

図書目録とセマンティック・ウェブの融合

安藤友晴

●要約

本論文では、セマンティック・ウェブの手法を使って図書目録を扱うことによる利点と問題点とを考察した。RDF/XML による Dublin Core を用いて図書目録を扱うことで、次の利点が得られる。

- 1) メディアの種類に関わらず検索できる。
- 2) データ検索の手法を同一にできる。
- 3) オントロジを用いて関連情報を含めた精度の高い検索を行うことができる。

また、次のような不具合もある。

- 1) Dublin Core の表記法に差異がある。
- 2) オントロジが未整備である。

●キーワード

セマンティック・ウェブ

RDF

Dublin Core

オントロジ

図書目録

件名標目

1. セマンティック・ウェブ

Tim Berners-Lee⁽¹⁾ が提唱した「セマンティック・ウェブ⁽²⁾」は、ウェブの情報に「メタデータ」を加え、プログラムがこれらのメタデータの内容を理解するとともに、自動的に問題解決のためのデータ処理を行うものである。

「メタデータ」とは、「データのためのデータ」と定義されることが多い。メタデータの具体例として、目録・索引・書評・辞書などを挙げることができる。

メタデータを記述方法としては RDF (Resource Description Framework)⁽³⁾ がある。また、RDF を XML で記述する RDF/XML Syntax⁽⁴⁾ が既に一般的に利用されている。例えば、ウェブなどのタイトル・作者・分類・作成日などを表すことのできる「Dublin Core⁽⁵⁾」や、ウェブの更新情報などを表す「RSS⁽⁶⁾」、名刺情報を表す「vCard⁽⁷⁾」などが RDF で表現できる。

現在は、RSS や Dublin Core の利用によって、セマンティック・ウェブの普及が緒に就いたという段階である。しかし、セマンティック・ウェブによって、ウェブの情報が有機的に結合し、より精度の高い検索やメタデータ間の関係に着目した推論を行うことも可能になる。

2. 図書目録データの検索

一方、現在では、有益かつ信頼性が高い情報が書籍・雑誌などに存在し、ウェブ上で情報そのものを入手できないことが依然として多い。

書籍や雑誌を通して必要な情報を得るには、当然のことながら現物を参照することになる。では、どのようにして必要な書籍や雑誌を探し当てるのだろうか？ 図書館や書店の書架を探すか、さもなくば目録を検索することになるだろう。例えば図書館の目録を検索する場合、最近ではインターネット上の OPAC (Online Public Access Catalogue) を利用できることが多い。

必要な情報がありそうな書籍や雑誌がすぐ見つければそれでよい。しかし、見つからなかった場合、別の図書館や書店の OPAC で検索を実行することになる。

ここで問題になるのは、個々の図書館や書店の OPAC で目録データの管理・アクセス方法が違っているため、目録を検索するための手段・方法がそれぞれ違ってしまうことである。もちろんこの場合、すべての図書館にまたがって横断的に検索するようなことは不可能である。

3. Dublin Core の利用

このような問題を解決するため、国立国会図書館などの図書館で、図書目録情報を Dublin Core で表記することの検証が進められている⁽⁸⁾。ここでは、図書目録情報に Dublin Core を用いることの利点と問題点を整理する。

3-1 利点

Dublin Core を用いることの利点は次の通りである。

●メディアの種類に関わらず検索できる

Dublin Core を用いてメタデータが記述されていれば、検索対象となるデータがウェブ・書籍・雑誌などのメディアにとらわれず、同一視して検索できる。

●データ検索の手法を同一にできる

「著者（作者）が安藤友晴のものを検索する」という処理の場合、Dublin Core で記述されたメタデータならば、Dublin Core の RDF 表記に従って「dc: creator の値が安藤友晴であるもの」という検索式を実行すればよい。

Dublin Core を用いることによってデータ検索の対象と手法をそれぞれ一元化できるのは利点となる。

3-2 問題点

Dublin Core を用いることの問題点は次の通りである。

●Dublin Core の表記法の差異

図書目録を表記する方法としては「日本目録規則（9）」がある。ただしこの目録規則は電子的な媒体でデータ処理をするための表記法にはなっていない。もともと Dublin Core で記述できるメタデータは特定のメディアに特化しないため、Dublin Core で図書目録を表記しようとしても、統一的な表記法が確立していないという問題がある。

例えば、先の国立国会図書館での Dublin Core の利用例では、複数の主題（キーワード）を付加するときにはひとつの dc: subject 要素の中に区切り記号を用いて記述するのに対し、国立情報学研究所の「メタデータ・データベース共同構築事業」では、複数の主題を表すときには同数の dc: subject 要素を用いる。

