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

Dimitrios Koutsomitropoulos, Georgia Solomou, Andreas Alexopoulos and Theodore Papatheodorou
University of Patras Institutional Repository

A mechanism for the efficient description, preservation, management, exploitation and distribution of the University’s educational and scientific material

- Built upon the open-source DSpace digital repository system
- Item description using the Dublin Core metadata schema

http://repository.upatras.gr/dspace

Articles, Books, Theses, Journal Papers, Images, Videos, Learning Objects, Data Sets, ...
University of Patras Institutional Repository

Additional features

- Multilingual support
  - User Interface (Greek, English, ...)
  - Metadata - Characterization of items in more than one language

- Advanced search service
  - Full text
  - Metadata
  - Semantic Search

- Advanced browsing
  - Semantic navigation
The Semantic Web

- New potential for the Web
  - Rich descriptions of resources + co-relations
  - Ability to reason about information
  - Knowledge acquisition and discovery (Inference based)

- Ontology Languages: OWL
  - Extensions to RDF(S)
  - Standard vocabulary for ontology representation
  - Decidable, sound & complete (! OWL Full)

- OWL 2: extension to OWL
  - More expressive constructs
    - Role-chains and characteristics
    - Negative assertions
    - Punning
 Metadata & Applications

- Metadata Standards
  - Capture a *level of meaning* of (web) resources
  - Predate SW standards

- The Dublin Core
  - 15 main elements
  - Many other *qualifications*
    - Sub-elements – correspond to *relations*
  - Popular in describing resources
    - In Digital Libraries / Repositories (like DSpace)
    - Supports interoperability
      - *Structural* consistency in information exchange
Metadata in DSpace

- Based on Dublin Core
  - Influenced by the Library Application Profile (DC-LAP)
  - A total of 66 elements (some invisible)
    - Including qualifications
- Includes non-standard elements
  - Cannot be mapped to DC
    - e.g. “author” and “sponsorship”
- Exportable through OAI-PMH
  - Provided a mapping is specified
The Problem

- Metadata in DSpace: Monolithic approach
  - Metadata flatly organized
  - Meaning lies *implicitly* in the **structure**
    - or in the (human-understandable) specifications!
    - Not machine “understandable”
  - Semantically **semi-structured** knowledge models
    - In contrast to fully-structured ontology models
Item 1987/96: Presentation about the 3D reconstruction of the Parthenon of Athens
**Item 1987/96**

