

WWW に見るメタデータの標準化動向

浦本 直彦 *1 津田 宏 *2 上田 隆也 *3 佐藤 研治 *4 野村 浩郷 *5

*1 日本 IBM(株) 東京基礎研究所 uramoto@trl.ibm.co.jp

*2 (株) 富士通研究所 htsuda@flab.fujitsu.co.jp

*3 キヤノン (株) 情報メディア研究所 takaya@cis.canon.co.jp

*4 NEC C&C メディア研究所 satoh@hum.cl.nec.co.jp

*5 九州工業大学 情報工学部 nomura@ai.kyutech.ac.jp

最近メタデータはインターネットの普及により、WWW を中心とするインターネット上のリソースを記述する枠組みとして注目を集めている。現在、いくつかのプロジェクトが、Web 技術の標準化団体 W3C を中心に行なわれている。本論文では、現在行なわれているメタデータプロジェクトの最新動向を紹介する。なお、本論文は筆者らが電子協ネットワークアクセス専門委員会で調査した内容をもとに、執筆時での最新の状況をまとめたものである。

Recent Trends in Metadata on World Wide Web

Naohiko Uramoto *1 Hiroshi Tsuda *2 Takaya Ueda *3 Kenji Satoh *4 Hirosato Nomura *5

*1 IBM Research, Tokyo Research Laboratory (uramoto@trl.ibm.co.jp)

*2 Fujitsu Laboratories Ltd. (htsuda@flab.fujitsu.co.jp)

*3 Canon Inc. Media Technology Laboratory (takaya@cis.canon.co.jp)

*4 C&C Media Research Laboratories NEC Corporation (satoh@hum.cl.nec.co.jp)

*5 Kyushu Institute of Technology (nomura@ai.kyutech.ac.jp)

This paper surveys the recent trends in “metadata” on World Wide Web. Metadata is “data about data”, although it is difficult to distinguish metadata from data itself. World Wide Web Consortium (W3C), which is the standard body for Web technology is working for standardization of metadata format used on Web. Resource Description Framework (RDF) is one of specification for metadata. It is based on XML, and is designed to be able to apply various application areas. This survey is based on the report from JEIDA networking access working group.

1 はじめに

Webの普及によって、HTMLは、分散された文書をハイパーリンクで連結し閲覧するという当初の設計をはるかに超え、様々な分野で用いられている。しかし、現在のHTMLの枠組みでは、情報過多や情報(データ)記述の限界などの問題が指摘されており、これらを解決するものとして、現在XMLに対して非常な(過剰ともいえる)期待が高まっている。一方、Web上のWWWを中心とするインターネット上のリソースを記述する枠組みとして、メタデータと呼ばれる情報の記述形式が注目を集めている。本論文では、W3Cで検討されているいくつかのプロジェクトを中心に、メタデータの最新動向を紹介する。なお、本論文は筆者らが電子協ネットワークアクセス専門委員会が調査した内容を元に、最新の情報を盛り込んだものである。

2 メタデータとは

メタデータは「データについてのデータ(data about data)」である。図書館学では文献記述の分野で古くから研究されている。最近のデジタル図書館や文書のデジタル化の研究においては、オンラインで得られる文献を記述するためにさまざまなメタデータ仕様が研究されている。その代表的なものとして、MARC(Machine-readable Catalogue)[MARC]やTEI(Text Encoding Initiative)[TEI]がある。MARCは1960年代後半から使われている、図書館の機械可読目録の仕様である。図書館の記述に関しては国ごとに個別のニーズがあるため、個々の国で仕様に違いがある。TEIは、ACH(Association for Computers and the Humanities)、ACL(Association for Computational Linguistics)、ALLC(Association for Literary and Linguistic Computing)の支援を受けて作成されたもので、文書データを研究者間で共有・交換するための記述の標準である。記述にはSGMLを使用する。

最近、メタデータはインターネットの普及により、Webを中心とするインターネット上のリソースを記述する枠組みとして注目を集めている。単に「データについてのデータ」というだけではなく、「計算機によって理解できる(machine-understandable)データ」という位置付けになってきている。すなわちメタデータを使って種々の機械処理ができることが重要である。また、対象として文書だけでなく種々のリソースを扱える必要がある。

