

1. 紙から電子へ

小町祐史 (松下電送システム)

文書技術の発達、しばしば“紙から電子へ”というキーワードによって表現される。文書を人の思いやことばを時間的に固定したものにとらえるとき、それは、文字と、図・表などの関連表現手段との集合（文書情報）として構成され、これまで紙というサブストレートの上に固定されることがほとんどであった。この場合、紙は文書情報の（1）記憶と（2）表示の機能を兼ねると共に、文書情報をのせた紙を移動することによって（3）交換の機能をも果している。特に（1）、（2）は、エネルギー供給に関して受動的に実行されるという優れた性質を持つため、今後とも広く利用されることは間違いない。

しかし、利便性の向上と環境保全という2大ユーザ要求の充足をめざして、文書の電子化は推進されてきた。まず文書の生成過程に電子的な手段が導入され、いくつもの過程での電子化情報の形態をとる文書が成立し、その結果、電子化文書の交換と記憶が容易になり、電子的な交換・流通形態をとる出版が電子出版と呼ばれるようになった。ここでは、この段階を第1世代の電子化と呼ぶことにする。

文書の電子化によって、文書情報とそれを記憶したり交換したりする媒体とを分離して扱うことが容易になり、さらに文書の論理構造と、それを読者に理解しやすく表すためのフォーマットとを分離して扱うことが明示的になったことは、この段階における文書の電子化に伴う大きな成果であった。文書が電子的な交換の対象となったことにより、文書情報の標準化の必要性が指摘され、国際と国内の多くの標準化組織がその検討を開始した。

電子化された文書は、それだけに止まらず、Internetの発達と共にWeb環境に大量に分散蓄積されることとなった。それらの多くは、ディスプレイを含むコンピュータによって表示されることを前提として、動画像や音声までもコンテンツの一部として取り込んだマルチメディア文書となり、互いにハイパーリンクで結ばれたハイパメディア文書となった。ここではこれを、第2世代の電子化と呼ぶ。

ネットワークを介して大量の文書に容易にアクセスできる環境が普及すると、文書は人が閲覧するためのものに止まらず、検索ロボットのようなマシンもが扱えるもの（document-like object）とか、閲覧者の対話的介入によってその内容を変化させる対話形文書へと変貌を余儀なくされている。これが、ここでいう第3世代の電子化である。

このような文書の電子化に伴う文書の変化を概観し、主要なトピックを紹介して、今後の動向を展望する。

■ 文書の生成過程 ■

生成過程は、電子化された情報処理がそこに介在するか否かにかかわらず、おおむね次の段階に区分される。各段階で得られる文書は、次の段階に受け渡す必要があり、これまでも当事者間での合意に基づく受け渡しが行われていた。

(1) 内容情報（contents）の生成

著者が文書によって表現したい主要情報の作成、原稿用紙への文字の書込み、イラスト書き、図示対象の写真撮影など。

(2) 構造指定

文書内容を理解しやすくするための、情報の論理的構造化（論理的要素の抽出と構造的配置）、本文から脚注、備考などを分離し、段落に区切って見出しをつけることなど。

ここで（1）と（2）は主として著者が担当し、この順序に実行する作業過程である。しかし場合によ

ては作業順序が逆になり、構造を規定する枠が与えられて、そこに著者が内容を書き込むこともある。

(3) フォーマット指定

文書構造を把握しやすくすると共に、内容情報の可読性を向上させるために、論理的要素に物理的な配置・体裁を指定する。見出しを左上に置いてゴシック書体で印字し、それに続けて本文を2段組みで明朝書体によって印字するなどの指定。

これは通常、編集者が印刷業者に対して指示する作業である。

(4) フォーマット指定の実行（フォーマテイング）

フォーマット指定に従って文書を構成する論理的要素を配置すること。以前は活字を組む職人の作業であったが、今ではフォーマタ（たとえばTEX）と呼ばれる処理系が実行することが多い。

(5) 校正

フォーマット指定に従って配置された文書構成要素の集まりをある程度レビューできるように可視化した

ゲラ刷りなどを、(1)、(2)、(3)の作業担当者(著者、編集者)にフィードバックして、それぞれのレビューを受ける。

(6) 印刷(清刷り)

