Navigating among Educational Resources in the Web of Linked Data

Dimitrios A. Koutsomitropoulos, Georgia D. Solomou and Aikaterini K. Kalou

High Performance Information Systems Laboratory (HPCLab),
Computer Engineering and Informatics Dpt., School of Engineering,
University of Patras, Building B, 26500 Patras-Rio, Greece

{kotsomit, solomou, kaloukat}@hpclab.ceid.upatras.gr

Abstract. Linked Data seem to play a seminal role in the establishment of the Semantic Web as the next-generation Web. This is even more important for digital object collections and educational institutions that aim not only to promote and disseminate their content but also to aid its discoverability and contextualization. Having already 'semantified' a popular digital repository system, DSpace, in this paper we show how repository metadata can be exposed as Linked Data, thus enhancing their machine understandability and contributing to the LOD cloud. Our effort comes complete with an updated UI that allows for reasoning-based search and navigation between linked resources within and outside the scope of the digital repository.

1 Introduction

Linked and Open Data (LOD) [2] appear to be the "silver-bullet" in the forming Semantic Web ecosystem, that promise to breathe new life to the latter's benefits for real-world web applications. This is often combined with lightweight semantics [3] so that known scalability problems and reasoning inefficiencies could be sidestepped and still to retain some essence of the knowledge discovery capabilities of ontologies. However, tried-and-true systems like digital repositories for educational and other institutions need a little more incentive to embark on such a migration and to get tempted to adopt this new paradigm.

In this paper we present our work for publishing Linked Data and navigating among resources of a popular digital repository system, DSpace. Semantics play a crucial role and this is exhibited by an OWL 2 inference-based knowledge acquisition mechanism that lies at the core of this implementation, aka *Semantic Search for DSpace* [6]. Challenges for imposing semantic searching over otherwise semantically-oblivious systems are well-known and have been discussed earlier (e.g. [5]). Further, Linked Data provision requires a careful replication design for existing resource descriptions; a data linking and resolution mechanism; and a content negotiation strategy to serve information both to end-users and machines.

Next, in Section 2 we describe the process of publishing and minting Linked Data out of DSpace resources; Section 3 presents the semantic querying interface and the Linked Data facility; and Section 4 summarizes our conclusions and future work.

Semantic Search is hosted and maintained as a Google Code project[1] and is listed as an official DSpace add-on[2]. A working demo of this implementation is also available[3].

2 Linking Data for DSpace

2.1 Publication and linking of entities

Linked Data principles are in essence a few simple rules that foster the idea of an interlinked 'Web of Data'. In our context this means that resource URIs need to be dereferenceable, to provide meaningful information for users and services alike and to give references (or links) to other related entities whenever possible.

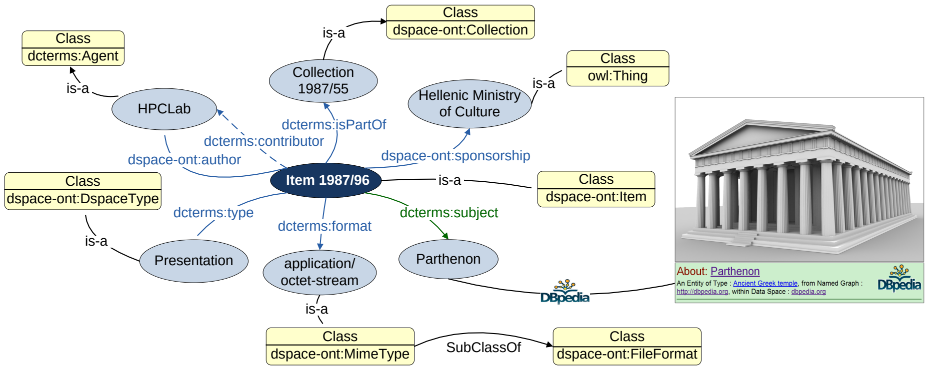

Fig. 1. Example repository item and its relationships to other entities

In DSpace the main unit of information is the 'item', i.e. a publication or learning object that is described with a set of metadata based on Dublin Core (DC). During the mapping to OWL however, we identify additional implicit entities and assign resolvable URIs to them too (see below, section 2.2). Further, these entities are linked together or refer to other external datasets like DBpedia (see section 3.2). Fig.1 illustrates a sample instance of the resulting DSpace ontology and the way it gets interlinked with other entities and/or datasets. Using the Jersey framework[4], the reference implementation of the Java API for RESTful services, both HTML as well as RDF/XML representations are accommodated and the following URI pattern has been established:

[1] http://code.google.com/p/dspace-semantic-search/
[2] https://wiki.duraspace.org/display/DSPACE/Extensions+and+Addons+Work#ExtensionsandAddonsWork-SemanticSearchforDSpacev2.0
[3] http://apollo.hpclab.ceid.upatras.gr:8000/dspace-ld
[4] https://jersey.java.net/

— http://{repositoryURL}/semantic-search/resource/{entity-id}
— http://{repositoryURL}/semantic-search/page/{entity-id}(HTML)
— http://{repositoryURL}/semantic-search/data/{entity-id}(RDF)

Moreover, these URIs are dereferenceable by using Jersey's content negotiation capability and performing HTTP *303 See Other* redirects.