**Full Metadata Record**

<table>
<thead>
<tr>
<th>DC Field</th>
<th>Value</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.contributor.author</td>
<td>HPCLab</td>
<td>en</td>
</tr>
<tr>
<td>dc.date.accessioned</td>
<td>2006-12-11T11:41:05Z</td>
<td>-</td>
</tr>
<tr>
<td>dc.date.available</td>
<td>2006-12-11T11:41:05Z</td>
<td>-</td>
</tr>
<tr>
<td>dc.date.issued</td>
<td>2006-12-11T11:41:05Z</td>
<td>-</td>
</tr>
<tr>
<td>dc.identifier.uri</td>
<td><a href="http://hdl.handle.net/1987/96">http://hdl.handle.net/1987/96</a></td>
<td>-</td>
</tr>
<tr>
<td>dc.description</td>
<td>Η παρουσίαση αυτή προβλήθηκε στην Κοπενχάγη στις 12/11/2006</td>
<td>el</td>
</tr>
<tr>
<td>dc.description.abstract</td>
<td>The presentation that is attached describes the virtual reconstruction and 3D Real-time walkthrough inside the Parthenon temple. Issues of objectivity, accuracy and detail are also described and as a conclusion innovations and benefits are presented too.</td>
<td>en</td>
</tr>
<tr>
<td>dc.description.abstract</td>
<td>Η παρουσίαση που επισημάντηκε είναι μια περιγραφή του αντίστοιχου έργου που σκοπεί να αποδώσει πραγματικούς χρόνους περιπάτους σε μια εικονική αναπαράσταση του Παρθενώνα. Κατά την παρουσίαση αυτή γνωστοποιείται πως οι ανιμετωπίζονται τα θέματα της αντικειμενικότητας, της λεπτομερείας και της ακρίβειας. Τέλος παρουσιάζονται οι καινοτομίες και τα πλεονεκτήματα του συγκεκριμένου έργου.</td>
<td>el</td>
</tr>
<tr>
<td>dc.description.sponsorship</td>
<td>Hellenic Ministry of Culture</td>
<td>en</td>
</tr>
<tr>
<td>dc.description.sponsorship</td>
<td>Υπουργείο Πολιτισμού</td>
<td>el</td>
</tr>
<tr>
<td>dc.format.extent</td>
<td>25177195 bytes</td>
<td>-</td>
</tr>
<tr>
<td>dc.format.mimetype</td>
<td>application/octet-stream</td>
<td>-</td>
</tr>
<tr>
<td>dc.language.iso</td>
<td>en</td>
<td>en</td>
</tr>
<tr>
<td>dc.subject</td>
<td>Parthenon</td>
<td>en</td>
</tr>
<tr>
<td>dc.subject</td>
<td>Παρθενώνας</td>
<td>el</td>
</tr>
<tr>
<td>dc.title</td>
<td>Virtual Reconstruction &amp; 3D Real-time Walkthrough Environment</td>
<td>en</td>
</tr>
<tr>
<td>dc.type</td>
<td>Presentation</td>
<td>en</td>
</tr>
<tr>
<td>dc.type</td>
<td>Παρουσίαση</td>
<td>el</td>
</tr>
</tbody>
</table>

Appears in Collections: Ημερίδα Πολιτισμού
Metadata Relationships

Metadata are flatly organized in the DB. Even structure is often unimplemented (exists only in the label)
Semantic Relationships
Approach

A. Create an **Ontology** for Dublin Core and DSpace
   ◦ up to OWL 2 level (non-standard inferences)
   ◦ Produce *meaning* out of *structure*
     • e.g. implement qualifiers as *sub-properties*
   ◦ Make *explicit* the spec and common-sense constraints
     • e.g. inverse relation between `dc:hasPart` and `dc:isPartOf`

B. Populate the ontology
   ◦ Transform and map *existing* DC metadata to a new ontological model

C. Semantics-aware services for DSpace
   ◦ Semantic Search
   ◦ Semantic Navigation
An Ontological Model for DC & DSpace

- Based on existing DC implementation in RDF
- Create incremental semantic profiles of DC
  - by applying the semantic profiling\(^1\) technique
- Gradually, tailor to the specific domain
  - University of Patras DSpace Installation
  - Based on DSpace
- Preserve the original DC model
  - Physically separate profiles
  - One owl:imports the other
  - Smoothly refine the original model

---

Ontological Data

- How to populate the ontology?
  - Harvest and map repository’s metadata
  - Through the standard OAI-PMH interoperability interface

- Minimum intervention
  - Not *altering* the database
  - Not *accessing* the database

- *Automated* population
  - Using standard XML-based technologies (XSLT)

...an interoperable *approach for ontology construction and population*
Semantic Relationships
Semantic Services: Architecture

- Ontology Population
- Inference Engine
- Semantic Search
- DSpace (Business Logic)
- Semantic Navigation
- XSLT Transformation
- DC Terms
- DCAM
- LOM
- Semantic Profiling and Namespace Separation
- Ontological Model
- Repository Metadata
- DB
- OAI Interface
- Ontology Population

DSUG '09, Gothenburg
The Semantic Search Interface
Semantic Search

- Queries are typed as simple text using the Manchester OWL syntax
- Type of accepted queries:
  - Valid ontological class names
  - Class expressions (existential qualifications, cardinality restrictions, ...)
  - Boolean combinations of class expressions
The Manchester OWL Syntax

- A user-friendly syntax for OWL
  - Maps Description Logics symbols to English words and phrases
- Designed for writing OWL class expressions or even complete OWL ontologies

