

6. 電子新聞における記事の標準化の現状

加藤博之, 森永裕美, 出本 浩 (NTTデータ通信 (株))

はじめに

世界的に広まったWWW (World Wide Web) をはじめとする情報通信技術の進展は、「電子新聞サービス」の可能性を広げるとともに、新聞業界自身にも変革を迫っている。本稿では、その変革の1つとして捉えられる記事の標準化の現状を明らかにしてみたい。

一言に「電子新聞サービス」といっても、その捉え方は一様でない。数年前、米国の新聞社を皮切りに始まったWWWによる新聞情報提供サービスの試みは、電子メールやPUSH技術によるニュース配信・PDF形式の紙面型ニュース情報提供などと、最新の技術を取り入れながら、今も変化し続けている。また、ISP (Internet Service Provider) ・ディレクトリサービス会社のような「サービスプロバイダ」が行っている、ニュース速報・株式情報サービスも多く見られるようになった。

ここでは「電子新聞」に、「電子的メディアで提供される新聞記事情報サービス」と広義の定義を与え、「電子新聞サービス」を取り巻く環境を整理した上で、標準化とのかかわりを明らかにし、国内外の新聞業界の標準化動向を探っていきたい。

まず、初めに、電子新聞を取り巻く環境を整理した上で、電子新聞サービスモデルを定義し、記事の標準化の目的を明確にする。次に、記事情報データの標準化動向として、記事のSGML化の現状を述べ、最後に、編集・表示に関連する標準化動向として、CDF, RDF, PDFについて適用事例を中心に簡単に触れる。

電子新聞における標準化の目的

電子新聞が購読者に提供されるまでの処理の流れ(サービスモデル)を整理すると図-1のようになる。

【電子新聞サービスモデル】

図-1に示すよう、通信社・新聞社から提供された記事情報は、サービスプロバイダ(新聞社自身の場合もある)に送られ、WWW・電子メール・PUSH技術といった各種方式に合わせた加工を行った後、購読者に提供される。この一連の処理の流れにかかわる、そ

それぞれの処理主体の立場から標準化の目的を考察する。

【通信社・新聞社】

通常、我々が新聞記事として目にするものには、各新聞社の記者が作成した記事と通信社が各新聞社に提供する記事情報「電文」との2つに分けられる(「電文」とは新聞業界用語であり、各通信社が提供する記事情報のことを指す。本稿では、以下「電文」と表現する)。現在の「電文」は、提供先である新聞社の契約内容および受信システムに合わせ、各社個別の形式となっている。このため、通信社および新聞社の設備・契約の更新ごとにシステム変更が必要であり、コストの面で不利な状況にある。

また、当初は、新聞社自身が「電子新聞サービス」を提供するという方式が中心であったが、WWW・メール・PUSH技術と情報提供手段が多様化した現在、すべてを新聞社自らがそれぞれに対応したシステムを開発するのは困難になりつつある。通信社はもとより、新聞社自身もコンテンツ提供に特化し、情報提供サービスそのものは「サービスプロバイダ」に任せるというスタイルも登場している。

【サービスプロバイダ】

多くの会社がひしめくISP・ディレクトリサービスにおいては、Internet接続・インデックスサービスといった本来のサービスだけでは集客効果は薄くなってきた。同業他社との差別化をはかるため、利用者へ情報提供サービスが盛んに行われており、その1つに「新聞記事情報」は位置付けられている。

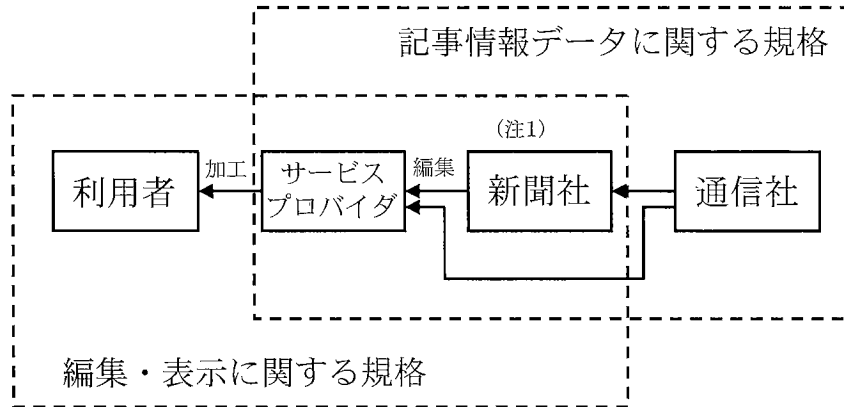
現在は、ニュース速報・チャンネル配信といった、記事情報の羅列に留まったものが多いが、今後は「記事の重要性などを加味した紙面」を電子メディア上で再現することが望まれると予想される。

