

セマンティックWeb適用システム

メタデータ付与によるセマンティック
Web技術の情報ポータル

佐藤宏之 (日本電信電話株式会社)
今村誠 (三菱電機株式会社)
伊藤山彦 (三菱電機株式会社)
小泉雄介 (株式会社NEC総研)
香取良和 (財団法人情報処理相互運用技術協会)
清水智公 (慶應義塾大学大学院 政策・メディア研究科)

概要

- セマンティックWebの基盤技術(RDF)の活用例
 - 身近なRDF:RSS
 - FOAF
- RDFを利用したセマンティックWeb技術のポータルシステムの紹介
- メタデータはどこに置くべきか(RDFメタデータの設置ガイド)

RDFはどこに？

- googleで「rdf filetype:rdf」を検索
 - rdfの拡張子を持つWeb上のファイル
 - 122,000件(2003年11月)
 - 日本語5,440件
 - 2003年5月頃は約7万件[1]
- アットマーク・アイティ(<http://www.atmarket.co.jp/>)に掲載された読者調査[2]によると
 - 「すでにRDF / RSSを作成・利用している」6%
 - 「今後作成・利用を予定 / 検討中」7%
 - 「作成・利用予定はないが興味はある」39%
 - 「RDF / RSSを知らない」42%

RSS (RDF Site Summary)

- 現在のインターネット上のRDFデータの多くはRSSとして存在
- 最新のニュースや記事、日記の要約などを配信するフォーマット
 - もともとはチャンネル情報を配信するためのもの
- 配信情報を利用するユーザは、Webブラウザで各サイトの更新状況を調べなくても新着情報を手にいれることができる
- ウェブログの発展とともに注目を集める

ウェブログ / ブログ

- ウェブログ (weblog)
 - 略してブログ (blog)
- 頻繁に更新される「個人の日記」や「気になったWebページに対するコメント」を載せたWebサイト
 - 米国では情報の生々しさや影響力が大手のマスメディアを脅かすものとして取り上げられることも

RSS1.0の構造

- INTAPセマンティックWeb委員会のRSSを例に説明

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns="http://purl.org/rss/1.0/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:sy="http://purl.org/rss/modules/syndication/" xml:lang="ja">
```

```
<channel rdf:about="http://www.net.intap.or.jp/INTAP/s-web/">
...
</channel>
```

channel要素: サマリー記述の対象となるサイトの
基本情報を記述

```
<item rdf:about="http://www.net.intap.or.jp/INTAP/s-
web/data/conference2003/index.html">
...
</item>
```

item要素: サイト内の個々の記事や文書に関する
情報を記述 (複数個記述可能)

```
</rdf:RDF>
```

(c)2003 INTAP. All rights reserved.

channel要素の例

```
<channel rdf:about="http://www.net.intap.or.jp/INTAP/s-web/">
```

```
<title>INTAP セマンティックWeb委員会</title>
<link>http://www.net.intap.or.jp/INTAP/s-web/</link>
<description>INTAPセマンティックWeb委員会ではセマンティックWeb技術の調  
査研究ならびに普及活動を行っています。</description>
```

タイトル、対象とす
るサイトのURL、
サイトの内容を記
述 (必須項目)

```
<dc:language>ja</dc:language>
<dc:rights>(c)2002 Interoperability Technology Association for  
Information Processing, Japan</dc:rights>
<dc:date>2003-06-05T00:00:00+09:00</dc:date>
```

Dublin Coreのボ
キャブライを用いて
記述された言語や
権利情報や日付に
関するメタ情報

```
<items>
<rdf:Seq>
<rdf:li rdf:resource="http://www.net.intap.or.jp/INTAP/s-
web/data/conference2003/index.html" />
<rdf:li rdf:resource="http://www.net.intap.or.jp/INTAP/s-
web/publicdata.html#gijutsu" />
</items>
```

item要素でサマ
リーを記述する
個々の記事や文
書の情報リソース
とその順番を示す
(必須項目)

