

Webを用いた記事配信のための同期型メタ情報配信システム A Synchronous Distribution System of Metadata

for a News Syndication on the WWW

児玉政幸[†] 伊藤正都[†] 大園忠親[†] 新谷虎松[†]

Masayuki Kodama, Masato Ito, Tadachika Ozono, Toramatsu Shintani

1 はじめに

通信社とは国内外のニュースを取材、編集して加盟社(全国の新聞社, NHK, 民間放送局など)に提供する組織であり, 代表的な通信社に共同通信社¹がある。共同通信社では出稿メモと呼ばれる特殊な記事を通信社社内, および加盟社に配信している。出稿メモとはその日の主要記事をまとめた出稿メニューであり, その用途としては, 大量の記事群からのフィルタリングや記事の整理などの目的に利用され, 通信社, および加盟社にとって欠かせない重要な記事となっている。出稿メモには各記事に関する様々な情報, たとえば要約文や記事への参照情報, 記事特有のキーワードなどが含まれている。本研究では, 上記の出稿メモに含まれる様々な情報を記事のメタ情報と定義する。出稿メモは NewsML²と呼ばれる XML フォーマットで記述され, 各記事のノードから構成されており, さらにその記事のノードの子要素にメタ情報を記述したノードが含まれる。また, 記事と出稿メモにはそれぞれユニークな ID が付与されており, その ID を頼りに記事と出稿メモを互いに参照できる。

通信社, および加盟社では, 出稿メモが記事の編集状況を確認する手段になっており, 配信されてくる記事のメタ情報を参照しながら記事の修正, 訂正を依頼したり, 紙面の構成を整えたりなどしている。このようなやりとりの中で, 記事が次々と作成され編集されれば記事のメタ情報も変わる。メタ情報の変化に柔軟に対応してそれを正確に配信するために, メタ情報が更新されるタイミングに合わせて自動的に再配信する環境が望ましい。現状のシステムで最新の出稿メモを閲覧するためには, 閲覧者がシステムに対して最新の出稿メモを受信する要求を出す必要がある。現状の問題点として, 閲覧者がたまたま要求作業を怠ると古いメタ情報を頼りに作業が進められ, 編集作業のトラブルに繋がることが挙げられる。また, 次々と配信されてくる出稿メモの変化をトレースしたいという要望があるにも関わらず, システムとして実現されていないのも問題である。

そこで本論文では, メタ情報のプッシュ配信機能, およびメタ情報のトレース機能をもつメタ情報配信システム TMemo を提案し, 実装する。我々は, ユーザが指定した時間帯に Web ページにコンテンツをプッシュ配信する技術 (Wisdom Ad Balloon:以下 WAB)[3]を開発している。本研究では, 上記で述べた問題点を解決するために, メタ情報配信信用に WAB を拡張した。

第2章で, 本システムの基盤技術となる WAB について述べる。次に第3章で, WAB を応用した出稿メモのリアルタイムプッシュ配信システムの概要, およびシステム構成について述べる。第4章では, 関連研究との比較をおこない本システムの有用性について議論し, 最後に第5章で本論文をまとめる。

2 TMemo の基盤技術

TMemo は Web ページをインタフェースとしたメタ情報のリアルタイムプッシュ配信システムである。本システムは,

文献 [3] の技術を基盤とした応用システムである。以下に, 本システムの基盤となる技術についての簡単な説明をおこなう。

上述したように, 我々は, 文献 [3] において Wisdom Ad-Balloon と呼ばれるシステムを開発した。本システムは Web エージェントを利用し Web 上でリアルタイムにコンテンツの制御を可能とする。Web エージェントは主に Javascript と CGI で実装されている。本システムの特徴は2点あり, 1点目はコンテンツの配信時間帯を自由に設定できる点である。2点目はエージェントを組み込んだ Web ページを閲覧することで, 動的にコンテンツが表示されていく点である。Web ページにあらかじめチャンネルと呼ばれるコンテンツ表示用の受け口を用意しておくことで, 指定した配信時間帯にコンテンツをプッシュ配信することができる。閲覧している Web ページ上にはエージェントが配置されており, コンテンツに変更があった場合エージェントが情報更新を確認しページのコンテンツを書き換えることでリアルタイムプッシュ配信を実現している。

