

国会図書館件名標目表 (NDLSH) の SKOS 化と そのグラフィカルブラウザの作成

永森 光晴, 杉本 重雄

{nagamori, sugimoto}@slis.tsukuba.ac.jp

茨城県つくば市春日 1-2

筑波大学図書館情報メディア研究科

抄録

デジタル図書館やセマンティックウェブ技術の普及に伴って、多種多様な情報資源のメタデータが作成され利用されるようになってきた。公共図書館や地方自治体の中には、それぞれが作成したメタデータの組織化を行うために、独自の語彙を定義しているところをも珍しくない。メタデータに付与する語彙を作成する場合、小規模であればその作成とメンテナンスは容易であるが、大規模な語彙を作成することは、コストやメンテナンスの観点から現実的ではない。そこで本研究では、国立国会図書館が 2006 年 9 月より提供を開始した国立国会図書館件名標目表 (NDLSH) テキストデータを SKOS 形式に変換し、ソフトウェアやネットワークを介して利用可能にした。また本研究では、SKOS 版 NDLSH テキストデータを対象とした、SKOS のためのグラフィカルなブラウザ HANAUI (Hybrid And Network-Assisted Vocabulary Interface) の作成を行った。

キーワード

件名標目表、メタデータ、SKOS、RDF、統制語彙

Representing National Diet Library Subject Headings (NDLSH) in SKOS and its Graphical Browser

Mitsuharu Nagamori, Shigeo Sugimoto

{nagamori, sugimoto}@slis.tsukuba.ac.jp

1-2 Kasuga, Tsukuba, Ibaraki

Graduate School of Library, Information and Media Studies, University of Tsukuba

Abstract

NDLSH (National Diet Library Subject Headings) is a list of subject headings applied to the catalog of the NDL. This paper describes how to represent NDLSH in SKOS (Simple Knowledge Organization System) which is a formal language for representing controlled structured vocabularies such as thesauri or classification schemes. This paper also describes HANAUI (Hybrid And Network-Assisted Vocabulary Interface) browser which visualizes NDLSH.

Keywords

Subject Headings, Metadata, Metadata Vocabularies, SKOS, RDF, controlled vocabularies

1. はじめに

デジタル図書館やセマンティックウェブ技術の普及に伴って、多種多様な情報資源のメタデータが作成され利用されるようになってきた。近年、メタデータは情報資源の組織化や検索のためだけでなく、保存や流通のための重要な役割を担うようになってきた。既に RSS を利用した情報資源の流通は、日々の情報収集において欠かせないものとなっている。また、公共図書館や地方自治体の中には、それぞれが作成したメタデータの組織化を行うために、独自の語彙を定義しているところをも珍しくない。例えば、各都道府県のウェブサイトでは、提供している各種情報資源を分類するために 30～60 語で構成される語彙を決め、情報資源を分類するためのディレクトリを構成している。岡山県立図書館が提供するデジタル岡山大百科では、岡山県に関連した情報資源を分類するための約 400 語からなる独自の語彙を作成している。

語彙を作成する場合、小さな規模であれば作成とそのメンテナンスは容易であるが、大規模な語彙を作成することは、コストやメンテナンスの観点から現実的ではない。そこで本研究では、様々なコミュニティで共通に利用可能な語彙として、国立国会図書館が 2006 年 9 月より提供を開始した国立国会図書館件名標目表 (NDLSH) テキストデータ [1] を SKOS (Simple Knowledge Organization System) [2] 形式に変換し、ソフトウェアやネットワークを介して利用可能にした。SKOS は、RDF を使った概念スキームを表現するためのモデルである。また本研究では、SKOS 版 NDLSH テキストデータを対象とした、SKOS のためのグラフィカルなブラウザ HANA VI (Hybrid And Network-Assisted Vocabulary Interface) の試作を行った。HANA VI は、NDLSH テキストデータを検索し表示するだけでなく、標目間の関連をグラフィカルに提示する。

2. NDLSH の SKOS 化

国会図書館件名標目表 (NDLSH) は、国立国会図書館の和図書・洋図書の目録において使用実績のある件名標目を収録した件名標目表である。今回提供されている NDLSH テキストデータは 2005 年度版で、昭和 24 年から平成 18 年 3 月末までを収録範囲としている。収録件数は、標目数 (主標目、細目、細目つき件名標目の総計) 16,879 件、50 音順見出し数 (標目および参照語の総計) 36,922 件である [3]。本研究では、NDLSH テキストデータをソフトウェアやネットワークを介してより使いやすくするために、シソーラスや件名標目など表現する RDF (Resource Description Framework) [4] を使ったモデルである SKOS (Simple Knowledge Organization System) への変換を行った。

