A Multilingual Metadata Schema Registry Based on RDF Schema

Mitsuharu Nagamori, Thomas Baker†, Tetsuo Sakaguchi
Shigeo Sugimoto, Koichi Tabata

University of Library and Information Science (ULIS)
GMD German National Research Center for Information Technology†

{nagamori, saka, sugimoto, tabata}@ulis.ac.jp, thomas.baker@gmd.de†

Abstract

This paper describes a multilingual metadata schema registry which has been developed at the University of Library and Information Science in Tsukuba, Japan. The registry currently has reference translations of Dublin Core Metadata Element Set written in 22 languages, DC Qualifiers and the DCMI Type Vocabulary written in English and Japanese, along with descriptive elements of the Nippon Cataloging Rules. The registry is realized using XML technologies; metadata schemas are expressed in the RDF Schema language, the RDF description is handled by a Java servlet using XSLT stylesheet. This paper shows the functions and the architecture of the registry.

Keywords: *Metadata Schema, Metadata Registry, RDF Schema, XSLT Stylesheet*

1 Introduction

Metadata schema registries (or simply, metadata registries) are widely recognized as an important tool not only to share information about metadata schemas but also to enhance reusability of metadata schemas. It is also known that metadata registries have important roles for metadata interoperability among communities speaking different languages and over time.

The ULIS Open Metadata Registry is a metadata schema registry which stores and provides metadata schemas written in multiple languages. The ULIS Open Metadata Registry provides reference descriptions of the basic 15 elements of Dublin Core Metadata Element Set (DCMES), Dublin Core qualifiers, and a controlled vocabulary for the Type element recommended by the Dublin Core Metadata Initiative (DCMI). As of June 2001, the 15 elements are expressed in 22 languages[1] and the qualifiers are expressed in English and Japanese. The registry also experimentally provides descriptive elements of Nippon Cataloging Rules[2].

The metadata schemas are given in the RDF Schema language. XML technologies, such as XSLT and DOM, are used to realize the registry. Unicode (UTF8) is used to encode reference descriptions in multiple languages in the current version, while local character encoding schemes have been used in our previous version. This paper briefly describes the technologies and functions of the ULIS Open Metadata Registry.

2 Background

Multilingual description of elements and qualifiers is recognized as an important issue for broader usage of Dublin Core in the global community on the Internet [1]. The ULIS multilingual metadata registry began as a prototype developed at AIT, Thailand by a research team led by Thomas Baker in 1998. The first ULIS multilingual registry was developed by the authors in 1999 based on the prior system. In this system, reference descriptions were encoded in RDF Schema language and the texts were encoded with local character sets, such as EUC-JP for Japanese, Big5 for Chinese, TIS for Thai, and so forth. In this prior system, a multilingual text browsing technology named MHTML was adopted to display multilingual texts on off-the-shelf WWW browsers. Since MHTML is an applet-based technology to display texts on an applet without using local fonts, users need no a-priori font loading to display multilingual texts on their Web browser[2]. In the latest MHTML-based registry, the characters in the texts are encoded in the ISO-2022-JP-2 standard, which is a character encoding scheme for multiple languages. Figure 1 shows a user interface to choose elements and languages, and Figure 2 shows a set of

[1] Arabic, Chinese, Taiwanese, Czech, Dutch, English, Finnish, French, German, Greek, Indonesian, Italian, Japanese, Korean, Maori, Norwegian, Polish, Portuguese, Spanish, Swedish, Thai,

