

# 「使える」ウェブを作ろう: Web の RDF について

吉賀夏子<sup>1</sup>、只木進一<sup>2</sup>

佐賀大学総合情報基盤センター<sup>3</sup>

何か分からないことがあれば、本や雑誌を読む前にとりあえずググってみる(=Googleで検索する)というのが、現代のモノの調べ方の主流ではないかを感じる今日この頃です。大学のような教育・研究機関ではなおさら利用しない訳にはいかないのではないのでしょうか？

実際に検索エンジンを使うと、大抵は、ある単語や小さな文に対して何万もの結果がでてきます。Google や Yahoo など大手の検索エンジンは、自分で世界中のウェブサイトからウェブサイトへと自動的にネットサーフィンして、データをまるごと自社のサーバに貯め込んでいきます。貯まったデータは利用頻度やウェブサイトのルール、スポンサーからもらった広告料などから、独自のランク付けをして、ユーザに表示するという方法を採用していますが、よく見ると、膨大な検索結果が出た割には全く同じ文章を貼り付けていたり、広告リンクが並んでいるだけだったりということがかなりあり、欲しい情報がなかなか見つけられません。

しかし、自分の研究分野である場合は現在過去、そして国内外と、なるべく広くアンテナを貼って毎日情報をこまめに収集していかなくてはならないはずです。こんな時は、一体どうしたら効率よく必要な情報を集められるのでしょうか？

逆にウェブサイトを作る側では、どのようにウェブを作っていけば多くの人に関心を寄せてもらえるようになるのでしょうか？

## 意味のあるウェブ: セマンティックウェブへの移行

今までのウェブページは、ある事柄に対して個人が不特定多数に対して「発表」した産物でした。しかし、そのページの内容、特に他のウェブリソース<sup>4</sup>とどのように関連しているのかという事までは、検索エンジンが使うデータ収集用のソフトウェアなどには理解できません。そのため、ある一定の「枠組み」を作ろうという動きが近年起こりました。

それが「**セマンティック(=意味のある、意味論の)ウェブ**」(Semantic Web, 略して SemWeb)というコンセプトです。セマンティックウェブに基づいたウェブリソースを作るには、あらかじめ決まったルールで分類し、構造化された「意味のある情報」をコンピューターが保持しておきます。すなわち **RDF(Resource Description Framework)**<sup>5</sup>というデータのためのデータ(=メタデータ)を今までのリソースにテキストデータで追加します。メタデータにはタイトルや URL、ジャンル、更新日、著者、具体的なデータなどを記述しておき、ウェブサイト同士で「このリソースは何について書かれているか」ということが明確にわかるようにしておきます。そのデータを利用者が自分で集め、カスタマイズします。

## RDF の概要

既に、世の中には大量のセマンティックでないウェブリソースが存在しています。RDF を個々のウェブページに追加したい場合は、その内容がどのようなものなのか、まずデータを RDF の枠組みに沿って整理する必要があります。

