Library Resource Construction and Service Based on Linked Data

Qi ZHOU1,2, Gang WU1,2, Chao-qun WEN1,2, Chao ZHAO2
Fan ZHANG2 and Hai-tao WANG2,*

1School of Software Engineering, Beijing Jiaotong University, Beijing, China
2China National Institute of Standardization, Beijing, China
*Corresponding author

Keywords: Linked data, Resource description framework, Information retrieval.

Abstract. Linked Data is widely used in the library fields recent years by connecting related data using Web. Because of the semantical variousness of different datasets and resources, it’s still an obstacle to construct them by Linked Data and provide better service. In this paper, we propose a technique which can combine the data from the Internet and other Linked Data sets with RDF (Resource Description Framework), implement an integrated platform for searching information. Through experiments, the results are satisfying.

Introduction

Linked Data, proposed by Tim Berners-Lee based on semantic web [1], describes a possibility to publish, share and connect various resources in the Web, to lower the barriers of linking and reusing data. Linked Data has so many advantages, such as concise framework, standardization, self-support and so on, that make Linked Data been widely used in many large libraries including American Library of Congress, Cambridge University Library, British Library, etc.

Actually, most libraries just build their resources as RDF files without making good use of the information from Internet and Linked Data sets from other libraries, which could be great supplement to improve the quantity and quality of their own services. When concerning data expansion, data representation and data searching become to be more complicated.

In this paper, we propose a method which could combine information from Web and other libraries, and provide a unique interface for service.

The rest of the paper is structured as follows. In Sect. 1, we introduce the related work of the use of Linked Data in libraries; in Sect. 2, the main logic of our technique is explained in detail; in Sect. 3, we present the design of the system used to verify our technique; the section 4 is mainly to show the results of the experiments.

Related Work

For years, libraries have accumulated pretty much knowledge information which is reusable and valuable. Linked Data is to build link between data from the Internet, which can solve the problem of heterogeneous data and improve the semantic link between data, so libraries begin to integrate their data resource with Linked Data. The library of congress and National Libraries of Sweden have tried to use the Linked Data. Since 2010, the W3C organized the library Linked Data incubator group to help the library of congress build Linked Data. [2]Since their resource is so large and the work to arrange the resource is very hard to accomplish, so they mainly use Linked Data to integrate their own data resource and just implement the link in several parts.

In our country, the several bigger libraries began to use Linked Data to rebuild their own resource to service the public better, but there is no big progress, either.

The final goal of the Linked Data is to link the whole Internet to provide more complete information. Therefore, the present use of the Linked Data in the library field is far from enough. So in this paper, under the help of technique that is to use the Linked Data to integrate local resource
and external resource which may be from other libraries or the Internet, we design an information-searching service platform for libraries to address the problem about searching data resource incompletely.

Algorithm
The main algorithm of the technique we propose focuses on the construction of multiple resource. The data resource which may be local, from other libraries or from the Internet have different structure and it is hard to combine them. Thus, in this section, we introduce the logic of construction of multi-resource data.

The core technology of Linked Data is RDF (Resource Description Framework) which specially services for the purpose of expressing the metadata of the resource which can be labeled from the Internet. In this paper, we will use RDF to integrate data resource to give the readers more complete searching service.

The main logic of our algorithm is shown as the figure 1.

![Figure 1. Main logic route.](image)

Resource Construction
Resource construction includes three parts – reconstruction of the local data resource, construction of incremental data resource and construction of data resource from other libraries.

**Reconstruction of the Local Data Resource.** In the reconstruction of the local data resource, the first step is to transform the MARC (Machine Readable Catalogue) data into data form of the relational database. And the attributes of tables include the key information which can describe a book, such as the name of book, the writers, press and so on.

The second step is to build the RDF Linked Data. The concrete operation procedure is as follows:

1. Define the corresponding predicate properties of namespace after constructing the RDF by analyzing the column field of tables
2. Define the relationship between the tables to confirm of the link between RDF resources.
3. Build the mapping files. Before transforming the relational database into RDF resource, we should build one-to-one mapping between the database and RDF files by using the D2RQ.
4. Construct the RDF Linked Data files by using D2RQ.

**Construction of Incremental Data Resource.** In the future, the local data resource may be updated, it is necessary to transform the modified relational data or the new data into the independent RDF files. There are two ways to construct the incremental data resource, which are complete incremental data construction method and independent incremental data construction method.
The complete incremental data construction method means when the existed data is updated, we need to construct new RDF files for the updated data resource.

Sometimes, it may add large amount of independent data, and the new data is independent to the existed data, which means they will not influence accessing the existed data. In this case, we can use independent incremental data construction method, which is to construct RDF files for the independent data resource.

**Using Data Resources from Other Libraries.** In this part, we construct data resource from other libraries by downloading the published RDF files or remote accessing API. If some libraries have published RDF files, we can download the RDF files directly, if not, we can only remote access the API provided by other libraries to get the data we want.

**Data Expansion**

To enlarge improve the completeness of the data resource, we not only reconstruct and the local data resource but also expand the local data resource through the Internet. The goal of data expansion is to grab corresponding data from the Internet to expand the local data resource and combine the local data and the data from the Internet. There are three main step in data expansion. The first one is data crawling, and the second one is structuration and verification of the data, and the third one is combination the data.

In this paper, we mainly crawl the data from the main website according to the ISBN of books in order to expand our local data resource. The some of the data we gather from the Internet are unstructured. We save the data into the local database according to the tag of the data. For image resource we gathered from the Internet, by building the accessing rule of the system path, we can give each image one specific URL, and the URL is structured and can be saved in the local database.

However, the data we get are not completely right or usable, so we need to verify and complete the data to improve the reliability, validity and usability. We can check if the content of the attribute is appropriate based on their format. For the lack of the data, the solution is to calculate the average value and use the average to fill the lack.

After structuration and verification of data resource, we can combine the data with local data resource on the basis of the ISBN of books, because the ISBN is the unique ID for each book. If there is no corresponding ISBN information for the book, the attributes, which includes the name of the book, the writer’s name and the publication, can be taken as the key value to locate the book we need too. The next step after finding the book is to compare the data with information we have had and combine the information with the data which is got from the Internet.

**Information Searching and Pushing**

After the whole resource building, we can use the technology of word segmentation in order to provide more convenient searching service. In this part, we support multi-source searching, which means the users can choose the source to get the information they need.

**Design of Service Platform**

The library information- searching service platform is implemented based on B/S. Figure 2 is the framework of the system. The main logic is on the server. The JENA is an API of java, which is mainly used to support the application about semantic web and Linked Data.
The architecture of the system is shown as figure 3. As we have mentioned, the main logic is on the server. To reduce coupling, we divided the server into several models to process specific function.

The process of searching function in the server is shown as the figure 4. First we should reconstruct the local existed data, which means transform the MARC data into data of relational database. The second is building RDF files and linking LOD data. And then, we combine the data from the Internet and the existed data to provide the searching service.

Experiments
The following picture is the achievement of the system, and he title is traditional Chinese medicine Linked Data service platform. The users can input some key word to search books in the text editor and choose the source of the result.
Then, the platform provides the result list of searching.

We can check the detail information of one book by clicking one of the results.

The right of figure 7 is the recommendation list of the book. And the left is the detail of the book the user searched.
Acknowledgement

This paper is supported by grants from Ministry of Science and Technology of the People’s Republic of China (2016YFF0204205) and China National Institute of Standardization (522016Y-4681, 522016Y-4486, 2015QK241, and 522014Z-3748-1).

References


