International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:12, 2012

A Semantic Assistant Agent for Digital Libraries

Mohamed Kholif, and Suaad Awad Mohamed

Abstract**—**In this paper we present semantic assistant agent (SAA), an open source digital library agent which takes user query for finding information in the digital library and takes resources' metadata and stores it semantically. SAA uses Semantic Web to improve browsing and searching for resources in digital library. All metadata stored in the library are available in RDF format for querying and processing by SemanSreach which is a part of SAA architecture. The architecture includes a generic RDF-based model that represents relationships among objects and their components. Queries against these relationships are supported by an RDF triple store.

Keywords**—**Digital Library, Dublin Core, Ontology, Semantic web.

I. Introduction

THE current web contains billions of documents and has many administrative problems and limitations; in addition to that the web content is still accessible only to humans. The solution to these problems is the Semantic Web. The Semantic Web is considered to be the extension for the current web. Digital libraries provide high quality and well-organised information but this data is often not connected with other pools on information such as personal homepages of authors. Within the Semantic Web, connecting the Semantic meaning of the pieces of information is the challenge.

The Semantic Web initiative of the World Wide Web Consortium (W3C) has been active for the last few years and has attracted users' interests. The main task of the Semantic Web is to add a meaning to the traditional web. In order to achieve that objective, layers of representational structures are needed [1].

A digital library is a library in which collections are stored in digital formats and accessible by computers [2]. In classical library interfaces the user mainly searches for title, author and keywords. With an electronic interface this is done with one text entry field. Results are presented in a short form, which could be extended to a longer description of the resource.

A large digital library may have many repositories of various types, including modern repositories, legacy databases, and Web servers. In digital libraries the user is more interested in the resource, which is available electronically or has at least the full description of the resource online.

Semantic Digital Libraries is a digital library system that applies semantic technologies to provide machine-support for the end users in their search for information [3]. Users in semantic digital library would be able to log in and create profiles that would translate to FOAF [4] files and user ontologies to keep track of semantic preferences. The search and retrieval functions would benefit from the ability to search semantically along with the traditional search methods the library currently employs. Semantic Web technology essentially teaches the computer the relationships between digital objects, their meaning, and how users would like to interact with them.

The paper is organized as follows. Section II gives a brief introduction about the use of Semantic Web in Digital Libraries and how semantically describe resources, Section III and IV introduces our proposed agent (SAA) including system architecture, storing digital objects and resource retrieval process. Finally section V conclusions and future work.

II. Semantic Web and Digital Libraries

The concept of ontology introduced by the Semantic Web is a promising path to extend Digital Library formalisms with the meaningful annotations [5]. Semantic Digital Library systems are building upon research on digital libraries, semantic web, social networking and human computer interaction. Here we will illustrate some famous digital libraries.

Talia, an innovative distributed semantic digital library, which is specifically, designed for meet the needs of scientific research in the Humanities. Talia is based on semantic web technology; it's being developed with the Ruby on Rails web framework and uses ontologies for the organization of knowledge, which can help the definition of a highly adaptable and state-of-the-art research and publishing environment [6].

JeromeDL social semantic digital library is an open source digital library system which takes users' profiles into account as one of the semantically enriched information sources to expand the queries. Semantic descriptions in JeromeDL are based on relations between different types of objects. Bibliographic information is enhanced by referencing entities defined with controlled vocabularies. The JeromeDL resource management component binds the bibliographic description component (MarcOnt Mediation Service) and manages the community resources (FOAFRealm component). On a more

Mohamed Kholif is a College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt.

Suaad Awad Mohamed is in Mathematics and Computer Science department, Faculty of Science, Alexandria University, Alexandria, Egypt.

library specific level there is MarcOnt, an ontology which mediates between the different legacy descriptions formats of the resources, e.g. MARC21, Dublin Core and BibTeX [7].

JeromeDL is the application most similar to Talia but the main difference lies in their primary target user group. JeromeDL's primary target audience is the generic user of a digital library while Talia will also include default user interfaces and tools that are targeted to Humanities Scholars.

A. Semantic Description of Resources in Digital Libraries

Text values are not enough to support machine based reasoning. To perform more intelligent interactions with readers, the knowledge base must be equipped with semantics.

Metadata is a key component of the provision of online catalogues that are searchable across the Web. In order to use the Semantic Web to its best effect, metadata needs to be published in RDF [8] formats. There are several initiatives involved with defining metadata standards in the library and publishing community such as Dublin Core Metadata Initiative [9] which provides a standard set of machine readable fields and guidelines for their use which is used in our agent. This now has a well-established RDF vocabulary. Such standards can be used across the Web in that they provide a common metadata vocabulary in XML or RDF which can be used to mark up and share library catalogues on the Web. PRISM and Dublin Core are usable now in the Semantic Web [10].

