Superiores de Monterrey –Centro de Investigación en Informática
Av. E Garza-Sada 2501
Monterrey, NL 64849, Mexico
lavariega@itesm.mx, lgomez@itesm.mx

Abstract

This is a position paper that describes a long-term ongoing research project that is being developed at ITESM. The project is called **Bymx**. The aim of **Bymx** is to integrate information, applications and services over the Internet. Among the **Bymx** characteristics is the use of technology that has not been fully assimilated by the web (such as distributed database concepts, component based software, and software product lines) and to incorporate technology in increasing demand such as mobile computing. The paper sketches our approach for achieving interoperability among heterogeneous data sources and software components and illustrates the issues that we are facing.

1. Introduction

In their origins information technologies were exclusive for a few specialized users. Nowadays, however, information technologies (IT) are part of our day-to-day activities, making practically impossible to think in activities where IT or products based on IT do not play a decisive role in our lives. Consider for example your life without a cellular telephone and the functionality that it offers, in addition to placing and receiving calls. It is also notorious the amount of information available through electronic media, being the Web the electronic publication media with the most growth in the last years. Together with this explosion of information, there is the need in many organizations to share information and to establish communication between computer applications within the same organization or with external entities. And of course there is the huge amount of public information accessible through the different search engines.

Most of the interoperability solutions in the organizations are ad-hoc strategies, where technologies involved in the solution include CORBA, COM and EDI, for example. However, the use of such technologies is tightly coupled with the domain of the application that we would like to communicate with each other. Another aspect that has been part of our motivation is the lack of use of Software Development Methodologies, when dealing with problems of interoperability among web based applications; although software development practices, is what organizations such as the IEEE Computer Society and the ACM have recommended among their members. New Internet

technologies such as XML, SOAP, WSDL, UDDI (Box 2001) and the concept of web services (Stal 2002) have a great influence in the way we will develop our future applications and on the way those applications will integrate information including services from other sources through the Web. What software development methods are appropriate for these new applications? Do we need special strategies? Can we reuse part or whole of the knowledge we have developed for other computer related areas such as distributed databases, component-based databases, or software product lines?. These are some of the questions that we are exploring in our research group. This paper addresses the issues we are facing and the contributions we are expecting to produce in the near future.

The rest of the paper is organized as follows. Section 2 presents the general research issues we are dealing with. Section 3 discusses the operational environment and architecture we are defining for applying our strategies and methods for integration of information, applications and service. And Section 4 concludes the paper.

2. Relevant Research Issues

Integration of information, applications and services implies also integration of several knowledge areas from databases for information handling, software product lines for the definition of applications grouped by knowledge domain, and development methodologies that include the technical aspects needed for supporting the integration of applications in the Web.

First we have to consider the source of the information available in the Web. We easily identify two types of information sources. One is information produced by other applications (i.e., reports, and forms which back end is a database, or results from an application with a Web interface). The other type of information is the one produced by humans (i.e., personal web pages based on html or plain text) that is un-structured (Abiteboul, Buneman, and Suciu 2000) by nature.

One of the major problems that we face while dealing with the Web is that most of the information freely available is of the unstructured or semistructured type making it difficult to querying it in a sophisticated way by using a language such as SQL. Also the issue of sharing information

between applications becomes a difficult task when dealing with unstructured information. XML has come as promising tool to provide the structure needed for querying and sharing web content. There is a great amount of excellent work related to the applicability of XML to the integration and interoperability problem, but we still need to work on issues such as identifying families of software products that can help us to get a better definition of ontologies. Other approach that we are exploring is to use XML for describing metadata models for including structured and unstructured sources in the integration. Other aspects related to the use of XML include the implications of having native XML databases, supporting query optimization for XML documents, and supporting transaction processing in XML databases.

We have mentioned before the need of development methodologies for Web environments mainly because the majority of methodologies are based on modular software architectures where every one of the components is part of a whole. Web has new challenges in software architectures because is a changing environment, it is also distributed, heterogeneous and loosely coupled. So methodologies for analysis, requirements specification, and design of web applications should be taken into account the Web characteristics (Arsanjani, 2002).

A completely new area is the development of software components for web services environments. Web service technology is going to play a major role in the integration of information, and applications in the web in the following years. So the immediate problem is to identify what are the components that can be adapted to the functional requirements for web services applications. Also we need to validate if the available technology (XML, SOAP, .Net, JINI) is going to support the in addition to the functionality of Web services, the functionality of components for web applications.

Even though our initial working environment includes the Web, we think that a natural extension to the strategies and tools we can develop is to include mobile computing, where the basic information providers to mobile devices are web services integrated with data repositories.

3. Bymx as Architecture Overview

In order to start developing and testing our proposal for solutions to the problems stated above, we proposed a large-scale mid term project with a limited budget. Our main strategy has been to involve graduate students from our university and focus all of their theses topics into our project that we have called **Bymx**. **Bymx** is formed by the first letter of the words Person, Net, Friend and Intelligence, in the Zapotecan language. Therefore Bymx can be thought as a net where the sharing of knowledge is friendlier. Our goal with Bymx is to produce an architecture that enables the integration of information, applications and services. So that every time a new application, service or piece of information decides to be part of the Web, the integration

and access to that component can easily be accomplished. However such an ambitious goal needs to produce some results in the short run to keep the interest of students and also show the functionality of the approach we are proposing. So we define a well-defined application that we plan to scale up as the project progress and starts producing results.

