Role of Ontologies in Semantic Digital Libraries

Sebastian Ryszard Kruk,
Bernhard Haslhofer, Piotr Piotrowski, Adam Westerski, Tomasz Woroniecki

Sebastian Ryszard Kruk
sebastian.kruk@deri.org

NKOS Workshop, 2006
21/09/2006
Alicante, Spain
Takeaway home message

• Ontologies play the key role in the semantic digital libraries
• We need ontologies supporting the main aspects of contemporary knowledge repositories:
  – Bibliographic descriptions
  – Extensible structure of resources
  – Community-aware features
Presentation overview

- Motivations
- What is a semantic digital library?
- Bibliographic Ontology
- Structure Description Ontology
- Community-aware Ontology
- Ontologies in JeromeDL
- Future – Mash-up Digital Libraries
- Conclusions
Motivations

• Networks of digital libraries linger for more expressive interoperability solutions than existing ones
• Community-aware solutions change the face of the Internet as we knew it digital libraries should be a part of these changes
• Multimedia play bigger and bigger role on the Internet, while there is a need for accessible and adaptive access solutions
What is a Semantic Digital Library?

Semantic digital libraries

- **integrate** information based on different metadata, e.g.: resources, user profiles, bookmarks, taxonomies
- provide **interoperability** with other systems (not only digital libraries) on either metadata or communication level or both
- delivering more robust, **user friendly** and adaptable **search and browsing interfaces** empowered by semantics
Metadata is the key concept

- the Web does not have metadata
  - the idea of a Semantic Web is nice but difficult to implement
- many digital libraries do have metadata in place

RDF:
- is a framework to model any kind of metadata
- it delivers certain level of technical interoperability
Application Areas for Semantic Web Technologies

• Thesauri & Controlled Vocabularies
  – qualified DublinCore
  – DMOz, DDC-based taxonomies
  – SKOS, WordNet and other thesauri

• Schema Mappings / Crosswalks
  – MarcOnt Ontology – aims to cover concepts from MARC21, BibTeX and DublinCore
  – MarcOnt Mediation Services – an open mediation framework between common legacy metadata standards

• Metadata Integration
  – RDF as a common data model for integrating metadata from various autonomous and heterogeneous data sources
  – OWL for modeling the data source’s semantics
  – SPARQL as a common query language
Semantic DL as Evolving Knowledge Space

• In state-of-the-art digital libraries users are **consumers**: Retrieve contents based on available bibliographic records

• Recent trends: user communities
  – Connetea, del.icio.us
  – Flickr

• In Semantic digital libraries users are **contributors** as well
  – Tagging (Web 2.0), Annotations
  – Social Semantic Collaborative Filtering

• Semantic Digital libraries enforce the transition from a static information to a **dynamic (collaborative) knowledge space**
Bibliographic Ontologies

• Build to capture the semantics of the legacy metadata

• Examples of bibliographic ontologies:
  – MarcOnt ontology aiming at capturing concepts from MARC21 and BibTeX
  – RDF Schema for FRBR
MarcOnt Ontology – Main Concepts

Material

Marc21Description

Domain

foaf:Agent

wnet:LexicalConcept

has...

hasDomain

hasCreator

hasKeyword
Generating various bibliographic descriptions

- All resources are described in MarcOnt Ontology, but user can access MARC21, BibTeX and DublinCore descriptions generated on the fly.
Mediation service used during searching

- User can select from wide range of description properties, defined in different metadata, during query building

DublinCore

BibTeX

MARC21

• User can select from wide range of description properties, defined in different metadata, during query building.
FRBR and Bibliographic Ontology

- FRBR was published in 1998 -> does not address virtual electronic resources except as downloadable copies of documents

- conceptual model based on entities-attributed-relationships

```
Work
  is realized through
  Expression
    is embodied in
    Manifestation
      is exemplified by
      Item
```

Group 1
• Semantic Web is based on “entity” metadata for resources (documents, people, concepts, etc.)
• FRBR and RDF
  – RDF Schema: http://vocab.org/frbr/core
  – 13 distinctive classes
  – 48 properties (most of them with coupled with their inverse counterparts)
Structure Ontology

- Describing structure of resources with RDF allows flexible content model
- Library resource can be decomposed into sub-resources
- Each part of the resource can be later additionally annotated to deliver:
  - Accessibility features (using e.g. WAI ontology)
  - Adaptive hypermedia (with an appropriate user client)
  - Rendering to different platforms
- Library resource can be:
  - Easily extended with new type of content
  - Versioned and internationalized
  - Decomposed to deliver fine-grained access control
Example of the structure description

Resource

- pdf
  - digitalType
  - abstract
  - contentType

- article

- chapter
- attachment
- content

- some description
- content

- position

- 1
- 2
Community-aware Ontology

- Bibliographic descriptions were always too complex for average user
- Seems that Semantic Web is suffering similar problems with understanding by larger group of users
- The notion of community-based computing (so called Web 2.0) gains larger and large group of users
Community Tagging