上に挙げたMARCやTEIの仕様はかなり大きいため、インターネット上の変化しつづける膨大なリ

ソースを記述するには不向きである。情報量は少なくても、簡単に作成でき、かつ検索等の機械処理に役立つようなメタデータが必要になる。

なお、メタデータ全般については、メタデータに関する網羅的なリソース集“DIGITAL LIBRARIES: Metadata Resources”[IFLA]を参考にされたい。

3 メタデータプロジェクトの概要

3.1 IAFA テンプレート

IAFA(Internet Anonymous FTP Archive)テンプレートはIETFのIAFAワーキンググループが設計したもので、FTPサイト上のリソースについての書誌的事項を表現する[Deutsch et.al 95]。

従来Anonymous FTPを目的として検索する際、archieというプログラムを使いファイル名で検索するしか方法がなく、ファイル名を知らない場合は検索が困難であった。IAFAテンプレートではこの問題を解決するために、リソースについて記述する構造化データの形式を決め、これを使ってインデクスを作成し、検索できるようにした。

IAFAテンプレートでは、対象として、文書、イメージ、ソフトウェア、メーリングリストのメールの蓄積、ネットニュースの記事の蓄積、音、ビデオ、FAQ(よく聞かれる質問とその答)などFTPで提供される種々の情報を扱う。

リソースに関する情報はコロン(:)で区切られた“属性名:属性値”ペアで表現する。属性名としては、Template-Type(情報の種類)、Title(タイトル)、Author-Name(著者)、Last-Revision-Date(最終更新日)、Category(カテゴリ)、Description(内容についての記述)、Copyright(著作権についての記述)、Keyword(キーワード)、URI(データの所在場所)、Language(言語)、Size(情報のサイズ)、Format(データフォーマット)、などが用意されている。

IAFAテンプレートはインターネット上のリソースを記述するためのメタデータとしては先駆的なものの一つである。

3.2 Dublin Core

Dublin CoreはOCLC(Online Computer Library Center)とNCSA(National Center for Supercomputing Applications)がスポンサーになって1995年以来開催している一連のワークショップで検討されているメタデータの仕様である[DUBLIN]。

先に述べたようにTEIやMARC等の従来のメタデータは仕様が大きく、誰でも記述できるわけではなかった。また、種々のメタデータ仕様が存在して、

各々の間の対応がとれないという問題が生じていた。Dublin Coreはこうした問題を解決するため、インターネット上のリソースに関して、一般の著者・コンテンツプロバイダでも作成できるぐらいに簡潔で、かつ検索等の役に立つメタデータ記述をすることを目的としている。さらに、さまざまなメタデータ間の互換をとるための統一的な表記方法を提供することを目的としている。

Dublin Coreの要素は下記の15種類だけである。

- ・内容に関するもの
 - Title: 名称
 - Subject: 扱っているトピック
 - Description: 内容に関する記述
 - Source: リソースのもとになった第二のリソース
 - Language: 言語
 - Relation: リソースと第二のリソースの関係
 - Coverage: 時間的・空間的な特性
- ・知的所有権に関するもの
 - Creator: 内容についての責任者(ex. 著者, 写真家)
 - Publisher: リソースをこの形態で提供している責任者
 - Contributor: Creator以外に貢献している人(ex. 編集者)
 - Rights: 権利に関する記述
- ・リソース自体に関するもの
 - Date: 作成した、もしくは入手可能になった日付
 - Type: リソースの種類(ホームページ, 技術報告, 詩)
 - Format: データフォーマット
 - Identifier: 識別子(ex. URL)

Dublin Coreは、メンバーがRDFの仕様作成にも参加しており、RDFのベースの一つになっている。

3.3 MCF (Meta Content Framework)

MCFは元々はApple Computer社で開発された、Webにおけるメタデータ記述形式である。これをNetscape Communications社が改良し、XMLで記述するようにして、"Meta Content Framework Using XML"[MCF 97]として1997年6月にW3Cに提案した。

