

## 連続メディア型情報提供に対するフィルタリング サービスシステムの構築

北嶋 玄樹†, 中村 友昭‡, 知念 賢一†, 山口 英†

†奈良先端科学技術大学院大学 情報科学研究科

‡松下電器産業株式会社 AVC 商品開発研究所

### 概要

テレビ放送や NetNews などの様々な種類の情報が人間の処理能力を越える量で提供されている。また、メディア毎に様々なデータフォーマットが存在し、統一的に扱うための手法が望まれている。そこで本研究では、これらの問題を解決するために、CMFS モデルを提案する。本モデルは、様々なデータフォーマットを統一フォーマットに変換し、スケーラビリティと通信の信頼性を考慮してユーザに配送する。そして、情報をフィルタリングすることによって必要な情報を得るためのシステムのモデルである。本モデルの有効性を検証するために CMFS モデルを文字放送と NetNews を情報源として実装し、評価を行なった。

## Design and Implementaion of the Filtering Service System for Continuous Media

Tuneki Kitazima†, Tomoaki Nakamura‡, Kenichi Chinen†, Suguru Yamaguchi†

†Nara Institute of Science and Technology

‡Matsushita Electric Industrial Co, Ltd. AVC Products Development Laboratory

### Abstract

Today, there are several information sources, such as TV or NetNews, which are providing "information" continuously in 24 hours. The amount of information provided by these sources is too large for users to check/retrieve byte by byte. In this model, since information come from several sources is converted from its won data format to the CMFS format, applications based or the CMFS model can handle several information sources in a single manner. Furthermore, the transmission mechanism for delivering information on the CMFS model provide scalability and reliability. In this article, we show CMFS model and sample applications for handling both TELETEXT service and NetNews.

## 1 はじめに

テレビ放送やラジオ放送のように、次々と情報が提供される情報を連続メディア型情報と呼ぶ。このような情報を提供するサービスは、大量に存在し、それらを通して人間の処理能力を越える量の情報が提供されている。このような状況下では、提供される全ての情報に人間が目を通すことは不可能である。さらに、提供されている多くの情報は特定の個人にとって必要のない情報である。

一方、情報のデジタル化によってコンピュータで情報を利用することが容易になってきている。また、コンピュータで情報を処理することにより、情報に新たな付加価値をつけることが可能となる。

このような状況から、必要とする情報だけを入力したいという利用者からの要求が強まっている。この要求を満たすために情報のフィルタリングの研究が行なわれているが完全な解が示されていない [1]。

さらに、様々な種類の連続メディア型情報は、それぞれのメディアによってデータフォーマットが異なっているという問題がある。

そこで、本研究ではこのような状況を改善するために Continuous Media Filtering Service (CMFS) モデル [2] を提案する。このモデルは、以下の機能を実現する。

- i) 多種の連続メディア型情報を統一フォーマットに変換し、効率的に取り扱う。
- ii) フィルタリングを行なうことで目的の情報のみを取り出す。
- iii) スケーラビリティとデータの信頼性を考慮して情報を配送する。

本研究では、このモデルの有効性を検証するために、CMFS モデルを文字放送と NetNews を情報源として実装し、評価を行なった。

## 2 連続メディア型情報

連続メディア型情報とは、提供者が情報を次々と提供し、その情報はある期間で失われてしまうような情報である。例えば、テレビ放送やラジオ放送や文字放送のように時間的に連続して提供される情報や、NetNews のように提供される量が膨

大なために、疑似的に連続メディア型情報と呼べる情報などがある。

連続メディア型情報の特徴として、以下のことがあげられる。

- i) 提供されている時にその情報を獲得しなければ失われてしまう。
- ii) 情報提供者から一方的に送られてくる。
- iii) 後から参照することが困難。
- iv) 時間経過ともなって変化していく情報がある。