最終的なフォーマティング結果を紙面などに可視化すること。印刷業者によって実行されることが多い。

これらの(1)～(6)の作業は通常、それぞれの作業場所で異なる作業者によって実行される。つまり空間的、時間的に分散した環境で文書は生成される。各作業の効率化のために導入された電子化処理に伴い、それぞれの段階の電子化情報が伝送の対象となり、その受け渡しを行う範囲を広げるために標準的な文書情報交換様式(document information interchange format)の検討が行われてきた。

■ 第1世代の電子化 ■

□ タイプセティングの自動化

タイプセティング(植字)は、文書処理の中でかなり早くから電子化が検討され、電子化植字装置の製品が作成された。そこでは、文書の文字列中に、タイプセッタ(植字機)用の命令コードが書き込まれる(図-1参照)。この命令コードを植字機が解釈し、フォーマティングを実行して、印刷出力を行う。

命令コードは機器に依存していたため、この命令コードを含む文書の交換はかなり限定される。命令は印刷技術に関する指定が多く、命令コードの書込みには印刷の専門技術を必要とした。

□ 論理構造のマーク付け

文書中に直接、印刷技術に関連する命令を書くのではなく、文書を構成する論理的要素を示すマーク付け(図-2参照)が行われるようになり、次の利点がクローズアップされた。

- マーク付けを機器非依存にできる。
- マーク付けに用いるタグの可読性が高く、しかも印刷の専門技術に関係しない。その結果、文書内容の作成者は意味内容の記述に専念できる。

このような論理構造のマーク付けの結果、タグを

```
<fm12><pr20><ch123> 政治対話の強調
<fm11><pr16><ch112> ロンドンで開かれていたアジア欧州会議は...
```

図-1 タイプセッタ用命令コードを含む文書情報の例

```
<見出し> 政治対話の強調
<段落> ロンドンで開かれていたアジア欧州会議は...
```

図-2 論理構造のマーク付けを施した文書情報の例

印刷用の命令コードに対応付けして、それを実行するシステム、つまりフォーマタが必要になる。ここに至って、文書の論理構造を扱う系とフォーマタとが明確に分離された[☆]。

□ マーク付けの一般化

マーク付けはさらに、ある文書クラスの論理的要素を共通に識別するようなタグ集合へと一般化され、共通マーク付け(generic markup)呼ばれた。要素に関する属性記述をもタグに含めて、多様なアプリケーションに対応できるようにしたマーク付けも行われ、一般化マーク付け(generalized markup)となった¹⁾。

このようなタグ集合の定義方法を国際的に取り決め、言語として体系付けたものがISO(国際標準化機構)によって承認され、標準一般化マーク付け言語(Standard Generalized Markup Language, SGML)として制定された²⁾。これによって、さまざまなタイプの文書やアプリケーションに対して、一般化マーク付けを定義でき、さらに各種の補助機能によってさらに利便性の向上が図られている。

□ フォーマット情報の交換

文書交換は文書を構成する論理的要素とその構造だけで十分というわけではない。表示機能を大きく異にする装置間の文書交換では、フォーマット情報はローカルに設定しなくてはならないため、交換の対象は論理的要素に限定されるが、十分なフォーマティング機能と表示機能を持つ環境では、再編集の可能性を維持したまま交換による版面の一致または最適近似が要求されることが多い。そこで文書の論理的要素の情報とそれに対するフォーマット指定情報とが交換の対象となった。

□ フォーマティング

フォーマティングは、カットシートとか本とかのページ上の制約^{☆2}の中に、文書を構成する論理的要素などを読みやすくマッピングすることである。この制約には、次に例示するような版面の制約と記述内容の意味的制約とがある。

(1) 版面の制約の例

- ノンブルなどの版面外にくる要素以外は、基本的に版面の中に位置付けられる。

[☆] 通常のワードプロセッサは、文書の論理構造を扱う系とフォーマタとの分離を行っていない。文書内容をキー入力しながら、字下げ用スペースなどのフォーマット指定用のキー入力も行う。これがワードプロセッサの操作を複雑にすると共に、異機種で作成された文書データの交換を困難にしている。異機種ワードプロセッサ間のデータ交換に際しては、結局フォーマット情報を除去したプレーンテキストを交換の対象にすることが多い。
^{☆2} 最近の電子化文書では、ディスプレイ画面における制約になる場合がある。