【利用者】

WWW・電子メールといった電子的情報提供の枠組みにおいて、新聞情報は利用者にとって情報メニューの1つでしかなく、汎用的なアプリケーションによって閲覧できることが重要である。

【標準化への流れ】

このように、通信社・新聞社・サービスプロバイダという、新聞業界内／外に渡り、記事情報を流通させ



(注1) 新聞社自身がサービスプロバイダの場合もある

図-1 電子新聞のサービスモデル

<文書の構造の例>

文書型定義DTD

```

<!ELEMENT MEMO -- ((TO & FROM), BODY, Sign?) >
<!ELEMENT BODY -- 0 (P*) >
<!ELEMENT P -- 0 (#PCDATA | REF) >
<!ELEMENT REF -- 0 EMPTY >
<!ATTLIST MEMO secrecy (top | pub) pub >
<!ATTLIST P id ID #IMPLIED >
<!ATTLIST REF refid REFID #REQUIRED >

```

文書実現例

```

<!DOCTYPE MEMO SYSTEM "MEMO.DTD">
<MEMO secrecy="pub">
<TO>あなた <FROM>わたし
<BODY>
<P id="メモ"> SGMLの説明会を開催予定です。
<P>説明会<REF refid="メモ">は、明日行います
<SIGN> わたし
</MEMO>

```

DTDの指定

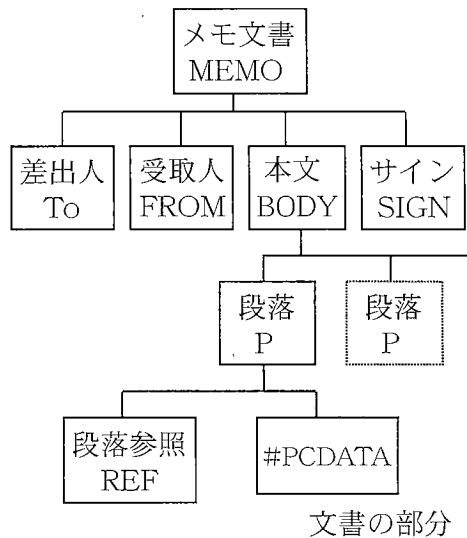


図-2 DTDとSGML文書の例

る仕組みを整備していく必要に迫られている。そして、その実現の第一歩として、記事情報形式の標準化が進められているのである。

そして、「電子新聞サービス」をより表現豊かなものにするためには、「サービスプロバイダ」を介在しても、閲覧クライアントが異なっても、新聞社の編集意図を忠実に再現できる仕組みが必要となってくるだろう。その中には、新聞記事情報の編集・表示に関する標準化が含まれると思われる。

まとめると、電子新聞に関する標準化は、通信社・新聞社・サービスプロバイダがかかわる記事情報データそのものに関する規格と、新聞社・サービスプロバイダ・利用者がかかわる記事情報データの編集・表示に関する規格との2つに分けて考えることができる。

■ 記事構造の標準化 ■

これまでに記事情報を流通させる仕組みを整備する必要があり、その第一歩として記事情報形式の標準化が進められていると述べた。ここでは、電子新聞における記事情報を実際にSGML形式で記述・管理（・配信）している新聞社、および記事情報SGMLの業界統一規格を制定しようとする新聞業界の動きに注目する。

□ 記事のSGML化

SGMLは、Standard Generalized Markup Languageの略で、ISO 8879およびJIS X 4151で制定された構造化文書を記述する国際標準規格であり、主

な適用事例として、米国国防総省の物資調達や保守支援のための文書、航空・自動車業界の整備マニュアル文書、また国内では各省庁の文書（白書など）が挙げられる。

SGML文書を一言で述べると、「構造化されたテキスト形式の文書」と言うことができる。構造化とは、具体的には対象文書の論理的構造の表現のことであり、論理構造をDTD（Document Type Definition）と呼ばれる形式で定義する。SGML文書はこのDTDに従って記述することになる。図-2にDTDとSGML文書の例を示す。

