Semantic Metadata Interoperability and Inference-Based Querying in Digital Repositories

Dimitrios A. Koutsomitropoulos, University of Patras, Greece
Georgia D. Solomou, University of Patras, Greece
Andreas D. Alexopoulos, University of Patras, Greece
Theodore S. Papatheodorou, University of Patras, Greece

ABSTRACT

Metadata applications have evolved in time into highly structured “islands of information” about digital resources, often bearing a strong semantic interpretation. Scarcely however are these semantics being communicated in machine readable and understandable ways. At the same time, the process for transforming the implied metadata knowledge into explicit Semantic Web descriptions can be problematic and is not always evident. In this article we take upon the well-established Dublin Core metadata standard as well as other metadata schemata, which often appear in digital repositories set-ups, and suggest a proper Semantic Web OWL ontology. In this process the authors cope with discrepancies and incompatibilities, indicative of such attempts, in novel ways. Moreover, we show the potential and necessity of this approach by demonstrating inferences on the resulting ontology, instantiated with actual metadata records. The authors conclude by presenting a working prototype that provides for inference-based querying on top of digital repositories.

Keywords: Knowledge Discovery, Metadata, Ontologies, Semantic Web, Semantic Matching

INTRODUCTION

Metadata are today one of the most widely adopted paradigms to facilitate description, integration, discovery and preservation of information and resources stored in remote databases or hosted in web-accessed portals and digital libraries. One of the reasons that the Dublin Core (DC) schema is widely adopted in such scenarios is probably its simplicity and its general applicability that makes it suitable for a number of different metadata-intensive applications.

DOI: 10.4018/jitr.2009062903
In many standard repository configurations (including the DSpace digital repository software) the DC Metadata Element Set (DCMES) is implemented as a flat aggregation of elements. This is also true for qualifiers, which are not always implemented as sub-properties of main elements; rather, they often appear at the same level as parent elements and the sub-element/qualifier relationship is maintained only in the label. This situation, evident also in the DSpace-based University of Patras institutional repository (http://repository.upatras.gr/dspace/) is depicted in Figure 1.

The semantic interpretation of the DC model that, as we see, is not always represented in applications, is formalized through the DC Abstract Model (DCAM) specification (Powell, Nilsson, Naeve, Johnston, & Baker, 2007) as well as the most recent recommendation for expressing DC in the Resource Description Framework (RDF) (Nilsson, Powell, Johnston, & Naeve, 2008). These documents virtually suggest an ontology of DC, expressed in RDF(S), a Semantic Web standard.

Such a DC ontology bears its own semantic structure that may be taken advantage of, in order to enable more refined descriptions of resources. This of course is reminiscent of the well-known Semantic Web “bootstrapping problem” (Dill, Eiron, Gibson, et al., 2003; Hendler, 2008): The availability of high-quality, complex and interconnected resource descriptions is a key aspect for the Semantic Web to be of some value; on the other hand, the burden to create a whole new set of rich annotations is too high, both from a conceptual (hard to conceive) as well as from an effort (too much time) point of view. It is unlikely that existing implementations, already employing flat descriptions on a large number of resources, would invest in reorganizing from scratch their underlying data model. Even if such a venture is undertaken, the cost for aligning and enriching existing descriptions can be prohibiting.

Having these in mind, we propose an implementation of the DC ontology that is to be carried out in terms of a most centralized approach. To do this we are based on the semantic profiling technique, well-applied previously on fully-structured knowledge domains, such as the CIDOC Conceptual Reference Model (CIDOC-CRM) (Crofts, Doerr, & Gill, 2003) and introduced in (Koutsomitropoulos, Paloukis, & Papatheodorou, 2007). Using this technique we try to better capture the intended semantics of the DC metadata domain, having the DC RDF(S) schema as a starting point.

Our goal is to upgrade this ontology up to OWL and OWL 2 level (Parsia, & Patel-Schneider, 2008), by incorporating new constructs and refinements, available only in these languages. At the same time, we build upon the initial model and do not require any alternations in its original specification. The resulting ontology, including the new refinements, is then populated in an automated way from metadata already existing within the live DSpace installation of the University of Patras.
15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/article/semantic-metadata-interoperability-inference-based/37408

Related Content

The Expert’s Opinion
www.irma-international.org/article/expert-opinion/50978/

Co-Creation between Food Allergy Customers and Restaurants using Mobile Devices
www.irma-international.org/chapter/co-creation-between-food-allergy-customers-and-restaurants-using-mobile-devices/113187/

Database Benchmarks
www.irma-international.org/chapter/database-benchmarks/13689/

Object-Oriented Software Reuse in Business Systems
www.irma-international.org/chapter/object-oriented-software-reuse-business/14577/

Information Systems Outsourcing in Large Companies: Evidences from 20 Ireland Companies
www.irma-international.org/article/information-systems-outsourcing-large-companies/59971/