

B-034

RDFa によるメタデータを活用した Web コンテンツ配信技術

Web contents distribution utilizing metadata expressed by RDFa

佐伯嘉康[†]
SAEKI Yoshiyasu

相田仁[†]
AIDA Hitoshi

1 序論

インターネットが普及し、Web 上には様々なコンテンツが存在している。

利用者が Web コンテンツを発見するためには、個人または団体が設置した検索エンジンを用いて、Web コンテンツに含まれるであろう単語を検索クエリとする。しかし、人手により検索データベースを構築しているディレクトリ型検索エンジン、エージェントに Web を巡回させ検索データベースを構築するロボット型検索エンジンのいずれも、膨大な量となった Web コンテンツを全て網羅しているとは言えず、Web 上に公開した情報、知識が等しく活用されるためには、新たな機構が必要となっている。

そこで本研究では Web コンテンツを、Web コンテンツに付与したメタデータを活用して分類、集約する事で Web コンテンツを配信し、利用者が Web コンテンツを活用する際の支援となるシステムを構築する。

2 RDFa によるメタデータの記述

2.1 RDF

現在の Web を見直し、次世代の Web を構築するプロジェクトとして、セマンティック Web [1][3] がある。1999 年に現在の Web の考案/開発者である Tim Berners-Lee が提唱して以降 W3C を中心とし、セマンティック Web を実現するための技術が、現在標準化に向けて研究されている。

セマンティック Web が目指すものの 1 つとして、メタデータを付与した Web コンテンツからデータの交換のための普遍的な媒体を形成し、データが持つ情報によってデータの(自動)機械処理を可能にするというものがある。セマンティック Web は、データの関係と分類を改善するために必要なプロトコルと技術を確立する事により、システム間やデータ間関係を生成したり推論を行ったりする能力を強化する。

そして、メタデータを構成するためのセマンティック Web 技術として、RDF(Resource Description Framework) が採用されている。RDF は主語 (Subject)、述語 (Predicate)、

目的語 (Object) の 3 つの要素のセット (三つ組, Triple) によって構成、記述されるデータモデルである。述語は RDF の中でタグとして使われ、語彙ともいう。述語は意味が重複 (衝突) しない様に、名前空間を定義、関連付ける事で、Web など分散環境での RDF データ構築をスムーズにする事を可能にしている。三つ組の関係は、文で書くと「[主語] の [述語] は [目的語] である」となり、図 1 のグラフとして表すことが出来る。

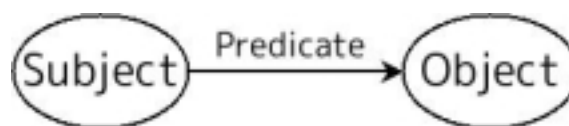


図 1 RDF Triple Graph

目的語には、定数値や空 (Empty) のデータの他にリソース (URI によって指し示されるもの) を定義する事が出来る。従って RDF によって記述されたメタデータは、ファイルによって完結せず、複数のメタデータによって意味を持つ事もある。

2.2 RDFa

RDF により構成されたメタデータの記述には、XML(Extensible Markup Language) による RDF/XML, 1 行に三つ組を記述する N-Triples などがあるが、現状に於いて Web 上に多く存在しているファイルは (X)HTML によって記述されたファイルであり、また RDF に対応したアプリケーションが少ない事から、RDF によるメタデータの普及のためには RDF と XHTML とを上手く折り返合わせる事が良策であると考えられる。

そこで本研究では、RDF メタデータの記述方式として、XHTML のタグ内に属性 (Attribute) として記述する RDFa(RDF/A, RDF with attributes) [4] による記述を採用する。RDFa は W3C によって提案され、仕様は現在 Working Draft の状態にある。

RDFa により、XHTML ファイル内にメタデータを記述した例を図 2 に示す。

3 メタデータと Web コンテンツ配信

Web コンテンツがメタデータを持つ事で、機械による Web コンテンツを処理が可能となり、人が Web コンテン

[†] 東京大学大学院新領域創成科学研究科基盤情報学専攻

```
<div
xmlns:exf="http://example2.com/family#"
xmlns:exp="http://example2.com/person#"
about="http://example.com/Alice">
  Description about Alice.
  Alice's father is
  <a href="http://example.com/Bob" rel="exf:father">
    this person
  </a>
  , and his name is
<span
about="http://example.com/Bob" property="exp:name">
  Bob
</span>.
</div>
```

