

Web 検索支援システムのための RDF エディタの開発

佐野公彦[†] 武井恵雄^{††} 荒井正之^{††}

あらまし 氾濫する WWW 上の情報の中から、求める情報を的確に探し出す方策として、Semantic Web が提案されているが、誰が Semantic web を書くのか、また書けるのかという重要な問題がある。我々は、これを部分的に解決することを目標に、意味論上の検索を可能とする RDF ベースの Web 検索支援システムを開発している。今回は、誰でも RDF 構文のメタデータを記述できるエディタの開発について報告し、学習システムの開発にも有用であることを述べる。開発する RDF エディタは、RDF 構文を表現するために用いられる XML やラベル付き有向グラフといったものを意識することなく、RDF 構文のメタデータを記述することができる。

Development of an RDF Editor for a Web Search Support System

Kimihiko SANO[†], Shigeo TAKEI^{††}, and Masayuki ARAI^{††}

Abstracts: Although the Semantic Web has been proposed as an ideal method for solving information retrieval on WWW, we have intrinsic questions concerning on who writes semantic web. We are developing an RDF based Web search support system which enables semantic searching the information on WWW. In this paper we report the design and development of an RDF editor, which is used for the Web search support system. This RDF editor can describe the meta-data of RDF structure without knowing the XML or the directed graph.

1. はじめに

WWW(World Wide Web)上に存在する Web サイトは爆発的に増え続けている。このような莫大な情報量の中から、目的とする情報を探し出すことは大変困難である。WWW 上から情報を探すための手段として Web 検索サイトがあるが、初期の頃に比べれば検索効率は向上したものの、カテゴリやキーワードだけに頼るような検索には限界がある。そのため、Web リソースの意味情報を考慮した

検索が必要である。

このような状況を解決するため、Tim Berners-Lee は Web リソースの持つ意味情報を機械に処理可能なデータにする一つの方式として、Semantic Web を提唱した[1]。しかし、その仕様の複雑さもさることながら、作成者自身による意味情報のカテゴライズ自体が高度な能力を要求するため、HTML のように誰もが記述できて、世間一般に普及するか疑問である。浦本[2]は、Semantic web

[†] 帝京大学大学院理工学研究科, Graduate School of Science and Engineering, Teikyo University; sano@mil.ics.teikyo-u.ac.jp

^{††} 帝京大学理工学部情報科学科, School of Science and Engineering, Teikyo University; {takei,arai}@ics.teikyo-u.ac.jp

が普及する条件として、オーサリング環境をはじめとするツールや、キラーアプリケーションの出現が必要不可欠だと述べている。

そこで、Semantic web の普及を待つことなく、Web リソースの意味情報を考慮した RDF(Resource Description Framework)ベースの Web 検索支援システムを開発し、検索効率の向上をはかることを考えている。本研究では、検索に用いる RDF を誰にでも記述することのできる RDF エディタの開発を行う。

2章では、Semantic Web の概要と有用性について述べ、3章では、本研究で活用する RDF と RDF スキーマについて述べる。4章では、開発する Web 検索支援システムの概要について述べる。5章では、開発中の RDF エディタの概要と、実際の機能について述べ、6章で、まとめと展望を述べる。

2. Semantic Web

Semantic Web とは、図1に示すような XML(eXtensible Markup Language)を利用したさまざまな技術によって構成されており、Web 上のリソースにメタデータ(データに関するデータ)を付与することで、機械によ

って処理可能なデータにし、Web の有用性を高めようという活動である[1]。

現在の Web においても、ある程度のメタデータは HTML のタグとして付与されている。例えば、作成者、作成日付、タイトルなどである。しかし、現在の Web コンテンツは、基本的には単なる書式付き文字列であり、検索エンジンによる Web 検索は文字列検索を行なっているに過ぎない。そのため、ユーザが得たい情報とは関係のない情報も検索結果として表示されてしまい、その結果を頭から調べて必要な情報を探し出すのは人間の仕事になる。これは、今の検索エンジンが、例えば「診療日」というデータを3文字の文字列としてしか見ておらず、その本当の意味を理解していないからである。

Semantic Web ならばメタデータが付与されているためにこのようなことはなく、「診療日」という文字列に含まれる意味情報を理解し、精度の高い Web 検索ができるようになると思われる。本研究では、Semantic Web に使用されている技術の中から、RDF と RDF スキーマだけを取り出し、Web 検索支援システム及び、RDF エディタの開発に活用する。

