

## NewsMLに基づく記事配信システムについて

青崎 保好<sup>†</sup> 佐野 隆<sup>†</sup> 内田 強<sup>†</sup> 黒田 義和<sup>†</sup> 大園 忠親<sup>‡</sup> 新谷 虎松<sup>‡</sup><sup>†</sup>共同通信社システム局 <sup>‡</sup>名古屋工業大学大学院情報工学専攻

## 1 はじめに

共同通信社は、国内外のニュースを取材、編集して全国の新聞社、NHK、民間放送局など200社以上に提供している日本を代表する総合国際通信社である。配信するニュースは、政治、経済、社会、海外の記事、写真だけでなく、金融証券情報、選挙開票情報、科学ニュース、各種スポーツ記録や戦評、地域の話題、文芸、レジャーからラジオ・テレビ番組、天気図まであらゆる分野を網羅している。

共同通信は2003年12月より、それまでのテキスト形式による漢字テレタイプフォーマット（漢テレ）でのニュース配信と並行して、ウェブ時代に対応するために、XMLのNewsMLフォーマットでの配信を開始した。

本論文では、NewsML採用の経緯と約4年間におよぶNewsMLによる記事配信システムの運用実績から得られた知見に基づき、NewsMLの利点や有効なNewsMLの利用方法について議論する。

## 2 NewsML採用の経緯

ニュース配信の計算機処理は1975年に遡る。当時、取材、編集した記事は漢テレキーボードでオペレーターが紙テープにパンチし計算機へ入力後、配信あて先選別処理をして専用オンライン・ネットワークを通じ新聞社に提供してきた。配信記事は漢テレと呼ぶ新聞紙面スタイルのテキストデータであった。

記事だけでなく株価などの経済情報やプロ野球のスポーツ記録などさまざまなデータも記事と同じフォーマットで配信していた。インターネット時代に入り、受信社側から多様な用途に形式変換し利用したいという要望も出てきた。テキストデータのみによる配信には限界が

見えていた。90年代後半に始まったシステム更改の検討では、経済情報やスポーツ記録などのデータに関して受信側で情報を加工しやすい形式で送ることとし、当初、CSV形式の採用が有力だった。しかし、より柔軟な加工ができることや情報の追加や削除がシステム全体に影響しないなどのXMLの特性を評価し、将来性を考えてNewsMLが採用された。

NewsML[1]は、2000年に国際新聞電気通信評議会（International Press Telecommunications Council）が、ニュースの制作、交換、ユーザーでの利用のすべての過程で活用できるXMLベースの標準フォーマットとしてリリースした。日本では翌年に日本新聞協会がNewsML使用ガイドライン[2]を公表した。

共同通信のNewsMLは、日本新聞協会NewsML使用ガイドラインに準拠し、新聞社などへニュース配信するために必要な独自のタグを追加した。

## 3 NewsML記事配信システムの利点

NewsML記事配信システムは、記事を作成する複数のシステム群と記事を配信するシステムから構成されている。

記事を作成するシステムには「記事編集システム」や「経済情報システム」、「スポーツ記録システム」など情報提供元によって処理を分散して約20のサブシステムで構成され、サーバー数は150台ほどである。記事や画像などのデータはすべてNewsMLで管理している。

作成されたNewsML記事データは、共同通信から受信社である報道機関へ、受信社のプロファイルに合わせたあて先選別処理後、専用ネットワーク経由で24時間リアルタイムに配信される。1日に作成しているデータにおける記事の総データ数は約2万ファイル、総データ量は約200Mバイトである。

受信社を含めたシステム間のインターフェースには非同期通信のメッセージキューイング（MQ）通信[3]を採用し、システム間の独立性と高信頼性を確保した。

NewsMLの採用によって①リビジョン管理②記事や画像、メモとのリンク管理③カテゴリ情報④多彩な加工—などの利点が明らかになった。

Advantages of NewsML for the on-line news editing and distributing system

<sup>†</sup>AOSAKI Yasuyoshi, KYODO NEWS, Systems Department, aosaki.yasuyoshi@kyodonews.jp

<sup>†</sup>SANO Takashi, KYODO NEWS, Systems Department, sano.takashi@kyodonews.jp

<sup>†</sup>UCHIDA Tsuyoshi, KYODO NEWS, Systems Department, uchida.tsuyoshi@kyodonews.jp

<sup>†</sup>KURODA Yoshikazu, KYODO NEWS, Systems Department, kuroda.yoshikazu@kyodonews.jp

<sup>‡</sup>OZONO Tadachika, Graduate School of Engineering, Nagoya Institute of Technology, ozono@ics.nitech.ac.jp

<sup>‡</sup>SHINTANI Toramatsu, Graduate School of Engineering, Nagoya Institute of Technology, tora@ics.nitech.ac.jp

