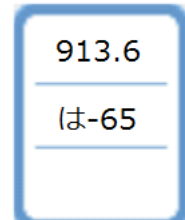


NDC を LOD 化することの意味と意義

-初心者のための、完全に正しくはないけど、ざっくりわかる解説-

1. NDC(日本十進分類法)

NDC(日本十進分類法)は、日本の図書館(公共図書館,大学図書館,学校図書館など)で広く使われている図書の分類法です。図書館にある本の背表紙に右のようなラベルが貼付されていることが多いですが、この1番上の記号(913.6)がNDCの分類番号となります。



NDCでは、世の中の資料で取り上げられているテーマ(以下、「主題」と呼びます)を、大まかな分類から細かい分類へと順次10ずつの項目に細分していく方式をとっています。細かい仕組みは省略しますが、主題を分類した本表以外に補助表と呼ばれる仕組みが整備されており、たとえば「長崎の路面電車」という図書に対して686.9183という番号が付与されるように、非常に細かいレベルまで分類番号を付与することが可能となっています。図書館のような大量の図書を取り扱う施設では、乱雑に図書を書架に配置した(配架)場合には探すのが極めて困難になることは予想に難くありません。そこでNDCを用いて類似の資料が近い場所に配架できるようにしているのです。NDCは、現在は主として冊子体として販売されたものが使用されています。

分類法は、図書館で資料を配架する時に使うというだけではありません。最近のWeb上で図書館の蔵書を探すシステム(WebOPAC)では、複数の分類番号を図書に付与しておき、分類番号を辿って、そのいずれもから図書館資料の検索システムでは探すことができる仕組みが整備されたりもしています。このように、NDCは図書館において資料を探す場合の手がかりとして重要な役割を果たしているのです。図書を探すという目的以外にも「網羅的な範囲の主題が分類されている」ということの意味は大きく、たとえば図書館の蔵書構成が適切かどうか検討する際には、まずNDCの分類番号ごとに所蔵数や貸出冊数を数えることが出発点となります。資料がきちんと検討された分類体系に基づいて整理されていることが、全ての図書館活動の基本であるとさえ言えるでしょう。

さらに近年では、図書館以外においても利用の範囲は広がってきています。たとえば、オンライン書店においても同じ主題の資料を探す場合にNDCの番号を入力して探す仕組みを持つところがあります。また、読書支援や図書の推薦を行うWeb上のサービスでは、利用者が求める情報にたどり着くために大きな概念から小さな概念に絞り込んだり、また逆により広い概念を提示する際にNDC番号の桁数を増減させることなどが可能となります。

NDCのような全ジャンルを網羅し、かつ詳細なレベルまで検討した分類法は日本でも希少な存在であり、個人や企業活動など利用範囲は無限に広がっているとも言えるでしょう。

2. Linked Data(LD) と LOD(Linked Open Data)

2.1 Linked Data

Web ページがリンクという仕組みを使って繋がれており、人間がこれを進んでいくことで様々な情報を得られることは良く知られています。Web ページには、文章や画像などだけではなく、様々なデータの公開を主な目的としたものが数多くあります。

このようなデータは人間が見て役に立つということもありますが、それだけではなく何らかの加工を施すなどの材料として使用することも少なくありません。たとえば、表の形で公開されている統計データをもとに、グラフ化したりクロス集計を行ったりして自分なりの新しいプレゼンテーション資料を作成するなど個人的にも行われているでしょう。このような公開データの再加工にとどまらず、自分の持つデータに公開データ中の情報を追

加したり、自分が公開しようとするアプリの中に公開データを取り込んで高度化するなども考えられます。さらには複数の公開されているデータを組み合わせて全く新しいサービスを生み出したりすることも盛んに行われています。たとえば、AED が設置されている場所のデータが公開されていることを受けて、現在地から目的地への移動時間を計算する仕組みと組み合わせて、事故が発生したときに最寄りの設置場所まで移動することができる方法を提供するようなアプリが公開されています。

このように公開されたデータの利用方法としては、データを手作業で Excel などに取り込んで加工するというようなことも可能ではありますが、特に大量のデータを取り扱う場合にはコンピュータプログラムを作成して、このプログラムによってデータを取得・利用することが一般的です。

コンピュータプログラムがデータを取得・利用するというと、何となく人間よりも賢く処理してくれるイメージがありますが、実はそんなことはありません。実際に、人間ならば多少の違いには柔軟に対応してくれるものでも、コンピュータでは理解できないということは数多く存在しています。

たとえば下のような2つの図書についてのデータがスペースを区切りとして表示されている場合、人間ならば1では著者名、書名、出版社名、出版年月が表示されており、2では全集なので著者名が省略されて書名、副書名、出版社名、出版年が表示されていると理解することができます。しかし、コンピュータにこれを理解させるのはかなり難しいのです。

1. 夏目漱石 我が輩は猫である 黒猫書店 1954.10
2. 夏目漱石全集第2巻 吾輩は猫である 白猫書店 1957

そこで、このようなデータをコンピュータできちんと処理するために、公開するデータの各項目ごとに、その項目が何を意味しているのかに関する情報(項目名のようなもの)をコンピュータが理解しやすい形で付加してあげるなどの工夫が必要とされます。