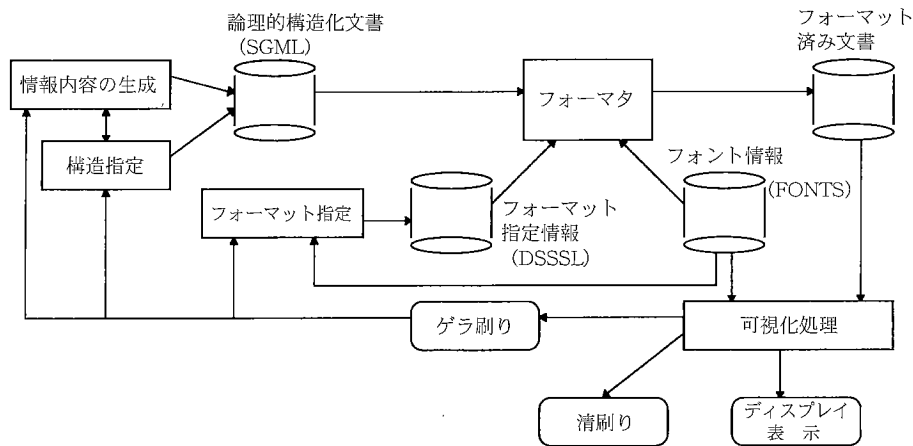


図-3 文書生成モデル

- 文字列は行末で改行され、次の行に続く。
- 図、表などが入るサブエリアは、分断されずに版面の中に位置付けられる。ただし見開きページでは、左右分断が許されることがある。

(2) 記述内容の意味的制約の例

- 改行に伴う分離禁止および禁則
- 図、表などが入るサブエリアは、原則としてそれを参照する参照元記述より後に位置付けられる。

フォーマティングの実行は、定式化されたこれらの制約のもとで解を求めるという問題に帰着するが、制約が厳しすぎると解がないこともあり、その場合には制約を緩和するための調停ルールを導入する必要がある。

ページものに関するフォーマティングルールは、出版・印刷業界で長い年月をかけて蓄積されたものがすでに存在し、ハウスルールとして運用されることが多い。電子化環境への適用も次第に進みつつある。

□ その他の文書情報

フォーマット情報に従ってフォーマットされた文書は、各種のプリンタやディスプレイによって表示される。それらの表示機能は多様であるため、表示機能に依存しないフォーマット済み文書情報が望まれる。この情報はプリンタやディスプレイの表示機能に応じた処理によって可視化される。

図形文字を含むフォーマティングや可視化処理では、図形文字を行に配置し可視化するためのフォント情報を必要とする。しかしすべてのフォーマタや表示装置が交換対象文書が必要とするフォント情報を含んでいるとは限らず、必要に応じてフォント情報を他のシステムから提供されなければならない。ここにフォント情報の交換が求められる。

■ 文書生成モデルと標準化

文書の生成過程における各処理と、処理間に受け渡される情報とを図示すると、図-3のようになる。ここで矩形が処理を示し、円筒形が情報を示す。それらの文書情報の交換には、その情報の表現形式、交換様式を送り手と受け手との間であらかじめ取り決めておくことが必要である。図中の円筒形に付した文字列は、すでに利用可能な情報表現形式または交換様式の名称の短縮形であり、正式な名称は次のとおりである。

SGML: Standard Generalized Markup Language²⁾

HTML: HyperText Markup Language³⁾

XML: Extensible Markup Language⁴⁾

DSSSL: Document Style Semantics and Specification Language⁵⁾

SPDL: Standard Page Description Language⁶⁾

PS: PostScript⁷⁾

FONTS: Font Information Interchange⁸⁾

これらの中のいくつかは、すでに広く利用されている。

この文書生成モデルは、文書処理の発達、特に各作業の電子化と各段階での文書情報交換への要求に伴ってしだいに明確化され、整備された。

■ 第2世代の電子化

□ マルチメディア文書

文字と、図・表などの関連表現手段との集合を文書情報とするとき、閲覧者に対してより理解しやすい文書情報表現が望まれる。表示媒体を紙に限定すると、関連表現手段は紙面にレンダリングされるものだけになるが、その限定を取り去った電子化環境では、文書内容に応じた多様な関連表現手段を導入することが可

