A methodology for semantic qualification of schemas

Vicente Palacios, Jorge Morato, Anabel Fraga, Juan Lloréns

Universidad Carlos III, Escuela Politécnica Superior
Av. Universidad 30 - 28911 Leganés, Spain
vicente.palacios@uc3m.es, {jorge,afraga,llorens}@ie.inf.uc3m.es

Abstract. The semantics of vocabulary elements use to be defined locally with scarce rigorousness and without consideration to shared definitions in other vocabularies, this approach produces a lack in the interoperability between resources. This document provides metadata vocabularies with a semantic map, articulated by a new qualifier set. The qualifier set is named semantic qualifiers, which permits to use semantics included in public reference resources. Moreover, a methodology is proposed. The solution improves previous initiatives concerning metadata management like Metadata Registries (ISO/IEC, 2006) or other DCMI proposals. In addition, it considers Modularity, Extensibility, Refining and Plurilinguistic criteria. A descriptive document is proposed with two views, one aimed to the user and an additional in RDF. The fields that would get benefits from the proposal are conceptual recovery of elements in the Semantic Web, the use of application profiles, and the friendlier use of vocabularies.

Keywords: Semantic Qualification, Schema Interoperability, Reference Ontology, DCAP (Dublin Core Application Profile), Dublin Core.

1 Introduction

The knowledge reuse discipline requests mechanisms in order to help identify semantics in concepts and guarantee the interoperability between models. This need is flagrant in areas like Semantic Web (SW), or even Domain Analysis in the Software Engineering field. As well as go on with Software Engineering Models, one of the major needs of SW is that documents must be structured and expressed in a reusable manner (it means not ambiguous, understandable, accessible and consensus). In this way, upcoming software applications could include and use in a standardized mode Web information coming from diverse sources. In order to insert the Web information, documents require to be delineated in base of elements with shared semantics and interoperable structures. This paper offers and carries out a methodology that makes possible the interoperability by adding a semantic layer and using a reference ontology.

In the SW to guarantee the understanding and interoperability between diverse schemes requests the adoption of a coding language like XML, for example; and at least this two elements: First, a common and interoperable structure is normally expressed through schemas (for example, W3C XML schemas). Second, a controlled

vocabulary used to define elements of the schema, thanks to public shared resources that integrate a vocabulary with a rate of structured and not ambiguous definitions. The resources have a variation in the complexity rate that goes from metadata vocabularies (like Dublin Core) to ontologies (like WebKB). At the moment, the structure and definition of the schema elements enclose some lacks akin to insufficient legibility and absence of search engines by meaning, it request a knowledge of schemas and it increase the complexity of reusing them. In consequence, the development of vocabularies and ad-hoc schemas starts off, but it depends on the needs of projects or applications, it makes hard the interaction between new and current schemas. The application profiles [1] includes metadata coming from diverse vocabularies, it allows just a local refinement of them; the metadata registries includes local definitions and just a set of recommendations regarding how to express it [2]; at last, the *crosswalks* or mapping one-by-one of elements [3] presents problems of consensus in the definition of elements' semantics and a scalable solution. These solutions are not conceived for reusing future metadata. The proposal of this work is the reuse of schemas, in an extensible frame able to include unambiguous definitions, understandable for users and accessible by Web. It is significant to point out that it is projected to reach the consensus using a collaborative Web resource; the resource could let users to include and manage models.

2 Semantic Qualifiers Schema

The semantic qualifiers schema creates the semantic layer that makes possible to interoperate diverse schemas, including elements in a similar way as DCMI [4], but the elements are designated to the semantic qualification of metadata schemas. The schema has two views hold in two documents, one aimed to users and one in RDF. It comprise of:

- Schema description: it includes the metadata that defines spaces for *name*, *title*, *description*, *location*, and *schema editors*. Thanks to the use of basic elements and Dublin Core (DC) terms.
- Semantic Schema Class (SemanticSchema): a common class to all the semantic schemes, for example ontologies. The representation used is Vocabulary Encoding Scheme [5]. It includes as reference attribute the element *hasSemantics*.
- The *hasSemantics* property represents the relations between schema elements and their meanings, including a reference to concepts of shared semantic resource(s). It refines the element *dc:relation* allowing incompatible applications to ignore the element using the Dumb-Down algorithm [6], [7].

3 Methodology for Semantic qualification of schemas

Besides the schema, a methodology for semantic disambiguation of terms and interoperability between schemes has been defined; it is based on the semantics of elements. The methodology must comply the requirements of: 1) the solution must be

valid to any set of schemas. 2) The compatibility between systems using the original schema must be guaranteed, it signifies the algorithm Dumb-Down must be applicable. 3) Due to its popularity, the naming recommendation used will be [8] and namespaces will be [9] of DCMI [7]. 4) The semantic assignment must allow the reuse of elements in the future, they must be consensus, unambiguous, understandable, and accessible. 5) Documentation for qualified schemas will be generated in order to facilitate the interpretation for users and computers. 6) A solution supporting modularity, extensibility, refinement, and plurilinguistic criteria is given [10].