現在MCFは、他のメタデータ仕様と統合して発展的にRDFになっているため、MCF自体は役目を終えているといえる。そこで、ここでは簡単に説明するにとどめる。MCFについては[Bray and Guha 97][日経IT 97]なども参照されたい。

メタデータの用途は、日付・著者等の目録情報、著作権・ライセンス等の知的所有権に関する情報、子供にふさわしい内容かどうかを保証する情報、プライバシーに関する情報、デジタル署名の分野での文書に関する情報、など多岐に渡っている。MCFは

従来適用対象ごとに存在していたメタデータを共通のデータモデルと語彙を使って統一的に記述することを目的としている。これによって、共通のソフトウェアでメタデータを扱うことができる。

MCFでは、オブジェクト(object)に属性(property)を付加する形でメタデータを表現する。この表現のもとになっているのは、DLG(Directed Labeled Graph; 有向ラベル付きグラフ)というネットワーク表現で、オブジェクトをノード、属性名を弧、属性値を弧につけられたラベルして表わす。属性名自体もオブジェクトであり、属性名も属性を持つことができるという点に特徴がある。また、オブジェクトのタイプを表わすものとしてカテゴリがある。

先に述べたようにMCFは記述にXMLを用いる。オブジェクトはXMLの要素(element)、属性はオブジェクト要素中の要素、カテゴリは要素(element)のタイプとして表わす。オブジェクトの表現の例を挙げる。

```
<WebPage>
  <url>http://www.jeida.or.jp</url>
  <author unit="nomura"/>
</WebPage>
<Person id="nomura">
  <name>Hirosato Nomura</name>
  <email>nomura@jeida.or.jp</email>
  <typeOf unit="Professor"/>
</Person>
```

これは、あるWebページのURLが

http://www.jeida.or.jp、著者がnomuraであること、nomuraという人物は名前が"Hirosato Nomura"であり、電子メールアドレスがnomura@jeida.or.jpであり、教授であること、を表わしている。WebPageやPersonがオブジェクト、url、author、name、emailが属性、Professorがカテゴリである。

MCFではWebPageやPersonが何を表わしているかという意味は表現できないが、これらの間の関係は記述することができる。例えば、

```
<PropertyType id="author">
  <domain unit="WebPage"/>
  <range unit="Person"/>
</PropertyType>
```

では、authorという属性が、WebPageの属性であり、その値はPersonに属するものであることを表わしている。domainによってどのオブジェクトの属性かを示し、rangeによって属性値の型や値の範囲を示す。このほか、オブジェクト間の継承関係、親子関係、順序関係等も記述できる。

MCF では基本的な語彙を提供して共通の枠組みで記述できるようにしている。しかし、適用分野ごとに独自のオブジェクトや属性があるので、語彙を拡張することもできる。MCF の応用によってこれまでよりも高度な検索が可能になる。従来のサーチエンジンではキーワードがどのような文脈であられるかを指定することができず、望まない検索結果が数多く出てしまう。メタデータを使うことにより、属性を利用した検索ができる。例えば、「野村教授が著者である論文」や「価格が30万円以下でCPUがPentium IIであるデスクトップPCのカタログ」といった検索が可能になる。

3.4 CDF (Channel Definition Format)

ここ1,2年プッシュ型情報配信サービスが増加している。このサービスは情報が更新された際に最新の情報をユーザに自動的に送り届けるサービスである。従来の Web ではユーザがアクセスしない限り情報の更新を知ることができなかつたが、プッシュ型情報配信はこの問題を解決しようとするものである。

こうしたプッシュ型情報配信サービスは、実際にはクライアントが定期的にサーバにアクセスしてコンテンツをとってくるという「スマート・プル」という方式で実現されていることが多い。この際、どのような種類の情報があり、どの程度の間隔で更新されるかをクライアント側で知る必要がある。サーバ側ではこうした情報を記述したファイルを用意するが、従来はこの記述仕様が各コンテンツプロバイダでバラバラで、互換性がなかった。しかし、これではユーザ側ではコンテンツプロバイダごとにブラウザを用意しなければならず、またコンテンツプロバイダ側でもコンテンツ作成のためのツールを個別に用意しなければならないため、記述仕様の標準化が望まれていた。