3 TMemo:Web を用いたメタ情報配信

3.1 WAB の拡張

現在 WAB は広告という媒体に絞ってシステムが設計されているため, 出稿メモ配信に直接適用しても十分な効果が得られない。出稿メモ配信信用に WAB の設計を拡張する必要がある。

出稿メモ配信の特徴として, 一度配信された出稿メモのメタ情報は記事の新規作成や編集, あるいは社会情勢や事件の展開などに伴い, 何度も変更されるのが一般的である。それに伴い, バージョン変更され内容が変わった出稿メモはその都度配信されていくことになる。そのような中, 出稿メモの内容を時系列に沿ってトレースしていく作業が頻繁に存在する。たとえば, 現バージョンの記事がいくつもの変更を伴っている場合, いつ, どの部分に変更されたのかを知るために, あるいは変更された原因が記者のミスであるのか, それとも事象の展開によって生じたのかを知るために上記作業をおこなう。WAB では, 一度配信したコンテンツは配信時間帯の間あまり変更されないと仮定しているため, 配信したコンテンツの時系列変化をトレースする機能は設けておらず, 本システム用に拡張する必要がある。

3.2 システム概要

TMemo は Web ページをインタフェースとした出稿メモのリアルタイムプッシュ配信システムである。本システムのユーザは出稿メモを編集し配信する配信者, および配信されてくる出稿メモを閲覧する閲覧者に分類される。

本システムの全体像を図1に示す。配信者はメモエディタで出稿メモを編集し, 何時から出稿メモを配信するかをメモエディタから入力する。配信設定を確認すると, 出稿メモの情報, および配信時間帯が WAB の番組表に登録される。番組表とは横軸にチャンネル一覧を, 縦軸に時間軸をとった二次元テーブルである。チャンネルにはプッシュ配信するコンテンツが登録される。本研究におけるコンテンツとは出稿メモに含まれる各メタ情報のことを指す。閲覧者が立ち上げている Web ページにはページエージェントが配置される。ペー

[†]名古屋工業大学大学院工学研究科情報工学専攻

¹<http://www.kyodo.co.jp/>

²<http://newsml.org/>

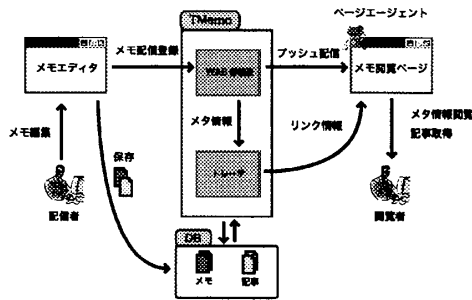


図 1: TMemo の構成

ページエージェントは番組表を監視し出稿メモの各メタ情報をクライアント側へダウンロードし、Web ページに配置するエージェントと定義する。閲覧者はページをリロードすることなしに、最新のメタ情報を取得できる。

配信されたメタ情報は時系列上にすべて保存される。つまり、リアルタイムに配信されてくる現時点でのメタ情報以外に、過去の任意の時点におけるメタ情報を閲覧者は確認することができる。従来方式では、通信社が定めた時刻にのみ出稿メモが配信されていたため、断続的にしか出稿メモは保存されていなかった。前節で述べたように、記事編集の変遷を辿るためなどにおいて、過去から現在までのメタ情報を継続的に閲覧しなければいけない状況が存在するため、従来方式では問題がある。本システムは上記問題を解決しており、任意の時間における出稿メモの内容や、ある記事におけるメタ情報の更新履歴(メタ情報のトレース)を出力することができる。

3.3 システム構成

本システムは、出稿メモを編集するためのメモエディタ、出稿メモのメタ情報をプッシュ配信するための WAB サーバ、およびプッシュ配信される出稿メモを閲覧するためのメモ閲覧ページから構成される(図 1)。本システムは Web アプリケーションであり、Web ブラウザのみを用いて出稿メモの編集、配信、および閲覧を可能とする。

メモ閲覧ページには、メタ情報をプッシュ配信するための受け口(以下、チャンネル)をあらかじめ用意しておく。図 2 で示した各メタ情報に対して、一つのチャンネルを用意することになる。各出稿メモに含まれる記事数、および 1 記事のメタ情報の種類を考慮した上で、チャンネル数を $3000(100(\text{記事数}) \times 30(\text{メタ情報の種類数}))$ 用意した。配信者がメモエディタを用いて図 2 のような出稿メモを作成すると、出稿メモがデータベースに保存される。配信者は、出稿メモを何時から配信するかをメモエディタから入力する。入力された配信時間帯、および配信する出稿メモが WAB サーバのデータベースに保存される。