能である。マルチメディア文書は、この電子化文書の利点を積極的に利用するものとして、早くから注目されていたが、動画画像や音声を含むマルチメディア文書が実用化するにはコンピュータの実行速度の向上を待つ必要があった。

なお、“多様な関連表現手段を導入”には、単に人間の各種知覚（視覚、聴覚、触覚、加速度に対する感覚、その他）に対応する各種表現手段の導入という意味だけでなく、情報技術を用いた活動によって生成される情報の生成段階に応じたさまざまな情報形態・符号化様式を持つ各種表現手段の導入という意味をも含めるのが適当であろう。たとえば本の生成過程には、符号化文字列の段階、フォーマット指定された段階、印刷され可視化された段階の各情報がある。聴覚に訴えるように音声に変換され朗読される本の情報形態もあり得る。音楽については、楽譜の段階の情報、それを演奏した情報がある。このような異なる段階に対応する情報形態を持ち、知覚可能な情報へのレンダリングを行えるようにした各種表現手段を含む文書がマルチメディア文書である。

◇ HTML

SGMLは外部実体としてどのようなファイルをも扱うことが可能であり、枠組みとしてはマルチメディア文書のサポートが可能である。しかし実際の環境でマルチメディア文書を扱うためには、各種表現手段に対応するファイルの内容を生成し、それをレンダリングして表示するシステムが必要であり、HTML文書とWebブラウザの普及によって多くのユーザがマルチメディア文書を手におさめることができた。HTMLではマルチメディアデータは外部ファイルとして扱われ、文書中へのその取り込みは、属性またはハイパリンクを指定する要素型によって指定する。データタイプはファイル名の拡張子によって、たとえば次のように示す。

GIF画像	.gif
TIFF画像	.tiff
AIFF音声	.aiff
MPEG動画像	.mpeg

◇ HyTime

HTMLはSGMLの1つのDTDであるが、マルチメディアデータをさらに柔軟に扱うために、SGMLの拡張であるHyTime⁹⁾が制定され、すでにその処理系¹⁰⁾も作られている。HyTimeは、文書内・文書間のリンク、他の情報オブジェクトとのリンクを指定し、時間的および空間的マルチメディア情報の配置を指定するための、標準化した機構を提供する。それによって、リンク情報やマルチメディア配置情報が、その情報を生成した環境とは異なるアプリケーションや実行環境による処理に対して利用可能になる。なお、このHyTimeについては、最近改訂版が出版された¹¹⁾。

HyTimeのマルチメディア対応機能の例を次に示す。

(1) 事象配置 (event schedule)

事象 (event) はオブジェクトを座標空間に割り付けるための要素であって、存在範囲指定と存在範囲調停 (オブジェクトが指定された存在範囲に満たないか、はみだす場合の処置を規定) の情報を持つ。隣接した事象は事象グループとして、共通の属性を関連付けられる。

事象配置は、事象の列を規定する。1つの座標空間には、複数の事象配置が可能であり、アプリケーションが各種の事象配置を定義できる。たとえば、複数種類のメディアに対応する事象をまとめて1つの事象配置を構成することもでき、メディアごとに事象をまとめて複数の事象配置を構成することもできる。

(2) オブジェクト修飾 (object modification) と事象投影 (event projection)

オブジェクト修飾と事象投影の機能によって、描出 (レンダリングの一般化) 処理を規定する引数を表現する。

オブジェクト修飾は、オブジェクトに対して、その描出の際に影響を与える要素 (たとえばグラフィクスに被せる色変換フィルタ) である。

事象投影は、ある座標空間から別の座標空間への (投影元から投影先への) 事象の存在範囲の変換である。これを使って、仮想時間で表現された音楽情報を実時間に写像することなどが可能になる。

□ ハイパメディア文書

マルチメディア情報へのアクセスには多様性が求められ、必ずしも論理時間の推移や事象列のシーケンスどおりにレンダリングを行う必要はない。遷移指示に従ってプレゼンテーションを行うとき、それを可能にする情報構造がハイパメディアであり、この構造を持つ文書がハイパメディア文書である。通常は2つ以上の情報オブジェクトの間の関係を表現するハイパリンクを用いて遷移先指示情報が記述される。

ハイパメディアは必ずしもマルチメディア環境だけで要求される技術ではない。たとえば仏典は多様な注を持ち、多くのバージョンに対する参照が多用される。そのため、以前からSGMLによるハイパメディア仏典記述²³⁾が各国の研究者達によって行われている。