2.1 NDLSH テキストデータの構造

NDLSH テキストデータの構造について述べる。NDLSH テキストデータは、TSV (タブ区切り) 形式のファイルとして提供されている。1 行が 1 レコード (件名標目または参照形) を表している。表 1 に各レコードのデータ項目を示した。例えば、件名標目「建築家」のレコードは以下のとおりである (紙面の都合上 1 行を折り返している)。

```
ケンチクカ 建築家<地理区分> ID:00565601 UF:アーキテクト BT:美術家<地理区分> [ビジュツカ]
{00560797} RT:建築<地理区分> [ケンチク] {00565563} ; 建築士<地理区分> [ケンチクシ] {00565615}
NDC(9):520.28
```

また、「建築家」の参照形である「アーキテクト」は以下のように記述されている。

```
アーキテクト アーキテクト USE:建築家<地理区分> [ケンチクカ] {00565601}
```

上記の「建築家」と「アーキテクト」の記述例からわかるように、NDLSH テキストデータでは、参照形に関する情報は件名標目のレコードに含まれている。そこで本研究では、NDLSH テキストデータのうち、SKOS 化の対象として件名標目のレコードのみを利用した。

表1 NDLSH テキストデータのデータ項目 (NDLSH テキストデータより引用)

区分	項目	導入句	記録内容
1	見出し語(件名 標目または参照 形)	件名標目または参照形 のよみ	必須。カナまたは英数字 細目がある場合は、一で連結した形で記録する(すべての項目で同様)
2	件名標目または参照形		必須 地理区分など細目が付加できるものについては<地理区分>等と付記 洋書にのみ使用している件名標目は、冒頭に※を付記 細目にのみ使用する件名標目は、《》で囲む <>及び※は、USE,UF,BT,NT,RTに記録する参照件名標目にも該当する場合は付記する
3	ID番号(ID)	ID	ID:の次にB桁の数字を記録する 参照形の場合は存在しない
4	を見よ参照	を見よ参照(USE)	USE
5	を見よ参照あり(UF)	を見よ参照あり(UF)	UF
6	をも見よ参照	上位語(BT)	BT
7		下位語(NT)	NT
8		関連語(RT)	RT
9	注記	をも見よ参照注記(SA)	SA
10		スコープノート(SN)	SN
11	分類	日本十進分類法第9版 NDC(9)	NDC(9)
12		国立国会図書館分類表 NDLC	NDLC

2.2 SKOS

SKOS (Simple Knowledge Organization System) は、シソーラスや件名標目などの概念スキーム (concept scheme) の構造や内容を表現するための RDF を利用したモデルである [2]。2006 年 10 月現在、SKOS の仕様のステータスは W3C Working Draft で、現在も議論が行われている。また、SKOS Extensions Vocabulary において、SKOS の拡張についても検討されている [5]。

図 1 に RDF を使った SKOS の記述例を示した。図 1 では、2 つの概念、ex:mammals と ex:animals を定義している (文献 [6] より引用)。図 1 は、ex:mammals を表現するラベルは 'mammals' で、その上位語は ex:animals であることを表している。また同様に、ex:animals のラベルは 'animals' で、その下位語は ex:mammals であることを示している。図 2 は、図 1 の記述をグラフで表現したものである。

2.3 SKOS への変換

表 2 は、本研究で NDLSH テキストデータを SKOS へ変換するために用いた、NDLSH テキストデータのデータ項目と SKOS エレメントとの対応付けである。SKOS では、NDLSH テキストデータのデータ項目全てを表現することができない。そのため、いくつかの項目は、Dublin Core [7]、または独自に定義したエレメントと対応付けている (例: dc:subject, slis:annotation, slis:transcription など)。NDLSH テキストデータの日本語の読みをどのようなエレメントを用いて表現するかは今後の課題である。本稿では、SKOS の仕様の編集者のひとりである Alistair Miles 氏と SKOS における日本語の読みの表現について議論した結果を反映している。

表 2 の対応表に基き、プログラミング言語 Ruby と XSLT を利用して SKOS 形式への変換を行った。図 3 は、件名標目「建築家」を SKOS で表現した例である。