The semantic qualified schemas are a set of two[1] documents: A formal document for users; and a RDF document for automatic processing.

The main steps of the methodology are along these lines:

P1. Analyzing the metadata schema to be qualified:

It is important to understand structure and substance of the schema, a RDF schema makes it easy.

P2. Generating the user document:

1. Include metadata of the state of the document, at least*: Title, author, document id, creation date, last version link, state, description and approval date*.
2. Include interesting references in order to help the understanding of the document.
3. Generate a schema description including at least: *title, description, editor, language, edition date, modification date, URL* and finally needed schemas for the schema interpretation.
4. Describe attributes that characterize each element in the schema, like: name, URL, tag, definition, etc.
5. Generate a semantic qualified element for each element in the original schema.

5.1. Naming elements (for example, "*name_element + Sem*"); the element must be labelled with a clear semantic and a short description. The type must be settled as *semantic-element* and the semantic parent term in case of (for example, *titleSem* makes a semantic refinement of *title* with the type *semantic-element*).

5.2. Include references in the meanings and a short description of it.

5.3. Include the element state in the approval process.

5.4. Include the creation date of the element (and modification if needed).

6. Include document references in case of typology and use of elements; it means, the grammatical of the document. If a new type of element must be included, then it is needed to generate a new document avoiding replication of metadata types. The new document must include only new types and important references.
7. (Optional) Include the semantic qualified schema in RDF.

P3. Generate the document in RDF format

By means of elements and terms from Dublin Core [11], RDF(s) [12] and semantic qualifiers schema, the substance in the user document will be represented.

---

[1] In case of new types of elements (in the same way we have defined a new type of qualifier, called *semantic-element*, to express the semantics) it will be needed to generate a third document defining the gramatical principles, as well as DCMI (2003).

4 Conclusions

The methodology located in this paper helps to improve the reuse of schemas, allows to include unambiguous definitions, makes it understandable for users and accessible by Web. The frame supports extensibility and precludes local definitions.

A result is the ability to retrieve elements by meaning, and bond elements of diverse schemas in a semantic way; it means a level of abstraction higher than syntactic.

The proposal increase the compatibility with current versions of schemas thanks to methods included in RDF(S) and following the established policies of DCMI in case of naming of terms and namespaces in metadata schemas. Therefore, not compatible applications with semantic qualification of elements could use its original elements and avoid the complexity in the original schema. Both aspects facilitate the use of semantic qualified schemas. Moreover, the method of qualification proposed allows the qualification of any available schema, it means a flexible and extensible semantic qualifiers schema. The methodology helps to develop the qualification process of any metadata schema. It is essential to elaborate some schemes version control for automatic processing, and it is important as well to make it user-friendly. This aspect will ease the understanding and use of developed schemas.

References

1. CEN 2004: Guidelines for machine-processable representation of Dublin Core Application Profiles: Final CWA 14885.
2. ISO/IEC: 11179: Metadata Registries (MDR) (2006). Available at: http://metadata-standards.org/11179/ [April 2006].
3. UKOLN: Mapping between metadata formats (2002). Available at: http://www.ukoln.ac.uk/metadata/interoperability/ [April 2006].
4. DCMI: Dublin Core Qualifiers. (2000) Available at: http://au.dublincore.org/documents/dcmi-terms/index.html [April 2006].
5. Powell, Andy; et al.: DCMI Abstract Model (2005). Available at: http://dublincore.org/documents/2005/03/07/abstract-model/ [April 2006].
6. Kokkelink, Stefan, Schwänzl Roland: Expressing Qualified Dublin Core in RDF/ML. (2002) Available at: http://dublincore.org/documents/dcq-rdf-xml/ [April 2006].
7. DCMI Usage Board: DCMI Grammatical Principles. (2003) Available at: http://dublincore.org/usage/documents/principles/ [April 2006].
8. Weibel, S.; Baker, T.: DCMI Policy on Naming Terms (2004). Available at: http://dublincore.org/documents/naming-policy/ [April 2006].
9. Powell, Andy; Wagner H.: Namespace Policy for the Dublin Core Metadata Initiative (DCMI) (2001). Available at: http://dublincore.org/documents/dcmi-namespace/ [April 2006].
10. Duval, E; Hodgins, W; Sutton, S; Weibel, S: Metadata Principles and Practicabilities. In D-Lib Magazine (2002), 8(4).
11. DCMI. Dublin Core Metadata Element Set (2004). http://dublincore.org/documents/dces/ [April 2006].
12. W3C: RDF Vocabulary Description Language 1.0: RDF Schema (2004). Available at: http://www.w3.org/TR/rdf-schema/ [April 2006].