2.2 Exposing educational metadata into OWL

The first step towards providing Linked Data is to unleash resource information and metadata that are hidden within databases. To maximize the semantic value of exported metadata as well as to maintain interoperability, a careful and elaborate process has to be conducted. This process includes an exhaustive mapping of the inherent DSpace metadata application profile to the DC metadata terms (DCTerms), part of which is shown in Table 1. This mapping provides for the following:

- Other than the default, additional elements are exported.
- Map everything under the DCTerms namespace, rather than mixing DC with DCTerms, which is generally not advisable [1].
- Provide additional Learning Object mappings to DCTerms (in case LOM metadata exist).
- Assign types to non-literal values.

Mapped metadata can then be exposed and harvested through the supported OAI-PMH interface [7].

Table 1. Snippet of the performed mapping of DSpace internal metadata to DCTerms.

DSpace internal metadata representation	Provided mapping	Notes
`dc.contibutor.author`	`dspace-ont:author` `dcterms:contributor`	*'author' is not compatible with QDC. However it is a subproperty of 'contributor'*
`dc.subject`	`dcterms:subject`	*We use the DCTerms namespace*
`dc.identifier.uri`	`dcterms:identifier type= http://www.w3.org/2001/XMLSchema#anyURI`	*Literals get typed when possible*
Bitstream metadata	`dcterms:format` `dcterms:extent`	*Not exposed by default*
`lom.intendedenduserrole`	`dcterms:audience type="lom:IntendedEndUserRole"`	*LOM specific mapping*

Having exported as much available DSpace metadata as possible, the next step is to translate them into full-fledged RDF/OWL triples. During this semantic translation, certain implicit entities (like items, collections, authors, sponsors) are reified and become individual nodes themselves instead of mere literals, thus resulting into an OWL 2 DSpace ontology [5]. Most of these entities are assigned resolvable identifiers, so that it would be easy for them to get dereferenced within the individual *Navigation Pane* (see below). In addition, this is when URLs to the DBpedia Lookup service[5] are injected, in order to enrich reified entities such as authors, contributors, sponsors and item types.

3 Semantic Search and Navigation

In what follows, we summarize the semantic search interface's main features (3.1) and then we detail the newly implemented Linked Data facility (3.2). A more thorough account of the architecture, methodology and design principles behind Semantic Search v2 can be found in [6].

3.1 An interface for semantic querying educational resources

The main idea used by our semantic search interface, lies behind the deconstruction of a semantic query into smaller building parts (query *atoms*) that are assigned to different fields of a dynamic UI. Query crumbs that are provided through these fields are then assembled by the underlying mechanism to create valid Manchester Syntax expressions [4] (see fig. 2). Each such expression is an atomic or anonymous class in the Web Ontology Language (OWL) [8] and its (both inferred and asserted) members are the answers to the query. Search results are presented as a browsable list of linkable entities/resources. Effective querying of the knowledge base is accomplished by interconnecting to an appropriate inference engine, capable of reasoning with OWL 2 ontologies.

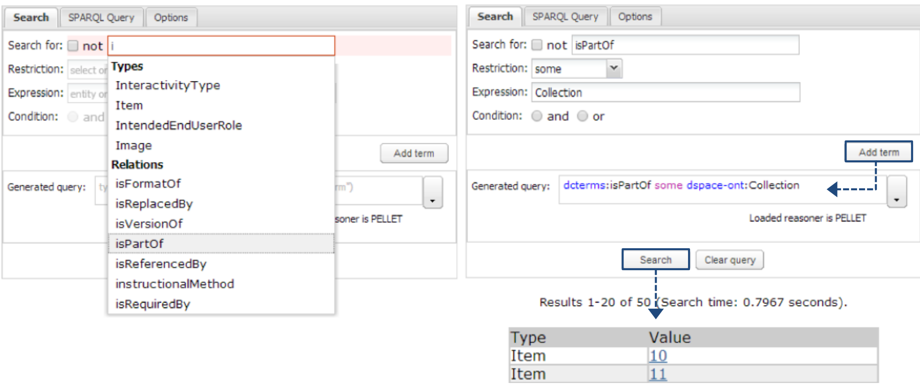

Fig. 2. The auto-complete and query construction mechanisms of the semantic search interface

New features in the upcoming version 2.4 include: *a)* syntax highlighting for Manchester Syntax, implemented through the CodeMirror Javascript component, *b)* a history subsystem that keeps track of the last ten input queries, *c)* a RESTful Linked

[5] http://wiki.dbpedia.org/Lookup

Data provider that exposes resources' metadata as resolvable entities, and *d)* a DBpedia URL injection facility.