⇒ Offers a simpler way to end users for formulating their queries

<table>
<thead>
<tr>
<th>OWL Expression</th>
<th>Description Logics Symbol</th>
<th>Manchester Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>someValuesFrom</td>
<td>some</td>
<td></td>
</tr>
<tr>
<td>allValuesFrom</td>
<td>only</td>
<td></td>
</tr>
<tr>
<td>hasValue</td>
<td>value</td>
<td></td>
</tr>
<tr>
<td>minCardinality</td>
<td>≥</td>
<td>min</td>
</tr>
<tr>
<td>cardinality</td>
<td>=</td>
<td>exactly</td>
</tr>
<tr>
<td>maxCardinality</td>
<td>≤</td>
<td>max</td>
</tr>
<tr>
<td>intersectionOf</td>
<td>□</td>
<td>and</td>
</tr>
<tr>
<td>unionOf</td>
<td>□</td>
<td>or</td>
</tr>
<tr>
<td>complementOf</td>
<td>¬</td>
<td>not</td>
</tr>
<tr>
<td>SubPropertyChain</td>
<td>◦</td>
<td>◦</td>
</tr>
</tbody>
</table>
Auto-complete Facility

Suggestion of a list of entities that belong to the knowledge base (classes, properties and individuals names)
Help for Semantic Search

Repository Help

Help Contents
- Browse
- Search
- Advanced Search
  - Semantic Search
- Communities
- Collections
- Sign on to Repository
- Submit
- File Formats
- Handles
- My Repository
- Compose a Lesson
- Edit Profile
- Subscribe to E-mail alerts

Repository captures, distributes, and preserves digital research products. Here you can find articles, papers, preprints, technical reports, conference papers and data sets in various digital formats as new communities and collections are added to Repository.

The Repository content is organized around Communities which can correspond to administrative units such as schools, departments, labs, and research centers. Within each community there can be subcommunities and an unlimited number of collections. Each collection may contain an unlimited number of items.

BROWSE
- Browse allows you to see a list of items in a given specified order:
- Oldest to newest

Repository Help

SEMANTIC SEARCH

Semantic Search is a novel feature that allows "intelligent" search in the repository's content, by taking advantage of Semantic Web Technologies. Semantic Search is activated by clicking on the link "Semantic Search" in the yellow search box at the top of the navigation bar on the left:

As a result, the semantic search query interface is displayed, allowing the submission of the query:

Query syntax follows a special syntax known as "Manchester Syntax", which maps Description Logics symbols to English words and phrases. More information on Manchester Syntax can be found here.

Accepted queries should be ontological class names, calls expressions or Boolean combinations of class expressions. For example, all of the repository items can be retrieved typing:

ds-ont:Item

All of the repository items that have at least one type can be retrieved typing:

ds-ont:Item and (dcterms:type min 1 owl:Thing)

Prefixes
The underlying ontology that is used to conduct queries on, is specific to the repository's content; DSpace main entities item, collection and community now map to classes, that are specified with the ds-ont: prefix. Dublin Core fields map to properties that use the prefixes dc: and dcterms:. Please note that properties with different prefixes are different properties! For example, dc:format is different than dcterms:format. It is recommended in general to use the dcterms: properties. Other prefixes used in the ontology include km:, specifying entities about educational metadata, and foaf: for the first and last names of authors (foaf:name),
Example—Conjunctive Queries

Show all items of type “Book” that are mainly intended for Students (audience of type “Student”)
Example – *Role Chains*

Show who draws sponsorship from the “Hellenic Ministry of Culture”

Sponsorship refinement:
```
inv(dcterms:contributor) o sponsorship SubPropertyOf sponsorship
```
Example – Role Chains (2)

Find co-authors of author with surname “Drake”

Co-author declaration:
inv(author) o author SubPropertyOf co_author
Example – *Cardinality Restrictions*

Authors of items that have at least two different formats

```
Semantic Search

DL Query: inv(dspace-ont:author) some
(dcterms:format min 2 owl:Thing)
```

Results 1-2 of 2.

