AMSL

Creating a Linked Data Infrastructure for Managing Electronic Resources in Libraries

Natanael Arndt[1], Sebastian Nuck[1], Andreas Nareike[1,2], Norman Radtke[1,2], Leander Seige[2], and Thomas Riechert[3]

[1] AKSW, Institute for Applied Informatics (InfAI) e.V.
Hainstraße 11, 04109 Leipzig, Germany
{lastname}@infai.org

[2] Leipzig University Library
Beethovenstr. 6, 04107 Leipzig, Germany
{lastname}@ub.uni-leipzig.de

[3] Leipzig University of Applied Science (HTWK)
Gustav-Freytag-Str. 42A, 04251 Leipzig, Germany
thomas.riechert@htwk-leipzig.de

1 Introduction

The library domain is currently undergoing changes relating to not only physical resources (e.g. books, journals, CDs/DVDs) but to the rapidly growing numbers of electronic resources (e.g. e-journals, e-books or databases) which are contained in their collections as well. Since those resources are not limited to physical items anymore and can be copied without loss of information, new licensing and lending models have been introduced. Current licensing models are pay-per-view, patron-driven-acquisition, short term loan and big deal. In addition, with changing markets, user expectations and publication formats, new models will be introduced in the future.

The existing infrastructure is not prepared for managing those new electronic resources, licensing and lending models. Software which is developed to meet the current requirements (cf. section 2) is likely to be outdated in a couple of years, due to the changing modeling requirements. To develop a future-proof managing system, a highly flexible data model and software which is adaptable to a changed data model is needed. The targeted user group are librarians with limited or no understanding of Semantic Web, Linked Data and RDF techniques.

We present AMSL [1], an *Electronic Resource Management System* which is based on the generic and collaborative RDF resource management system OntoWiki [2, 3]. Using RDF as data model makes the application flexible and thus future-proof, while the changing modeling requirements can be met by adjusting the used RDF vocabulary resp. application profile. To adapt the generic OntoWiki to the needs of the domain experts, we have added some components to support domain specific use cases, while still keeping them agnostic to

the used RDF vocabulary. The current status of the project including links to the source code and a live demonstration is available at the project web-page: http://amsl.technology/.

2 State of the Art

Software currently in use in libraries is mainly concerned with managing print resources, such as the *Integrated Library System* LIBERO (http://www.libero.com.au/), which is used at Leipzig University Library. But such software is not prepared for managing electronic resources with complex licensing and access models. Recently, some commercial vendors have started providing *Next Generation Library Systems* which constitute a new approach to Electronic Resource Management. Due to their commercial nature, it is difficult to evaluate how flexible the used data models are and which requirements are met [4, 5]. Since the data models are closed, they can not be extended by the customer as requirements change, whereas the contract and subscription terms often change with every new business year. Further, it is more difficult to integrate external knowledge bases in contrast to Linked Data and the Linked Open Data Cloud [6].

Another approach is to avoid specific library software but to use generic resource or document management systems, such as Wikis. OntoWiki was developed as a Wiki system for collaboratively creating semantic data in RDF and OWL knowledge bases. Over the time, OntoWiki has evolved towards a highly extensible development framework for knowledge intensive applications [3].

3 The AMSL Electronic Resource Management System

The AMSL Electronic Resource Management System is based on the OntoWiki Application Framework. It provides the basic functionality for managing resources i.e. creating, editing, querying, visualizing, linking and exporting RDF/OWL knowledge bases. With the Linked Data Server and Linked Data Wrapper/Importer components, it can publish and consume Linked Data according to the rules [6]. An Application Programming Interface allows the development of powerful third party extensions.

A core part of our application is to use an expressive **data model** as application profile which comprises different RDF and OWL vocabularies. The expressiveness of the data model helps to move design decisions from the program code to the easily adoptable vocabulary definitions, which increases the flexibility of the whole system. For expressing the data model we combine well known existing vocabularies, such as DCMI Metadata Terms (http://purl.org/dc/terms/), Dublin Core Metadata Element Set (http://purl.org/dc/elements/1.1/), The Bibliographic Ontology (http://purl.org/ontology/bibo/), Academic Institution Internal Structure Ontology (http://purl.org/vocab/aiiso/schema#) and The Friend of a Friend RDF vocabulary (http://xmlns.com/foaf/0.1/), to keep the data compatible to already existing components. With AMSL, we further introduce the *Vocabulary for Library ERM*

(BIBRM, `http://vocab.ub.uni-leipzig.de/bibrm/`). It provides terms for expressing licensing and access models and is aligned to the ideas of the Electronic Resource Management Initiative (ERMI) [4]. If new requirements arise in the future, the data model has to be changed and if necessary new vocabulary terms can be added.