このように、Dublin Core での表記法がそれぞれの組織で違っていると、メタデータの活用に支障が出る恐れがある。統一する必要があるだろう。

4. オントロジを使った高精度な検索

4-1 オントロジの利用

セマンティック・ウェブの技術で図書目録を表記することの利点には、オントロジを利用できることがある。Tim Berners-Lee はオントロジについて『分類体系』と『推論ルール集』である、と述べている。⁽¹⁰⁾ ここでは、オントロジの「分類体系」の部分について着目してみる。

ある書籍のメタデータに dc: subject 要素があり、"JSF" という値が入っていたとする。この "JSF" と同一の意味を表す言葉に "JavaServer Faces" があり、また "JSF" の上位語に "J2EE" という言葉がオントロジで定義されていたとする。この場合、dc: subject 要素が "JSF" である情報を検索すると、オントロジの働きによって、dc: subject が "JavaServer Faces" であるものを検索結果に含めることができる。また、上位語を検索結果に含めることで dc: subject が "J2EE" である情報を検索結果に含めることもできる。

このように、オントロジを利用することで、関連情報を含めた精度の高い検索を行うことができる。

4-2 オントロジの未整備の問題

オントロジ記述言語である OWL が W3C 勧告となったのが2004年2月ということもあり⁽¹¹⁾、セマンティック・ウェブで利用すべきオントロジの整備は十分とは言えない。現在いちばん整備が進んでいるもので、英英辞典かつシソーラスの機能を持った "WordNet"⁽¹²⁾ が約100,000語の語彙を持っている。

オントロジの体系的な整備は専門的な知識を必要とし、作業量も膨大である。ここで筆者は、これまで図書館の件名索引で使われている「件名標目表」をオントロジに変換することを提案する。件名標目表は、図書目録で件名標目（主題、いわゆるキーワード）を付加するときの語彙集である。オントロジでも利用されるシソーラスでの概念のように、上位語・下位語・関連語といった表記もある。日本国内でもっともよく使われている「基本件名標目表（第4版）」^(BSH)⁽¹³⁾ で8,000語、「国立国会図書館件名標目表」^(NDLSH)⁽¹⁴⁾ で18,000語、「米国議会図書館件名標目」⁽¹⁵⁾ に至っては260,000語が含まれている。これらの件名標目は、図書館で利用することを前提とした体系的な標目表であり、これらの変換作業によって基本的なオントロジが整備されることが十分期待できる。

5. まとめ

セマンティック・ウェブの手法を使って図書目録を扱うことによる利点と問題点とを考察してきた。

本論文では、小規模なサンプルプログラムを用いて検証を行った。今後の研究展開として、本論文で扱った内容を本学の図書館システム全体に実装し、セマンティック・ウェブを用いた検索精度に関する詳細な評価を行う予定である。

●参考文献

- (1) <http://www.w3.org/People/Berners-Lee/>
- (2) Tim Berners-Lee, James Hendler, Ora Lassila 「自分で推論する未来型ウェブ」 日経サイエンス, 31(8), 2001.8. pp.54-65
- (3) Resource Description Framework (RDF): Concepts and Abstract Syntax
<http://www.w3.org/TR/rdf-concepts/>
- (4) RDF/XML Syntax Specification (Revised)
<http://www.w3.org/TR/rdf-syntax-grammar/>
- (5) Dublin Core Metadata Initiative
<http://dublincore.org/>
- (6) RDF Site Summary (RSS) 1.0
<http://web.resource.org/rss/1.0/spec>
- (7) Representing vCard Objects in RDF/XML
<http://www.w3.org/TR/vcard-rdf>
- (8) 国立国会図書館：図書館員のページ：書誌データの作成及び提供
http://www.ndl.go.jp/jp/library/data_make.html#syoshi_doc
- (9) 「日本目録規則 1987年版 改訂 2 版」, 日本図書館協会, 2002
- (10) Tim Berners-Lee, James Hendler, Ora Lassila 「自分で推論する未来型ウェブ」 日経サイエンス, 31(8), 2001.8. pp.54-65
- (11) OWL Web Ontology Language Use Cases and Requirements
<http://www.w3.org/TR/webont-req/>
- (12) <http://www.cogsci.princeton.edu/~wn/>
- (13) 日本図書館協会件名標目委員会「基本件名標目表 (BSH) 第 4 版」, 日本図書館協会, 1999
- (14) 国立国会図書館件名標目表2004年版 (暫定版)
http://www.ndl.go.jp/jp/library/data/ndl_ndlsh.html
- (15) <http://www.loc.gov/cds/lcsh.html>

● 英文タイトル

A fusion of catalogue and semantic web

● 英文要約

In this paper, I try to consider advantage and disadvantage to treat catalogue using Semantic Web. There are advantages to treat catalogue using Dublin Core in RDF/XML.

- 1) We can retrieval data without minding kind of media.
- 2) We can retrieval data by the same mean.
- 3) We can retrieval data and related with high degree of accuracy using Ontology.

On the other hand, there are disadvantages.

- 1) There are different notations of Dublin Core.
- 2) There are few Ontology which can use.