CDF はこうした背景のもとに、Microsoft 社が作成し、1997年3月にW3Cに提案した「プッシュ型情報配信におけるチャンネル記述のためのメタデータ記述の枠組み」である[CDF 97]。前項のMCFと同様にXMLで記述する。チャンネルとはここでは「コンテンツに関するメタ情報」を意味する。

以下、CDFの記述について仮想的な例(図1)を使って説明する。

3.4.1 チャンネルの構成

CDF では root 要素として“CHANNEL”要素が用いられる(1)。

はチャンネルのURLを表わす(2)。<TITLE>はチャンネルのタイトル(3)、<ABSTRACT>はチャンネルについての簡単な説明(4)を表わす。チャンネルは階層的に構成することができる。サブチャンネルがある場合は再帰的に<CHANNEL>要素を置く。階層構造の末端の構成要素は<ITEM>によって表わす。このチャンネルには3つの構成要素が存在する(10, 14, 17)。

3.4.2 チャンネルの更新

チャンネルの更新スケジュールは<SCHEDULE>によって表わす(6)。この例では、1998年2月18日から1998年3月31日までの間、0時から6時間おきに(7)更新することが示される。スマート・プル方式の問題として、各クライアントが更新時にいっせいにサーバにアクセスした場合、ネットワークやサーバに対する負荷が大きいということがある。これを回避するために、サーバへのアクセスを一定期間内のランダムな時点に行なわせることができる。これを指定するのが<EARLIESTTIME>(8)と<LATESTTIME>(9)である。この例では更新時から1時間後の時点から3時間後の時点までの間(2時間)のランダムな時点にサーバにアクセスして更新することが指定される。すなわち、1時と3時の間、7時と9時の間、13時と15時の間、19時と21時の間に各1回更新する。

3.4.3 コンテンツの表示

チャンネルの構成要素をクライアント上でどのように表示するかは<USAGE>によって指定することができる。チャンネルの一部としてクライアントプログラムで表示したり(“Channel”, 12)、スクリーンセーバで表示したり(“ScreenSaver”, 13)、PCのデスクトップに窓を作って表示したり(“DesktopComponent”, 16)、更新を電子メールで通知したり(“Email”)することができる。

3.4.4 ログの管理

チャンネルへのアクセスログを管理することができる。<LOGTARGET>はログの保存先を示す(5)。<LOG VALUE="document.view" />(11, 15, ...)というタグのつけられた要素についてログが記録される。

W3Cでは、CDFを次節で述べるRDFを用いて記述することを検討している。しかし、CDFは既に

```

<?xml version="1.0" encoding="x-sjis" ?>
1) <CHANNEL HREF="index.htm" BASE="http://www.jeida.or.jp/" SELF="channel0.cdf">
2)   <A HREF="http://www.jeida.or.jp/index.htm"></A>
3)   <TITLE>ネットワークアクセス技術専門委員会</TITLE>
4)   <ABSTRACT>知的情報アクセスにおける言語処理技術・検索技術の動向についての調査</ABSTRACT>
   <LOGO HREF="images/logo.gif" STYLE="Image" />
   <LOGO HREF="images/icon.gif" STYLE="Icon" />
5)   <LOGTARGET HREF="http://www.jeida.or.jp/log.htm" Method="Post" Scope="All"></LOGTARGET>
6)   <SCHEDULE STARTDATE="1998.02.18T00:00+0000" ENDDATE="1998.03.31T00:00+0000">
7)     <INTERVALTIME HOUR="6" />
8)     <EARLIESTTIME HOUR="1" />
9)     <LATESTTIME HOUR="3" /></SCHEDULE>
10)  <ITEM HREF="NetAccess/gengo.htm" Precache="No">
   <A HREF="http://www.jeida.or.jp/NetAccess/gengo.htm"></A>
   <TITLE>自然言語処理技術</TITLE>
   <ABSTRACT>知的情報アクセスにおける自然言語処理技術について述べる</ABSTRACT>
11)  <LOG VALUE="document.view" />
12)  <USAGE VALUE="Channel"></USAGE>
13)  <USAGE VALUE="ScreenSaver"></USAGE>
   </ITEM>
14)  <ITEM HREF="NetAccess/kensaku.htm" Precache="Yes">
   <A HREF="http://www.jeida.or.jp/NetAccess/kensaku.htm"></A>
   <TITLE>検索技術</TITLE>
   <ABSTRACT>知的情報アクセス技術における検索技術について述べる</ABSTRACT>
15)  <LOG VALUE="document.view" />
   <USAGE VALUE="Channel"></USAGE>
16)  <USAGE VALUE="DesktopComponent">
   <HEIGHT VALUE="50" /><WIDTH VALUE="100" /></USAGE>
   </ITEM></CHANNEL>

