A service of the National Diet Library, Japan, to the semantic web community

Oshiba Tadahiko
National Diet Library
Tokyo, Japan

Abstract:

The National Diet Library, Japan (NDL) launched the "Web NDLSH", the web version of the National Diet Library Subject Headings (NDLSH) on June 30, 2010. "Web NDLSH" is a service that provides Japanese subject headings as Linked Open Data and makes them accessible to the Semantic Web. This paper will provide an overview of the "Web NDLSH". The NDL is currently preparing a service to offer its name authority records in RDF/XML. This paper will also describe a blueprint of the new service which will be called the "Web NDL Authorities". The "Web NDL Authorities" will be released in January 2012.

Background and introduction

The National Diet Library Subject Headings (NDLSH), whose first edition dates back to 1964, is a Japanese controlled vocabulary list compiled and maintained by the National Diet Library as a subject access tool, and is also used for search by "subject heading" in the NDL's catalog.

The turning point was in 2004. The NDLSH had had many problems such as shortage of subject headings, lack of "see also" references, vague citation order and delay of release. It had not been easy to use it when searching the NDL's catalog. Then we launched the revision of the NDLSH in 2004 and started to solve the above problems. Major revisions compared with the previous editions were the following:

- Added new headings;
- Introduced "see also" references;
- Reviewed the way to apply subject subdivisions;
- Enhanced scope notes;
- Introduced class numbers of the Nippon Decimal Classification (NDC) newly revised 9th edition.

We applied a thesaurus structure (BT, NT and RT) for "see also" references introduced by the revision. We decided to use a fixed citation order: Topical term – Topical subdivision – Geographic subdivision – Chronological subdivision – Form subdivision (in some cases, a geographic term is used as main heading). Catalogers of the NDL have positively created subject headings since the revision (see Table 1).

Table 1: Increase in the number of newly established subject headings

Oct. 1998 – Sep. 1999	51
Oct. 1999 – Sep. 2000	58
--Revision 2004--	
Apr. 2008 - Mar. 2009	414
Apr. 2009 - Mar. 2010	817

The NDL holds a "Conference on Bibliographic Control" every year to discuss with related institutions matters related to bibliographic data production and provision, and to promote national bibliographic control and standardization of bibliographic services. The theme discussed in 2004 was "Subject Headings Now and in Future: Subject Access in Network Environment". The conclusion of this conference encouraged the revision of the NDLSH.

The NDLSH is now used as an important means of subject access, especially when searching the National Diet Library Online Public Access Catalog (NDL-OPAC). But there was one problem area: the NDL had formerly been providing the NDLSH in book form;

since 2005, in PDF and since 2006, in text format on the NDL website, but those were far from satisfactory for the Web environment. Then we embarked on a new project to provide our controlled vocabulary for the Semantic Web community.

The "Web NDLSH"

The "Web NDLSH" is the web version of the NDLSH. When we started to develop it in 2009, we decided to use the SKOS (Simple Knowledge Organization System) for the "Web NDLSH" to make our subject headings easier to use in the Web environment, aiming to have it applied to various applications or systems on the Web. The SKOS data model adopted in the Web NDLSH is suitable for expressing controlled vocabularies such as subject headings or classifications within the framework of the Semantic Web, and has already been used for providing subject headings in other countries, as in the Library of Congress Subject Headings (LCSH) provided by the Library of Congress Authorities and Vocabularies.

We launched the "Web NDLSH" on June 30, 2010 (see Figure 1).

セマンティックウェブ	
件名標目詳細	グラフィカル表示
ID	01017771
件名標目 (xl:prefLabel)	セマンティックウェブ
同義語 (xl:altLabel)	セマンティクウェブ; Semantic Web [LCSH]
上位語 (skos:broader)	インターネット
関連語 (skos:related)	オントロジー(情報科学); メタデータ; 情報検索
分類記号 (skos:relatedMatch)	547.483 (NDC9); 694.5 (NDC9); DK341 (NDLC); ND633 (NDLC)
出典 (dct:source)	000007786457 (書誌ID); goo辞書(200511)
作成日 (dct:issued)	2005-11-16
最終更新日 (dct:modified)	2005-11-28T14:34:25
参考	セマンティック・ウェブ (Wikipedia)
他形式のデータ	RDF/XML形式, RDF/Turtle形式, JSON形式

Figure 1: The "Web NDLSH" result set for the search term "Semantic Web" (in Japanese).