We have developed **data templates** to provide a user interface for resource creation to the domain experts. Since our system does not require technical RDF knowledge of its users, the data templates provide a form based editing interface, which further supports the created resources to be compliant to the defined application profile. The template definition itself is expressed in RDF as well, to achieve the required extensibility, without need to change the software.

To support the work-flows for managing meta-data coming with electronic resources (such as contacts, contracts, packages, agreements and licenses), special **import and integration components** are developed. Publicly available Linked Data and SPARQL services, which have evolved in the library domain in the past years (e.g. title information, ISSN-history and authority files), are necessary for the electronic resource management and are imported using the existing Linked Data import process.

Being able to restore changes made on triples is one of the existing features of OntoWiki. In the AMSL project further requirements for reproducibility, consistency and increased clarity of resource changes where formulated. To meet these requirements the present **versioning** system was extended towards a ChangeSet ontology[4]. The versioning metadata is stored in a SQL database containing the ChangeSet elements. Hence, the underlying data model is now capable of expressing the same amount of information as a ChangeSet. Additions and removals concerning the same triple are aggregated to a change statement. Even though the versioning metadata is not stored in form of triples, the information could be queried and transferred into a ChangeSet knowledge base for further purposes. Moreover multiple retrieval capabilities for querying extended versioning information have been added in form of an OntoWiki extension.

The present search function of the OntoWiki is based on a conventional SPARQL search, using a `bif:contains` filter on labels. To improve search speed and provide additional features like a fuzzy search, the Elastic Search search engine was integrated as an OntoWiki extension. This **full-text search** makes use of a class based index structure. Hence, it provides a faster access to pre-indexed resources. The underlying index structure is built up by indexing classes of the knowledge base which contains information that need to be accessed frequently. That is, classes containing properties such as `dc:title` are more meaningful to be searched with a full-text search than classes that only include properties like e.g. `bibo:issn` which contain numerical sequences. Additionally to the autocompletion features search function, an extended search supports an enhanced fuzzy search which provides the possibility of restricting the result set to the previously defined classes and is more robust against typing mistakes.

[4] http://vocab.org/changeset/schema.html

4 Conclusion and Future Work

We demonstrate a novel approach to build up an electronic resource management system for libraries by using generic RDF resource management technology. The used generic components are extended and complemented by some domain adaptable components to provide a customized interface to domain experts. Consequently using Linked Data and RDF further enables and supports libraries to build up a Linked Data infrastructure for exchange of meta data across libraries as well as across institutions using the services provided by library.

5 Acknowledgments

The presented software system was developed in the AMSL project for developing an Electronic Resource Management System based on Linked Data technology (`http://amsl.technology/`). We want to thank our colleagues Lydia Unterdörfel, Carsten Krahl and Björn Muschall from Leipzig University Library, Jens Mittelbach from Saxon State and University Library Dresden (SLUB) and our fellows from Agile Knowledge Engineering and Semantic Web (AKSW) research group especially Henri Knochenhauer for their support, helpful comments and inspiring discussions. This work was supported by the European Union and Free State of Saxony by a grant from the European Regional Development Fund (ERDF) for the project number 100151134 (SAB index).

References

1. Nareike, A., Arndt, N., Radtke, N., Nuck, S., Seige, L., Riechert, T.: Amsl – managing electronic resources for libraries based on semantic web. In: Workshop on Data Management and Electronic Resource Management in Libraries (DERM 2014) : INFORMATIK 2014. (September 2014)
2. Heino, N., Dietzold, S., Martin, M., Auer, S.: Developing semantic web applications with the ontowiki framework. In Pellegrini, T., Auer, S., Tochtermann, K., Schaffert, S., eds.: Networked Knowledge - Networked Media. Volume 221 of Studies in Computational Intelligence. Springer, Berlin / Heidelberg (2009) 61–77
3. Frischmuth, P., Martin, M., Tramp, S., Riechert, T., Auer, S.: OntoWiki - An Authoring, Publication and Visualization Interface for the Data Web. Semantic Web Journal (2014)
4. Jewell, T.D., Anderson, I., Chandler, A., Farb, S.E., Parker, K., Riggio, A., Robertson, N.D.M.: Electronic resource management – report of the dlf erm initiative. Technical report, Digital Library Federation, Washington, D.C. (2004) `http://old.diglib.org/pubs/dlf102/`.
5. Jewell, T., Aipperspach, J., Anderson, I., England, D., Kasprowski, R., McQuillan, B., McGeary, T., Riggio, A.: Making good on the promise of erm: A standards and best practices discussion paper. Technical report, NISO ERM Data Standards and Best Practives Review Steering Committee, One North Charles Street, Suite 1905, Baltimore, MD 21201 (January 2012) `http://www.niso.org/apps/group_public/document.php?document_id=7946&wg_abbrev=ermreview`.
6. Berners-Lee, T.: Linked Data. Design issues, W3C (June 2009) `http://www.w3.org/DesignIssues/LinkedData.html`.