[Dimitrios Sourlas](mailto:dimitrios.sourlas@uni.lu)

[Δημήτριος Σουρλάς](mailto:dimitrios.sourlas@uni.lu)
Semantic Navigation Pane
## Semantic Navigation – Example

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

<table>
<thead>
<tr>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dspace-nt:Item</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dcterms:creator</code></td>
<td>Αργυρίου, Αλέξανδρος</td>
</tr>
<tr>
<td><code>dcterms:format</code></td>
<td>pdf</td>
</tr>
<tr>
<td><code>dcterms:isPartOf</code></td>
<td><a href="http://1987/117">http://1987/117</a></td>
</tr>
<tr>
<td><code>dcterms:type</code></td>
<td>Book</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Property</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dcterms:abstract</code></td>
</tr>
<tr>
<td><code>dcterms:available</code></td>
</tr>
<tr>
<td><code>dcterms:identifier</code></td>
</tr>
<tr>
<td><code>dcterms:language</code></td>
</tr>
<tr>
<td><code>dcterms:provenance</code></td>
</tr>
<tr>
<td><code>dcterms:provenance</code></td>
</tr>
<tr>
<td><code>dcterms:provider</code></td>
</tr>
<tr>
<td><code>dcterms:subject</code></td>
</tr>
<tr>
<td><code>dcterms:title</code></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dspace-nt:dspace:type</code>, <code>nis:Class</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dcterms:creator</code></td>
<td>Αργυρίου, Αλέξανδρος</td>
</tr>
<tr>
<td><code>dcterms:format</code></td>
<td>pdf</td>
</tr>
<tr>
<td><code>dcterms:isPartOf</code></td>
<td><a href="http://1987/117">http://1987/117</a></td>
</tr>
<tr>
<td><code>dcterms:type</code></td>
<td>Book</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>label</code></td>
</tr>
<tr>
<td><code>language</code></td>
</tr>
<tr>
<td><code>label</code></td>
</tr>
<tr>
<td><code>language</code></td>
</tr>
</tbody>
</table>
# Semantic Navigation – Example

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

<table>
<thead>
<tr>
<th><strong>Property</strong></th>
<th><strong>Value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>dspace-nt:inc</td>
<td></td>
</tr>
<tr>
<td>dspace-nt:prop</td>
<td></td>
</tr>
<tr>
<td>dspace-nt:term</td>
<td></td>
</tr>
<tr>
<td>dspace-nt:atitem</td>
<td></td>
</tr>
<tr>
<td>dspace-nt:dt</td>
<td></td>
</tr>
<tr>
<td>dspace-nt:int</td>
<td></td>
</tr>
<tr>
<td>dspace-nt:typ</td>
<td></td>
</tr>
</tbody>
</table>

### Classes

- dspace-nt:term
- dspace-nt:atitem

### Object Property

<table>
<thead>
<tr>
<th><strong>Property</strong></th>
<th><strong>Value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>dspace-ont:author</td>
<td>Árhondas Árgoúrakí</td>
</tr>
<tr>
<td>dspace:format</td>
<td>pdf</td>
</tr>
<tr>
<td>dspace:upartOf</td>
<td>bhl:1987.42</td>
</tr>
<tr>
<td>dspace:atype</td>
<td>Book</td>
</tr>
</tbody>
</table>