The "Web NDLSH", providing about 90,000 Japanese subject headings as Linked Open Data, allows the following services:

・Reference function to the URIs of subject headings by assigning a URI to each of them;

・ Download of each datum in three formats: RDF/XML, RDF/Turtle and JSON;

・Mechanical coordination with external systems that enables searching from external applications by using the SPARQL;

・Search by the National Diet Library Classification (NDLC) and the Nippon Decimal Classification (NDC) as well as by headings or references;

・Graphical display of the NDLSH thesaurus structure (see Figure 2);

・Creation and display of links to the LCSH and the Wikipedia when they have subject headings or terms corresponding to each subject heading of the NDLSH.
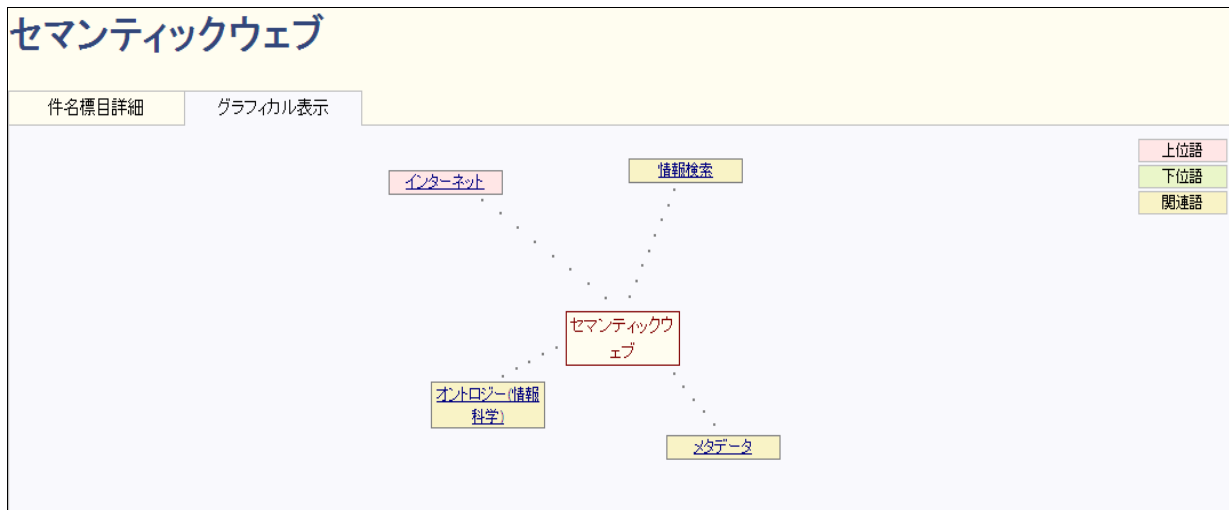

Figure 2: Graphical display of NDLSH thesaurus structure for the heading "Semantic Web" (in Japanese).

According to the revision 2004, each NDLSH heading has an equivalent LCSH heading. Since 2009, we have also started to enter LC Control Number (LCCN) into our subject authority file. The links are generated by LCCN to corresponding LCSH headings, and by using dynamic queries with literal values in the case of authority records without LCCN. Certain subject heading lists hold equivalent LCSH headings for their headings, as le Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU) and le Répertoire de vedettes-matière (RVM) de la bibliothèque de l'Université Laval, Québec. Therefore, it may be possible to create a list of French-English-Japanese multilingual subject headings.

On July 27, 2009, the NDL held a lecture "Semantic Web and libraries: toward an era when machines read information" which dealt with the concept of the Semantic Web introducing the "Web NDLSH".

The "Web NDL Authorities"

We have provided the JAPAN/MARC, a machine-readable form of the Japanese National Bibliography, since 1981. We have also provided our author name authority file as the JAPAN/MARC (A) since 1997. We will adopt the MARC21 format for the

JAPAN/MARC from January 2012.