```
</channel>
```

(c)2003 INTAP. All rights reserved.

item要素の例

```
<item rdf:about="http://www.net.intap.or.jp/INTAP/s-web/data/conference2003/index.html">
  <title>セマンティックWebコンファレンス2003 開催案内 (11/17開催)</title>
  <link>http://www.net.intap.or.jp/INTAP/s-web/data/conference2003/index.html</link>
  <description>INTAPと慶應 ... セマンティックWebコンファレンス2003を開催致します。</description>
  <dc:subject>セマンティックWebに関するお知らせ</dc:subject>
  <dc:creator>INTAP事務局</dc:creator>
  <dc:date>2003-10-14T00:00:00+09:00</dc:date>
</item>
```

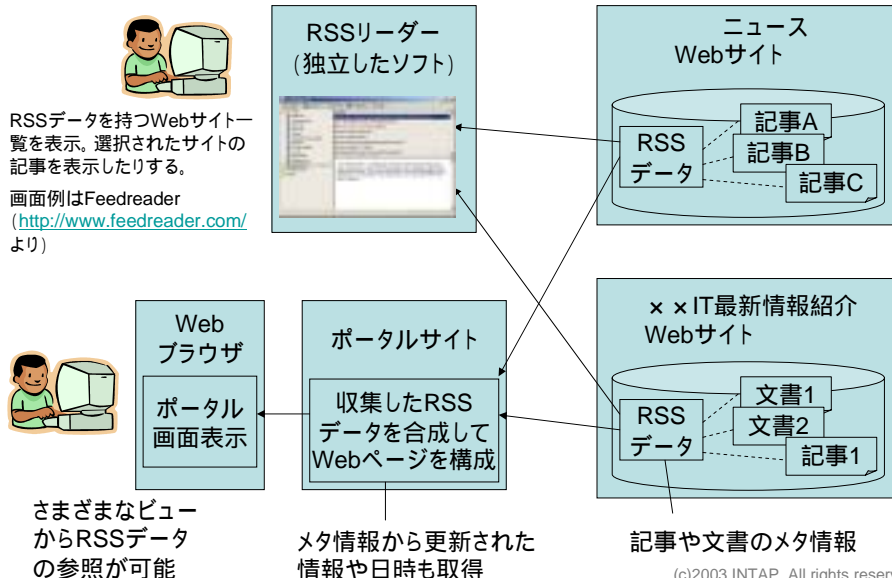
```
<item rdf:about="http://www.net.intap.or.jp/INTAP/s-web/data/14-semanticweb-report.pdf">
  <title>平成14年度セマンティックWeb技術の調査研究報告書</title>
  <link>http://www.net.intap.or.jp/INTAP/s-web/data/14-semanticweb-report.pdf</link>
  <description>本報告書はセマンティックWeb委員会の平成14年度の活動...</description>
  <dc:subject>セマンティックWebに関する文献</dc:subject>
  <dc:creator>INTAPセマンティックWeb委員会</dc:creator>
  <dc:date>2003-05-01T00:00:00+09:00</dc:date>
</item>
```

タイトル、対象とする情報リソース(記事や文書)のURLは必須項目

それ以外にDublin Coreのボキャブラリを用いてトピック(dc:subject)、作者(dc:creator)、日付(dc:date)に関するメタ情報を記述している。

(c)2003 INTAP. All rights reserved.

RSSデータのユーザ利用



RSSを利用したポータルサイトの実例

- セマンティックWeb技術の情報ポータル
 - INTAPセマンティックWeb委員会のWebサイト上に構築し開設
 - セマンティックWebに関連する団体・企業のRSS(RDF)ファイルを収集し一括的に表示することを目的
 - <http://www.net.intap.or.jp/INTAP/s-web/sweb-portal/>

ポータルサイトのユーザインタフェース

- 登録されたURLに存在する複数のRSS(RDF)データからWebページを動的に生成して提示
- メタ情報を利用した検索が可能
 - rss:descriptionとrss:titleのプロパティ値を対象にした全文検索
 - プロパティdc:creatorを利用して同じ著者の記事や文書を表示
 - プロパティdc:subject、rss:categoryを利用して同じテーマやカテゴリの記事や文書を表示

セマンティックWeb技術の情報ポータル(1) トップ画面

WebサイトのURLとリンク

クリックするとWebサイトの全てのitemを表示(次のスライド)

itemのtitleを表示 実際の記事のURLとリンク

RSSデータを持つ各Webサイトの一覧と、各サイトの最新の記事や文書のタイトルと要約の表示

(c)2003 INTAP. All rights reserved.

セマンティックWeb技術の情報ポータル(2) item一覧表示画面

選択されたWebサイトの全てのitemを表示

(c)2003 INTAP. All rights reserved.

セマンティックWeb技術の情報ポータル(3) メタ情報を利用した検索機能

dc:subjectのプロパティ値が
同じ記事を検索するボタン
(この他にrss:categoryを利用した同じ
カテゴリの記事を探すボタンの表示が可能)

rss:descriptionと
rss:titleを対象に
した検索を利用
するための
入力フォーム