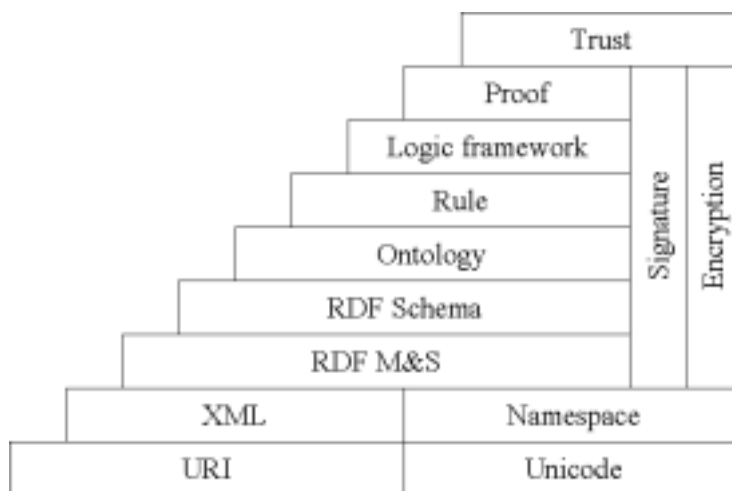


図1 Semantic Web の階層構造

3. RDF と RDF スキーマ

RDF[3]とは、Web リソースに対するメタデータを記述するための枠組みである。RDFの基本データモデルは次の3つから構成される。

リソース

RDF で記述されるものはすべてリソースと呼ばれる。リソースは Web サイト全体でも個々の HTML 文書でも構わない。リソースは必ず URI(Uniform Resource Identifier)およびフラグメント id で指定される。

プロパティ

リソースを記述するために用いる特徴、関係、属性のことをプロパティと呼ぶ。それぞれのプロパティは、その意味、許される値の範囲、記述できるリソースのタイプ、他のプロパティとの関係などがスキーマによって定められており、URI で参照される(つまりプロパティもリソースの一種ということになる)。

ステートメント

あるリソースとプロパティ+値の組み合わせを結びつけたものを、RDF の文(ステートメント)という。この3つの要素はそれぞれ主語(subject)、述語(predicate)、目的語(object)と呼ばれる。ステートメントの目的語は、別のリソースもしくはリテラル(文字列あるいはXMLの基本データ型)となる。

RDF では、これらの関係をラベル付き有向グラフ(directed labeled graph)を用いて表現する。RDF のグラフでは、リソースを楕円で、プロパティをアーク(矢印)で、リテラルを長方形で表現し、アークの矢印は主語を始点とし、目的語を終点としている。このような RDF の構造は、意味ネットワーク

(Quillian's Semantic Network)の基本要素と同じだが、述語表現に使用しているところが機能的である。

例えば、「http://test.com の作者は佐野公彦です」という文の場合、図2のように表すことができる。また、「http://test.com の作者はhttp://sano.com で参照される人物で、名前は佐野公彦、メールアドレスはsano@sano.comです」という文の場合は図3のように表せる。

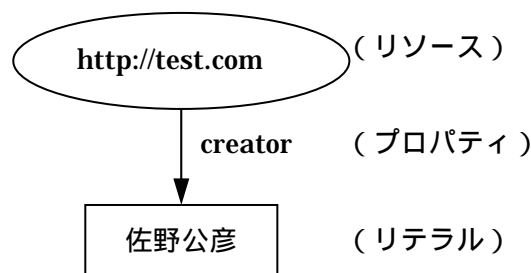


図2 RDF データモデル1

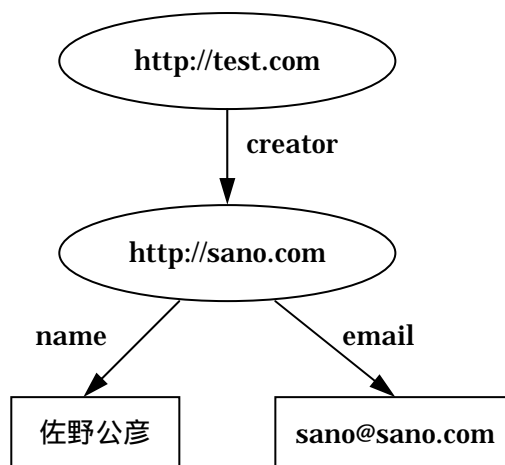


図3 RDF データモデル2

このデータモデルをXMLとして表現すると、図4のようになる。この構文では、Description要素が「文」を示し、about属性が「主語」となるリソースのURIを、要素の内容が「述語」にあたるプロパティ(creator, name, email要素)と「目的語」となるリテラル値(佐野公彦, sano@sano.com)を記述している。本研究で使用するのは、このXML表現のRDFである。