In addition, we are now developing a service to offer an RDF/Linked Data version of our name authority file to make it more user-friendly on the Web. In order to do this, we intend to expand the "Web NDLSH" for the new service which will be called the "Web NDL Authorities". The "Web NDL Authorities" will be able to provide both our subject authority file (Subject heading and Subdivision) and our name authority file (Personal name, Family name, Corporate name, Uniform title and Geographic name) in the same interface. That is to say:

Web NDL Authorities = Web NDLSH + NDL's Name Authorities

The "Web NDL Authorities" will be fully released in January 2012 and before that, the BETA version will be available. About one million headings will be provided as Linking Open Data.

The feature of the "Web NDL Authorities" is almost the same as that of the "Web NDLSH":
・Assignment of a URI for each of the headings;
・Download of each datum in RDF/XML, RDF/Turtle and JSON;
・Search using the SPARQL
・Search by NDLC and NDC as well as by headings or references;
・Links to the LCSH, the Wikipedia, and in case of name headings having LCCN, to VIAF.

The vocabularies used are SKOS, RDFS, Dublin Core, FOAF and others. We use "DC-NDL" terms for Yomi (Japanese transliteration) in authority records. The "DC-NDL" is our own metadata schema based on the Dublin Core, aiming to contribute to the interoperation of metadata between libraries and related institutions in Japan. For Personal name, Family name and Corporate name, the RDF model consists of two levels of resources, one corresponding to the authority record, and the other to the real world entity (see Figure 3).

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:xl="http://www.w3.org/2008/05/skos-xl#" xmlns:rda="http://RDVocab.info/ElementsGr2/"
  xmlns:frbrent="http://RDVocab.info/uri/schema/FRBRentitiesRDA/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:ndl="http://ndl.go.jp/dcndl/terms/"
  xmlns:dct="http://purl.org/dc/terms/">
- <rdf:Description rdf:about="http://id.ndl.go.jp/auth/ndlna/00104237">
  - <foaf:primaryTopic>
    - <foaf:Person rdf:about="http://id.ndl.go.jp/auth/entity/00104237">
        <foaf:name>村上春樹</foaf:name>
        <rda:dateOfBirth>1949</rda:dateOfBirth>
      </foaf:Person>
    </foaf:primaryTopic>
    <dct:modified>1999-09-24T13:55:00</dct:modified>
    <dct:created>1981-03-16</dct:created>
  - <xl:prefLabel>
    - <rdf:Description>
        <xl:literalForm>村上, 春樹, 1949-</xl:literalForm>
        <ndl:transcription xml:lang="ja-Kana">ムラカミ, ハルキ, 1949-</ndl:transcription>
        <ndl:transcription xml:lang="ja-Latn">Murakami, Haruki, 1949-</ndl:transcription>
      </rdf:Description>
    </xl:prefLabel>
    <rdfs:label>村上, 春樹, 1949-</rdfs:label>
    <dct:source>風の歌を聴け</dct:source>
    <dct:source>文化人名録</dct:source>
    <skos:inScheme rdf:resource="http://id.ndl.go.jp/auth#personalNames" />
  </rdf:Description>
</rdf:RDF>
```

Figure 3: RDF model of the authority record "MURAKAMI, Haruki" (in Japanese).

The URI assigned to the real world entity redirects to the URI of the authority record in the same heading.

The "Web NDL Authorities" will also allow users to download the list of NDLSH in RDF/XML or TSV, and provide RSS feeds of new and changed subject headings.

Conclusion

Authority records are valuable assets for the library community, and for the Semantic Web community, too. We are launching the "Web NDL Authorities" to make our assets available to the Semantic Web. We are also preparing to provide bibliographic records, another of our valuable assets, in RDF on the new integrated search system "NDL Search" whose BETA version was released in 2010, and which will become fully operational in January 2012. The "Web NDL Authorities" works together with the "NDL Search". We will continue our contribution to the Semantic Web community as well as to library users.

Reference:

All references checked on June 2nd, 2011.

・Web NDL Authorities: BETA version (Japanese only)   http://id.ndl.go.jp/auth/ndla

・Web NDLSH (Japanese only)   http://id.ndl.go.jp/auth/ndlsh   (transferred to the URL of Web NDL Authorities")

・DC-NDL (Japanese only)   http://www.ndl.go.jp/jp/library/data/meta.html

・NDL Search: BETA version   http://iss.ndl.go.jp/?locale=en