

1 はじめに

現在の Web コンテンツは、ユーザが様々な形で利用する目的で作られており、検索エンジンによってユーザが求める情報を早く、正確に抽出するのは難しい。ユーザが認識、判断する前に機械が理解して文章の解釈やユーザの為に役に立つ情報を抽出できないのが現状である。そこで、あるデータに対してメタデータを付与して検索率の向上を図ることを目的としたセマンティック Web 技術が今、注目を集めている。これは機械が理解するための Web 技術であり、RDF や OWL などのオントロジー記述言語でオブジェクトとオブジェクトの概念間の関係をグラフ・データ形式で表現する。しかし、RDF/XML 文書による構文表現の正誤性を確認、修正するツールはあるが、記述した構文の意味表現が必ずしも正しいと判断できる基準が無いので、RDF/XML 文書の新たな構成方法を考えることが必要となる。

2 研究目的

RDF/XML 構文も XML 構文と同様に、与える情報が増えるとタグの量が増えるので文法構造が複雑になり、ユーザにとって構文の意味を理解する時間がかかる。そこで、RDF/XML 文書の中のオントロジーを表現する時にメタデータの共通認識語彙となるダブリン・コア要素を RDF/XML 構文で扱う必要性があるかについて考える。

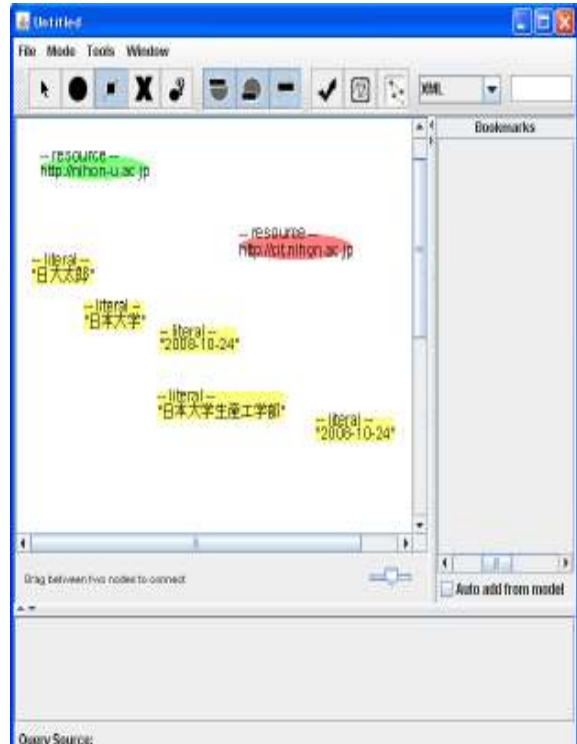
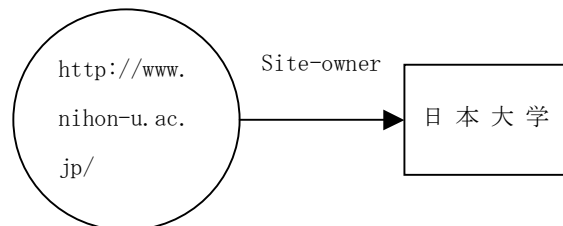


図1 RDF Author による語彙の記述



Web ページ <http://www.nihon-u.ac.jp/>
の所有者は、日本大学である。

図2 RDF トリプル(二項関係)と
ステートメント表現例

```

<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-r
df-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1
/"><rdf:Description
rdf:about="http://www.nihon-u.ac.jp/">
<dc:creator>日大太郎</dc:creator>
<dc:title>日本大学</dc:title>
<dc:date>2008-10-24</dc:date>
<dc:relation>http://www.cit.nihon-u.ac.jp
/</dc:relation>
</rdf:Description>
<rdf:Description
rdf:about="http://www.cit.nihon-u.ac.jp/"
>
<dc:title>日本大学生産工学部
</dc:title>
<dc:date>2008-10-24</dc:date>
</rdf:Description>
</rdf:RDF>

```

図 3 RDF/XML 文書の設計
(ダブリン・コア要素あり)

```

<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-r
df-syntax-ns#"
xmlns="http://www.nihon-u.ac.jp/">
<rdf:Description
rdf:about="http://www.nihon-u.ac.jp/">
<作成者>日大太郎</作成者>
<大学名>日本大学</大学名>
<更新日>2008-10-24</更新日>
<リンク>
http://www.cit.nihon-u.ac.jp/
</リンク>
</rdf:Description>
<rdf:Description
rdf:about="http://www.cit.nihon-u.ac.jp/"
>
<名前>日本大学生産工学部</名前>
<更新日>2008-10-24</更新日>
</rdf:Description>
</rdf:RDF>

