

RDF APPLICATIONS IN DIGITAL LIBRARY

MARCELLO BAX

bax@ufmg.br

Escola de Ciência da Informação
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - Campus UFMG
CEP: 31270-010, Belo Horizonte - MG - Brasil

Despite the dissemination of user-friendly search engines which allow for the retrieval of scientific information in computer networks, researchers do not always obtain expected results in a simple and efficient manner. There is increasingly evidence that the scientific community, in particular Social and Humanities researchers, finds it difficult to accurately retrieve their information requirements through the Web. Some of the difficulties are magnified by the fact that the Web organization is not based on commonly adopted standards. Therefore one needs to gather, organize and better disseminate the freely available scientific information on the Web through the use of open standards in digital libraries. Resource Description Framework (RDF) is one important standard from the World Wide Web Consortium (W3C). In our talk we introduce some of the standards aiming to better understand how digital library could benefits from them.

The number of potential research resources (namely articles, reports, presentations, course notes, etc.) significantly grows on a daily basis, however in a quite poor organization format given the Web's inherent characteristics. As a consequence, researchers are exposed to informational overload and are required to spend considerable amounts of time to eventually retrieve useful information. This leads to a non-efficient deployment of time and efforts. This issue has motivated many research groups to investigate and to propose tools that aim to assist the scientific communities in (i) coordination of research efforts; (ii) improvement in overall satisfaction in retrieving their information needs; (iii) avoiding the costs related to poor process performance; (iv) improving the investment return in science and technology programs. In this context, we propose that digital library research in the semantic web context indeed provides significant contribution towards a more efficient organization of scientific contents in the Web.

In the article "A Framework for Digital Library Research", D-Lib Magazine, December 2002, one finds the theme we want to explore here: "Digital libraries need semantic structure". The autor says: "To convey meaning to users to support search and retrieval, to provide knowledge-based support in the user interface, and to support agents that perform work for the user, a DL needs semantic structure, an ontology both in the broad sense of a conceptual schema for a domain (which includes metadata standards) and in the more narrow sense of a classification of subjects and values of other entity types." (Soergel, 2002).

In fact, the state of art in Knowledge Organization already makes it possible to use sets of shared metadata (formally named ontologies) in annotation of science resources. Nevertheless, one should note three important challenges in the effort of building a semantic web of scientific resources, to mention a feel:

(1) how to build and maintain extensible ontologies which would allow for the organization of scientific knowledge in different areas (with similar effect as the use of Decimal Codification in libraries);

(2) how to make as convenient and transparent as possible for an end user to mark-up its resources with a view to make them available to the community; and

(3) how to build infrastructure tools capable to integrate the diverse semantic webs that may be suggested, as well as tools to allow for the navigation in these nets.

In a semantic-based distributed digital library, each user would mark-up its resources and make them available on the net. Such resources would be automatically integrated by software agents that would be able to understand the expressed ontologies in Resource Description Framework (RDF), forming in this manner an extensive web of decentralized knowledge, however organized in a standard format.

This would allow for dissemination of selective information and other new services (yet to be proposed/designed) that could together overcome the information overload problem and the difficulties related to keeping up-to-date with the fast growing knowledge in any research area.

Important to emphasize that, in our opinion, an overall approach should not be building a big centralized digital library or link server, but an extensible federated digital library, composed of peers agents. Thus, each member of the community would mark-up its resources and put them on the net. Such resources would be, after that, integrated automatically for “conscientious” agents of software that could understand the expressed ontologies in RDF, forming then a great net of knowledge decentralized, however organized. Then, we would see selective information dissemination and other innovative services yet to appear, that could together overcome the information overload problem and the problem of keeping up-to-date with the fast growing knowledge in a research field.

The Resource Description Framework (RDF) is a language for representing information about resources in the Web. It is intended for representing metadata about Web resources, such as the title, author, and modification date of a Web page, copyright and licensing information about a Web document, or the availability schedule for some shared resource. However, by generalizing the concept of a “Web resource”, RDF can also be used to represent information about things that can be identified on the Web, even when they cannot be directly retrieved on the Web. Examples include information about items available from on-line shopping facilities (e.g., information about specifications, prices, and availability), or the description of a Web user’s preferences for information delivery.

RDF Applications in Digital Library Field

DC metadata is used to identify, describe, or locate information resources, whether these resources are physical or electronic. While structured metadata processed by computers is relatively new, the basic concept of metadata has been used for many years in helping manage and use large collections of information. Library card catalogs are a familiar example of such metadata.

The Dublin Core is a set of “elements” (properties) for describing documents. The element set was originally developed at the March 1995 Metadata Workshop in Dublin, Ohio. The Dublin Core has subsequently been modified on the basis of later Dublin Core Metadata workshops, and is currently maintained by the Dublin Core Metadata Initiative. The goal is to provide a minimal set of descriptive elements that facilitate the description and the automated indexing of document-like networked objects, in a manner similar to a library card catalog. The Dublin Core metadata set is intended to be suitable for use by resource discovery tools on the Internet, such as the “Webcrawlers”

employed by popular Web search engines. In addition, the DC is meant to be sufficiently simple to be understood and used by the wide range of authors and casual publishers who contribute information to the Internet. Dublin Core elements have become widely used in documenting Internet resources. The current elements of the DC contain definitions for the following properties:

Title: A name given to the resource.

Creator: An entity primarily responsible for making the content of the resource.

Subject: The topic of the content of the resource.

Description: An account of the content of the resource.

Publisher: An entity responsible for making the resource available.

Etc...

Information using the DC elements may be represented in any suitable language (e.g., in HTML meta elements). However, RDF is an ideal representation for DC information. The examples below represent the simple description of a Web Page in RDF using the DC vocabulary.

A Web Page Described using Dublin Core Properties

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://www.dlib.org">
<dc:title>D-Lib Program - Research in Digital Libraries</dc:title>
<dc:description>The D-Lib program supports the community of people
with research interests in digital libraries and electronic
publishing.</dc:description>
<dc:publisher>Corporation For National Research Initiatives</dc:publisher>
<dc:date>1995-01-07</dc:date>
<dc:subject>
<rdf:Bag>
<rdf:li>Research; statistical methods</rdf:li>
<rdf:li>Education, research, related topics</rdf:li>
<rdf:li>Library use Studies</rdf:li>
</rdf:Bag>
</dc:subject>
<dc:type>World Wide Web Home Page</dc:type>
<dc:format>text/html</dc:format>
<dc:language>en</dc:language>
</rdf:Description>
</rdf:RDF>
```

In our talk we explain the description on the above example aiming to clarify how digital library could benefits from RDF resource description.

References

(Soergel, 2002) Dagobert Soergel. A Framework for Digital Library Research - Broadening the Vision, in D-Lib Magazine. V8, N12.