Structuring information on the Web from below: The case of Educational Organizations in Chile

Ernesto Krsulovic-Morales and Claudio Gutiérrez
Computer Science Department
Universidad de Chile
Blanco Encalada 2120
Santiago, Chile
{ekrsulov,cgutierr}@dcc.uchile.cl

Abstract

We present an ongoing work to help populate the Web with metadata by structuring and integrating information of organizations at a small scale. This is a natural complement to big scale projects to build the Semantic Web infrastructure We show an implementation for Computer Science departments in Chile, and present current work on educational organizations generalizing previous experience.

1. Introduction

One of the most attractive aspects of the Semantic Web (SW) is the promise to be able to make inferences over the information on the Web resulting from the integration of applications [2]. This promise will not be realizable until the basic architecture proposed by the W3C is implemented and we have a critical mass available to allow its utilization [3]. While waiting for these days to come, we can take advantage of the most basic and stable part of this architecture, namely the metadata layer (standardized by the RDF model), to create applications which require more structured information, but with distributed providers. Projects of this kind, besides the functionality provided by themselves, help greatly to build the critical mass of metadata required to leverage the next levels of the SW.

In the educational area, metadata initiatives focus mainly on the integration of educational contents. There are good examples at a national level in the area of e-learning, like TheGetaway (USA), Plash (Canada), Edutella and Universal (Europe), which aim at the use of standard metadata for learning objects [10] to be retrieved in portals or over P2P networks. Unfortunately, there are no such initiatives for educational organizations. In Chile today there are 11.066 schools and 236 universities and colleges distributed in 13 regions. Most of them have a Web site.

There are several tools that "structure" information, stores it, and allows users to query that information. We can identify three paradigmatic models: Classical databases, directories, search engines. There are other which have particular features of one of the above: yearbooks (constraint of time added), intranet search engines (constraint on the set of URIs visited), etc. Table 1 is a comparison among different solutions.

On the Web, at the educational level in Chile, a first integration step is syntactic. For example, use a standard local search engine, e.g. *TodoCl.cl*, with all the limitations involved. Another approach currently used is to build communities around portals (global gateways), an approach which has several drawbacks compared to the use of metadata [7]. Two examples are *EducarChile.cl*, focused on school level, and *Universia.cl*, focused on the university level. Almost all of them are built using LDAP or relational databases. Our work presents an alternative distributed approach using metadata.

2. Proposal and Ongoing Work

Our work focuses on organizations with small number of components: We are targeting educational organizations in Chile (Schools, Universities, Institutes, etc.) We chose the educational area because, besides the intrinsic value for users, it has many advantages for our research: a known environment, no direct commercial interest involved, people are open to adopt and test new technologies, and permanence over time.

We assume two important weaknesses of today's SW: (1) the fact that several of the language specifications involved in the SW are not yet stabilized, and (2) the lack of accessible tools to facilitate the manipulation of metadata, especially regarding markup of resources by non-expert users.

Features	Database	Directory	Search Engine	Ont-driven
data model	schema fixed	schema distr. edit.	none	ontology dist. edit.
data location	centralized	centr/distr	centralized	distributed
access level	views	views	views	data source
data load	man./ aut. schema ass.	manual schema ass.	crawler	manual ont. ass.
suscrib.	requested	submitt.	web opt-out	web opt-in
search	query SQL-like	browsing patt. match.	patt. match.	RDQL patt. mat
model ext.	suffic.	bad	good	very good
trust level	very good	good	bad	bad

Table 1: Comparison of different forms of structuring information on the Web

Seems that we are still far from a common platform between applications for the SW, like "*libwww*" and the browser "*mosaic*" in the beginnings of the Web. There are some interesting attempts, for example Jena [8], a toolkit to build applications with RDF in Java, and Haystack [5] to help markup and retrieval of resources.

The system follows a standard architecture for several SW applications: An *ontology*, a *site markup tool*, a *collect of metadata system*, and a *query system*. The detailed structure is discussed in [6]. The whole system must be lightweight and portable to allow its deployment even in components which could not have the infrastructure necessary to host big systems. To cope with the goals proposed, tools should be web-based and oriented to users with no experience in knowledge representation techniques.

2.1. Depmark Project

As test bed we are targeting CS departments at Chilean universities (See project at `http://purl.org/net/depmark`). This area has several advantages for testing such a tool: besides the intrinsic value for users, a known environment, no direct commercial interest involved, people open to adopt and test new technologies, and permanence over time. Moreover, all of them have Web sites, constantly updated, and with rich and diverse information in content and format. The central problem we faced implementing the above architecture was the non-existence of a markup tool. Without such a tool it would be very difficult for people that generate contents (professors, researchers, students, directors, employees, etc.) to generate the metadata about current information on Web pages.