RDF スキーマ[4]は、RDF を使って記述されたメタデータの意味的な制約を定義するものである。対象に対してどのようなプロパティのメタデータを付けるのか、また、その

値としてどのような型になるのかなどを規定できる。さらに、プロパティ間のクラス階層なども指定できる。

```

<rdf:Description rdf:about=" http://test.com">
  <creator>
    <rdf:Description rdf:resource =" http://sano.com " >
      <name>佐野公彦</name>
      <email> sano@sano.com </email>
    </rdf:Description>
  </creator>
</rdf:Description>

```

図4 図3のXML 構文での表現

4. 開発中の Web 検索支援システム

検索は、Web リソースと XML 表現で記述された RDF 構文のメタデータを用いて行い、これらは、WWW 上から Web クローラ(Web Crawler)であらかじめ収集しておく。

検索の中核となる Web リソースの RDF 構文文化は、Web サイトの作成者に行ってもらい、しかし、RDF 構文の記述は HTML に比べると複雑なため、誰にでも記述できるような RDF エディタを用意する。

検索の流れは図5に示すように、まず、ユーザは入力ボックスにキーワードを入れる。そのキーワードを RDF ファイルに照らし合わせ、一致したものの属性を順に表示させる(例えばキーワードが「宮沢賢治」ならば、属性は **人名 作家 小説家** などとなる)属性をクリックすると、その属性が付加されている Web リソースの一覧が表示される。また、キーワードや属性との一致が不完全だったものは、完全だったものの後に順次表示される。

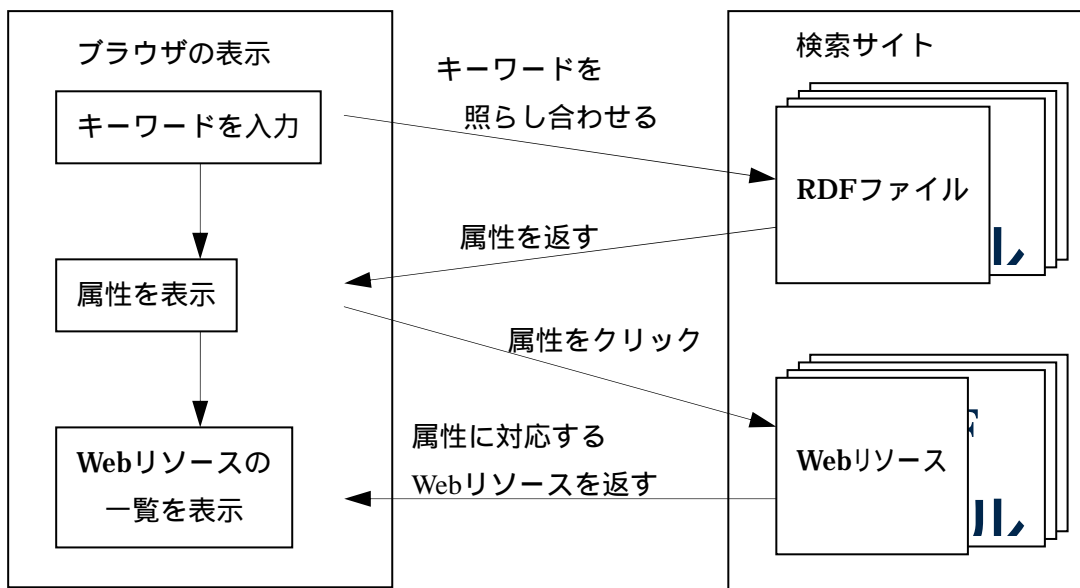


図5 RDF ファイルからの検索

5. RDF エディタ

Semantic Web や Web 検索支援システムにおいて, RDF 構文を XML 表現で書く作業は労苦が多い.そこで,誰にでも簡単に RDF 構文を XML 表現で書くことのできるエディタが必要であり,本研究ではこのエディタを開発する.

5.1 今までの RDF エディタ

現在ある RDF エディタとして, W3C が発表した IsaViz[5]がある. IsaViz は, ラベル付き有向グラフを用い視覚的に RDF を記述できるツールである.そのため, RDF を XML 構文で直接手書きするよりはメタデータを作成し易い.しかし,図 6 のように RDF の構造が小さいうちは良いが,図 7 のように構造が大きく複雑になるにつれ全体像が把握しにくくなり,記述がかえって困難になる.また,ラベル付き有向グラフの書き方の詳細を理解しなければならない.