```

図 4 RDF/XML 文書の設計
(ダブリン・コア要素なし)

3 検証方法

RDF/XML 文書を作成してそれを検証する方法の流れを 3.1～3.5 に示す。

3.1 RDF 作成ツールによる語彙の記述

3.2 ダブリン・コア要素の利用

3.3 多項関係の RDF は二項関係の RDF へ分解

3.4 RDF/XML 文書の作成

3.5 RDFVS による検証

3.1 RDF 作成ツールによる語彙の記述

RDF トリプルと有効ラベル付きグラフは RDF Author というツールで作成する。RDF Author は RDF エディタであり、RDF トリプルの作成が

しやすく、有効ラベル付きグラフを画像ファイルとして出力できるのが利点である。図 1 に RDF Author を使用して語彙の記述をする様子を示す。また、RDF Author はテキストエディタによる編集も可能であるが、テキストエディタの編集は Java 言語で作成した簡易エディタを使用し、ソースの編集を行う。

3.2 ダブリン・コア要素の利用

Web 上のリソースのメタデータを記述する語彙として開発されたものである。ダブリン・コア要素は 15 の基本要素で定義されている。さらに 15 の基本要素の意味を限定して、より明確な記述を可能にするダブリン・コア精密化

表1 図3によるRDFトリプルの検証結果(一部省略)

Subject	Predicate	Object
http://www.nihon-u.ac.jp/	http://purl.org/dc/elements/1.1/creator	“日大太郎”
http://www.nihon-u.ac.jp/	http://purl.org/dc/elements/1.1/title	“日本大学”
http://www.nihon-u.ac.jp/	http://purl.org/dc/elements/1.1/date	“2008-10-24”

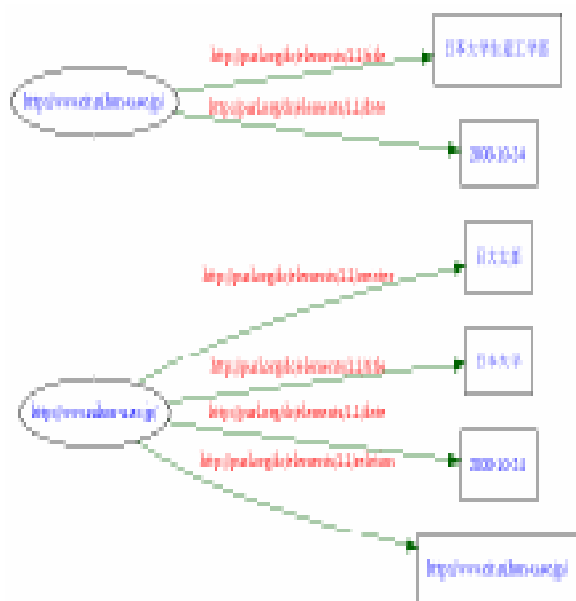


図5 図3の有効ラベル付きグラフの検証結果

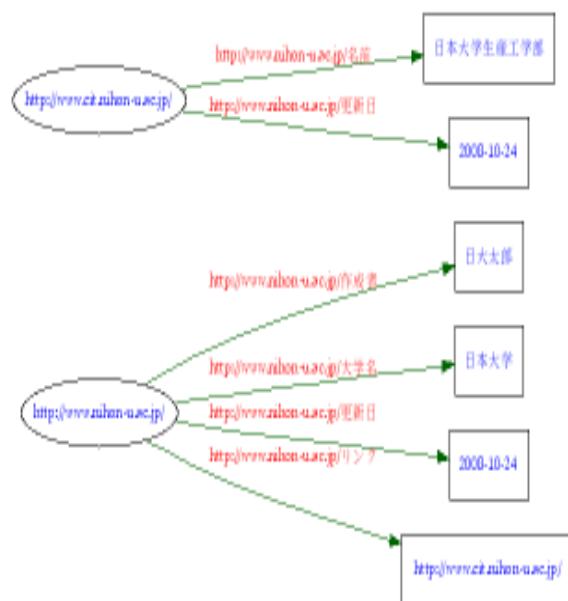


図6 図4の有効ラベル付きグラフの検証結果

要素が用意されている。例えば、基本要素名の data だと何を示す日付であるか意味が曖昧だが created(作成日)、modified(更新日)などの精密化要素名を用いることにより、意味を明確にすることができる。