DCMI Metadata Ontology [11] has been modified by adding some resources descriptions.

This ontology contains all metadata terms maintained by the Dublin Core Metadata Initiative [9] including the fifteen terms of the Dublin Core Metadata Element Set, which have also been published as IETF RFC 5013 [12], ANSI/NISO Standard Z39.85-2007 [13], and ISO Standard 15836:2009 [14].

SAA uses the modified DCMI ontology as the base ontology for describing resources. Fig. 1 presents the key components that are used to manage important information about a resource which are the Dublin Core common elements (creator- description- format- title…etc) and Fig. 2 presents an instance of the class Resource which is a book.

The main classes defined in the DCMI ontology are: Resource which represents resources described by this ontology, Author (defines a class of creators of resources) and ResourceType (instances of this class are used to describe a resource type).

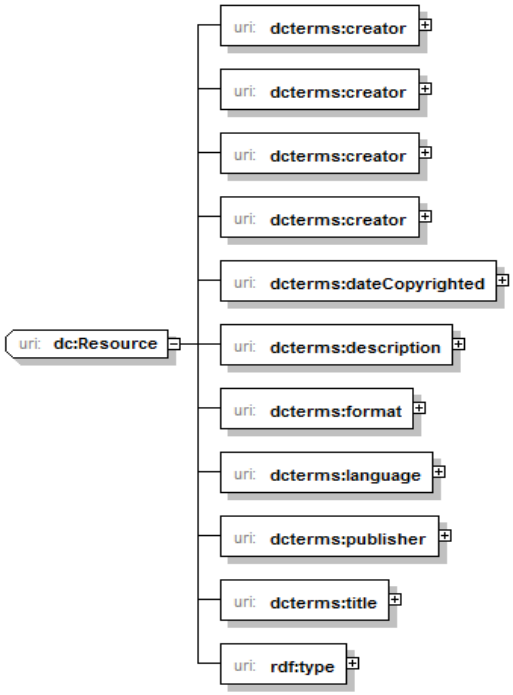

Fig. 1 Resource description components

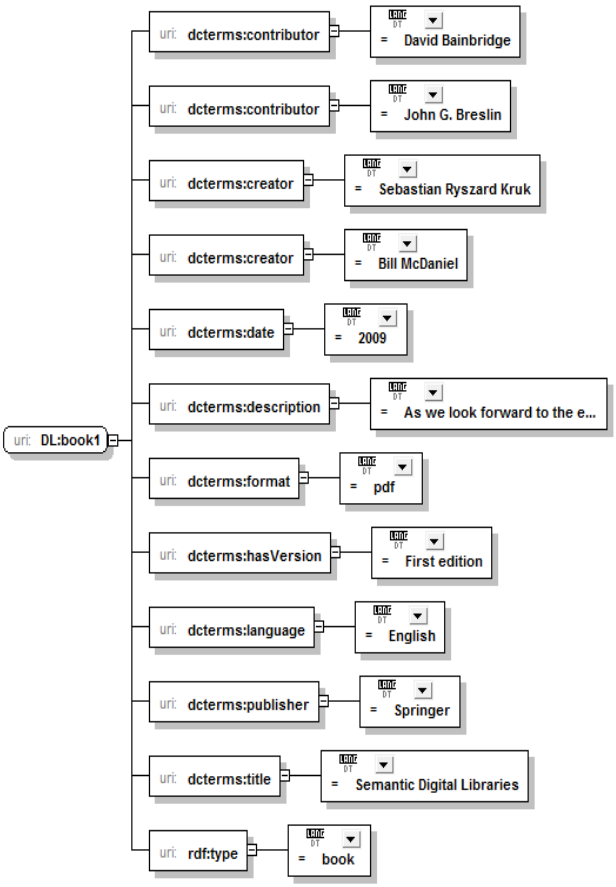

Fig. 2 An instance of the class Resource

III. Proposed Work

We have presented SAA, a full open-source digital library Semantic Assistant Agent, that exemplifies how digital libraries benefit of semantic technologies. SAA uses Semantic technologies to improve browsing and searching for resources in digital library to avoid keyword search problems discussed in [15]. All metadata stored in the library are available in RDF format. Querying the library is performed by SemanSreach (part of SAA architecture). The architecture includes a dynamic RDF-based model that represents relationships among objects and their components which updated frequently using user input metadata or importing XML metadata files.

IV. System Architecture

The key components of the agent are shown in Fig. 3. Repository store manage digital objects and other information. The repository store stores both the content and the metadata of the digital objects. Searching the system is divided into two parts; standard search and advanced search. Fig. 4 shows a detailed architecture of the agent.