When an entity is selected, the corresponding individual's navigation pane that gathers the resource's semantic metadata is produced on the fly. This ontological information is formed as Linked Data and is presented in either an HTML or an RDF format, depending on whether the request was made by a person or a service, respectively. It is important to notice that this holds not only for DSpace items themselves, but also for every other (implicit) entity in the repository model that gets reified during the semantic translation phase.

3.2 Navigation and Data Linking

The main objective of the navigation pane is not only to give a detailed reference to the resources' ontological information (semantic metadata) but also to allow users to further explore and navigate among interlinked information in the LOD cloud. To achieve this, information is structured in the form of resolvable URIs, as much as possible.

All non-literal metadata values are denoted as URIs, which can be dereferenced on the web. In particular, each class redirects back to the semantic search page with the specific class already predefined and its members appearing on the result list. In the case of object properties, the corresponding values are resolvable entities that lead to the particular entity's navigation pane. And even for data properties, where mere text values mostly apply, a datatype of `xsd:AnyURI` is rendered as a resolvable link. This is useful for example to maintain context with the original DSpace item view or with external references, such as the DBpedia Lookup service.

Fig. 3. The navigation pane - The DBpedia lookup service is triggered for author "Tom Scott"

More specifically, a DBpedia icon is automatically placed next to specific reified entities (contributor, author, type, sponsor) and next to the object property values (within an item's navigation pane) that correspond to these entities. When the icon is clicked, the DBpedia Lookup service is triggered for this entity, leading to a keyword-based search against DBpedia. This label matching process inevitably includes a certain extent of ambiguity. In order to resolve this, a dynamic tooltip is presented to the user, including up to five matching DBpedia resource URIs (see fig. 3). Moreover, the `dcterms:subject` and `dcterms:publisher` literal values are also linked to DBpedia in the same way.

All information gathered in the navigation pane can also be obtained in a machine-readable RDF format. An RDF icon next to the entity points to our REST Linked Data service and requests the RDF representation of the entity's data.

4 Conclusions and Future Work

We soon intend to release this feature-set in the upcoming version of our Semantic Search plugin for DSpace. The combination of a reasoning-based knowledge acquisition mechanism with a Linked Data service can help educational institutions to provide new discovery capabilities for their content and to be part of the greater LOD cloud effortlessly. DBpedia is naturally a nodal point of the latter, but interconnecting with other data sources would also be useful, like for example DBLP. What is more, data from these sources can be brought back into our model, so that we could reason with them and reveal a whole new set of correlations between repository assets and the outside world.

References

1. DCMI Usage Board: DCMI Metadata Terms. DCMI Recommendation. (2008) http://dublincore.org/documents/2008/01/14/dcmi-terms/
2. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136 (2011)
3. Hendler, J.: "Why the Semantic Web will Never Work". Keynote at the 8th Extended Semantic Web Conference. Heraklion, Greece (2011)
4. Horridge, M., Patel-Schneider, P.S.: Manchester Syntax for OWL 1.1. In: 4th OWL Experiences and Directions Workshop. Gaithersburg, Maryland (2008)
5. Koutsomitropoulos, D., Solomou, G., Alexopoulos, A., Papatheodorou, T.: Semantic Metadata Interoperability and Inference-Based Querying in Digital Repositories. Journal of Information Technology Research, 2(4), 36-52 (2009)
6. Koutsomitropoulos, D., Solomou, G., Papatheodorou, T.: Semantic query answering in digital repositories: Semantic Search v2 for DSpace. International Journal of Metadata, Semantics and Ontologies, 8(1), 46-55 (2013)
7. Lagoze, C., Van de Sompel, H., Nelson, M., Warner, S.: The Open Archive Initiative Protocol for Metadata Harvesting (2002) http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm
8. Motik, B., Parsia, B., Patel-Schneider, P.F. (eds.): OWL 2 Web Ontology Language XML Serialization (2nd ed.). W3C Recommendation (2012)