The tool was implemented in the language JAVA. As API to manage RDF models we used JENA. The storage is implemented on the RDB interface model provided by JENA using MCKOI as relational database (a database implemented in JAVA that can be embedded in the application). The schema is RDFS serialized in XML. The markup interface is implemented in JSP.

2.2. The next step: Educational organizations

The plan is to select one level from the educational network in Chile, and to develop a framework to achieve semantic integration of resources already found in Web pages, as well as to implement an extended test-bed in the selected level. There are three key steps in the development of this plan:

- To build a general framework by creating an extension of *profile* [4] of one of the existing standards. To develop a tool that facilitates the composition of ontologies from existing standards and allowing the creation of profiles and extensions of ontologies. We named this tool RdfWiki and is introduced briefly in the next subsection.
- To create *ad-hoc* ontologies for areas of particular interest at regional level.
- To design and implement the infrastructure to facilitate creation, storing and query of metadata, and developing of a prototype.

2.3. RdfWiki Tool

DepMark worked with a single ontology: this fact presents an ideal setting for implementation, but lacks desirable properties such as composition and reusability of ontologies. We identified the necessity to develop a content management system driven by multiple ontologies. Projects with a similar aim are OntoWeb Portal [9] and ODESeW [1], but they are big-scale projects not aimed at small organizations.

We chose OWL as the natural language for driving this system, mainly because of the rich features to specify restrictions compared to RDF Schema. The system is in the design phase. We called it *RdfWiki* to remind designed features it has, like public writable pages and rollback history. This application should not only permit the creation of metadata; Additionally, it may work like a collaborative space for creation of ontologies, by facilitating their composition from existing standards and allowing the creation of profiles and extensions of ontologies.

3. Conclusions

Our experience so far (with Depmark and in the design of RdfWiki) has highlighted some issues we consider relevant.

1. Markup must be made by people close to the data. It is not possible to bridge the semantic gap using only automatized extraction of metadata or centralized markup.
2. Proliferation of online markup tools would make it possible to mark by hand in a distributed manner.
3. Projects like this show that small scale markup efforts is a natural complement to big scale projects of massive automatic extraction of data from existent sources.
4. The limitations of single-based ontology projects like Depmark makes it desirable to work with several ontologies. This direction is also closer to the spirit of the SW vision.

Acknowledgments Supported by Millennium Nucleus "Center for Web Research" (P01-029-F), Mideplan, Chile.

References

[1] O. Corcho, A. Gómez-Pérez, A. López-Cima, V. López-García, and M. Suárez-Figueroa. ODESeW. Automatic Generation of Knowledge Portals for Intranets and Extranets. http://webode.dia.fi.upm.es/sew/, 2003.

[2] J. Handler, T. Berners-Lee, and E. Miller. Integrating Applications on the Semantic Web. *Journal of the Institute of Electrical Engineers of Japan*, 122(10):676–680, October 2002.

[3] S. Haustein and J. Pleumann. Easing Participation in the Semantic Web. In *Workshop at WWW2002, International Workshop on the Semantic Web*, May 2002.

[4] G. Hodge. Metadata Made Simpler. http://www.niso.org/news/ Metadata_simpler.pdf, Annapolis Junction, MD, 2001.

[5] D. Huynh, D. Karger, and D. Quan. Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF. In *The Eleventh International World Wide Web Conference, WWW2002*, May 2002.

[6] E. Krsulovic and C. Gutiérrez. Building Yearbooks with RDF. In A. Abraham et al., editor, *Soft Computing Systems: Design, Management and Applications*, pages 593–601. IOS Press, Dec 2002.

[7] S. Lissonnet. A proposal for using metadata to support the building of an educational community . In *The Eleventh International World Wide Web Conference, WWW2002 Alternate Tracks*, 2002. May.

[8] B. McBride. Jena: Implementing the RDF Model and Syntax Specification. In S. Decker, D. Fensel, A. Sheth, and S. Staab, editors, *Proceedings of the Second International Workshop on the Semantic Web - SemWeb'2001*, Hongkong, China, May 2001.

[9] Ontoweb. Ontoweb portal. http://www.ontoweb.org/.

[10] D. A. Wiley. *The Instructional Use of Learning Objects*, chapter Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. Agency for Instructional Technology and the Association for Educational Communications and Technology, 2000.