Digital Repositories and the Semantic Web: Semantic Search and Navigation for DSpace

Dimitrios A. Koutsomitropoulos, Georgia D. Solomou, Andreas D. Alexopoulos and Theodore S. Papatheodorou

High Performance Information Systems Laboratory, School of Engineering, Computer Engineering and Informatics Department, University of Patras,
Building B, 26500, Patras-Rio, Greece
{kotsomit, solomou, aalexopoulo, tsp}@hpclab.ceid.upatras.gr

Abstract. In many digital repository implementations, resources are often described against some flavor of metadata schema, popularly the Dublin Core Element Set (DCMES), as is the case with the DSpace system. However, such an approach cannot capture richer semantic relations that exist or may be implied, in the sense of a Semantic Web ontology. Therefore we first suggest a method in order to semantically intensify the underlying data model and develop an automatic translation of the flatly organized metadata information to this new ontology. Then we propose an implementation that provides for inference-based knowledge discovery, retrieval and navigation on top of digital repositories, based on this ontology. We apply this technique to real information stored in the University of Patras Institutional Repository that is based on DSpace, and confirm that more powerful, inference-based queries can indeed be performed.

1 Introduction

In this paper we present and document a process that builds upon the well-known *digital repositories* paradigm and enhances it with the Semantic Web's features. In other words, the main goal that drives our efforts is not to re-implement a digital repository system using Semantic Web APIs and technologies, but to provide inference-based knowledge discovery, retrieval and navigation *on top* of such a system, based on existing metadata and other semi-structured information.

To prove our concept, we describe a concrete, working prototype that provides for inference-based search and navigation on top of the DSpace digital repository system. DSpace metadata follow the Dublin Core (DC) specification by default, while it is possible to import and use other metadata schemata as well. Our work and results are based on real-world data and applied on the official University of Patras institutional repository that is based on DSpace (http://repository.upatras.gr/dspace/). Its metadata are based on the original DSpace schema extended with learning object (LOM) metadata.

A partial description of this work and source code are publicly available at: http://wiki.dspace.org/index.php/User:Kotsomit.

2 Extracting an ontology out of metadata records

In many standard repository configurations (including the DSpace digital repository software), resources are described based on the Dublin Core Metadata Element Set (DCMES) which is often implemented as a flat aggregation of elements. The semantic interpretation of the DC model that is not always represented in applications, is formalized through the DCMI Abstract Model (DCAM) specification [5] as well as the most recent recommendation for expressing DC in RDF [3]. 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. However, as pointed out in [1], the burden of producing from scratch a whole new set of richer descriptions can be prohibiting.

Fig. 1. Partial view of the DSpace ontology class hierarchy (excluding some imported axioms).

As a potential solution to this problem, within our work we have implemented a DC ontology 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 [1]. Our goal is to upgrade this ontology up to OWL and OWL 2 level [4], 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. In this process, we also take into account the LOM metadata, with which we have extended the original DSpace schema. 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 institutional repository, through its OAI-PMH interface.

Part of the resulting ontology is depicted in Figure 1.

2 Semantic Search and Navigation

The most important modules and interfaces that enable semantic services in our digital repository are the following:

- *Semantic Search* interface (Fig. 2), which, in collaboration with the appropriate inference engine, allows for the construction, submission and evaluation of a semantic query. Retrieved results are displayed here in the form of a list.
- *Semantic Navigation* interface (Fig. 3) is where detailed ontological information about a selected entity (individual) is presented.
- *Ontology Population* refers to the dynamic construction of the ontology, which comes from DSpace's OAI harvested metadata, after applying the appropriate XSLT transformation on them.
- The *Inference Engine* is responsible for processing the ontological documents and for performing reasoning over them.

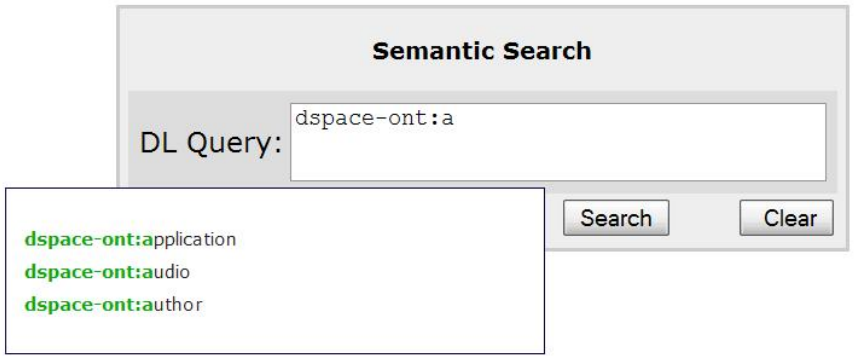
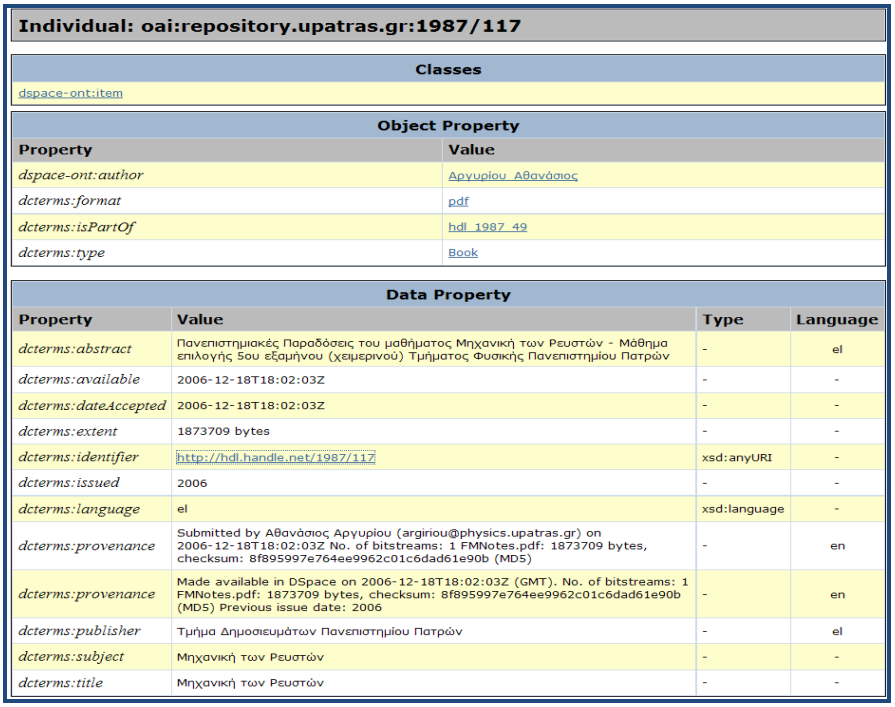