```

図 1: CDF ファイルの例

Internet Explorer 4.0 の Active Channel や Point-Cast 2.0 で実際に利用されており、50 社以上の会社が支持を表明している。プッシュ型情報配信サービスを記述するメタデータに関しては CDF が標準としての地位を確立していくであろう。

現在の CDF で表現されているのはコンテンツに関する基本的な情報のみである。今後、プッシュ型情報配信サービスにおいては、ユーザごとに必要な情報のみを届けるニーズが出てくると予想されるが、現在の CDF ではこれに対応することはできない。チャンネルやサブチャンネルに対して個人化 (personalization) やフィルタリングを行なえるようにするために、チャンネルやサブチャンネルのカテゴリを記述したり、重要度・緊急度等を表現したりする枠組みが必要になってくると考える。

3.5 RDF

3.5.1 RDF の概要

RDF (Resource Description Framework) は、WWWtics におけるメタデータの仕様であり、現在 W3C で標準化が進められている。RDF におけるメタデータの定義とは、「Web リソースを記述するためのデータ」であり、もっとも単純な形式では、ある HTML 文書に対する言明 (例. 更新日時, 作成者) を、属性とその値の集合で記述するものである。しかし、RDF の表

現力はそれにとどまらず、構造化された複雑なデータ構造を表現するための仕組みを有しており、様々なアプリケーション分野での応用が期待されている。

RDF のゴールは大きく分けて 2 つある。そもそも、RDF は Web リソースへの rating を行うための情報記述の仕様 PICS と、Web リソースへの電子署名の仕様 DSig の 2 つのプロジェクトから生まれたものである。両者とも「特定の Web リソースを記述するためのデータ」であり、これらを一般化することで汎用的なメタデータのモデルおよび構文を規定することが RDF の一つの目的である。W3C で標準化が行われている HTML や XML と同様に、Web 上でのメタデータの仕様が標準化されれば、異なるアプリケーション間での情報交換が可能となりその意義は大きい。

もう一つのゴールは、Web における (いわゆる AI 的な) 知識表現のセマンティクス (semantics) の確立である。そもそも HTML ではデータ構造の seman-

tics を規定することは不可能であり、XML もまた基本的には DTD によって syntax を規定する言語であるので、例えば、<date>タグで囲まれたデータは日付を表すものでなければならないといった意味的な制約を記述することはできない。またデータオブジェクトに階層関係を定義することができれば、オントロジのような意味的な知識体系を構築できる。RDF

では、データモデルに対するセマンティクスをスキーマ (schema) と呼んでいる。

W3C における RDF の標準化作業は、RDF Model and Syntax, RDF Schema という 2 つのワーキンググループ (WG) で行われており、2 つの WG のミッションは上の 2 つのゴールに対応している。日本からの参加者は両 WG とも 2 人ずつであり、浦本が両 WG に参加している。