また, RDF ではないが, RDF スキーマに基づいて HTML やテキスト文書にメタデータを半自動的に付与する,文献[6]のような研究も行われているが,実用化にはまだ時間が掛かるようだ.

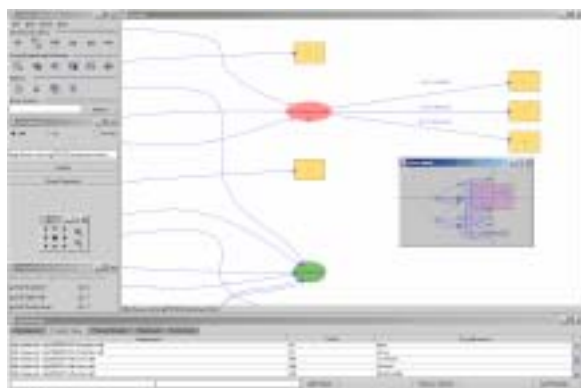


図 6 IsaViz の作成画面 1

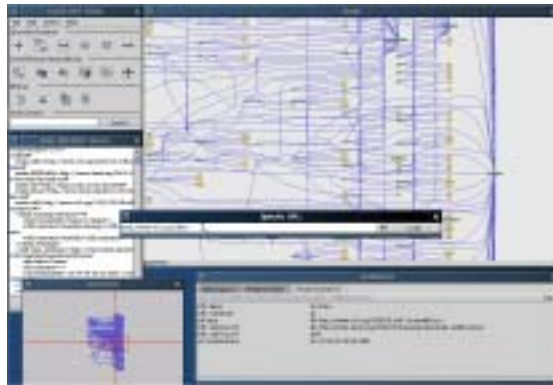


図 7 IsaViz の作成画面 2

5.2 開発した RDF エディタの構成

本研究で開発する RDF エディタは, IsaViz のように RDF 構文をラベル付き有向グラフで表現せず,図 8 のようにフレーム状にした入力ボックスと選択ボックスで表現する.

実際にユーザがメタデータを付与する画面は図 9, 図 10 のようになる. 図 9 では, 選択ボックスから Web リソースのファイルタイプや種別などを選択し, 入力ボックスに URI やリテラルを入力することにより, RDF 構造を意識することなく Web リソースにメタデータを付与することができる. ボックスの左側には RDF スキーマで定義したプロパティのラベルを表示し, ユーザが間違っただメタデータを書かないように誘導する.

図 10 では, 入力ボックスと XML 表現の RDF 構文を表示している. 両方を同時に表示することで, 入力ボックスへの入力でどのように RDF 構文が記述されるかわかる. また, この画面では, RDF 構文を XML 表現で直接手書きすることができ, その記述は入力ボックスに反映される.

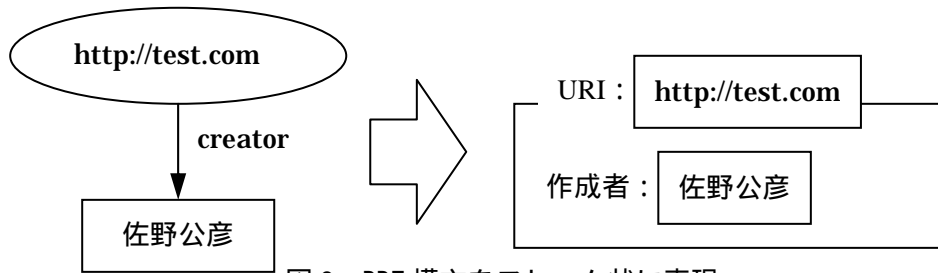


図 8 RDF 構文をフレーム状に表現

図 9 RDF エディタの作成画面

```

<?xml version="1.0" encoding="Shift_JIS" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www
<rdf:Description rdf:about="http://test.com">
  <site>
    <rdf:Alt>
      <rdf:li xml:lang="ja">テスト</rdf:li>
      <rdf:li xml:lang="en">Test</rdf:li>
    </rdf:Alt>
  </title>
  <content-type>top</content-type>
  <creator rdfs:type="Resource">
    <name xml:lang="ja">佐野公彦</search:name>
  
