

## メタ記事記述言語 KDML における記事リンク整合性管理機構の試作

伊藤 正都\* 児玉 政幸\* 大園 忠親\* 新谷 虎松\* 黒田 義和† 青崎 保好‡

名古屋工業大学大学院工学研究科情報工学専攻\* (社)共同通信社システム局システム技術部†

(社)共同通信社システム局システム企画室‡

e-mail: {itomasa, kodama, ozono, tora}@ics.nitech.ac.jp

## 1 はじめに

近年, ニュース配信・管理フォーマットとして, 国際新聞電気通信評議会 (IPTC)<sup>1</sup>により策定された, NewsML(News Markup Language)<sup>2</sup>が利用されている [1]. NewsML とは XML ベースのフォーマットであり, 柔軟に多くのデータを格納することが可能である. ニュース配信・管理における NewsML 文書の内部は, 大きく次の二つに分けられる. 一つ目は, ニュース記事自体の項目 (コンテンツ) であり, 二つ目はニュース記事間のリンクを表す項目である. ニュース記事自体の項目とは, 実際に配信されるニュースの本文, およびそのニュースに付帯する地域, 分野などのメタ情報である. ニュース記事間のリンクを表す項目とは, 配信したニュース記事間のリンク情報である. ニュース配信において, 重要なニュースはニュース記事間のリンク構造からリスト化されて表示される. そのため, ニュース記事間のリンクを表し, その情報を維持することは, ニュース配信社, 加盟社にとって非常に重要である.

ニュース配信の特徴として, ニュース記事の作成, および更新頻度が非常に高いことが上げられる. この結果, ニュース配信において, ニュース記事間のリンクを維持することは容易ではない. この理由として, ニュース配信社は配信済みの記事を操作することが出来ないことが挙げられる. 例えば, ニュース記事の内容を訂正する場合, 訂正部を示した新たな NewsML 文書が配信される. そのため, 前回送られた NewsML 文書と訂正部を示した NewsML 文書間のリンク構造を作成・維持する必要がある. さらに, 訂正だけでなく, ニュース記事が複数の記事から構成される場合もあり, これらも NewsML 文書間のリンク構造を維持する必要がある. このように, NewsML によるニュース配信・管理

Building a Consistency Maintenance Mechanism of News Article Links by Meta-Article Markup Language KDML

Masato ITO, Masayuki KODAMA, Tadachika OZONO, Toramatsu SHINTANI, Yoshigazu KURODA, and Yasuyoshi AOSAKI

Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya, 466-8555 JAPAN

<sup>1</sup><http://www.iptc.org/>

<sup>2</sup><http://newsml.org/>

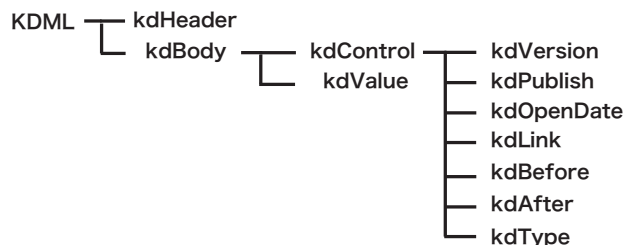


図 1: KDML におけるメタデータ構造

において, 配信される NewsML 文書は増加する一方であり, またどのような状況においても NewsML 文書間のリンク構造を維持することが必要である.

本稿では, NewsML 文書間でのリンクの維持手法として NVML(News Value Markup Language)[2] を基にした, 大園らの KDML(Knowledge Description Markup Language)[3] を用いる, NewsML 文書間のリンク整合性機構について述べる.

## 2 メタ記事配信言語 KDML

KDML とは, ニュース配信における NewsML 文書から, リンク管理情報とコンテンツ管理情報を切り離れたものであり, NewsML 文書での配信・管理を独立して制御するための言語である.

## 2.1 KDML を用いた記事管理

ニュース記事は頻繁に作成され, その都度配信される. しかしながら, NewsML 文書を用いたリンク管理では, NewsML 文書間のリンク情報がそれぞれの NewsML 文書に含まれるため, リンク整合性を管理するコストが大きくなる. また, ニュース記事の更新時には, 更新前のリンク情報, および関連記事へのリンク情報を更新する必要があり, リンク情報は複雑になり, 整合性を保つのは容易ではない. そのため, 本稿ではニュース配信・管理において, NewsML から配信・管理に特化した KDML を用いることで記事間のリンク整合性を管理する手法を提案する.