RDF の仕様は現在も議論中であり、最終的な仕様は、本解説で述べたものと異なる可能性があることに留意されたい。本論文の執筆時において、モデルおよび構文 (Model and Syntax) に関しては、1998 年 2 月 16 日版の working draft [RDF-MS 98] が最新である。このドラフトは数回の改訂を経ており、現在、最終的な勧告として公開されるためのプロセスが進行中である。スキーマに関しては、1998 年 4 月 9 日に最初の working draft が公開された [RDF Schema 98]。

3.5.2 RDF の特徴

RDF の特徴の一つは、内部表現としてのデータモデルとネットワークを介して交換される外部表現を明確に分離していることである。データモデルとして有向ラベル付きグラフ (directed labeled graph) を用いている内部表現をグラフにすることで (基本的に木モデルである XML と比べて) より強力な記述力を実現している。

外部表現としては、XML を用いている。データモデルは、XML で表現されたデータに変換されアプリケーション間で交換される。この過程を serialization と呼ぶ。RDF プロセッサは、データモデルに基づく内部表現と XML を用いた外部表現間の変換を行なえなければならない。また、RDF では、名前空間 (namespace) と呼ばれる手法を用いて、複数の schema を一つのデータモデルの中で混在して用いることができる。これによって、既存の schema を組みあわせることによって必要なデータ表現を構築することが可能となる。

3.5.3 RDF のデータモデル

RDF データモデルの基本要素は、property type, source node, target node と呼ばれる 3 つ組 (triple) である。

この 3 つ組は、図 2 に示すように、2 つのノードとそれを連結するアークとして表現される。



図 2: RDF の基本データモデル

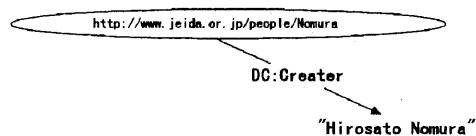


図 3: RDF の基本データモデル

それでは簡単な例を見ていくことにしよう。次のような言明を考える。

「Web リソース

<http://www.jeida.or.jp/people/Nomura>

の作者は、「Hirosato Nomura」である」

これをグラフ表現したものを図 3 に示す。ここで、DC:Creator というのは、Creator という Property type が Dublin Core の語彙であることを示している。ここで例は、target node が文字列の場合であるが、target node がさらに別のリソースノードを指している場合がある。例えば、上記 URL の作者に対する情報も URI を用いて参照できる場合である。

RDF で規定されているその他の特徴としては、以下のものがある。

- source node として表されたリソースに対する言明だけでなく、3 つ組全体への言明を表現することもできる。このグラフ操作は reification と呼ばれている。
- リソースの集合を記述することができる。RDF では Bag (順序なし), Seq (順序付き), Alt (択一) と呼ばれる 3 種類のタイプが用意されている。
- 2 ノード間の関係だけでなく、多ノード間の関係を記述できる。
- リソースに対する電子署名のための枠組が検討されている。

3.5.4 RDF の外部表現としての XML

前節で述べたデータモデルを XML 文書に変換されアプリケーション間で交換される。データモデルを XML 文書に写像する課程を serialization と呼ぶ。例えば、図 3 で示されたデータモデルは図 4 のように写像される。

この例では、RDF のスキーマと Dublin Core のスキーマを用いて記述されている (タグの prefix が

```

<?xml:namespace name="http://www.library.com/bibliography-info/" as="BIB">
<?xml:namespace name="http://www.w3.org/RDF-schema" as="RDF">
<RDF:RDF>
  <RDF:Description RDF:HREF="http://www.jeida.or.jp/people/Nomura/">
    <BIB:Author>Hirosato Nomura</BIB:Author>
  </RDF:Description>
</RDF:RDF>

```

図 4: RDF の外部表現

名前空間を表している)。(RDF:RDF)タグは、これが RDF の外部表現であることを示し、(RDF:Description) タグで、データモデルから写像されたデータ表現を表している。この serialization syntax は、ターゲットの HTML 文書中に埋め込まれたり、別々に提供されたりといったいくつかの形態でネットワーク上を流通する。