連続メディア型情報を扱う上での問題点として、以下のことがあげられる。

- i) ユーザの必要とする情報を得るためには必要な情報が流れてくるのを待ち続けなければならない。
- ii) ユーザに配送された情報を管理・加工することが考えられていない。

現在、連続メディア型情報に関しては、(株) インプレスの Internet TELETEX などの連続メディア型情報をインターネット上に流す試みや NetNews などの特定の情報源に対するフィルタリングサービスなどが行なわれている。

## 3 CMFS モデル

本研究で提案する CMFS モデルとは以下の機能を提供するサービスのモデルである。

- 様々な種類の連続メディア型情報を統一フォーマットに変換し、効率的に扱う。
- スケーラビリティと信頼性を考慮した情報の配送。
- フィルタリングによる必要な情報のみの抽出。
- 抽出した情報のユーザの視点での整理・利用。

CMFS モデルは、CMFS 情報源・CMFS サーバ・CMFS クライアント・CMFS データベースから構成される。(図 1)

CMFS モデルによって、大量に流れてくる情報から必要な情報を取り出して、有効に活用できるようになる。

以下で、それぞれについて述べる。

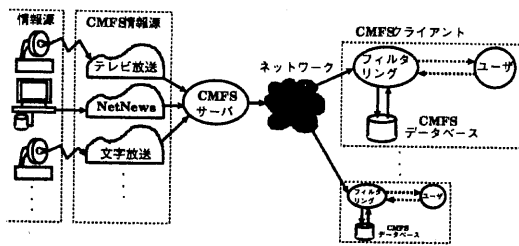


図 1: CMFS モデル図

### 3.1 CMFS 情報源

様々な種類の連続メディア型情報を発信する情報源の集合である。ここで、各種データの取り扱いを簡単にするために統一フォーマットに変換する。そして、いわゆるストリームとして届けられる情報を独立した一つの話題毎に分割する。

この分割された情報を CMFS モデルでは article と呼び、article 単位で情報を扱う。

連続メディア型情報の中には文字放送のように同じ内容の情報が繰り返し送られてくるものがある。CMFS モデルでは、CMFS 情報源で同じ内容の情報が繰り返されることを除去する。

情報源から情報は、同じ内容のものが除去された、article の列のストリームとして CMFS サーバに送られる。

### 3.2 CMFS サーバ

複数の CMFS 情報源から送られてくる複数のストリームを一つのストリームにまとめて、CMFS クライアントに配送する。情報の配送方式としては、システムのスケーラビリティを考慮して 1 対多通信 (マルチキャスト) を行なう。1 対多通信では、通信路で情報の一部が失われることがある。通信の信頼性を確保するために CMFS サーバ内でしばらくの間 article を保持しておき、再送要求を受け付けることで、通信の信頼性を確保している。

### 3.3 CMFS クライアント

CMFS サーバから送られてきたストリームを article に分割して、article 単位でユーザの必要と

する情報のみをフィルタリングする。そして、フィルタリングされた article を CMFS データベースに登録し、CMFS データベースを制御する。

CMFS クライアントは、ユーザからの指示で、フィルタリングを行う部分とフィルタリングされた情報をユーザに提示し、ユーザが article を管理するためのインタフェースの役割も持つ。

また、配送された情報に欠落が発見された場合には、CMFS サーバに再送要求を行ない、欠落した情報を入手する。

### 3.4 CMFS データベース

CMFS クライアントから送られた article を蓄積し、情報の管理を行なう。CMFS クライアントからの命令で article をユーザの観点で再分類する。

## 4 CMFS モデルの実装

CMFS モデルの有効性を検証するために、文字放送と NetNews を CMFS 情報源とするシステムを実装した。

### 4.1 文字放送

文字放送はデジタル連続メディア型情報であり、以下のような特徴がある。

文字放送はテレビ放送で未使用の 4 本の走査線を使って放送されている。1 本の走査線は 30 分の 1 秒づつ 1 フレームという単位で送出されている。1 フレームは、22 バイトで構成されているため、1 秒間の転送レートは 2640 バイト/秒となる。