3.3 多項関係のRDFは二項関係のRDFへ分解

図2のRDFトリプルは、xをリソース、Pをプロパティ、yをオブジェクトとすると式(1)で表現できる。

$$(x, P, y) \rightarrow P(x, y) \quad (1)$$

すなわち、図2のRDFトリプルを式(1)の論理式を使って表現すると式(2)になる。

$$\text{Site-owner}(\text{http://}\sim, \text{日本大学}) \quad (2)$$

一般的に表現されている二項関係のRDFトリプルは式(1)や式(2)である。しかし、RDFは、二つ以上の項を持つ多項関係を使用する場合の方が多い。RDF/XML文書の設計でリソースに対してプロパティとオブジェクトが追加されてRDFが多項関係になるのであれば、RDFを二項関係で表現する必要がある。

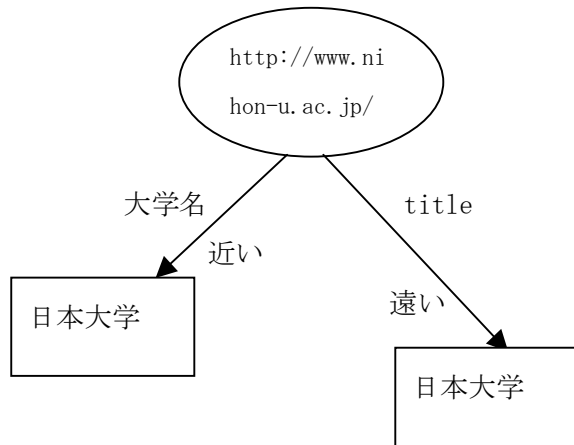


図 7 RDF による距離の表現

3.4 RDF/XML 文書の作成

ダブリン・コア要素を使用した RDF/XML 文書と使用していない RDF/XML 文書の二つの RDF/XML 文書を図 3、図 4 に示す。

3.5 RDFVS による検証

RDFVS は W3C の RDF Validation Service と呼ばれる検証ツールである。RDF/XML 文書を解析し、リソース、プロパティ、オブジェクトを列に属性として指定して RDF トリプルや有効ラベル付きグラフをユーザが選択して出力することができる。この RDFVS を用いて 3.4 で作成した二つの RDF/XML 文書の整合性検証を行う。

4 検証結果

二つの RDF/XML 文書を検証した結果、表 1 は図 3 を RDF トリプルで検証した結果であったが、同様に図 4 も同じく検証できた。また、どちらの RDF/XML 文書も図 5、図 6 のように有効ラベル付きグラフとして出力できた。図 3 の〈dc:title〉と図 4 の〈大学名〉のプロパティを比較すると、〈大学名〉の方が〈dc:title〉よりも距離概念が近く感じられた。それは〈大学名〉が〈dc:title〉よりも抽象化を突き詰めた概念を持っているからである。このことから、ダブリン・コアは抽象化を突き詰めた基本

属性をトップで決めており、メタデータをトップダウンアプローチしていくものであると思われる。しかし、リテラルの属性を考えて上位概念へ向かってメタデータを構築していくことを考えるとボトムアップアプローチの点も考えられる。RDF/XML 文書を作成する際はリソースとリテラルの距離概念を測ると同時に、トップダウンとボトムアップの両方からのアプローチを RDF/XML 文書の設計の時に考える必要があることが分かった。

5 まとめと今後の課題

このようにオントロジーを構築していくには、図 7 のようにプロパティの設定だけではなく、オブジェクトとオブジェクト間の距離を客観的に分析して考えていく必要がある。また、RDF/XML 文書を構築する時にはダブリン・コア要素の属性を規定する使い方だけではなく、ダブリン・コア要素の記述方法を決めていくことが重要である。今後の課題は、ダブリン・コア精密化要素を利用した RDF/XML 文書を作成し、RDF の距離概念をベースにした新たな RDF/XML 文書の再構成方法について考える。

「参考文献」

- 1) 神崎正英,「セマンティック・ウェブのための RDF/OWL 入門」, 森北出版, (2005), pp. 124~133.
- 2) AIDOS,「オントロジ技術入門」, 東京電機大学出版局, (2005), pp. 32~36.
- 3) 溝口理一郎,「オントロジー工学」, 人工知能学会, (2005), pp. 52~61.
- 4) 北海道情報専門学校,「改訂 XML 入門」, SCC 出版局, (2006), pp. 29~39.
- 5) 高橋勝敏, 佐藤俊也, 原島一郎, Web コンテンツからのオントロジーの再構成方法の提案と試作, 人工知能学会論文誌, 20 巻 6 号, (2005), pp. 419~420.