プロパティdc:creatorの
プロパティ値が同じ記事
を検索するボタン
「同じ著者の記事を探す」

クリック

検索結果
itemの表示

(c)2003 INTAP. All rights reserved.

RSS(RDF)の生成支援ツール

- プロパティ値などの必要項目を入力するだけ
 - Headline-Editor(Lite版) <http://www.infomaker.jp/editorlite/index.htm>
 - 新規記事のタイトルを列挙したヘッドライン情報や、Webサイトの更新情報などを要約したRSSファイルの作成作業を支援するソフト
- 既存のホームページから生成
 - JMdRdf (Java Midori Rdf)
 - <http://www003.upp.so-net.ne.jp/midori/JMdRdf.ja.html>
 - HTMLからRDF/RSSを自動作成するツール
 - HTML中のタイトルやdescriptionなどから自動的に情報を抽出
- Webサイトのファイル管理と連携して生成
 - Legacy Web Manager
 - <http://kamicup.at.infoseek.co.jp/webman/index.html>
 - HTMLファイルや画像ファイルを一元管理し、ファイルの更新に合わせてRSSファイルも更新できる
- 日記から生成
 - tDiary <http://www.tdiary.org/>
 - Web日記システム上で更新及びコメントが入った日の日記をRDFに吐き出すプラグインを備える

RSSのバージョン

基盤技術	正式名称	バージョン	策定団体	策定時期
RDF	RDF Site Summary	0.9	Netscape社	1999年5月
		1.0	RSS-DEV WG	2000年12月
XML	Rich Site Summary	0.91	Netscape社+ UserLand Software社	1999年7月
		0.92	UserLand Software社	2000年12月
		0.93	UserLand Software社	2001年4月
	Really Simple Syndication	2.0	UserLand Software社	2002年8月

- 汎用のメタデータとしても利用できるように考慮するか、コンテンツ配信に特化するかで策定の方針に違いがある
- セマンティックWeb委員会としてはRDFベースのRSS1.0を推奨

FOAF (Friend of a Friend)

- RSSの次にインターネットユーザがRDFを活用している例
- 人に関するメタ情報(名前、興味、どこの近くに住んでいるなど)をRDFで記述して公開するプロジェクト
 - <http://www.foaf-project.org/>
- FOAFのメタ情報は、RSSのように個々のWebサイトなどここに置いても良い
- 知っている人を記述できるので、個々のRDFから知人のRDFを連鎖的に辿れるようになる
- googleで「foaf filetype:rdf」を検索
 - 10,900件(2003年11月)
 - 日本語41件

FOAFのビューワの実例

FOAF (RDF) で記述された
近くの空港のメタ情報を利用
して、人がいる場所をノード
として地図上にマッピング

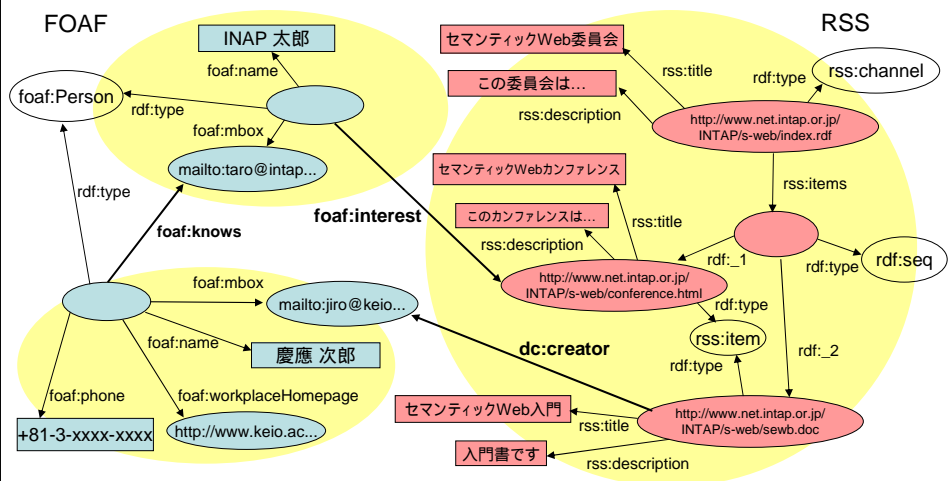
RDFで記述された
知人に関する情報を利用
して、人と人との
関係をグラフ構造で
表示 (SVGを利用)

ノードを
クリック

FOAF people map
<http://www.jibbering.com/foaf/foaf-people-map.svg>

foafnaut
<http://www.foafnaut.org/>

FOAFとRSSのデータをグラフで表現



- 複数のWebサイトからこれらのRDFデータを収集すると、例えば以下のような情報をグラフ構造を辿って機械的に取得できる
 - 「慶應 次郎」の知人が興味を持っているページの要約
 - 「セマンティックWeb入門」の著者の電話番号や働いている場所

メタ情報の記述にRDFを用いる利点

- 機械がRDFデータのみから、どの情報リソースに対するどんなメタ情報であるか明確に識別可能
 - RDFのデータモデルに従って、情報リソースとそのプロパティとプロパティ値(主語・述語・目的語からなるRDFの文)を抽出可能
 - 情報リソース、ボキャブラリの定義場所はURIで指定される
 - 誰がメタデータを記述しても構わない
 - RDFデータはどこに置いても構わない