### ①記事のリビジョン管理

通信社は速報が第一であり、作成する記事は事象の発生を伝える「一報」から、次第に事象の中身が判明していくに従って記事の内容を厚くして配信することが多い。そのため記事のリビジョン管理が大切である。

### ②メモによる記事や画像のリンク管理

共同通信では「朝刊メモ」や「夕刊メモ」と呼ばれるその日の主だったニュースの一覧を記載したメニューを記事として配信しており、受信社側ではこの「メモ」を頼りに配信されてくる記事の状況を見ながら新聞紙面などの制作を行っている。これらの運用にとって、NewsMLのタグを用いたメモと記事のリンク情報の管理も大きな特色で、便利になった点である（図1）。

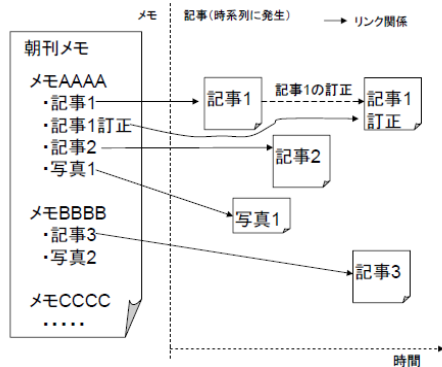


図1. メモと記事リンクの概略

### ③カテゴリ情報

記事をジャンルごとに分類付けすることで、データ出力先の制御や検索、再利用が容易になった。

### ④多彩な加工

漢テレフォーマット時代、受信側では共同通信が紙で提供していた記事のフォーム集を参考にテキストデータのフォーマット解析を行って紙面を制作していた。特にスポーツ記録のプロ野球の打撃成績テーブルなどの表データは、横表記の縦送信であったため変換処理が複雑だった。情報の構造化は必要な情報へのアクセスを容易にし、多彩な加工を実現した。

これらの利点を活用して、NewsMLデータの受信社では、地域の出身者をニュースから拾い出し掲載したり、地元のプロ野球やサッカーのチームの記録を他のチームの記録より詳しく掲載したりと独自の新聞紙面作りができるようになった。また、DBやウェブとのデータの共用、迅速なページ変更など大きな可能性が見えている。

メタ情報の活用で、送信側と受信側ともに効率的な編集活動ができるようになった。

## 4 NewsMLの課題

オンラインシステムで取り扱うデータには、動的に変化するデータと、永続的に変化しないデータがある。

共同通信のNewsMLでは、記事やメモの中に記事のリビジョン管理やメモのリンク管理を行うための動的情報をタグとして持っている。

運用では、記事の追加や訂正などがあればその都度、新規に作成し配信する。メモにリンクされる記事は主要ニュースだけで、メモは定時に配信した後、記事に動きがあればその度に追加部分をメモとして配信する。記事が存在しないうちに予定として記事をメモに記載し配信するケースや記事配信後に格上げしてメモとリンクされるケースなど、複雑な運用が多くある。

メモのリンク管理には2つのIDを使用して相互リンクを送信側と受信側で同期させているが、実データ（記事とメモ）内のメタ情報で同期を管理する方法は受信側に負担を強いている。

負担解消には、文書間のリンクの整合性維持のための情報を適切に配信する必要がある。

ユーザーの声を反映し課題を解決しながら、さらに使いやすいニュース交換フォーマットとして、サービスを中心とした考え方 [4]に基づいたNewsMLも念頭に置く必要があるだろう。

## 5 おわりに

国内最大規模の記事配信システムにおけるNewsMLの採用は、送信側と受信側ともに大きな利点をもたらし、将来へのさらなる発展の可能性をも示している。

今後、既存のサービス形態を堅持しながらも、XMLの特性を生かしたウェブ配信などの新しい配信方式の検討も必要である。また、データを配信する仕組みだけでなく、知識社会に向けた記事を作成するプロセスへのナレッジ・マネジメントの導入も検討が必要であり、XMLの高度利用のためのさらなる研究が望まれる。

### 参考文献

[1] International Press Telecommunications Council (IPTC), "News Markup Language : NewsML", <http://www.newsml.org/pages/index.php>, Accessed on 2007/Jan/03

[2] (社)日本新聞協会, "NewsML.jp", <http://www.newsml.jp/>, Accessed on 2007/Jan/03

[3] 都築 裕也・久保田 整一郎, "TP1/Message Queueにおけるメッセージキューイング(MQ)機能の実現およびMQの適用分野", 情報処理学会全国大会講演論文集, Vol.第52回 No.1, pp.175-176(1996)

[4] 日高一義, "サービス・サイエンスについての動向", 情報処理 Vol.47(5), pp.467-472(2006)