メモ閲覧ページのページエージェントは WAB サーバに対し、定期的にデータの更新があるかをチェックしに行く。データの更新の際にはタイムスタンプを用い、このタイムスタンプを比較することでデータの更新がされたかどうかのチェックをおこなう。配信者が出稿メモの一部を書き換えるなどしてデータが更新されている場合、ページエージェントは最新のデータを取得している。そして、取得したデータからメモ閲覧ページのどのチャンネルに何時の時間帯にメタ情報が表示されるべきかというデータの整理を行う。整理を行った結果、現在の時刻とメタ情報の表示時刻を監視し、表示されるべきメタ情報が手元にあった場合はメモ閲覧ページにメタ情報をプッシュする。このように、ページエージェントがバックグラウンドで情報の最新性をチェックし Web ページを書き換えるため、Web ページに表示されるメタ情報をリロード

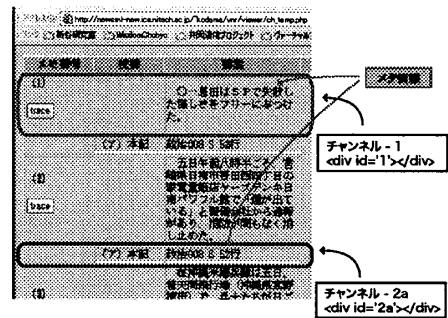


図 2: 出稿メモ配信例

無しに最新に保つことできる。

閲覧者がメモ閲覧ページのトレースボタンをクリックすると、トレーサを通じてエージェントがその記事の ID をもとに番組表からその記事のすべてのメタ情報を取得してくる。取得してきたメタ情報を時系列に沿って整理し、新たな Web ページ上にメタ情報の一覧を表示させる。

4 関連研究と考察

通信社では記事の絶え間ない編集に伴い出稿メモの更新頻度も高く、リアルタイムに情報が変更されるインターフェースは非常に効果的だと考えられる。本システムのメモ閲覧ページにアクセスすれば、すべての閲覧者が同じメタ情報を閲覧できるため、誰かが古い情報を頼りに編集を進めるといった状況を回避できる。

リアルタイム情報配信に関する研究は数多く存在しており、たとえば文献 [2] は、Web の情報を複数のユーザ間でリアルタイムに共有できる点で本研究と類似している。しかし、出稿メモ配信においては、ただリアルタイムに情報が配信されるだけでは不十分であり、過去の出稿メモをインタラクティブに出力できる機能が必要とされる。たとえば、何度も修正、訂正、追記された記事がどの時点で、どの箇所が変化したのかを確認する状況において、上記の機能がないと非常に不便である。従来システムでは、このような機能を実装しておらず、人間の目で比較する必要がある。本システムでは、番組表をもとに過去のメタ情報の変化の過程を提示することができる。

5 まとめ

本論文では、メタ情報のプッシュ配信機能、およびメタ情報のトレース機能をもつメタ情報配信システム TMemo について述べた。本システムは、WAB の技術を基盤としたシステムであり、メタ情報配信用に WAB を拡張し応用させたものである。本システムの成果として、すべての閲覧者に対して同じメタ情報を提供し編集作業の食い違いを解消できることがまず挙げられ、さらにメタ情報の変化、つまり編集状況の変化を過去にさかのぼってトレースすることも可能とした点が挙げられる。

参考文献

- [1] 青崎保好, 佐野隆, 内田強, 黒田義和, 大園忠親, 新谷虎松, "NewsML に基づく記事配信システムについて", 第 69 回情報処理学会全国大会論文集, Mar. 2007.
- [2] 木村訓康, 田村浩一郎, "Put It On Web - Web ページに情報を付加することによる情報共有の手法-", 情報処理学会研究報告, Vol.2006, No.9, pp.105-110, 2006.
- [3] 向井康人, 大園忠親, 伊藤孝行, 新谷虎松, "WisdomAd-balloon: Push 型情報配信技術に基づく動的ページ構成システム", 第 4 回情報科学技術フォーラム, 情報処理学会, Sept. 2005.