https://doi.org/10.23106/dcmi.952106607

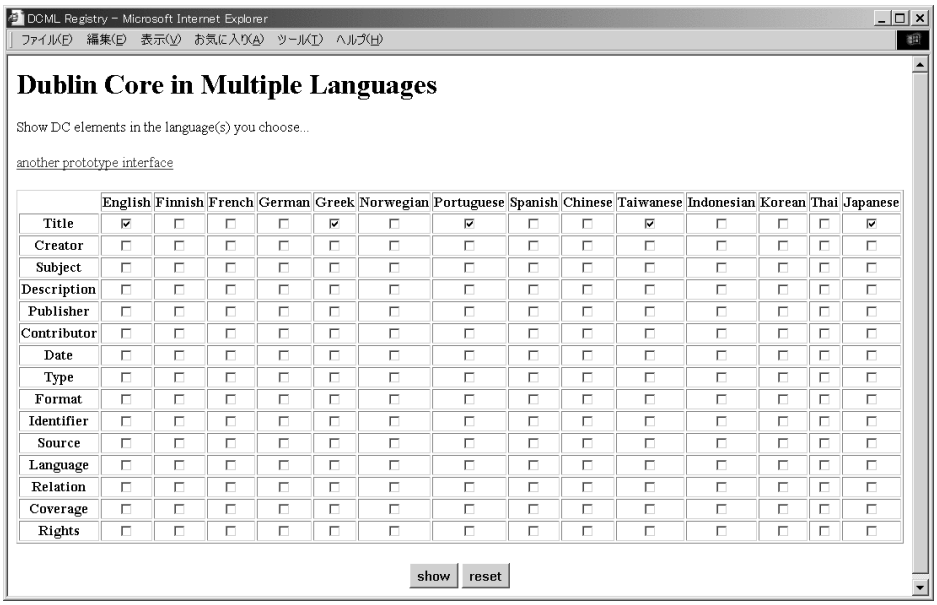

Figure 1. A Interface to Choose Elements and Languages

Figure 2. Reference Descriptions of "Title" in Portuguese, English and Greek

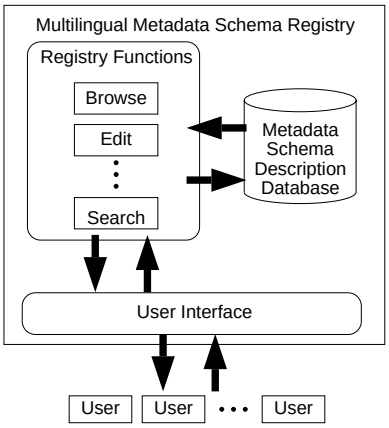

Figure 3. System Architecture of the Metadata Schema Registry

texts displayed on a browser.

3 ULIS Metadata Schema Registry - System Architecture and Examples

Figure 3 shows the overview of the system architecture. The system is composed of three major components. These components are (1) metadata schema description database, (2) a servlet which provides registry functions such as editing, browsing and searching, and (3) user interfaces defined in accordance with functions provided by the servlet.

Figure 4 shows English and Japanese reference descriptions of the Title element of DCMES encoded in the RDF Schema language. Figure 5 shows a browser interface which shows reference descriptions in Thai. Figures 6 and 7 show a search interface and search result, respectively. A word "title" is given for search in Figure 6. A list of descriptions which include "title" in its definition are displayed in Figure 7. Figure 8 shows an editing tool where metadata schema design-