◇ HyTime

HyTimeでは、単一文書中に、または1つのハイパ文書を構成する文書と他の情報オブジェクトとの間に、ハイパリンクを定義できる。

(1) ハイパリンクの生成

ハイパリンクは2つ以上のリンク端を持つ要素で表現する。リンク端は、端点 (anchor) と呼ぶオブジェクトをハイパリンクに関連付ける (図-4参照)。

ハイパリンクのもう1つの特性にリンク型があり、それは意味的役割をハイパリンクに関連付ける。リンク型はアプリケーションが規定し、ハイパリンク要素の要素型で表現する。

(2) ハイパリンク要素型形式

HyTimeは、次の要素型形式を規定している。

- 文脈独立リンク (independent link) : 柔軟性の高いハイパリンクであって、任意個数のリンク端を持ち、どこにでも端点を置ける。

- 文脈依存リンク (contextual link) : 2つの端点を持ち、端点の1つは文脈依存リンク要素の内容である。

- 記法固有文脈依存リンク (notlink) : HTMLのタグ<A>による簡潔な要素記述に対処して、HTML文書のHyTime文書への変換を容易にする。

◇ XLink

XLink (XML Linking Language)²¹⁾は、XMLリソースの中でオブジェクト間のリンクを記述するための構成子を規定する。シンタックスとしてXMLを用い、HTMLのリンク機能に加えて、次のような機能拡張が用意されている。

- (1) <A>で表される要素以外にもリンクを表現できる。
- (2) リンク処理方法を指定できる。
- (3) 論理構造をロケータで利用できる。
- (4) 3つ以上のリンク端を持てる。
- (5) リンクとリソースとを分離できる。

これらは、当初、XLL (Extensible Linking Language)¹²⁾として、XML言語規定と併行して開発されてきたものである。

■ 第3世代の電子化 ■

□ 対話形文書

対話形文書における対話形とは、利用者または他のソースからの入力 (stimulus) に対してアプリケーションが実時間ベースで応答することを意味する。この応答は、コンテンツと後の情報表現とを変化させることが多い。

対話形文書を生成する際には、ページベース文書の生成や簡単なハイパメディア文書の生成のための規準とは異なる規準が必要になる。つまり、文書の中にbehaviorをプログラムしなければならない。behavior要素を含むことは、対話形文書の開発システムと配布システムとの間で非互換性を引き起こす可能性を与える。

従来の受動的な文書とハイパメディア文書 (つまり第2世代の電子化までの文書) については、その構造、内容、およびフォーマット指定を規定する規格が存在する。しかし、対話形文書中のbehaviorを曖昧性な

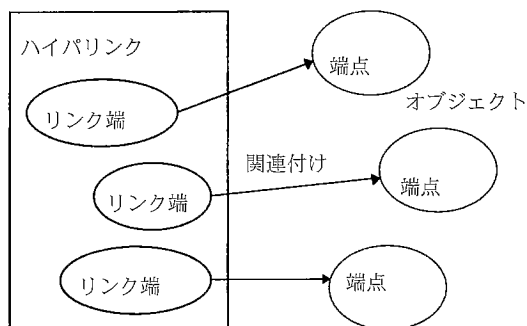


図-4 HyTimeのハイパリンク構成

く記述する規定はいまだない。異なるシステム間での対話形文書交換のためには、この規定を制定する必要がある。

プログラミング文脈の中にGUIオブジェクトを実装するための規格はすでに制定されている。しかし対話形文書を作る者は、装置非依存の方法で、どのようにしてこれらのインタフェースオブジェクトを利用しようとするかを表現可能にする必要があり、これを満たす規定への要求もある。

最近これらのユーザ要求が高まり、ISO/IEC JTC1に対してISMID¹³⁾という規格開発の作業課題が提案された。

□ 機械可読文書

最近のXML開発の背景には、少なくとも次の2項目がある。

- (1) SGMLツールの進歩と普及により、SGML文書をハンドマークアップで作成することがなくなり、そのためのSGML機能が不要になった。
- (2) HTMLの普及の結果、文書をブラウザによって閲覧するだけでなく、ネットワーク環境に分散した多くの文書をアプリケーションプログラムなどの処理系で処理したいという要求が高まった。そこでHTMLという固定したDTDだけでなく、SGMLと同様にユーザがDTDを指定することが必要になった。