```

図 10 入力ボックスと RDF 構文を同時に表示する画面

5.3 RDF エディタの活用

RDF ファイルの作成は、図 11 に示すような流れになる。まず、ユーザはメタデータを付与したい Web リソースを用意し、それを RDF エディタに読み込ませる。RDF エディタは、読み込んだ Web リソースに対する RDF ファイルを生成し、その Web リソース

が HTML だった場合、Web リソースのタイトル、作成者、他の Web リソースへのリンクなどのような簡単なメタデータの付与を、meta タグや title タグなどを見て自動で行う。

その後、自動では付与できないような属性などのメタデータをユーザが付与することになる。

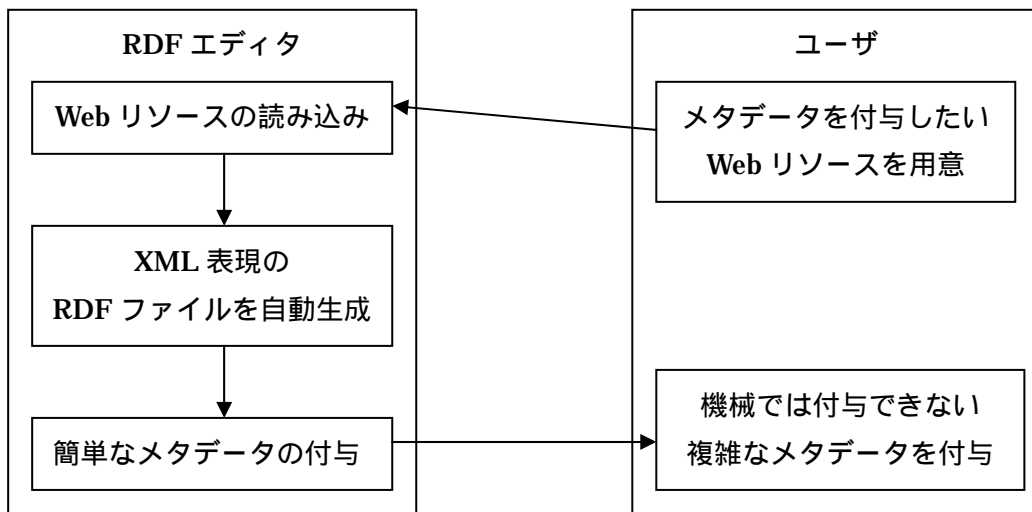


図 11 RDF ファイルの作成手順

6. おわりに

本研究では、意味情報を考慮した Web 検索支援システムの開発を進めており、今回は RDF エディタの開発を行った。開発中の Web 検索支援システムは、RDF により必要なメタデータが付与されるため、カテゴリ検索やキーワード検索よりも効果的な意味情報の検索が可能になる。また、開発した RDF エディタならば、ラベル付き有向グラフや XML 表現を理解できないようなユーザでも、手軽に RDF 構文を書くことができる。

この RDF エディタを使うことで、学習システムの開発にも RDF を活用することが容易になると考えている。人間の知識構造や多重概念の表現に RDF を用いれば、効率的な学習システムの開発が可能になるからである。

今後の課題として、RDF エディタの有用性の検証を行う必要がある。このようなシステムは今までに例がないので、分野の異なる人たちにメタデータを付与してもらおう必要がある。知識構造が異なれば、メタデータの付け方も変わり、ある人にとっては有用でも、ある人にとっては有用ではないということがわかると考えている。この結果を RDF エ

ディタの改良にぜひ反映させたい。また、付与したメタデータが本当に検索に有用かどうか、開発中の Web 検索支援システムを用いて大規模な評価実験を行う必要を感じている。

参考文献

- [1] W3C : Semantic Web, <http://www.w3.org/2001/sw>
- [2] 浦本 直彦 : Semantic Web 機械のための Web , 人工知能学会誌, 16 巻, 3 号, pp.412-419(2001) .
- [3] W3C : RDF/XML Syntax Specification (Revised) , <http://www.w3.org/TR/rdf-syntax-grammar/>
- [4] W3C : RDF Vocabulary Description Language 1.0: RDF Schema, <http://www.w3.org/TR/rdf-schema/>
- [5] W3C : IsaViz: A Visual Authoring Tool for RDF, <http://www.w3.org/2001/11/IsaViz/>
- [6] 大沼宏行, 松平正樹, 淵上正睦, 森田幸伯 : Web コンテンツの分析に基づくオントロジ構築および属性抽出の試み , 情報処理学会 研究報告 第 157 回自然言語処理研究会 2003-NL-157, pp.49-54(2003)