1 natsukoy@cc.saga-u.ac.jp

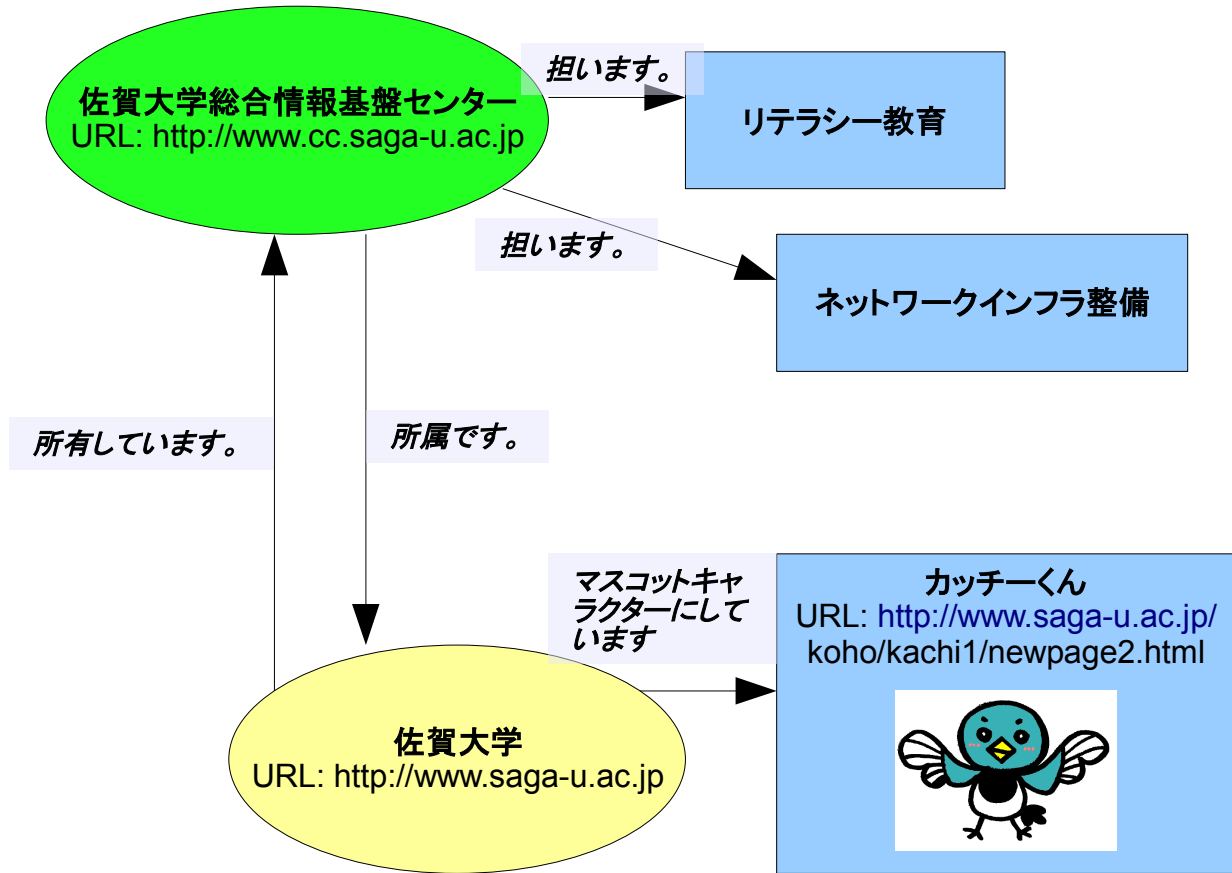
2 tadaki@cc.saga-u.ac.jp

3 <http://www.cc.saga-u.ac.jp/>

4 インターネット上にある情報資源全般(ホームページ、データベースなど)

5 <http://www.w3.org/RDF/>

RDF では、ある情報を主語・目的語・述語の三つの要素から組 (Triples) になるようにデータを作るのが基本です。例として「佐賀大学」と「佐賀大学総合情報基盤センター」についての情報をタグを使って RDF 的に整理してみます。



図中の楕円がトピックの主題(主語)、矢印がどういう関係をもつか(述語)、四角が主題に対して何を目的にしているか(目的語)にそれぞれ対応しています。

上図の中から一部を文章にしてみます。

【主語】	【目的語】	【述語】
a. 佐賀大学(は)	佐賀大学総合情報基盤センター(を)	所有しています。
b. 佐賀大学総合情報基盤センター(は)	リテラシー教育(を)	担います。
c. 佐賀大学(は)	カッチーくん(を)	マスコットキャラクターにしています。

このような文章は a,b,c 以外にも、このデータを作る人によってどんな風にも作れます。自由に作れる代わりに、多くの人々がどれだけ正確に、迷いなく欲しいデータにたどり着けるのか手腕が問われるところです。したがって、これらの文は互いに関連があるべきで、その情報(ここでは「佐賀大学」や「総合情報基盤センター」)が確かにその情報であり、同じ名前の別物でないように表現されなくてはなりません。

3つの文を RDF/XML フォーマットで表記すると、以下のようになります。

<rdf:RDF> ----->①

```
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:edu="http://www.saga-u.ac.jp/">
```

---->②

```

<rdf:Description rdf:about="http://www.saga-u.ac.jp/">

  <edu:Faculties rdf:resource="http://www.cc.saga-u.ac.jp/" dc:title="佐賀大学情報基盤セン
ター"/>
  <edu:Mascot rdf:about="http://www.saga-u.ac.jp/koho/kachi1/newpage2.html">
    <dc:title="カッチーくん" />
    <edu:img rdf:about="http://www.saga-u.ac.jp/koho/kachi1/21224kachi1.gif" />
  </edu:Mascot>
</rdf:Description>

<rdf:Description rdf:about="http://www.cc.saga-u.ac.jp/">
  <edu:Works rdf:about="http://www.cc.saga-u.ac.jp/outline/works.php">
    <dc:title="リテラシー教育" />
  </edu:Works>
</rdf:Description>

