A hierarchical one-to-one mapping solution for semantic interoperability

David Portolés-Rodríguez
Universidad de Zaragoza
Tel: +34 976 762134
Fax:+34 976 761914
dporto@unizar.es

Rafael Tolosana-Calasanz
Universidad de Zaragoza
Tel: +34 976 762134
Fax: +34 976 761914
rafaelt@unizar.es

Javier Nogueras-Iso
Universidad de Zaragoza
Tel: +34 976 762358
Fax: +34 976 761914
jnog@unizar.es

Rodolfo Rioja
Universidad de Zaragoza
Tel: +34 9767621234
Fax: +34 976761914
rodolfo@unizar.es

Pedro R. Muro-Medrano
Universidad de Zaragoza
Tel: +34 976 761950
Fax:+34 976 761914
prmuro@unizar.es

Fco. Javier Zarazaga-Soria
Universidad de Zaragoza
Tel: +34 976 762358
Fax: +34 976 761914
javy@unizar.es

Abstract:

The importance of interoperability among computer systems has been progressively increasing over the last years. The tendency of current cataloguing systems is to interchange metadata in XML according to the specific standard required by each user on demand.

According to the research literature, it seems that there exist two main approaches in order to tackle this problem: solutions that are based on the use of ontologies and solutions that are based on the creation of specific crosswalks for one-to-one mapping. This paper proposes a hierarchical one-to-one mapping solution for improving semantic interoperability.

Keywords:

Metadata standards, metadata servers, catalogue servers, SDI, interoperability, semantic interoperability, crosswalks, XML, Dublin Core, RDF, ISO 19115.

1. Introduction

The importance of interoperability among systems, the ability of two or more systems or components to exchange information and to use the information that has been exchanged (5), has been progressively increasing over the last years.

One of its specific aspects, which is concerned with the ability to access, consistently and coherently, to similar (though autonomously defined and managed) classes of digital data, objects and services distributed across heterogeneous repositories (9), is known as semantic interoperability.

With little doubt, the most obvious way to broaden the opportunities for interoperability is by making the information stored, which is known as metadata (or simply descriptors), exchangeable. This may be carried out by promoting a commonly understood set of descriptors that helps to unify other data content standards.

The tendency of current cataloguing systems is to interchange metadata in XML according to the specific standard required by each user on demand. Furthermore, metadata schemas from different domains are not usually semantically distinct but overlap and relate to each other in complex ways. As a consequence, the semantic interoperability has to deal with the equivalences between those descriptions.

According to the research, it seems that there exist two main approaches in order to tackle this problem: solutions that are based on the use of ontologies and solutions that are based on the creation of specific crosswalks for one-to-one mapping. This work proposes a hierarchical one-to-one mapping solution for improving the semantic interoperability.

The remainder of this paper is structured as follows: in section 2, related work from the semantic interoperability domain is reviewed. In section 3, our proposal for improving the semantic interoperability is described. The paper closes with some conclusions.

2. Related work

Since the emergence of the Internet, a great deal of effort has been invested in the development of metadata vocabularies to enable the exchange and discovery of information across different applications and domains. Metadata vocabularies such as Dublin Core (3), MARC (10), FGDC (2), provide standardized sets of descriptive elements to enable the exchange of resources for specific applications or domains. Although these standards enable interoperability within domains, they introduce the problem of incompatibility between disparate and heterogeneous metadata schemas or schemas across domains.

On the other hand, there are three main scenarios in which interoperability among metadata schemas is required, according to (6):

- To enable a single search interface across heterogeneous metadata schemas
- To enable the integration or merging of descriptions which are based on complementary but possibly overlapping metadata schemas or standards
- To enable different views of the one underlying and complete metadata schema, depending on the user's interest, perspective or requirements.

In the next subsections, the two approaches, commented on above, will be reviewed in detail.

2.1. The ontology-based approach for semantic interoperability

In the information systems and knowledge representation field, the ontology concept denotes a knowledge model that represents a particular domain of interest. This kind of solutions are based on these models since they may help to define a common ground between different information communities.

In this sense, the work developed in the OBSERVER system (4) provides an architecture for query processing in global information systems that supports interoperation using ontologies.

Another interesting research work (6) implements the ontology by means of a thesaurus (MetaNet) applied to the ABC model. Its main objective is to provide the semantic knowledge required in order to enable machine understanding of equivalence and hierarchical relationships between metadata terms from different domains. Furthermore, its thesaurus has been implemented by using technology, RDF (Resource Description Framework, (11)) and RDFS (RDF Schema (11)), borrowed from the semantic web field which is in fact a closely related conception. However, other proposals (8) remark the limitations of RDFS, since it does not provide mechanisms for specifying general axioms (rules that allow additional reasoning) which appear at most artificial intelligence ontologies. The result of this research is the SHOE language (7) which attempts to solve this lack of functionality.