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://www.example.com/concepts#mammals">
    <skos:prefLabel>mammals</skos:prefLabel>
    <skos:broader rdf:resource="http://www.example.com/concepts#animals"/>
  </skos:Concept>
  <skos:Concept rdf:about="http://www.example.com/concepts#animals">
    <skos:prefLabel>animals</skos:prefLabel>
    <skos:narrower rdf:resource="http://www.example.com/concepts#mammals"/>
  </skos:Concept>
</rdf:RDF>

```

図1 SKOS の RDF 記述例 (SKOS Core Guide より引用)

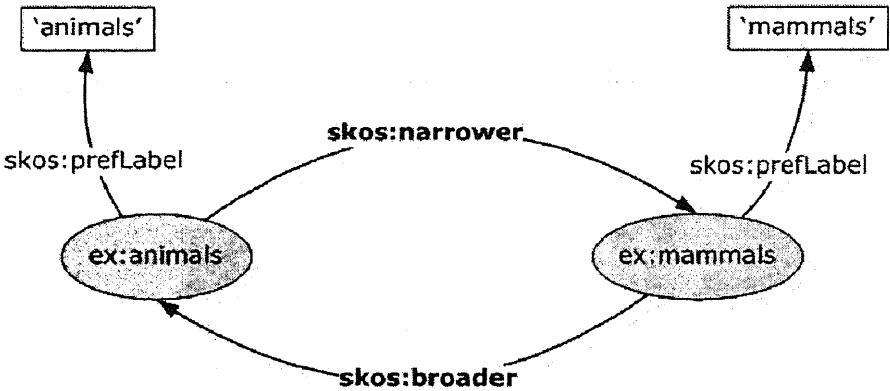


図2 図1 に示したSKOS のグラフ表現

3. ネットワークを介したNDLSHの利用

本研究ではソフトウェアからSKOS形式に変換したNDLSHテキストデータを利用するため、以下の2つの機能を作成した。

1) 典拠IDを指定した件名標目データの取得

各件名標目には典拠IDが割り当てられている。この機能では、指定した典拠IDを持つ件名標目のデータをSKOS形式で取り出すことができる。<http://raus.slis.tsukuba.ac.jp/ndlsh/> に続けて典拠IDを指定したURLにアクセスすると、SKOS形式での記述が返される。例えば、以下のURLにアクセスすると、図3に示した件名標目「建築家」(典拠ID 00565601)のデータを取得できる。

例) SKOSで記述した件名標目「建築家」の取得

<http://raus.slis.tsukuba.ac.jp/ndlsh/00565601>

2) NDC第9版またはNDLCを指定した件名標目の検索

NDLSHの個々の件名標目には、その件名標目の内容を表す分類記号(代表分類記号)が付与されている。代表分類記号は、その件名標目が表す主題に対し、一般的に付与される分類記号を示したものである。NDLSHテキストデータで使われている代表分類記号には、日本十進分類法第9版(NDC(9))と、国会図書館分類法(NDLC) [8]の2種類がある。

表2 NDLSH データ項目と SKOS エLEMENTの対応付け

NDLSH データ項目	SKOS
件名標目	skos:Concept
典拠 ID	dc:identifier
を見よ参照 (USE)	skos:note
を見よ参照あり (UF)	skos:altLabel, slis:annotation, slis:JapanesePronunciation, slis:label, slis:transcription の組み合わせ
上位語 (BT)	skos:broader
下位語 (NT)	skos:narrower
関連語 (RT)	skos:related
をも見よ参照注記 (SA)	skos:note
スコープノート (SN)	skos:scopeNote
NDC(9)	dc:subject
NDLC	dc:subject

本研究では、与えられた NDC 第 9 版または NDLC の値を使って件名標目の持つ代表分類記号を検索し、適合した件名標目の典拠 ID と件名標目のリストを RSS 形式で返す機能を作成した。NDC 第 9 版の場合は、<http://raus.slis.tsukuba.ac.jp/ndc/> に続けて NDC の分類記号を指定する。NDLC の場合は、<http://raus.slis.tsukuba.ac.jp/ndlc/> に続けて NDLC の分類記号を指定する。以下に、NDC 第 9 版と NDLC それぞれを使って検索する例を示した。

例 1) NDC 第 9 版で 520 (建築学) を持つ件名標目のリストを取得する

<http://raus.slis.tsukuba.ac.jp/ndc/520>

例 2) NDLC で KA31 (建築 - 建築史) を持つ件名標目のリストを取得する

<http://raus.slis.tsukuba.ac.jp/ndlc/KA31>

図 4 は、例 2 の検索結果である。

4. HANAUI: グラフィカルな NDLSH ブラウザの作成

SKOS に変換した NDLSH テキストデータを利用し、NDLSH を対象とした SKOS のためのグラフィカルなブラウザ HANAUI (Hybrid And Network-Assisted Vocabulary Interface) の試作をおこなった [9]。今回試作した HANAUI は以下の機能を持つ。

- 「を見よ参照 (USE, UF)」と「をも見よ参照 (BT, NT, RT)」の情報を使った、標目間の関係の可視化
- キーワードによる件名標目と参照語の検索
- NDC 第 9 版を使った件名標目の検索