- One of the key aspects of Web 2.0 apart from collaboration is the simplicity of descriptions (so called tagging)

- What do people tag:
  - Resources (URLs): del.icio.us, connotea.org
  - Photos: flickr.com
  - Events: upcoming.org

- How do people tag:
  - Free tagging – any keyword goes
  - Controlled/suggested vocabulary - based on established folksonomy
  - Geo-tagging – drag&drop (Flickr Maps), GPS info (Google Maps)
Folksonomies

- A tag itself has no meaning
- A tag within a context of other tags or some actions/states has meaning
- We can come up with an ontology based on:
  - Groups of tags users use
  - Keywords in users' queries
  - Explicitly defined groups of tags (e.g. bundles in del.icio.us)
- Folksonomies are reverse-engineered ontologies of users' tagging/querying actions
What is Social Semantic Collaborative Filtering?

- Goal: to enhance individual bookmarks with shared knowledge within a community
- Users annotate catalogues of bookmarks with semantic information taken from taxonomies or thesauri
- Catalogs can include (transclusion) friend's catalogues
- Access to catalogues can be restricted with social networking-based polices
- SSCF delivers:
  - Community-oriented, semantically-rich taxonomies
  - Information about a user's interest
  - Flows of expertise from the domain expert
Example of Social Semantic Collaborative Filtering

foaf:knows

xfoaf:include

xfoaf:bookmark

Alice

Bibliographic Ontologies Mediation

Bob

Artificial Intelligence

Caroline

Digital Libraries

Distributed Systems

Damian

Libraries

Eric

P2P Systems

Semantic Web
Social Semantic Collaborative Filtering
JeromeDL – Social Semantic DL

- Digital Library build with semantics and communities in mind
- Build to reflect requirements of:
  - Librarians
  - Researchers
  - Average users
- Ultimate goal – accessiblity achieved through
  - Interface design
  - Search and browsing technologies
  - In-depth internationalization effort
Ontologies in JeromeDL

- **Structure (system administrators):**
  - JeromeDL structure ontology

- **Bibliographic and legacy descriptions (domain experts and expert users):**
  - MarcOnt bibliographic ontology
  - Extensible MarcOnt suggestions

- **Communities (normal users, expert users with restricted vocabulary):**
  - FOAF and FOAFRealm identity management ontology
  - Social semantic collaborative filtering (SSCF) ontology
  - Semantical Interlinked Online Communities (SIOC) ontology
Ontologies in JeromeDL

FOAF

Alice

0.8

P2P systems

Annotation

foaf:annote

foaf:eval

foaf:interests

Libraries

foaf:in

foaf:includers

Alice[2,3]

FOAFRealm

Very Interesting

JeromeDL

Resource

text/rtf

jeromedl:binaryType

JeromeDL

Sebastian
Kruk

JeremyDL

dc:title

dc:creator

DublinCore

DESA 2005
Proceedings

arx:h:book

BibTeX

Springer
Verlag

MARC21

Chapter_1

jeromedl:contains

marc21:260b

Chapter_2

jeromedl:isAfter

MarcOnt

MarcOnt Mediation Service

Computer Science

Digital Libraries

Marcont:category

Marcont:keyword

Making Semantic Web real.
JeromeDL – Delivering Semantic Content

- Providing semantic annotations during uploading process:
  - open module for handling any taxonomies
  - keywords based on WordNet and free tagging
  - defining structure of resources in the JeromeDL ontology
- Lifting legacy metadata to MarcOnt ontology
- Community maintained annotations
  - social semantic collaborative filtering
  - semantic descriptions based on the FOAF metadata
JeromeDL – Semantic Information In Use

- Keyword-based search with semantic query expansion
- Semantic search:
  - Direct RDF querying
  - Natural language templates
- Social Semantic Collaborative Filtering
- Heterogeneous communication:
  - Bibster
  - A9
  - OAI
Mash-Up Digital Libraries

• Business world aims towards SOA – to easily integrate, choreograph and orchestrate existing services
• Users tend to mash-up various Web 2.0 services to deliver solutions answering their needs
• Community-aware semantic digital libraries can easily become one of the mashed-up services
• But the real challenge is to build mash-up features directly into the digital library to provide users with completely new experience of browsing beyond the resources of DLs.
Conclusions

• Ontologies play a key role in the evolution of digital libraries
• Current streams of Semantic Web and Web 2.0 should and can be combined into the development of digital libraries
• We have identified 3 applications areas for ontologies: legacy, structure and community-aware descriptions
• However the future of DLs seems to lay beyond integration of information, reaching towards integration with other services
Gracias
[thank you]

Questions & Answers

Sebastian Ryszard Kruk
sebastian.kruk@deri.org
http://www.corrib.org/
DERI, NUI Galway, Ireland