Curricula Contents Project

As a short-term project within the context of Bymx we have defined an application consisting in the development of a software tool for the integration of academic curricula contents, specifically for Computer Science and Computer Engineering programs in Mexico. The ANIEI is the National Association of Higher Education Institutes in Informatics; this organization has the task of certify the national academic programs in informatics and computing. Additionally there are other national institutions that certify the level of knowledge that a student once graduated from one of the Institutions recognized by ANIEI should possess. Knowledge certification is acquired by applying to the general graduation exam for informatics and computing (EGEL). However, depending of the own needs determined by the geographical area where a university is located or, the body of knowledge from a particular program, or the professional characteristics that a graduate must have, each institution can decide to emphasize some of the curricula contents defined in the recommendation given by the ANIEI and evaluated in the EGEL.

The correct satisfaction of the requirements imposed by the standardization efforts mentioned above require for each institution to have access to the appropriate bibliographical material, however specialized human resources (professors) and economical resources are difficult to obtain for the majority of Mexican Institutions. The work of Levine (Levine 2001) presents a general coverage of the knowledge required by the ANIEI and the EGEL, but there is still the need for support academic institutions with detailed didactic material for the covering of each one of the curricula contents involved in the academic programs. Therefore a reduce number of professor do have to prepare all the material for all the computer related courses offered at their institutions.

In this subproject (called BECCA) we are proposing the creation of the software tools for managing the curricula contents for the bachelor's degree programs in computing and informatics that follow the recommendation given by the ANIEI. The main objective is to offer the means by which the academic institutions (offering bachelors' degree in computing and informatics) may locate appropriate curricula contents that is compliant with the standardization efforts of the ANIEI. Location, creation and manipulation of curricula contents must be possible from any geographical site through the Web and eventually through mobile devices.

Curricula content can be any type of document that supports the leaning interaction between students and professors such as lecture notes, problems, exercises, cases of study, that cover a topic or one section for a particular area of computer science or informatics (i.e., data modeling in the entity relationship model for database applications)

So this type of material is unstructured by nature, may or may not have all the elements of a textbook or guide but should be organized in some way that the access and querying of the material is possible.

Also the material available in such environment should be validated for correctness and most likely an expert in the area produces it. Other professors with minor experience and not enough time to prepare the curricular material may use the material that an expert agreed to export in BECCA.

Initially, we are planning to have a central repository of documents but the documents not necessarily require to be located at a central site, as in the web information can be disperse in remote homepages. Also as in distributed databases, information for curricula contents may be fragmented at different sites (i.e. we can use the problems defined by Professor A in topic B, but the lecture notes of Professor C). We also face the problem of producing an integrated document that serves as guide for students and that such a document has editorial standards.

Due that the project will be conducted as part of theses for graduate students it is very important to apply rigorous controls during the development of BECCA, as a form to guarantee continuity and uniformity. Usually, what we would do is to follow one of the several development methodologies for construction of software products (Sommerville, 2002). However the unstructured nature of the information, the heterogeneous origins of the sources, and the distributed complexity of the web push the creation of novel methodologies and tools that can be applied to the development of the applications as BECCA. In fact one of the projects is to evaluate the current development methodologies and their applicability to BECCA.

Also because BECCA is an application in a particular domain we want to use it as a test bed for identifying components that can be scaled up to a more general architecture. During the development of BECCA we will follow the concepts of component-based software products (Dittrich and Geppert 2001) (Crnkovic et al. 2002).

We believe that development for Web applications with characteristics similar to BECCA imposes the creation of new tools, methods and strategies, but we want to validate until what extend current development technology can be reused for the creation of BECCA and Bymx.

4. Conclusion

In this short position paper, we have presented an overview of our efforts for creation of tools that support the major problem of integrating information, applications and services in the Web. We described an scaled down

application that can produce results in the short term by integrating curricula contents for academic institutions offering bachelor' degree programs in computing and informatics.

It is our intention to initiate collaboration efforts with other institutions that are facing similar problems.

References

- Abiteboul, S., Buneman, P., and Suciu, D., eds, 2000. *Data on the Web*, Morgan Kaufmann.
- Arsanjani, A., 2002. Developing and Integrating Enterprise Components and Services. *Communications of The ACM* 45(10): 30-34.
- Box, D., 2001. XML, XSD and SOAP as a Better Component Model. *Tutorial #15 presented at the Object-Oriented Programming, Systems, Languages and Applications, (OOPSLA 2001) Conference*, Tampa Bay, Florida
- Crnkovic, I., Hnich, B., Jonsson, T., Kiziltan Z. 2002. Specification, Implementation and Deployment of Components. *Communications of The ACM*. 45(10):35-40.
- Dittrich, K., and Geppert, A.. 2001 *Component Database Systems*, Morgan-Kaufmann
- Levine G. 2001. *Computación y Programación Moderna*. Addison Wesley
- Schilit, B.N., Trevor, J, Hilbert, D.M., Koh, T.K. 2002. Web Interaction using Very Small Internet Devices", *IEEE Computer*, 35(10):37-45
- Sommerville, I., *Ingeniería de Software 6a Edición*, Addison Wesley, 2002.
- Stal, M., 2002 Web Services: Beyond Component-based Computing, *Communications of The ACM*, 45(10):71-76.