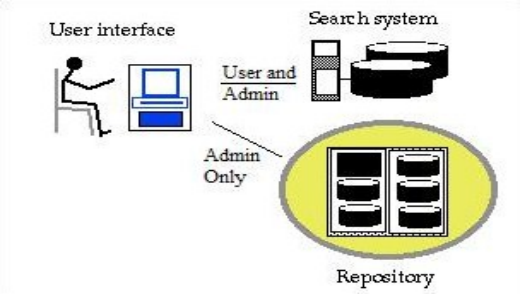

Fig. 3 The system key components

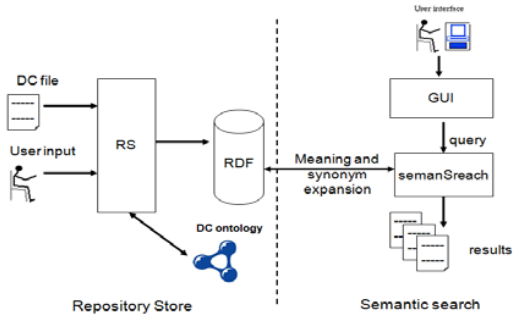

Fig. 4 Detailed architecture of the agent

A. Storing Digital Library Objects' Metadata Semantically (RS)

There are two ways to add metadata for a digital object:

1. Using user interface: librarians add the metadata using GUI with the available information about the objects. Our agent takes these data and stores it in RDF model (See Fig. 5).

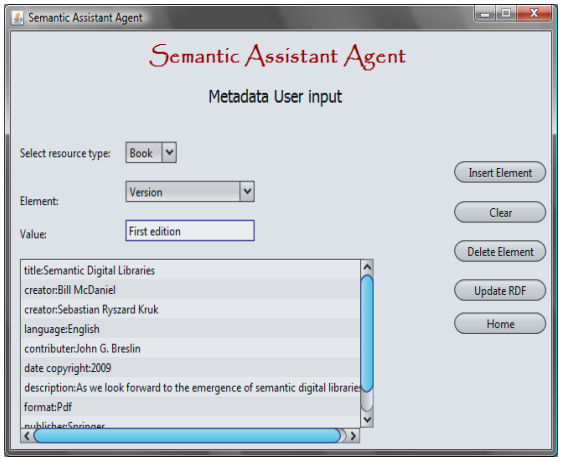

Fig. 5 Snapshot of user input metadata form

2. Using Dublin Core (DC) XML-file:
   Our agent converts the Dublin Core metadata file to RDF format then appends it to the RDF model of the system.

B. Semantically Enhanced Resources Retrieval in Digital Library (SemanSreach)

To find resources, a search over the complete content of the library is possible. The search algorithm of SAA (see Fig. 6) consists of two type of search; standard search and advanced search.

```
procedure SEMANTIC SEARCH ( File RDFfile ,  String query)
model.read(RDFfile, null);
queryString = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-
syntax-ns#> "
            + "PREFIX dcterms:<http://purl.org/dc/terms/> "
            + "SELECT    ?title ?type"
            + " WHERE { ?person dcterms:title ?title .  ?person
rdf:type ?type.  ";
        + " { ?person dcterms:<criteria> ? <criteria>   FILTER
regex(?<criteria>, " <query>",i ) } }";
Query query1 = QueryFactory.create(queryString);
QueryExecution qe = QueryExecutionFactory.create(query1,
model);
ResultSet results = qe.execSelect();
End procedure SEMANTIC SEARCH
```

Fig. 6 Searching algorithm

In Standard search, user enters search query, selects the criteria and then start the search. The results will include all resources of any type that semantically match the query based on the semantic description of the resources including synonyms the keywords in the query. Fig. 7 shows the searching results for the query "Sebastian" with criteria "creator". The results are five resources; a book and four papers. Clicking on one of them displays more details about the resource.

In Advanced search, user enters search query and selects the criteria and select the type of the resource then start the search. Fig. 8 shows the searching results for the query "Sebastian"

with criteria “creator” and type is “Paper”. The results here are four which are papers. Clicking on one of them displays more details about the resource.

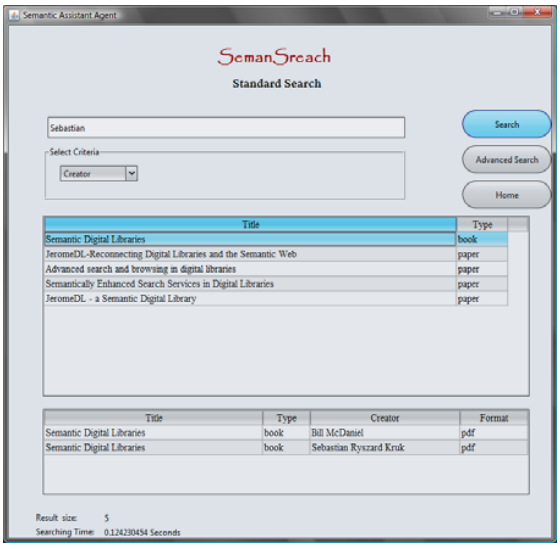

Fig. 7 Standard search results for search query “*Sebastian*”