### Data Property

<table>
<thead>
<tr>
<th><strong>Property</strong></th>
<th><strong>Value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>dspace:abstract</td>
<td>Πανεπιστημικές Παραδοσές του μαθήματος Μηχανική των Επιλογής &amp; Εξειδίκευσης Κειμένων (έμφαση στην) Τμήματος Φυσικής Πανεπιστημίου Πάτρας</td>
</tr>
<tr>
<td>dspace:available</td>
<td>2006-12-18T18:02:03Z</td>
</tr>
<tr>
<td>dspace:dateAccepted</td>
<td>2006-12-18T18:02:03Z</td>
</tr>
<tr>
<td>dspace:extent</td>
<td>1873709 bytes</td>
</tr>
<tr>
<td>dspace:identifier</td>
<td><a href="http://hdl.handle.net/1987/117">http://hdl.handle.net/1987/117</a></td>
</tr>
<tr>
<td>dspace:issued</td>
<td>2006</td>
</tr>
<tr>
<td>dspace:language</td>
<td>el</td>
</tr>
<tr>
<td>dspace:provenance</td>
<td>Submitted by Árhondas Árgoúrakí (<a href="mailto:argoun@physics.upatras.gr">argoun@physics.upatras.gr</a>) (2006-12-18T18:02:03Z) No. of Streams: 3 FileNames: pdf: 1 FileHash: 8895997e74ee8962c01c60dd0a8e989 (MD5) Previous Issue date: 2006</td>
</tr>
<tr>
<td>dspace:provenance</td>
<td>Made available in DSpace on 2006-12-18T18:02:03Z (GMT) FileNames: pdf: 1873709 bytes, Checksum: 8895997e74ee8962c01c60dd0a8e989 (MD5) Previous Issue date: 2006</td>
</tr>
<tr>
<td>dspace:publisher</td>
<td>Τμήμα Δημοσιογραφίας Πανεπιστημίου Πάτρας</td>
</tr>
<tr>
<td>dspace:subject</td>
<td>Μηχανική των Ψηφίσματος</td>
</tr>
<tr>
<td>dspace:title</td>
<td>Μηχανική των Ψηφίσματος</td>
</tr>
</tbody>
</table>

### Semantic Search

**DL Query:** dspace-out:item

**Results 1-20 of 41.**

- oai:repository.upatras.gr:1987/102
- oai:repository.upatras.gr:1987/103
- oai:repository.upatras.gr:1987/104
- oai:repository.upatras.gr:1987/111
- oai:repository.upatras.gr:1987/112
- oai:repository.upatras.gr:1987/113
- oai:repository.upatras.gr:1987/114
- oai:repository.upatras.gr:1987/115
- oai:repository.upatras.gr:1987/116
Semantic Navigation – Example


Please use this identifier to cite or link to this item: http://hdl.handle.net/1987/117

Choose Metadata Display Language (N/A for all): English •

Title: Μηχανική των Ρευστών

ISSUE DATE: 2006

Appears in Collections: Μετάδοση Γερμότητας - Μεταφορά Μόρια

Files in This Item:

<table>
<thead>
<tr>
<th>File</th>
<th>Description</th>
<th>Size</th>
<th>Format</th>
<th>View/Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>FHNotes.pdf</td>
<td></td>
<td>1.83 MB</td>
<td>Adobe PDF</td>
<td>View/Open</td>
</tr>
</tbody>
</table>

Show full item record
Recommend this item
Possible Improvements

- A more user-friendly interface
  - Provision of a fixed list of common queries expressed in physical human language
  - Facility that will guide (non-familiar with OWL) users in creating queries in Manchester OWL syntax

- Integration with controlled vocabularies/thesauri
  - Expressed in SKOS (OWL)
  - Extend semantic search to include controlled vocabulary/thesaurus concepts
  - Augment subject search

- *Federated* Semantic Search
Conclusions

- Semantics for DSpace metadata
  - Knowledge discovery \((high, OWL 2 expressivity)\)
  - Automatic model population
    - Alleviate the “bootstrapping” problem
- Novel Semantic Services
  - Augment traditional search and navigation
  - Intelligent retrieval and discovery
  - “Plug-in” philosophy
- Interoperable design
  - Easy to integrate in any digital repository (OAI-PMH)
  - Straightforward integration with other schemata (e.g. LOM)
  - Semantic interoperability
Questions? Comments?

Dr. Dimitrios Koutsomitropoulos
kotsomit@hpclab.ceid.upatras.gr
http://www.hpclab.ceid.upatras.gr

University of Patras Institutional Repository:
http://repository.upatras.gr/dspace

Semantic Search for DSpace:
http://www.dspace.org/add-ons-and-extensions/addons/
(code available)