ただし、ここで使う項目名が利用する人全員にとって共通に理解できるものとなっていないければ、Web のように複数のデータを組み合わせて利用する環境では使いものにならないことには注意が必要です。実際に使われている例では、標準化を担う組織が中心となって合意が得られている名前が項目名に付与されることとなります。たとえば、Dublin Core などはその例です。

上記の例で Dublin Core を用いて項目名に関する情報を付与して、<dc:author>夏目漱石</dc:author>のように追加することが行われるわけです。もちろん、このような工夫は項目名だけではなく様々な観点から行われます。

このように、Web 上で公開するデータのうち、コンピュータで利用しやすい形に加工し、標準化された項目名を決まった形で付与した上で公開するものを **Linked Data** と呼びます。Linked Data として公開されているものは、コンピュータプログラムを用いて利用することが簡単、逆にいくらデータが Web 上で公開されていても **Linked Data** でなければ(人間が見て手作業であれば簡単に利用できても)コンピュータプログラムでは利用が難しいということができるよう。

2.2 Linked Open Data

前節のように、さまざまなデータが Linked Data 化されることにより、アイデア次第で新しいサービスが次々と生まれてくることが期待できます。

しかし、これはあくまで「技術的に利用可能」という話であって、存在するからといって使ってよいとは限らないというのは人間社会の通例です。たとえば、アイドルの写真は各芸能事務所では公開されていますが、これを自分の Web ページに持ってきて利用することは法的にもできません。

営利活動のためのデータなどでは公開していても利用は不許可ということは当たり前の

ことでもありますが、世の中には多くの人々が自由に使うことが社会にとって有用であるデータも多数存在しています。たとえば、国や地方公共団体、公益団体が作成するデータなどは、このような自由な公開が望ましいものが多数含まれています。近年、国内外で **Linked Data** を公開しようという動きが活発化してきており、2013年6月のG8首脳会議でも「政府のデータをまとめて機械判読できるようにし、オープンなライセンスで公開すること」などを原則とする「オープンデータ憲章」に各国首脳が合意しています。ここで、「オープンなライセンス」とは、「誰でも自由に利用することができるようにする」ということを意味しており、その多くは無償で公開されます。

Linked Data のうち、オープンに(「誰でも自由に利用して良い」という許諾を与えた上で)公開されるデータを **Linked Open Data** と呼びます。ただし、「自由に」という部分は「非営利ならば」などの条件がつけられるケースもあります。

3. NDC と Linked Open Data

さて、ここで NDC です。

「1. NDC(日本十進分類法)」に書いたように、NDC は図書館の資料組織化の中心的なツールであり、日本中の図書館で利用されているだけではなく、図書館外も含めて図書の管理を行うさまざまな局面での利用も広がっています。

たとえば、Web 上で公開されている図書館の蔵書検索システム(WebOPAC)で検索結果として表示される図書1冊1冊のデータの中には、ほとんどの場合 NDC の番号が含まれています。しかし、これら WebOPAC の中でこの番号が何を意味しているのかを「番号ではなく言葉で」表示している図書館は数多くありません。利用者の中には「この番号は何?」と思っていたのではないのでしょうか。また、ある図書を検索した時に、関連する図書を見つけることか便利だと思われたことはありませんか? NDC が LOD 化されていれば、これらの問題は既に解決し、図書館の WebOPAC はもっと便利になっていたかもしれません。

さらに、読書管理システムなどで、同じ主題の図書をまとめて表示したければ NDC ごとに集約すれば今でも実現可能ですが、集約した結果を示すのが分類番号だけでユーザに理解してもらえないと思いませんか? 大学の図書館司書課程などで分類番号付与の自習システムを作った時、学生が入力した番号に対応する語句さえ表示されないシステムなんて使ってもらえないと思いませんか?

上記は、ほんの一例にすぎません。人間にとって図書は最も身近な情報メディアのひとつであり、人間生活・社会生活にとって必要不可欠な存在でもあります。当然、図書と関わり、図書を取り扱うシステムは山と存在しています。図書には人類の知恵と知識が詰まっていることを考えれば、NDC は図書だけではなく人間生活全般に関わるツールになりえるとも言えるでしょう。

もちろん、NDC が LOD 化されたからといって出来ないこともあります。たとえば、受け入れた図書に対する分類番号の付与を LOD 化された NDC だけを元に行うことは実際問題としては困難でしょう。上記の「図書館司書課程における分類番号付与自習システム」においても、学生が入力した番号がどのように間違っているかを表示するのに LOD 化された NDC は有効でしょうが、学生が番号を入力する段階では冊子体の NDC を使うことが必要になるでしょう。さまざまな例外もありますが、**多くの場合において NDC の番号を付与する際には冊子体が有効、NDC の Linked Data は NDC を表示したり利用したりする際に有効ということができるかもしれません。**

日本の社会全体を考えた場合、最終的には NDC の冊子体自体がデジタル化され、自由に再利用できる形でオープンにされることが望まれますが、それが時期尚早だとしても利用のための、とりわけインターネット上での利用のための **Linked Data** のオープン化(公開)は待ったなしの状態にあると言っても過言ではないでしょう。