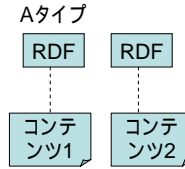
RDFはどこに置くべきか？

- RDFは見つかりさえすればOK
 - しかし、情報リソースに対するメタデータの場所は一律ではない
 - 現状
 - RDFメタデータのURLを登録するか、
 - ロボットによってクロールされる必要がある
- ガイドラインに相当するものが必要では？
 - 置き方ガイド
 - 探し方ガイド
- 以降にセマンティックWeb委員会で議論中の「RDFメタデータの設置場所ガイド」について紹介
 - 詳細は委員会のWebページに掲載予定

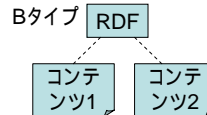
RDFメタデータの設置場所

- 次の3つのタイプが考えられる

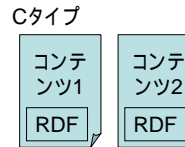
A. 1つのコンテンツに対応するRDFメタデータをコンテンツとは独立に記述する場合



B. 複数のコンテンツのRDFメタデータをコンテンツとは独立に一括して記述する場合



C. コンテンツ内にRDFメタデータを埋め込む場合



(c)2003 INTAP. All rights reserved.

Aタイプ(メタデータを独立して個々に管理) の例

- 既存のコンテンツの拡張子を.rdfとする

	元コンテンツのURL	RDFメタデータファイルのURL
例1	http://www.intap.or.jp/s-web/ file1.html	http://www.intap.or.jp/s-web/ file1.rdf
例2	http://www.intap.or.jp/s-web/ file2.txt	http://www.intap.or.jp/s-web/ file2.rdf

- 拡張子を除いた部分が同じURLになるコンテンツがある場合を考慮して、file1.html.rdf、file2.txt.rdfというようにURLの最後に.rdfという拡張子を加える方法もある

メリット： シンプル
 デメリット： 例外も考慮する場合、1つのコンテンツに2つの拡張子

(c)2003 INTAP. All rights reserved.

Aタイプの例(その他のアイディア)

- RDFメタデータ用にmetaというディレクトリを用意する方法

	元コンテンツのURL	RDFメタデータファイルのURL
例1	http://www.intap.or.jp/s-web/ file.html	http://www.intap.or.jp/s-web/ meta/file.html
例2	http://www.intap.or.jp/s-web/ file.pdf	http://www.intap.or.jp/s-web/ meta/file.pdf

メリット： シンプルで例外を考慮する必要がない
 デメリット： メタデータファイルの拡張子が.rdfでない

- RDFメタデータファイル専用のサーバ(名)を作成し、そこにすべてのRDFメタデータファイルを置く

	元コンテンツのURL	RDFメタデータファイルのURL
例1	http://www.intap.or.jp/s-web/file.html	http:// rdf .www.intap.or.jp/s-web/file.html
例2	http://www.intap.or.jp/s-web/abc.cgi?parml=abc	http:// rdf .www.intap.or.jp/s-web/abc.cgi?parml=abc

メリット： 動的なコンテンツにも容易に対応が可能

Aタイプのメタデータ参照
 シンプルだが、以上のようなルールに関して提供者と参照者でコンセンサスがとれないと機能しない

Bタイプ(メタデータを独立して一括管理)の例

- Webサイトのwell-knownな場所のファイルにRDFメタデータを一括して記述

	元ファイルのURL	RDFメタデータファイルのURL
例	http://www.intap.or.jp/s-web/ file1.html http://www.intap.or.jp/s-web/ file2.html http://www.intap.or.jp/s-web/ file2.pdf	} http://www.intap.or.jp/s-web/ index.rdf

メリット： 既存の大量のWebコンテンツに対してメタデータを一括管理するのに適している
 デメリット： 単一のRDFメタデータファイルのバージョン管理や変更権限などに注意を要する

Bタイプのメタデータ参照
 Aタイプと同様にメタデータの場所に関するコンセンサスが必要だが推測はし易い