3.5.5 RDF スキーマ

上であげたデータモデルは、RDF の構文を規定するものである。それに対し、schema はデータモデルに対して意味的な制約を与えるものである。その詳細は現在 RDF Schema WG で議論中であるが、現在公開されているスキーマの working draft [RDF Schema 98] によると以下のような基本的なセマンティクスが規定されている。

1. node に対する型付けシステム。スキーマでは、各ノードが object 指向に基づくクラスとして定義されている。クラスは階層を持ち、sub-ClassOf という property type を用いて関連づけられる。
2. クラスやノード間の制約。例えば、あるノードに許される property type の集合や、その個数(例えば、"last modified" という Property type は document を表すクラスには一回しか出現してはならない)。

4 個人情報や情報の格付け

4.1 Platform for Internet Content Selection(PICS)

PICS は、情報発信を制限することなく、情報に対するラベル付けに基づく設定によって、「見たくない、見せたくない」情報を排除できるフィルタリング方式のプラットフォームである。PICS 自体はフィルタリングの機能は有してはならず、選別の機能が作動できるようにプラットフォームを提供するに過ぎない。

情報の受け手は、PICS 準拠のソフトを利用し、利用者自身が主に格付け機関を指定することによって、情報の選別が行われる。PICS の標準化を検討している人たちは、これにより情報の検閲なしに、適切な情報選別を行うことが可能としている [小倉 97]。

PICS の標準化グループでは、以下の3点に対して標準を提供している [PICS]。

(1) 自己格付け

コンテンツプロバイダーが、自主的に発信しているコンテンツに対し、格付けを行い、その格付け情報をラベルとしてコンテンツと共に発信する方法

(2) 第3者機関による格付け

第3者機関が、他者の発信しているコンテンツに対し格付けを行い、その格付け情報を発信する方法

(3) 容易な利用

両親や先生が、子供がアクセス可能なコンテンツを、格付けを利用して容易にコントロールする方法

PICS を用いた情報の選別に対する例として、HTML ファイル中に PICS ラベルを META タグを用いて埋め込む方法を以下に示す。(コンソーシアムでは、より良いラベルの提供法として、HTTP サーバの出力する HTTP ヘッダを拡張して PICS ラベルを埋め込む方法、および HTTP サーバでラベル発行サービスプログラムを走らせる方法を推奨している。)

```

<HEAD>
<META http-equiv="PICS-Label" content='
(PICS-1.1 "http://www.gcf.org/v2.5"
  labels on "1994.11.05T08:15-0500"
    until "1995.12.31T23:59-0000"
      for "http://w3.org/PICS/Overview.html"
  ratings (suds 0.5 density 0 color/hue 1)
'>
</HEAD>

```

この例では、PICS に準拠したクライアントがこのコンテンツにアクセスを行うと、クライアントは同時に格付けサービスである <http://www.gcf.org/v2.5> にアクセスし、コンテンツのラベル全体を受け取る。

そのラベルを参照し、クライアントが許可していないレベルであれば、クライアントはこのコンテンツの内容表示を行わないことで、情報選別を行う。

4.2 Platform for Privacy Preferences Project(P3P)

P3Pは、ユーザがネットワーク情報をアクセスするソフトウェア(例えばウェブブラウザ)に設定した個人情報を、どの程度までアクセスしたサーバへ公開するか、また、公開された情報を利用することでサーバ側がどの程度までコンテンツを利用してサービスを行えるかと言ったことを考慮したウェブ上のプライバシー情報伝達のためのプラットフォームである [P3P]。

P3PはW3Cで行われている標準化のプロジェクトであるが、このプロジェクトは、Netscape社を中心としたOpen Profile Standard(OPS)[OPS 97]およびMicrosoft社のPrivacy and Profiling on the Web(Web Privacy)[PPW 97]の2つのW3Cに対する提案が基となっている。この2つのW3Cに対する提案は、別々に、どちらも97年7月2日に行われたが、その直後の7月11日に、Netscape社およびMicrosoft社は、OPSをウェブ上のプライバシーの標準として共同で制定して行くことに同意した [Netscape 97]。