このように、記事情報データをSGMLで記述することにより、以下のようなメリットが生じる。

- 文書はすべてテキスト形式で表現され、かつ文書が構造化されているため、他メディアや他社情報との流通性に優れている。
- 構造化することで、見出し、概要、本文、,,、といった新聞記事に必要な情報の洩れを防ぐことができる。
- メディアに合わせた変換が可能であり、さまざまなメディアに出力することができる。
- 文書構造を識別するタグ（項目）を活用した部分検索、関連記事検索など、多様な検索が可能である。

□ 毎日新聞社の適用事例

毎日新聞社は電子新聞を事業化するにあたり、複数のメディアおよび複数のサービスプロバイダに情報を配信することを想定し、それに適した記事形式として、SGMLを採用した。

毎日新聞社は1996年8月より開始した毎日デイリークリックサービスにおいて、ジャストシステムのジャストネット、ニフティサーブ、ジーサーチ、ぶらら、凸版、OBERON、US INTERNETなどに配信しているが、毎日新聞社からすべて同じ記事（SGML形式）が配信されており、配信先の各社が記事構造を保持したままさまざまな表現形式に変換してサービスを実施している。

記事をSGML化するには、まずDTDと呼ばれる文書型定義を決める必要がある。DTD定義の検討作業には、NTTデータ通信と東芝が協力、初版は1995年11月に完成している。検討内容については、文献1)、文献2)を参照いただきたい。

SGML形式で記事を配信することは、電子新聞サービスの運営側・利用者側、いずれにとっても有効であると、毎日新聞社は、毎日デイリークリックでの実運用の経験から結論づけている²⁾。それぞれの利点は次のとおりである。

(1) 新聞社側

- 個々のサービス形態に合わせた異なる複数の素材を作成、管理する必要がない。

(2) サービスプロバイダ側

- 記事の表現方法は自由であり、サービスプロバイダ独自の表現方法やサービスが可能。
- 同一DTDによるSGMLテキストであれば、1つの変換プログラムで複数新聞社からの受信にも対応できる。

(3) 利用者側

- SGML記事の配信サービスが普及すれば個人で好みの表示が可能だし、複数新聞社の記事を同じユーザインタフェースで閲覧することも可能である。

□ WSJ（Wall Street Journal）の適用事例

米国のダウ・ジョーンズ社は、インターネットによる電子新聞サービスである「ウォールストリートジャーナル・インタラクティブ」の記事作成と管理にSGMLを採用している。

SGML採用の理由は、将来別の言語が必要になっても簡単に対応できる、その拡張性である。実際、最近アフターダークオンラインというサービスで別のマークアップ言語が必要となったが、変換プログラムを1つ開発するだけで簡単に対応できている。今後はWSJ本紙でもSGMLを採用する計画も進んでいるようである。

記事のSGML化では、その複雑さ故に素人では記述ができないという運用上の問題点があるが、ここでは「DJML」と名付けたSGMLのサブセットを作成するとともに、マイクロソフトワードを使った入力システムを開発し、SGML（DJML）を意識せずに記事を書けるという工夫がされていることも特記すべき事項である³⁾。

□ 日本新聞協会の取り組み

日本新聞協会の製作部会では、1996年2月から協会の会員の新聞社21社が参加する「SGML等利用研究会」を発足して、新聞記事情報の構造化作業（SGMLのDTD作成）などの検討を行っている^{4)、5)}。

この研究会では、1996年度の活動の中で通信社からの配信電文をターゲットとし、記事情報の構造化を行ってきた。ここで作成されたDTD（k-news）では、現状使用されているプレーンテキストとして配信している電文の中の情報を分析し、用途別に以下の3つの情報に区分して構造化した形となっている。

- 記事の属性情報
配信や記事の整理などに使用
- 記事本体の情報
この部分が新聞記事となる
- 課金のための情報
料金関連（再販権など）

この形は、配信会社から新聞社への情報伝達形態を簡潔に表している。

さらに、SGML等利用研究会では、1997年度に、共同通信社の電文を実験的に配信するテストを開始している。この実験の中で、新聞製作システムとの連携や、インターネットなどの電子媒体への新聞記事情報の提供などを検討している。

今後は、プレーンテキストでは表現が難しかった表などの特殊形態の電文（選挙投票結果の表や野球のスコア）を表示表現ではなく、データ表現で記述する方法や、DTDをXML（Extensible Markup Language）対応することによるインターネットへのコンテンツ提供実施などの検討を予定している。

その他にも、この研究会では国際新聞電気通信評議会（IPTC）が中心となって提案している新聞記事情報の流通フォーマット（後述）の日本国内での利用に対する検討や、（財）日本規格協会のG-DTDの分析などをテーマに幅広く活動している。