Cタイプ(コンテンツ内埋め込み)の例

- (X)HTMLの場合
 - メタデータを直接埋め込み可能(注1)
 - メタデータの場所を埋め込み可能
 - コンテンツ内の<link>タグを使用して、
 <link rel="meta" type="rdf/xml" href="http://www.intap.or.jp/s-web/file.rdf">
 のようにRDFメタデータファイルのURLを指定(注2)

メリット： コンテンツの本体とメタデータの結びつきが最も強い

デメリット：メタデータ情報を変更するのに、コンテンツ自体の変更権限が必要になる
メタデータを埋め込むことができない
コンテンツには適用できない
コンテンツを扱うアプリケーションによってはコンテンツ本体とメタデータを区別できない場合がある

(注1)HTMLの<head>要素に入れるなどの方法がある

RDF/XML記述のXMLスキーマはXHTMLのスキーマには準拠していないため、ValidなXHTMLではなくなる
W3Cでは以下のように問題回避を検討中
RDF in XHTML Task Force Document 27 May 2003
<http://www.w3.org/2003/03/rdf-in-xml.html>

(注2)rel="meta"という属性値が、HTML4.0勧告で規定されていない属性値だという問題あり

(c)2003 INTAP. All rights reserved.

Cタイプの例(HTML以外)

- 画像ファイル
 - RDFPic
 - 写真のメタデータ登録ツール
 - RDFのメタデータ記述をデジタル写真のイメージそのものに埋め込むことが可能
 - SVGのXML仕様はRDFメタデータを埋め込むことを考慮
 - <http://www.w3.org/TR/SVG-access/#Metadata>
- Adobe XMP (Extensible Metadata Platform)
 - Adobeのアプリケーションが生成するPDFなどのファイルにRDFのメタデータの埋め込みが可能



RDFPicの画面例
<http://jigsaw.w3.org/rdfpic/>より

Cタイプのメタデータ参照
アプリケーションがメタデータを抽出できれば、参照は容易

(c)2003 INTAP. All rights reserved.

RDFメタデータの参照について

- Cタイプのメタデータを最初に参照することが推奨される
- ただし、コンテンツの作成者とメタデータの管理者が異なる場合などサイトの事情に応じたRDFメタデータの設置場所を考慮する必要がある
- 今後、RDFメタデータの収集・検索エンジンによる各タイプへの対応(注)とその参照順序を実装する際の参考となる設置場所ガイドを発行したい

(注)Webサイト内で複数のタイプによるメタデータを用意する場合には矛盾などが生じないように注意が必要

ご協力をお願い

- セマンティックWebに関する情報を提供しているサイトのメタデータをご提供ください
 - RSS1.0 (<http://web.resource.org/rss/1.0/spec>) 準拠のデータ
 - Dublin Coreなどのモジュールの利用を歓迎
 - (例)INTAPセマンティックWeb委員会のRSS
 - <http://www.net.intap.or.jp/INTAP/s-web/index.rdf> にRSSファイルを設置
- メタデータの設置場所ガイドについてご意見をください
- 連絡先
 - 情報処理相互運用技術協会 セマンティックWeb委員会
 - info-semanticweb@intap.or.jp

参考

- [1] 津田宏 「セマンティックWeb」入門第3回 日経インターネットソリューション2003年6月号
 - RSSに関する解説など

- [2] 小柴 豊 アットマーク・アイティ マーケティングサービス担当 「XML & Web Services 第10回 読者調査結果発表 ～ RDF / RSSの活用状況と情報統合への取り組み～」
<http://www.atmarkit.co.jp/fxml/survey/survey10/xml10.html>

- [3] RDF Site Summary (RSS) 1.0 <http://web.resource.org/rss/1.0/spec>

- [4] The Web KANZAKI (メタ情報とセマンティック・ウェブ)
<http://www.kanzaki.com/docs/sw/>
 - RSS、FOAFを含めてセマンティックWeb全般に詳しい解説