2.2. The Crosswalk-based approach for semantic interoperability

This set of solutions use software components which map the relationships and equivalences between two or more metadata schemas. This software components are called crosswalks. Interesting collections of links to metadata crosswalk initiatives can be found through the Web sites of the UK Office of Library and Information Networking and the Metadata Architecture and Application Team of the National Digital Archives Program in Taiwan. There, it is possible to find several mappings (specially those used for library metadata): from MARC standards to Dublin Core; from Dublin

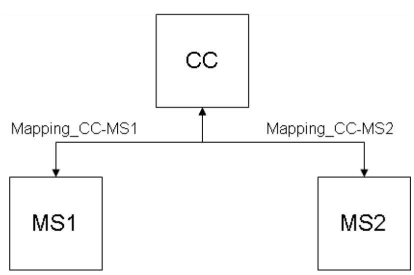

Figure 1: correspondences between CC and the metadata schemas MS1 and MS2.

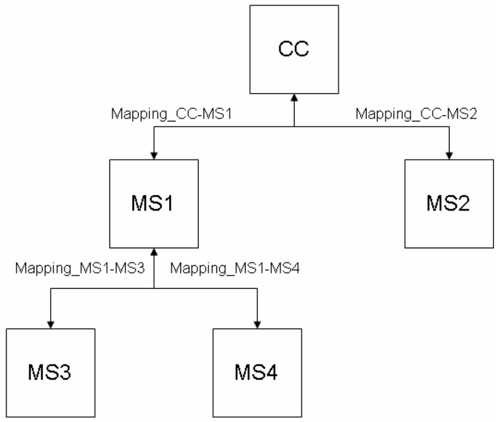

Figure 2: a simple metadata profile hierarchy

Core to EAD (Encoded Archival Description) (12); from Dublin Core to GILS (a Z39.50 metadata profile for the US Government Information Locator Service); or from Dublin Core to GCMD DIF (Directory Interchange Format (2)).

Other works, such as the CORC (Cooperative Online Research Catalogue) project (1), have also proposed the conversion of CSDGM towards more generic standards like MARC or Dublin Core.

On the other hand, there exists an interesting proposal (8) in which the attention is focused in the crosswalk process creation.

3. Our hierarchical one-to-one mapping solution for semantic interoperability

The three required scenarios, mentioned in the previous section, guided our proposal for the interoperability improvement. Its main idea consists of defining or choosing a general purpose metadata schema as a common core.

Next, one-to-one semantic correspondences between the system metadata schemas and such a general description have to be established. The system, at this point, can build as much functionality as possible through this common core view. Therefore, new metadata schemas introduced to the system do not necessarily have to involve code rewriting, providing that appropriate correspondences can be established between those new schemas and the common core. On the other hand, new metadata schemas do not have to be directly mapped to the common core, but to other schemas which were introduced previously. The related set of metadata schemas can be seen as a metadata profile hierarchy.

3.1. A simple example

Let us suppose a system with two metadata schemas, MS1 and MS2, in which MS1 presents a very rich description with hundreds of elements and MS2 presents a more general description with a considerably smaller number of elements. Our solution would consist of:

- defining the common core (CC).
- establishing semantic correspondences between the existing metadata schemas and this new general one (Mapping_CC-MS1, Mapping_CC-MS2) (see Figure 1).

CC has to be concrete enough in semantics in order to provide the system with a useful data view. Besides, for this same reason, it should be desirable that the schemas of MS1 and MS2 are semantically richer than the information of CC. Were it not the case, the services of the system through the common core view would probably not provide the quality results desired.

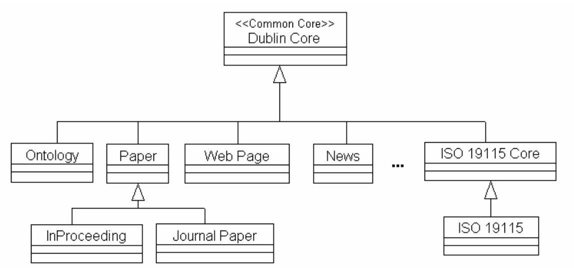

Figure 3: a metadata profile hierarchy based on Dublin Core

Let us suppose now that two new metadata schemas MS3 and MS4 have to be integrated in the system and that they are based on the metadata schema MS1, since they extend its elements. As a consequence, MS3 and MS4 might be mapped directly to MS1 and thus, the metadata schema hierarchy represented in Figure 2 is obtained.
Another remarkable aspect is the way in which the semantic correspondences can be implemented. They will clearly depend on the metadata schemas involved. There are, however, two main possibilities:

- As reviewed in the previous section, there are a lot of experienced techniques in mapping some standards to others by using crosswalks. This may be helpful when dealing with heterogeneous metadata schemas which may represent the most frequent situation.
- In case that some bottom-hierarchy metadata schemas extend a top-hierarchy schema, there will be no need for crosswalks.