この(2)の要求こそが、最近の電子化文書に求められる機械可読文書としての機能である。(2)で求められているアプリケーションは、表面的に次の3タイプに分類される¹²⁾。

- (1) いくつもの文書データベースをWebによって連携させる。
- (2) クライアントで文書情報に対する処理を行う。
- (3) WWWエージェントが文書情報を選択する。

(1)のアプリケーションのためには、連携させる文書情報をあらかじめ合意した互いに整合し得るDTDによって記述する必要がある。もともと連携させたい文書データは類似の構造を持ち、類似した

DTDで表現されることが多いが、完全には一致していないことがほとんどである。そこで、共通の上位DTDからの派生による各アプリケーションでのDTD開発¹⁴⁾、または共通するDTD部品を用いた各アプリケーションでのDTD開発¹⁵⁾などが検討されている。

(2) のアプリケーションの例として、半導体データシートを配布してそのデータを使ってクライアントで設計支援ツールを実行するシステムを挙げることができる。(1) の場合と同様に、クライアントの処理系と半導体メーカーが提供する半導体データシートとは、データの構造と記述方法とに関してあらかじめ合意しておく必要がある。半導体データシートについては、ECIX (Electronic Component Information Exchange) というプロジェクトが活動を行い、DTDの定義などを行っている¹⁹⁾。

(3) のアプリケーションのためには、文書情報選択を可能にするための、その文書に本質的な特徴を記述する文書関連情報 (メタデータ) を文書ごとに記述する必要がある。以降にメタデータに関するいくつかの活動を紹介する。

◇ Dublin Core

Dublin Core^{16), 17)} は、ネットワーク環境に分散した文書の本質的特徴を記述してインデックス作成を自動化するための、簡単かつ有用なメタデータ要素の集合を定式化するための試みであり、オブジェクトの記述に焦点を当て、著者、タイトル、主題などの記述データを含む13の要素 (図-5参照) を規定している。その規定に際しては、要素と、図書分類 (AACR2/MARC) やFGDCメタデータスキームなどの既存の専門化された記述システムとの間のマッピングが考慮され、対象範囲、関連性などについては、熟練した分類専門家を必要とするような細かな分類を行わないための記述的分類方法の一般化が試みられた。なお、シンタックスの問題は意図的に避けている。

これらのデータ要素の列挙に加えて、Dublin CoreをまとめたDublin Workshopは中核をなすメタデー

Subject : オブジェクトが扱うトピック
Title : オブジェクトの名前
Author : オブジェクトの知的コンテンツに主として責任のある人
Publisher : オブジェクトを利用可能にしているエージェントまたは機関
Date : 公表日時
ObjectType : 小説、詩、辞書といったオブジェクトのジャンル
Form : PostScriptファイルなどのオブジェクトの記法
Identifier : オブジェクトを一意的に定義するために用いられる文字または数字
Relation : その他のオブジェクトに対する関連事項
Source : このオブジェクトが派生したオブジェクト
Language : 知的コンテンツの言語
Coverage : オブジェクトの空間的配置および継続時間の特徴

図-5 Dublin Coreの要素

タ集合全体に適用される次の基本原理を明確にしている。

(1) Coreメタデータ集合は、サイト固有またはドメイン固有のデータ要素の追加を可能にするために、拡張可能である。

(2) Coreメタデータ集合のすべての要素は、Coreを用いたオブジェクトのどの特定記述においてもオプションである。

(3) すべての要素は繰り返し可能である。

(4) 各要素のすべてのセマンティクスは、次のいずれかによって変更できる。

- もっと詳細または固有のセマンティクスをあらかじめ存在していた外部のメタデータスキームからDublin Coreに取り込むことを許容する、他の既存のメタデータスキームから借りてきたqualifierの使用

- coreデータ要素の通常の意味を修正するために特にcoreを用いて行った特定に専門化と拡張

◇ Warwick Framework

Warwick Framework^{17), 18)} は、文書内容の提供者・分類作業・インデックス作成者と、自動的にリソースを見つけて記述するシステムとの間の相互運用性を拡大することを目的として、具体的で利用しやすいDublin Coreの定式化を求めるために、Warwick Workshopが行った成果であり、コンテナアーキテクチャの提案である。これは、異なるメタデータスキームによって記述されたメタデータを含むパッケージと、複数のパッケージを論理的に集めるためのメカニズムであるコンテナを主な構成要素とする。