Ukrainian

[2] Anglo-American Cataloging Rules' counterpart in Japan

https://doi.org/10.23106/dcmi.952106607

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<rdf:Peoperty xml:lang="en" rdf:ID="title">
  <rdfs:label>Title</rdfs:label>
  <rdfs:comment>A name given to the resource.
  </rdfs:comment>
</rdf:Property>
<rdf:Description xml:lang="ja"
  rdf:about="http://purl.org/dc/elements/1.1/#title">
  <rdfs:label>著者（あるいは作者）</rdfs:label>
  rdfscomment>情報交換の知的内容の創造に主たる責任を持つ人
  あるいは組織。たとえば、著述された文書の場合の著者、視覚的資料の
  場合の画家や写真家イラストレータ。</rdfs:comment>
</rdf:Description>
   :
</rdf:RDF>
```

Figure 4. English and Japanese Reference Descriptions of "Title" Element of DCMES

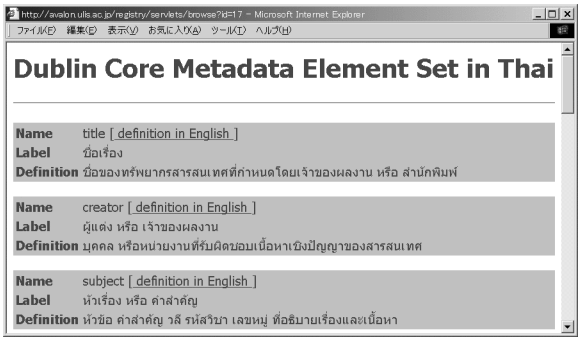

Figure 5. Browsing Metadata Schema: Dublin Core in Thai

ers can edit a reference description of a vocabulary, i.e. elements, qualifiers and other terms. The editor converts the text to/from RDF Schema description in the database.

The registry system has a generic architecture shown in Figure 9. It has a servlet that has basic functions such as edit, browse and search, which are defined independently of metadata schemas. Every HTML document that comprises the user interface of the registry is dynamically created from a user interface specification and a metadata schema description(s) using an XSLT stylesheet. The user interface specification is defined based on a metadata schema, and all metadata schema descriptions are given in the RDF Schema language. Each XSLT stylesheet is defined based on its corresponding registry function. Is is not necessary to modify the servlet in order to add new schemas to the registry. Thus, the registry is easily extended to realize new user interface in accordance with new metadata schemas and new user interfaces.

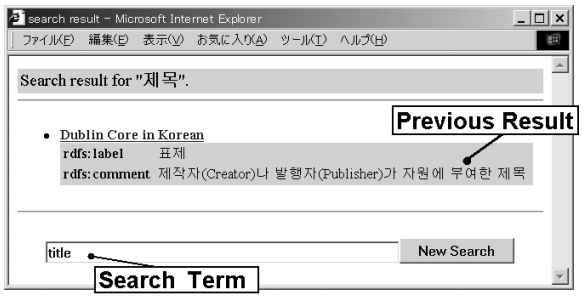

Figure 6. Search for "title"

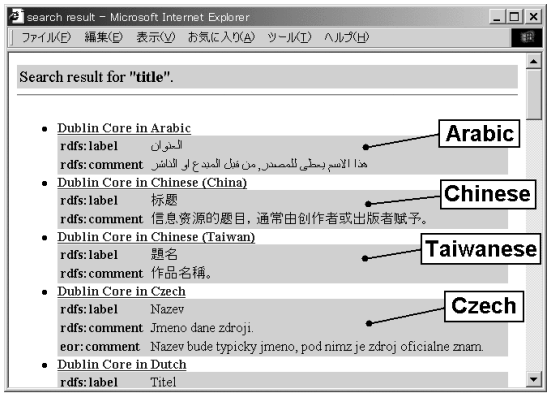

Figure 7. Search Result for "title"

4 Discussion and Future Extension

Metadata schema registries have been recognized as important for the development and maintenance of metadata vocabularies. UKOLN has developed a registry which has a crosswalk function [3]. One of the Max Planck Institutes has a German Metadata Registry [4]. The SCHEMAS Forum funded by the IST program of EU provides access to many metadata schemas [5]. The registry of the SCHEMAS Forum is built on top of using an open source software toolkit for metadata registries called EOR [6].

As of June 2001, the ULIS registry stores 27 entries in the schema database. The Search function is one of the crucial issues for the scalability of the registry. EOR, for example, uses MySQL for this purpose. However, a text search technology which can handle RDF Schema descriptions in multiple languages is needed. In this study, XPath is being used to realize the search function both because XPath is useful for realizing XML text string searching for multilingual texts encoded in UTF8 and because it is portable. Our next step is to use a database management software tool which accepts UTF8 as the basis for the data management.

The authors discussed the versioning issue for

https://doi.org/10.23106/dcmi.952106607

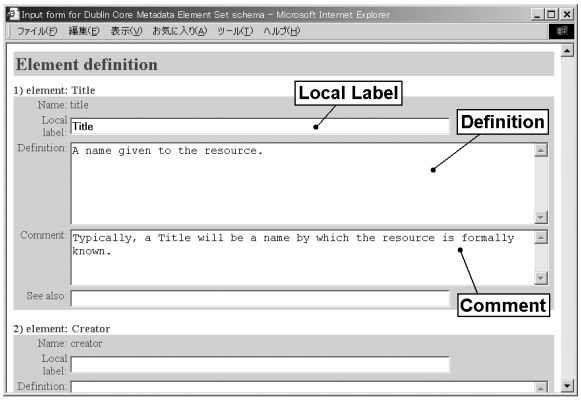

Figure 8. Input Form for Creating New Schema

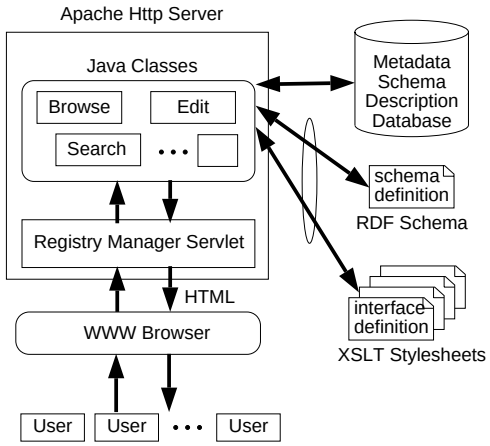