□ 国際新聞電気通信評議会（IPTC）⁶⁾の取り組み

IPTC（International Press Telecommunications Council）では、通信社や新聞記者から新聞社のみならず直接他業種（プロバイダ、DB業者）に販売するための新聞記事情報フォーマット形式NITF（News Industry Text Format）を提案している。NITFは、SGMLを使用して構造化されており、配信や記事の属性にかかわる部分としてHEAD部と、記事の内容を記述するBODY部に分かれている。

主な特徴として、BODY部に一般に広まっているHTMLのタグを採用することで、記者に受け入れやすいようにしている。また、内容記述の中では、登場するキーワード・日付などの情報にマークアップをつける形となっているため、データベース化時に検索しやすいような記事データとなっている。

IPTCでは、NITFを全世界で共通のフォーマットとして提案しており、各国の通信社や新聞社の事情や仕様を採り入れて行く方向を打ち出している。また、NITFはインターネットでの利用を対象としているため、XMLへの対応も今後の検討課題に入っている。

■ 編集・表示に関連する標準化動向 ■

今後、「電子新聞」の編集・表示に関連すると思われる、注目すべき標準をいくつか取り上げる。

• RDF（Resource Description Framework）⁷⁾

情報に関する情報（metadata）、すなわち、情報の作者・作成日時・キーワードといったものを記述する枠組みを規定している。XMLの表現を用いており、新聞記事情報のXML化が進む際は、記者名・記事日付・取材場所といった情報の記述はRDFの表記が採り入れられるだろう。また、このような情報が標準規格に則って表記されることにより、情報検索といった

分野に大きく影響を与えると思われる。

• CDF（Channel Definition Format）⁸⁾

CDFは、米国Microsoft社が提唱するpush型情報配信に関するフォーマットであり、情報更新のタイミング制御・掲載する情報の階層構造などを記述することができる。現在CDFに対応したPushクライアント（ActiveChannel・PointCastなど）は、この記述に従ったコンテンツを受信し、表示を行っている。

現在は、PUSH型クライアントのみを対象としているが、今後放送Channelに掲載する情報記述を発展させた形で、新聞記事の面を構成するための標準規格が登場するかもしれない。

• PDF（Portable Document Format）⁹⁾

PDFは米国Adobe Systems社が開発した電子文書配布用のデータ形式であり、現在、新聞紙面そのものを電子的に提供するための表現形式としては最有力候補である。米国ではLos Angeles Times、New York TimesがPDFによる記事提供を行っており、日本でも、朝日新聞社の1997年の甲子園速報号外配布・読売新聞社の一部記事提供といった試みが行われている。

■ 最後に ■

これまで「電子新聞サービス」は、新聞社・通信社そしてサービスプロバイダが各社独自に進めてきたものだったが、今後は本稿で述べたような標準化方式を適用することにより、ますます多様なサービスが提供されるであろう。

謝辞 最後に、この原稿を執筆する機会をくださった毎日新聞社メディア事業局長蓑田正彦様、新聞業界の標準化動向について、快く情報提供いただいた毎日新聞社メディア事業局情報開発部の小野寺尚希様、W3Cの標準化動向に示唆いただいたNTTデータ通信（株）技術開発本部マルチメディア技術センタの皆様に感謝します。

参考文献

- 1) 小野寺尚希, 吉田保雄, 蓑田正彦, 新村敦子: 情報流通を考慮した新聞記事のSGML化, 新聞技術, No.154, pp.73-80 (Apr. 1995).
- 2) 大槻亨介, 吉田保雄, 蓑田正彦, 加藤博之: 電子新聞のサービスモデルと事業化について, 新聞技術, No.160, pp.40-53 (Feb. 1997).
- 3) USA FRONT LINE 2, INTERNET Magazine (July 1997)
- 4) 技術委員会活動報告, 新聞技術, No.163, pp.111-112 (Jan. 1998).
- 5) 技術委員会活動報告, 新聞技術, No.160, pp.95 (Feb. 1997).
- 6) IPTC (International Press Telecommunications Council), <http://www.xe.net/iptc/>
- 7) RDF (Resource Description Framework), <http://www.w3.org/RDF/>
- 8) CDF (Channel Definition Format), <http://www.microsoft.com/standards/cdf.htm>
- 9) PDF (Portable Document Format), <http://www.adobe.com/prodindex/acrobat/AdobePDF.html>
(平成10年3月31日受付)