メタデータ集合からなるパッケージは、Dublin CoreやMARCレコードなどによって記述されたネットワークオブジェクトの属性情報を直接保有する。HTMLの<META>タグによって、本体の一部として記述されたメタデータなどがこれに該当する。

インダイレクトのパッケージは、属性情報を直接保有せず、外部に存在するメタデータ集合を参照するポインタからなるパッケージである。インダイレクトのパッケージの参照先は、メタデータ集合からなるパッケージである場合と、別のインダイレクトパッケージをさらに参照したインダイレクトパッケージである場合とがある。

コンテナは、それ自体がコンテナであるパッケージである。再帰上の限界は定義されていない。これら3つのパッケージの関係を、図-6に示す。

メタデータは、パッケージとしてスキーマごとに独立して管理される。そのためスキーマの管理はスキーマを作成した主体に完全に委ねられ、管理の分散化が図られている。この枠組みでは、メタデータコンテナがオブジェクトによって内部的に参照され

る場合と、オブジェクトの外部から参照される場合の両方が可能である。内部から参照されるメタデータコンテナは、コンテンツオブジェクトの著作者または管理者がオブジェクトを記述するために選んだメタデータを指す。外部から参照されるメタデータコンテナは、コンテンツオブジェクトの作成者や管理者とは別の管理者によって作成され、維持されるメタデータを指す。

この仕組みをどう実現するかは、実装に委ねられているが、文献17)では、HTML, MIME, SGMLおよび分散環境での実装例を示している。

◇ RDF

RDF (Resource Definition Framework)^{20), 22)}は、Apple Computerで開発されたMCF (Meta Content Framework)をもとに、W3C (World Wide Web Consortium)で検討されたメタデータのためのフレームワークで、Web上の機械可読情報を交換するアプリケーション間の相互運用を目的とする。人が介在せずにWebリソースを自動処理できるように配慮され、そのアプリケーションには、検索エンジンによるリソース探索、電子図書館などの内容のカテゴリライズ、知識共有のためのソフトウェアエージェント、内容のランク付けなどが考えられている。

RDFの情報モデルでは、メタデータをリソース、プロパティ型、値という三つ組んで表現する。同一リソースの三つ組みの集合であるアサーションとしてメタデータを記述し、記述にXMLを用いている。

■ むすび ■

文書情報は紙から解き放たれることにより、急速にその活性度を高め、これまでの文書流通系からネットワーク環境に進出して、ハイパメディア化、マルチメディア化、機械可読化などの変貌を遂げてきた。当分の間この傾向は継続するであろう。その結果予想されることは、ネットワーク環境における膨大な情報量の文書の氾濫である。氾濫とは言っても、高度に発達したツールによって適切に作成され、適切にアクセスされる文書の氾濫である。すでにその兆候は、日々の作業の中で占めるEmail対処時間の増大に見ることができる。

結局問われるのは、それを最終的に処理する人の能力であり、今後期待される技術は、人の興味と能力に応じて電子化文書情報をフィルタリングし、それを適切なフロー制御のもとに人の脳に送りどけるツールということになりそうである。