## 2.2 KDML の仕様

KDML におけるメタデータの階層構造は図 1 である. KDML はヘッダと内容部があり, ヘッダ (kdHeader) にはその KDML 文書自身のメタ情報が含まれる. 内容

表 1: kdControl の仕様

タグ	用途
kdControl	KDML 文書管理情報
kdVersion	KDML 文書バージョン
kdPublish	KDML 文書配信日付
kdOpenDate	ニュース記事解禁日時
kdLink	KDML 同士のリンク情報 , KDML-NewsML 間のリンク情報
kdBefore	修正前情報
kdAfter	修正後情報
kdType	KDML の種類

部 (KdBody) には KDML 文書間のメタ情報 (kdControl), および KDML 文書で管理している NewsML 内の記事に関するメタ情報 (kdValue) が含まれる。本稿において最も用いる情報は KDML 文書間のメタ情報 (kdControl) であり, 表 1 に仕様を示す。

kdControl におけるバージョン (kdVersion) とは KDML 文書の配信数である。常に最新のバージョンを持つ KDML をリンク構造の整合性管理するために利用する。KDML-NewsML 間のリンク管理情報 (kdLink), は KDML 文書同士, もしくは KDML 文書と NewsML 文書を結びつける情報である。kdLink はリンクの構造を示しており, リンクの整合性を保つために最も重要な情報である。kdControl では, その他に配信日時 (kdPublish), 解禁日時 (kdOpenDate), KDML 間, 新規・更新の種別 (kdType), 修正前情報 (kdBefore), 修正後情報 (kdAfter) などの情報が含まれる。これらは KDML を用いてニュース配信を管理するためのメタデータ群である。

### 3 記事リンク整合性機構

KDML を用いて記事リンクの整合性を管理する利点は, NewsML 文書リンク構造分離し, 扱い易くすることである。NewsML 文書のみでのニュース配信では, NewsML を利用した記事の管理と記事が混在しており, リンク構造の取得が容易ではない。記事リンク整合性機構を利用し, リンク情報を NewsML 文書から KDML に分離する事で, ニュース記事ツリーの作成を従来の NewsML 文書から作成するのと比較し, 容易にすることが可能となる。この結果, ニュース配信時にリンク構造を用いた, ニュース記事のリスト表示が容易になる。

#### 3.1 リンク整合性管理の比較

表 2 において, KDML を用いたニュース配信と従来通りの NewsML でのリンク整合性管理方法のメリッ

表 2: リンク整合性管理の比較

	メリット	デメリット
NewsML	・NewsML のみで 配信が可能	・記事と管理情報が混在し煩雑 ・検索コストが大きい
KDML	・文書間のリンク 管理が容易 ・リンク構造の表示が容易	・KDML が必要 (データ量の増加)

ト・デメリットを比較した。

KDML によりリンクの整合性を維持する事は, 配信コストを考える場合, NewsML 文書のみでのリンク整合性管理と比較し不利である。しかしながら, 昨今のネットワークインフラ, ストレージの発展を考えるに, 大きな問題ではないと考えられる。逆に, KDML を用いてリンク整合性を維持する事により, より容易にニュース記事間の相関をリンクとして表現する事が可能となる。

### 4 おわりに

本稿では KDML を用いた記事リンク整合性機構について述べた。KDML による記事リンク整合性の利用は, NewsML でのリンク整合性管理より容易にリンクの管理が可能となる。その結果, NewsML を用いた, 今までの同一記事間でのリンク構造の利用に留まらず, KDML のリンク構造を用いる事でニュース記事のトピック, 登場人物, および内容等による相関を表す事が可能となる。また, KDML のリンク構造を用いる事で, ニュース配信社・加盟社において NewsML 文書の取り扱いが容易になる。

### 参考文献

- [1] 井上明, 猪狩淳一, 金田重郎, "ニュース配信のための国際データフォーマット NewsML: その概要と現状について" 情処学情報システムと社会環境論研報, Vol.2002, No.056, pp.1-8, 2002.
- [2] 児玉 政幸, 伊藤 正都, 大園 忠親, 新谷 虎松, "次世代記事編集システムにおける NewsML を用いたメタデータ配信について" 合同エージェントワークショップ&シンポジウム 2006 論文集, 2006.
- [3] 大園 忠親, 新谷 虎松, 青崎 保好, "NewsML 記事のためのメタ記事記述言語 KDML について" 第 69 回情報処理学会全国大会論文集, 2007.