```

<skos:Concept rdf:ID="http://raus.slis.tsukuba.ac.jp/ndlsh/00565601">
  <dc:identifier>00565601</dc:identifier>
  <skos:prefLabel xml:lang="ja">建築家 (地理区分) </skos:prefLabel>
  <slis:annotation>
    <slis:JapanesePronunciation>
      <slis:label xml:lang="ja">建築家 (地理区分) </slis:label>
      <slis:transcription xml:lang="ja">ケンテカ</slis:transcription>
    </slis:JapanesePronunciation>
  </slis:annotation>
  <skos:altLabel xml:lang="ja">アーキテクト</skos:altLabel>
  <skos:broader rdf:resource="http://raus.slis.tsukuba.ac.jp/ndlsh/00560797" />
  <skos:related rdf:resource="http://raus.slis.tsukuba.ac.jp/ndlsh/00565563" />
  <skos:related rdf:resource="http://raus.slis.tsukuba.ac.jp/ndlsh/00565615" />
  <dc:subject rdf:datatype="http://juror.slis.tsukuba.ac.jp/ns/tsukubaterms#NDC">
    520.28
  </dc:subject>
  <dc:subject rdf:datatype="http://juror.slis.tsukuba.ac.jp/ns/tsukubaterms#NDLC">
    KA31
  </dc:subject>
</skos:Concept>

```

図3 NDLSH テキストデータをSKOSで記述した例

- 複数の件名標目定義を同時表示

図5は、HANAVIで件名標目「建築家」を表示した画面例である。グラフのノードは件名標目または参照語を示している。ノードが件名標目の場合は、ノードをクリックすることにより、さらに関連する件名標目をたどって閲覧することができる。図では、「建築家」の「を見よ参照」と「をも見よ参照」に関する情報から、関連する件名標目と参照語をグラフ表示している。さらに図では「建築家」と「建築史」の件名標目に関する情報を同時に表示している。この機能は、実際の目録作業において、複数の件名標目を比較する必要があることから作成した。

図6は、NDCから件名標目を検索するためのディレクトリである。ディレクトリには第2次区分までの分類項目名が表示されている。分類項目名をクリックすると、その項目の分類記号を持つ件名標目を検索した結果が表示される。検索結果の利用方法は、図5と同様である。

HANAVIの実現には、Ruby on Rails 1.1.6, MySQL 5.0.26, lighttpd 1.4.14, prototype.js, effects.js, wz_jsgraphics.js を利用した。グラフ表示にはSVGやFlashなどを用いておらず、HTMLとCSSを用いて実現している。そのため、通常のウェブブラウザであれば、特別なプラグインをインストールすることなくグラフを表示することができる。

5. 関連研究

件名をグラフィカルに提示する関連研究には、村上による Subject World がある [10]。Subject World では、OPACにおいて主題や件名による情報検索をより使いやすくするために、用語体系の統合と可視化を行っている。概念体系としては、件名に BSH 4、分類に NDC 第9版を利用している。

NDLSHをSKOSで表現した例には、神崎の「シソーラスとクラス階層とRDF/OWL」がある [11]。ここでは、件名標目をRDF/OWLで表現し主題へのアクセスに利用する際に、クラス階層を使うことの利点と欠点について検討している。文献 [11] では、読みに関する表現については言及されていない。

Protégéはスタンフォード大学SMI (Stanford Medical Informatics) が開発したオントロジエディタである [12]。ProtégéではXML/RDFを意識することなく、対話的にオントロジを作成することができる。

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns="http://purl.org/rss/1.0/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<channel rdf:about="http://raus.slis.tsukuba.ac.jp/ndlc/KA31">
  <title>search results for NDLC: KA31</title>
  <link>http://raus.slis.tsukuba.ac.jp/</link>
  <description>2 item(s) found.</description>
  <items>
    <rdf:Seq>
      <rdf:li resource="http://raus.slis.tsukuba.ac.jp/ndlsh/00565589"/>
      <rdf:li resource="http://raus.slis.tsukuba.ac.jp/ndlsh/00565601"/>
    </rdf:Seq>
  </items>
</channel>
<item rdf:about="http://raus.slis.tsukuba.ac.jp/ndlsh/00565589">
  <title>建築 (西洋)</title>
  <link>http://raus.slis.tsukuba.ac.jp/ndlsh/00565589</link>
  <dc:identifier>00565589</dc:identifier>
</item>
<item rdf:about="http://raus.slis.tsukuba.ac.jp/ndlsh/00565601">
  <title>建築家 (地理区分) </title>
  <link>http://raus.slis.tsukuba.ac.jp/ndlsh/00565601</link>
  <dc:identifier>00565601</dc:identifier>
</item>
</rdf:RDF>