図2 Example of RDFa

ツを理解する前に、機械が Web コンテンツの意味を解釈することが出来る。

本研究では、Web コンテンツの配信を行う。配信とは即ち、テレビあるいは雑誌の様に、内容が関連した Web コンテンツを集約し、機械あるいは人の手による編集を加え、利用者に対してプッシュ型で内容を示しながら提供する事である。

プッシュ型の提供をする事の利点として、Web コンテンツの作成者が他の Web コンテンツと結び付け、利用を促す事が出来る、Web コンテンツのタイトルの列挙でなく、関連した Web コンテンツの内容を同時に見せる事ですぐに Web コンテンツを利用する事が出来る、という事が挙げられる。

現在、本研究で用いている Web コンテンツが持つメタデータ語彙を以下に列挙する。

- title (Web コンテンツの題名)
- creator (Web コンテンツの作成者)
- publisher (Web コンテンツが属する Web サイト)
- description (Web コンテンツの概要, 本文)
- date#created (Web コンテンツの作成日時)
- date#updated (Web コンテンツの更新日時)
- related#common (同一のテーマの Web コンテンツへのリンク)
- related#reference (参考とした Web コンテンツへのリンク)
- related#postscript (追記元に当たる Web コンテンツへのリンク)
- related#reedit (再編集元に当たる Web コンテンツへのリンク)

ここで使用するメタデータの語彙の名前空間は <http://rdf.aida.k.u-tokyo.ac.jp/rdfs/webc> とし、この URI によって示される空間で語彙が定義されているものとする。語彙の選択については、RSS と書誌情報の公開のための語彙集である Dublin Core を参考にした。

Web コンテンツの集約として、ハイパーリンクのメタデータ (related#) を活用する。出来れば、Web コンテンツ

の閲覧の際には、ビジュアルとして直感的に関連が分かる様なためのインタフェースを提供したい。

RDFa によるメタデータを付与した Web コンテンツの一部の例を図 3 に示す。

```
<div
xmlns:webc="http://rdf.aida.k.u-tokyo.ac.jp/rdfs/webc"
about="http://example.com/~saeki/rdfabook">
<span property="webc:title">RDFa Book review</span>
<span property="webc:creator">Saeki</span>
<span property="webc:publisher">Example Co.</span>
<span property="webc:description">
  This is a "RDFa Book" Review by Saeki.</span>
<span property="webc:date#created">2007-07-02</span>
<a href="http://example.com/~yoshi/rdfabook"
rel="webc:related#common">
  Yoshi's RDFa Book review is here.</a>
</div>
```

図3 Example of Web contents with Metadata

Web コンテンツを作成すると同時に、Web コンテンツの内容に関するメタデータを記述する手間が、Web コンテンツの作成者に増える事を考えれば、出来るだけその負荷は軽減するべきである。そのための手法としては、適切なインタフェースの提供、他者との連携、機械と連携したメタデータの自動作成/協同付与が考えられる [2]。

厳密なメタデータを始めから人の手で充実させるのではなく、例えば Folksonomy や Web コンテンツ内の頻出語句の解析など、セマンティック Web の分野の外で行われている研究と上手く組み合わせる事で、メタデータの素を作り、Web 上に RDF と互換のあるメタデータを充実させる事が、セマンティック Web の普及のために求められている事だと考える。

4 結論

以上で、RDFa によって記述されたメタデータを活用した、Web コンテンツ配信技術について述べた。

今後は、Web コンテンツが増えた場合のスケーラビリティの調査を行い、より効率的な Web コンテンツとメタデータ管理機構を探る予定である。

参考文献

- [1] 斎藤信男・荻野達也 (監修), 財団法人 情報処理相互運用技術協会 (編), “セマンティック Web 入門”, オーム社, 2004
- [2] 岡田謙一, “情報共有空間における協同”, 情報処理 2 月号 2007 Vol.48 No.2 通巻 504 号, pp.123-127, 2007/2
- [3] SICO P (Semantic Interoperability Community of Practice), “Introducing Semantic Technologies and the Vision of the Semantic Web”, <http://colab.cim3.net/file/work/SICO P/WhitePaper/>, 2005/2
- [4] W3C, “RDFa Primer 1.0”, <http://www.w3.org/TR/xhtml1-rdfa-primer/>, 2007/3