By establishing those semantic levels, a metadata profile hierarchy was obtained. In this hierarchy, top schemas are semantically more general than bottom ones.
The benefits of such an approach are considerable, regarding the requirements which guided the design of our proposal:

- The single interface across heterogeneous metadata schema could be obtained by designing the search interface through the common core schema.
- The integration of descriptions that are based on complementary but possible overlapping metadata standards are achieved by introducing the schemas into the hierarchy.
- The different views of the information may be achieved by designing specific crosswalks across standards. Nevertheless, it can be somehow useful the fact that all the metadata schemas have the general description in common.

3.2. A metadata profile hierarchy based on Dublin Core

Dublin Core seeks to promote a commonly understood set of descriptors to help facilitate interoperability across disciplines (3). Some of its well-designed features are its simplicity and extensibility as well as its objective of facilitating discovery of electronic resources. These features make Dublin Core an ideal candidate to be incorporated in our metadata profile hierarchy as the common core.
Certain mechanisms provided by the Dublin Core Metadata Initiative (DCMI) such as application profiles (3) permit describing a wide range of heterogeneous resources by extending and adapting the semantics of Dublin Core.
However, at certain situations, there is no possibility to extend the semantics and the correspondence has to be carried out by using a crosswalk.
Figure 3 shows a possible metadata profile hierarchy based on Dublin Core. UML was used to represent it graphically and the inheritance relation should be interpreted in terms of semantic description. Dublin Core is the standard at the top of the hierarchy. Then, as it can be seen, several DC application profiles have been defined to describe web pages, papers,

news and even ontologies. All these schemas were created by means of the mechanisms provided by Dublin Core (application profiles). However, in order to establish the mapping between Dublin Core and ISO 19115 core a crosswalk had to be designed and implemented, since those standards are extremely different. Several guidelines from (13), (14) and (15) were followed in this process.

4. Conclusions

This paper described a technique for the improvement of semantic interoperability in systems with heterogeneous metadata schemas. This technique consists of erecting a semantically general metadata standard, called common core, and building the most general functionality of the system through this general view. Then, semantic correspondence is established between other metadata schemas, which have to be integrated and managed by the system, and the common core. Those schemas, at the same time, can be general descriptions for other new schemas and so a metadata profile hierarchy is generated.

Besides, a metadata profile hierarchy, in which the root schema is Dublin Core, was proposed.

Acknowledgements

The basic technology of this work has been partially supported by the Spanish Ministry of Science and Technology through the project TIC2003-09365-C02-01 from the National Plan for Scientific Research, Development and Technology Innovation.

References

1. A. Chandler et al. Mapping and converting essential federal geographic data committee (FGDC). Metadata into MARC21 and Dublin Core. Towards an alternative to the FGDC clearinghouse. D-Lib magazine 6, 1. January, 2000,
2. Content standard for digital geospatial metadata (GSDGM). http://www.fgdc.gov/metadata
3. Dublin Core Metadata Initiative (DCMI). http://dublincore.org
4. E. Mena. OBSERVER: An approach for query processing in global information systems based on interoperation accross pre-existing ontologies. Tesis doctoral, Universidad de Zaragoza, 1998.
5. Institute of Electrical and Electronic Engineers. IEEE standard computer dictionary: A compilation of IEEE standard computer glossaries. New York, 1990.
6. J. Hunter. MetaNet-A Metadata Term Thesaurus to enable semantic interoperability between metadata domains. Journal of digital information 1, 8. 2001
7. J. Helfin and J. Hendler. Semantic interoperability on the web. In Proceedings of extreme markup languages, 2000. Alexandria, VA, 2000. G. C. Association, Ed., pp. 111-120.
8. J. Nogueras-Iso et al.. Geographic information metadata for Spatial Data Infrastructures. Resources, interoperability and information retrieval. Springer-Verlag. Berlin Heidelberg 2005. ISBN 3-540-24464-6
9. J. Nogueras-Iso et al. Metadata Standard interoperability: application in the geographic information domain. Computers, Environment and Urban systems. November 2004. Volume 28, Issue 6, pp. 611-634.
10. MARC STANDARDS. http//www.loc.gov/marc/marc.html
11. RDF. Resource Description Framework. W3C. http://www.w3c.org/RDF/
12. US Library of Congress. Encoded Archival Description (EAD) DTD. http://lcweb.loc.gov/ead/, 1998
13. CWA 14856. Guidance material for mapping between Dublin Core and ISO in the Georgraphic Information domain. CEN Workshop Agreements.
14. CWA 14857. Mapping between dublin core and IS 19115, "Georgraphic Information - Metadata". CEN Workshop Agreements.
15. CWA 14858. Dublin Core Spatial Application Profile. CEN Workshop Agreements.