Figure 9. Registry Architecture

Dublin Core in our previous paper[7]. The primary role of metadata registries is to serve as an authoritative provider of metadata schemas over the long term. In addition, metadata registries are expected to support the development and customization of metadata vocabularies for various local- and/or domain-specific communities. Registries should provide software tools or services to enhance the re-usability and interoperability of metadata schemas among communities and to help a community develop metadata-base services in accordance with their local requirements.

At ULIS, the authors are developing a subject gateway for network resources which are useful for public library users and written in Asian languages. This is a collaborative activity with the Internet Public Library at University of Michigan [8]. As of June 2001, Chinese, Korean and Japanese resources have been collected. It uses a metadata schema based on IMS and Dublin Core. For the development of this subject gateway, we have noticed that the XML-based technology shown in Figure 9 is applicable for developing software tools to support the development of the service, and also subject gateways in general, even though there is a fundamental difference between them; i.e., a registry handles metadata schemas and a subject gateway does metadata instances. We are developing a support tool to create metadata-based application systems based on the XML-based technology developed in this study.

5 Conclusion

For the development of metadata schemas in the global network environment, we have to cope with the dilemma that we need globally approved metadata standard for global interoperability of metadata and, on the other hand, local and/or domain specific communities need metadata schemas which are well-suited to their requirements. The authors believe that metadata schema registry is a key component in the network environment to cope with this dilemma. In addition, long-term maintenance of metadata schemas, which include traceability of version and translation history, is also an important issue for metadata registries. These issues are left for further research.

acknowledgements: This research is supported by Grants-in-Aid for Scientific Research #13558039 in part.

References

[1] Baker, T., Dublin Core in Multiple Languages: Esperanto, Interlingua, or Pidgin?, Proceedings of the International Symposium on Research, Development and Practice in Digital Libraries 1997, 1997, pp. 8–15.

[2] Sugimoto, S. et al., Experimental Studies on an Applet-Based Document Viewer for Multilingual WWW Documents - Functional Extension of and Lessons Learned from Multilingual HTML, Proceedings of ECDL98 (LNCS 1513, Springer), pp 199-214, 1998.10

[3] DESIRE Metadata Registry. http://desire.ukoln.ac.uk/registry/index.php3

[4] German Metadata Registry, http://www.mpib-berlin.mpg.de/dok/metadata/gmr/gmr1e.htm

[5] SCHEMAS: Forum for Metadata Schema Designers and Implementers. http://www.schemas-forum.org/

[6] The EOR Toolkit. http://eor.dublincore.org/

[7] Sugimoto S., et al. Versioning the Dublin Core across Multiple Languages and over Time Proceedings of SAINT 2001 workshop IEEE pp.151-156 2001.1

[8] Internet Public Library, http://www.ipl.org/

https://doi.org/10.23106/dcmi.952106607