```

図4 NDLCのKA31（建築 - 建築史）を検索した結果

6. おわりに

本稿では、NDLSH テキストデータの SKOS 化と、SKOS のためのグラフィカルブラウザ HANA VI について述べた。SKOS において NDLSH テキストデータの日本語の読みをどのように表現するかは、今後の課題である。

HANA VI を作成する中で件名標目間の関連の扱いについて問題があった。それは、いくつかの件名標目では他の件名標目との関係が明示的に記述されていないことである。例えば、件名標目「工学」では、「をみよ参照注記 (SA)」の項目に「各種の工学の件名をも見よ (例:教育工学)」と記述されているのみで、約 60 件ある各種工学の件名標目と明示的に「をみよ参照 (BT, NT, RT)」によって関連付けられていない。ソフトウェアから NDLSH を利用する際には、このような SA に記述されている件名標目間の関連を処理する必要がある。

現在、様々なコミュニティが扱う情報資源の特徴にあわせた独自の語彙を決め、メタデータの作成を行っている。今後は、NDLSH テキストデータを中心 (ハブ) にコミュニティが作成した語彙を結び、メタデータの相互利用性を高めるための方式を検討したい。

謝辞

国立国会図書館書誌部書誌調整課のみなさまには、NDLSH テキストデータの変換と HANA VI の開発において多大なるご協力と貴重なご意見をいただきました。この場を借りて深く感謝いたします。

参考文献

- [1] 国会図書館件名標目表, <http://www.ndl.go.jp/jp/library/data/ndl.ndlsh.html>
- [2] SKOS, <http://www.w3.org/2004/02/skos/>

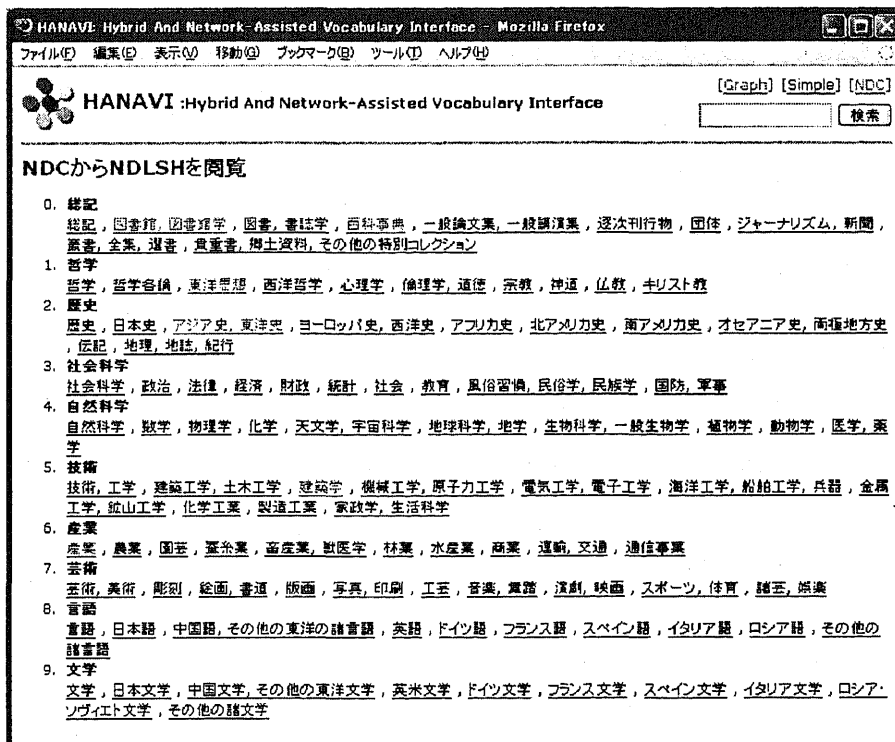


図6 NDCを使ったディレクトリからの検索インタフェース