ここでは、P3Pの核として採用されているOPSについて、以下解説を行って行く。

OPSでは、会社やサービス団体が個人に特化した情報や娯楽をインターネットを通じて顧客に提供したがっているのに対し、ユーザは個人のプライバシーの侵害を恐れて個人情報を公開することに対して懸念している現状に対し、安全に必要な情報のみを必要なサービスへ公開できる枠組みを提供する。OPSの提供する個人情報の伝達は、1)ユーザとサービス間、2)サービスとサービス間(ユーザによって仲介されている)、3)ユーザとユーザの間、の3点について行われる [OPS 97]。

5 おわりに

本論文では、現在インターネット上で行なわれている様々なメタデータについて概説した。ここであげたもののほとんどが、XMLを用いて記述されており、少なくともメタデータの構文記述にXMLを用いるという方針は、今後引き継がれていくと思われる。しかし、データの意味を記述するための枠組(スキーマ)は、まだ開発途上であり、今後さらなる議論が必要であろう。現在は、RDFスキーマが議論されているが、XMLそのもののスキーマに関する

議論、データ記述のための新しい提案 [XML-Data] など、今後の動きは興味深い。

本論文で参照した仕様

[CDF 97] Channel Definition Format(CDF)

<http://www.microsoft.com/standards/cdf.htm>

[DUBLIN] Dublin Core Metadata

http://purl.oclc.org/metadata/dublin_core/

[IPWG 97] Internet Privacy Working Group Privacy Practices for the Web

<http://www.w3.org/TR/NOTE-IPWG-Practices>

[MARC] MARC Standards

<http://lcweb.loc.gov/marc/>

[MCF 97] Meta Content Framework Using XML

<http://www.w3.org/TR/NOTE-MCF-XML/>

[OPS 97] Proposal for an Open Profiling Standard

<http://www.w3.org/TR/NOTE-OPS-FrameWork.html>

[P3P] Platform for Privacy Preferences Project(P3P)

<http://www.w3.org/P3P/Overview.html>

[PICS] Platform for Internet Content Selection

<http://www.w3.org/PICS/>

[PPW 97] Privacy and Profiling on the Web

<http://www.w3.org/TR/NOTE-Web-privacy.html>

[RDF MS 98] Resource Description Framework (RDF) Model and Syntax

<http://www.w3.org/TR/WD-rdf-syntax>

[RDF MS Schema] Resource Description Framework (RDF) Schema

<http://www.w3.org/TR/WD-rdf-schema>

[TEI] Text Encoding Initiative

<http://www.uic.edu:80/orgs/tei/>

[XMLdata 98] XML-Data

<http://www.w3.org/TR/1998/NOTE-XML-data>

参考文献

[Bray and Guha 97] Tim Bray, R.V. Guha: An MCF Tutorial

<http://www.textuality.com/mcf/MCF-tutorial.html>

[Deutsch et.al 95] P.Deutsch, A.Emtage, M.Koster and M.Stumph: Publishing Information on the Internet with Anonymous FTP, IETF IAFA WG Internet Draft, 1995.

(<http://www.nlc-bnc.ca/ifla/documents/libraries/>

[cataloging/metadata/iafa.txt](http://www.nlc-bnc.ca/ifla/documents/libraries/cataloging/metadata/iafa.txt))

[IFLA] IFLA(International Federation of Library

Associations and Institutions):

"DIGITAL LIBRARIES: Metadata Resources",

<http://www.nlc-bnc.ca/ifla/II/metadata.htm>

[Netscape 97] Firefly, Netscape and Microsoft Cooperate to Build Upon Previously Proposed OPS Standard for Personalization With Privacy (Netscape news release)

<http://search.netscape.com/newsref/r/newsrelease423.html>

[小倉 97] 小倉利丸: PICSとは

<http://www.jca.or.jp/toshi/NetCensor/WhatPICS.html>

[日経 IT 97] MCF(Meta Content Framework)「コンテンツを体系化するMCF、検索や管理など情報の操作性を変える」、日経インターネットテクノロジー1997年11月号, pp.122-129, 1997.