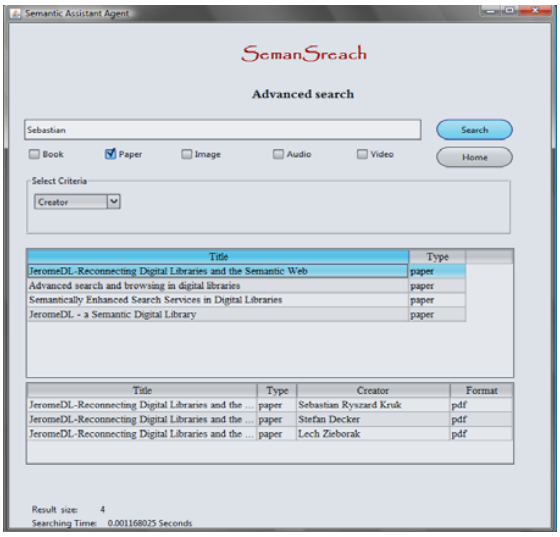

Fig. 8 Advanced search results for search query “Sebastian”

V. Conclusion and Future Work

Many of digital library problems may be solved with the help of Semantic Web technologies. In this paper we have presented SAA, a full open-source digital library Semantic Assistant Agent that exemplifies how digital libraries benefit from emerging semantic technologies. This paper also showed a quick glimpse of SAA semantic web features and demonstrated how semantic web development can be made a breeze by combining the power of digital library framework with a dynamic RDF store API.

The aim of the search algorithm presented in the previous section is to reflect the readers’ expectations and to reduce the time required to find the specified resources in a digital library. In order to test the semantic enabled search algorithm, the database of the prototype system has been filled with 60 resources. For testing, a set of 30 queries have been processed. SAA provides one of future work of Talia semantic digital library discussed in [16] which is a user-created metadata. More semantic web features will be included in future version, like semantic links between remote libraries and integration with online digital library.

References

[1] P. Rajiv, Dr. D. Sanjav, *Interoperability between semantic web layers: A communicating agent approach*, international journal of computer application, Vol. 12-no.3, pp.28-32, November 2010.
[2] D. Greenstein, S. Elizabeth, *The Digital Library: A Biography*, Digital Library Federation Council on Library and Information Resources 1755 Massachusetts Avenue, NW, Suite 500 Washington, ISBN 1933645180, 2002.
[3] S. Kruk, B. Haslhofer, P. Piotrowski, A. Westerski, and T. Woroniecki, *The Role of ontologies in semantic digital libraries*, In Proceedings of NKOS 2006 Workshop at ECDL2006, 2006.
[4] The Friend of a Friend (FOAF) project: www.foaf-project.org/ [last accessed: January, 2010].
[5] S. Kruk, S. Decker, and L. Zieborak, *JeromeDL: Adding Semantic Web Technologies to Digital Libraries*, In Proceedings of the 16th International conference on database and expert systems applications, Copenhagen, Denmark, 2005.
[6] M. Barbera, M. Nucci, D. Hahn, and C. Morbidon, *A Semantic Web Powered Distributed Digital Library System*, In ELPUB2008. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing held in Toronto, Canada 25-27 June 2008 / Edited by: Leslie Chan and Susanna Mornati, pp. 130-139, 2008.
[7] S. Kruk, K. Zimmermann, and B. Sapkota, *Semantically enhanced search services in digital libraries*, In Telecommunications, AICT-ICIW ’06, International Conference on Internet and Web Applications and Services, 2006.
[8] Resource Description Framework (RDF): http://www.w3.org/RDF ; [last accessed: June, 2011].
[9] Dublin Core Metadata Initiative: http://dublincore.org [last accessed: May, 2012].
[10] B. Matthews, *Semantic Web Technologies*, JISC Technology and Standard Watch, 2005.
[11] The Dublin Core Metadata Initiative ontology: http://dublincore.org/documents/2012/06/14/dcmi-terms/
[12] RFC Standards: http://www.ietf.org/rfc/rfc5013.txt
[13] NISO Standard: http://www.niso.org/standards/z39-85-2007/
[14] ISO Standards: http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on.
[15] G. Antoniou, F. Harmelen, A Semantic Web Primer, The MIT Press Cambridge, Massachusetts London, England, 2008.
[16] M. Nucci, D. Hahn, and M. Barbera, The Talia Library Platform - Rapidly Building a Digital Library on Rails, In SFSW2008, 4th Workshop on Scripting for the Semantic Web, European Semantic Web Conference 2008.