```

---->③

```

</rdf:RDF>

```



RDF 形式でメタデータを作成するには Notation 3 Format や XML フォーマットなどいろいろありますが、やることはほぼ同じです。①の部分で RDF 形式のメタデータであることを宣言し、②で名前空間、③で上記 a,b,c の関係をタグで階層化します。③の<rdf:Description rdf:about="http://www.cc.saga-u.ac.jp/">というパートで主語を宣言し、そのタグの中で述語+目的語(具体的な URL と主語に関連する情報)を記述していきます。

例えば edu というラベルのついたウェブサイト(=<http://www.saga-u.ac.jp/>)があり、そのなかに施設一覧を参照できる場所があるとします。(=faculties) そのページに総合情報基盤センター(=<http://www.cc.saga-u.ac.jp/>)という場所(URL)があると記述することができます。文 a,c のように主語が同じであれば、まとめて記述できます。このような RDF 形式のテキストデータをウェブサイトを追加すれば、セマンティックウェブサイトになります。

詳細は、この記事の末尾にある参考 URL・文献を参照してください。

## RDF 形式のデータを使ったウェブサービスの紹介

RDF の規格に沿ったサービスのひとつに、RSS(RDF Site Summary<sup>6</sup>または Really Simple Syndication<sup>7</sup>)フィード機能というものがあります。RSS を使えば毎日いろんなサイトを手で巡回しに行かなくても、専用のソフトウェアやウェブサービス上で、新しく更新されたニュースだけをまとめてリスト表示できます。

まず、「RSS」(  や  というアイコンで URL の横やホームページに表示されています。)という表示のあるサイトやブログを探します。

例えば、総合情報基盤センターウェブサイトでは「総合情報基盤センターニュース」のタイトル・緊急連絡・最新ニュース・佐賀大学ネットユーザ向けセキュリティ対策メッセージとニュースが何種類もあります。現在、これらのニュースをまとめて RSS ファイルにして配信しています。利用者は、その場所を RSS フィーダ(リーダー)に登録すれば、情報が更新されると、すぐにそのニュースを読めます。

<sup>6</sup> <http://web.resource.org/rss/1.0/>

<sup>7</sup> <http://cyber.law.harvard.edu/rss/rss.html>

具体的には Firefox<sup>8</sup>の場合、メニューの「ブックマーク」から「このページを購読...」を選びます(ホームページに RSS 機能がない場合は選べません)。あとはダイアログが出てくるので、ブックマークを保存するように設定すればできあがりです。ブックマークのフォルダを長押しすると、自動的にニュースのタイトルの一覧が作成されます。興味のある記事を選ぶと、その記事の載っているページに移動できます。記事のリストは、ニュースサイトでそれぞれ更新時間が設定され、定期的に更新されています。このようなニュースフィードは、ブラウザ以外にフリーウェアで電光掲示板のようにティッカーになっていてデスクトップに表示できるものや、ブラウザの機能拡張で、各サイトのニュースを1ページにまとめて表示するサービスがたくさんあります。さらに Google アラート<sup>9</sup>や理系でよく使われている論文検索サービス「PubMed<sup>10</sup>」では、キーワードで論文を検索し、RSS フィードとして定期的に検索結果を受け取れます。

ぜひあなたに必要なウェブサイトの RSS フィードを見つけて、研究や情報収集にご活用ください。

## 本当に使える情報は実は少ない

RDF にはデメリットもあります。データを自動的に自分のサイトに取り込めるため、同じ情報の一部や全体を自動的に大量に貼り付けているだけのブログやアフィリエイト系のサイト(= 広告リンクをクリックするとサイト管理者に謝金が支払われるようになっている)が氾濫して、サーチエンジンによる検索結果を無駄に増やしています。特に携帯電話用のサイトはひどい状況です。そのため、ウェブサイト上に RDF 規格のデータを追加し、本当に情報を必要としている人が能動的に、迷いなく情報を受け取れるようにするのは、これからのウェブサイトには必須の条件になるでしょう。

## 参考文献

1. Joshua Tauberer: “What is RDF?”, XML.com, <http://www.xml.com/pub/a/2001/01/24/rdf.html>
2. 新納浩幸: “入門 RSS”, 毎日コミュニケーションズ, ISBN4-8399-1649-7
3. W3C: “RDF/XML Syntax Specification (Revised)”, <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>
4. “既存のデータをセマンティック Web で利用する”, <http://www.ibm.com/developerworks/jp/xml/library/x-semweb.html>

---

8 <http://mozilla.jp/firefox/>

9 <http://www.google.co.jp/alerts>

10 <http://www.pubmed.gov>