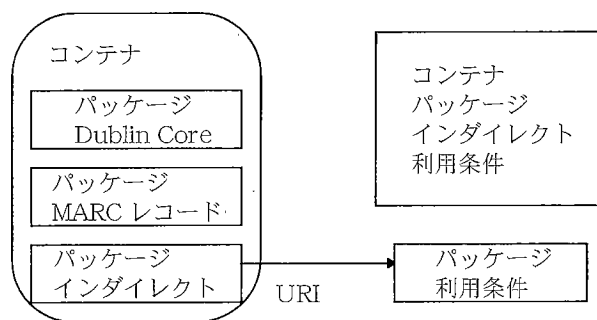


図-6 3つのパッケージを持つメタデータコンテナ

参考文献

- 1) Tucker, H. A. and Bogh, T.: SGML and ODA, Dansk Standiseringsrad (1989).
- 2) ISO 8879, Standard Generalized Markup Language (SGML) (Oct. 1986) およびISO 8879/Amendment 1 (July 1988) [JIS X 4151がこれらに対応].
- 3) Raggett, R. (W3C): HTML 3.2 Reference Specification, <http://www.w3.org/TR/REC-html32> (Jan. 1997).
- 4) Bray, T., Paoli, J. and Sperberg-McQueen, C. (W3C): Extensible Markup Language (XML) 1.0 Specification, <http://www.w3.org/TR/REC-xml> (Feb. 1998).
- 5) ISO/IEC 10179, Document Style Semantics and Specification Language (DSSSL) (Apr. 1996) [JIS X 4153がこれに対応].
- 6) ISO/IEC 10180, Standard Page Description Language (SPDL) (Dec. 1995) [JIS X 4154がこれに対応].
- 7) Adobe Systems: PostScript Language Reference Manual, Addison-Wesley (1990).
- 8) ISO/IEC 9541, Font Information Interchange (Sep. 1991) [JIS X 4161~4163がこれに対応].
- 9) ISO/IEC 10744, Hypermedia/Time-based Structuring Language (HyTime) (Nov. 1992) [JIS X 4155がこれに対応].
- 10) SGML SIGhyper, Interim Newsletter (July 1992).
- 11) ISO/IEC 10744 2nd ed., Hypermedia/Time-based Structuring Language (HyTime) (Aug. 1997).
- 12) 村田 真: XML入門, 日本経済新聞社 (Jan. 1998).
- 13) ISO/IEC JTC 1/WG4 N1949, Interchange Standard for Modifiable Interactive Documents (ISMID) (Dec. 1997).
- 14) INSTAC: 分散形情報共有環境における文書記述用の基準文書型定義 (G-DTD), 平成7年度電子文書処理システムの標準化に関する調査研究報告書 (Mar. 1996).
- 15) INSTAC: SGML要素型定義の部分的な交換・再利用・共有のための支援機構, 平成9年度電子文書処理システムの標準化に関する調査研究報告書 (Mar. 1998).
- 16) Weibel, S., Godby, J., Miller, E. and Daniel, R.: OCLC/NCSA Metadata Workshop Report, http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html (1995).
- 17) Lagoze, C., Lynch, C. and Daniel, R.: The Warwick Framework, A Container Architecture for Aggregating Sets of Metadata, <http://cs-tr.cs.cornell.edu:80/Dienst/Repository/2.0/Body/ncstr1.cornell%2fTR96-1593/html> (July 1996).
- 18) グローバル・コミュニケーション・センター: OEDC委員会/電子商取引消費者インタフェース委員会報告書, <http://www.y-adagio.com/public/reports/oedc/1997/oedc97.htm> (Mar. 1998).
- 19) Jeffery, T. and the ECIX Project: The ECIX Project (Electronic Component Information Exchange), <http://www.cfi.org/pinnacles/techwrite.html> (Jan. 1996).
- 20) Swick, R. (W3C): Resource Description Framework (RDF), <http://www.w3.org/RDF/Overview.html> (Feb. 1998).
- 21) Maler, E. and DeRose, S. (W3C): XML Linking Language (XLink), <http://www.w3.org/TR/1998/WD-xlink-19980303> (Mar. 1998).
- 22) Lassila, O. (W3C): Introduction to RDF Metadata, <http://www.w3.org/TR/NOTE-rdf-simple-intro-971113.html> (Nov. 1997).
- 23) 謝 清俊, et al.: 電子辞典中処理中文本的方法, 5th CJK DOCP (日中韓文書処理会議), CJK-DOCP/94-13 (May 1994). (平成10年4月7日受付)

訂 正

本誌第39巻6号（1998年6月号）の中で次の個所に誤りがありました。お詫びして訂正いたします。

会誌編集委員会

誤

正

- | | |
|---|--|
| (1) p.505, 右段, 上から20行目: DVD (Digital Video Disk) | DVD *特定用語の略語ではない |
| (2) p.515, 左段, 下から5行目: EDI (Electric Data Interchange) | EDI (Electronic Data Interchange) |
| (3) p.515, 左段, 下から6行目: EC (Electric Comers) | EC (Electronic Commerce) |
| (4) p.515, 左段, 下から22行目: HTML (Hyper Text Markup Language) | HTML (HyperText Markup Language) *HyperTextは 語 |