Fig. 2. The Semantic Search interface.

These facilities have been implemented in Java servlets, using the OWL API. These servlets extend `DspaceServlet`, an extension of the Java `HttpServlet` class. This is the only (and necessary) reference to the DSpace API. Reasoning is performed by the FaCT++ inference engine, but any other DL reasoner may be used. FaCT++ is interfaced by the appropriate abstract class of the OWL API (`Reasoner`), through JNI.

The populated ontology is dynamically constructed and silently fed to the reasoner over HTTP. In fact, our semantic search and navigation services are designed and implemented in such a way, that they can work with any OWL document, not just the one populated with the repository's metadata: Since the ontology URI is passed as an HTTP parameter, it is easy to parameterize the user interface to ask for an ontology URL as well, making our implementation totally independent of the specific ontological model.



Individual: oai:repository.upatras.gr:1987/117

Classes

dspace-ont:item

Object Property

Property	Value
dspace-ont:author	Αργυρίου Αθανάσιος
dcterms:format	pdf
dcterms:isPartOf	hdl_1987_49
dcterms:type	Book

Data Property

Property	Value	Type	Language
dcterms:abstract	Πανεπιστημιακές Παραδόσεις του μαθήματος Μηχανική των Ρευστών - Μάθημα επιλογής 5ου εξαμήνου (χειμερινού) Τμήματος Φυσικής Πανεπιστημίου Πατρών	-	el
dcterms:available	2006-12-18T18:02:03Z	-	-
dcterms:dateAccepted	2006-12-18T18:02:03Z	-	-
dcterms:extent	1873709 bytes	-	-
dcterms:identifier	http://hdl.handle.net/1987/117	xsd:anyURI	-
dcterms:issued	2006	-	-
dcterms:language	el	xsd:language	-
dcterms:provenance	Submitted by Αθανάσιος Αργυρίου (argiriou@physics.upatras.gr) on 2006-12-18T18:02:03Z No. of bitstreams: 1 FMNotes.pdf: 1873709 bytes, checksum: 8f895997e764ee9962c01c6dad61e90b (MD5)	-	en
dcterms:provenance	Made available in DSpace on 2006-12-18T18:02:03Z (GMT). No. of bitstreams: 1 FMNotes.pdf: 1873709 bytes, checksum: 8f895997e764ee9962c01c6dad61e90b (MD5) Previous issue date: 2006	-	en
dcterms:publisher	Τμήμα Δημοσιευμάτων Πανεπιστημίου Πατρών	-	el
dcterms:subject	Μηχανική των Ρευστών	-	-
dcterms:title	Μηχανική των Ρευστών	-	-

Fig. 3. The Semantic Navigation interface (viewing item 1987/117).

References

1. Koutsomitropoulos, D., Paloukis, G., Papatheodorou, T.: From Metadata Application Profiles to Semantic Profiling: Ontology Refinement and Profiling to Strengthen Inference-based Queries on the Semantic Web. International Journal of Metadata, Semantics and Ontologies, 2 (4), 268-280 (2007).
2. Koutsomitropoulos, D., Solomou, G., Papatheodorou, T.: Semantic Interoperability of Dublin Core Metadata in Digital Repositories. In: 5th International Conference on Innovations in Information Technology, Al Ain, UAE (2008)
3. Nilsson, M, Powell, A., Johnston, P., Naeve, A.: Expressing Dublin Core metadata using the Resource Description Framework (RDF), DCMI Recommendation (2008) http://dublincore.org/documents/dc-rdf/
4. Parsia, B., Patel-Schneider, P. F.: OWL 2 Web Ontology Language: Primer. W3C Working Draft (2008) http://www.w3.org/TR/owl2-primer/
5. Powell, A.,Nilsson, M., Naeve, A., Johnston, P., Baker, T.: DCMI Abstract Model. DCMII Recommendation (2007) http://dublincore.org/documents/abstract-model/