文字放送は、複数のチャンネルで構成されている。一つのチャンネルは複数の番組で構成され、一つの番組は複数のページからなる。一ページのデータは、文字・画像・音などの情報が混在している。各情報は、圧縮され伝送効率を高めている。

文字放送は番組によって、随時更新・一日数回更新・一日毎更新というように、番組内容が更新される時間は異なる。内容が更新されるまで、同じ内容が繰り返し放送される。

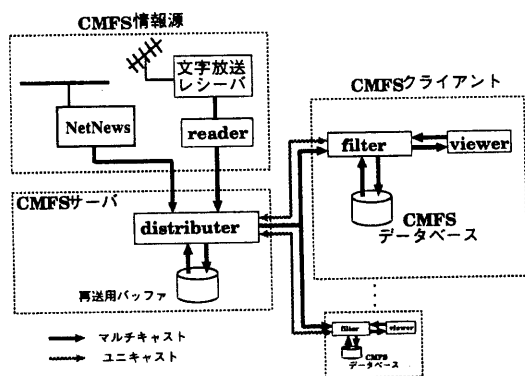


図 2: CMFS モジュール構成

## 4.2 NetNews

NetNews は、利用者が発信した記事を、世界中の NetNews サーバが協調して配送することにより、世界中に記事を伝えるシステムである。主に文字情報だけの記事が提供されているが、マルチメディア情報も扱える。

NetNews は、世界中から記事が配送されるため、膨大な量の記事が配送されている。そのため、NetNews は、疑似的な連続メディア型情報とみなせる。

## 4.3 実装

本研究では、インターネット環境において CMFS モデルの実装を行なった。実装した CMFS モデルのモジュール構成を図 2 に示す。以下、それぞれのモジュールについて解説する。

### 4.3.1 reader (CMFS 情報源)

文字放送では一つの独立した話題を article とした。これは、文字放送では一つの話題が複数のページにまたがるものや、1 ページに複数の話題が入っているものなどがあるため、ページを article とするのは不適切だからである。NetNews では、一つの記事を一つの article とする。

統一フォーマットとして、文字コードは iso-2022-jp を使い、画像フォーマットは、XPM を用いた。iso-2022-jp コードを用いた理由は、情報交換用の

符号として、標準的に使用されているためである。XPM を用いた理由は、全てテキストで表現されるフォーマットなので取り扱いが容易であるためである。

reader は、article を CMFS サーバに流すために一つのストリームにまとめる。本研究では、article を MIME[3][4] の Multipart の part とし、その MIME の Multipart をストリームとして distributor に流す。これは、MIME はマルチメディアメッセージ交換用フォーマットとして幅広く使用されており、様々なデータを構造化して送ることが可能であることから採用した。

### 4.3.2 distributor (CMFS サーバ)

複数の reader から流れてくる MIME のストリームを受け取り、article 毎に分割する。分割された article を一つの MIME ストリームに組み直してクライアントに IP マルチキャストを用いて配送する。その際に、最近の 100 article を再送用に保持し、通信の信頼性を高めている。

### 4.3.3 filter (CMFS クライアント・CMFS データベース)

本システムで実装したフィルタリング手法は、ユーザが指定するキーワードによるパターンマッチの and/or を用いた手法である。本手法を用いる際に article 中のテキストデータを日本語形態素解析システムの JUMAN[5] を用いて単語に分解することにより、不適切なマッチングを排除している。フィルタリングされた article は、キーワードに基づいてデータベースに蓄積する。ユーザの好みで、新しい分類を作成し、蓄積されている article を再分類することができる。

### 4.3.4 viewer (CMFS クライアント)

フィルタリングのためのキーワードの登録や蓄積されている article を見たり、新たな article の集合をつくり、article を再分類するための GUI を提供するプログラムである。図 3 には、キーワードの登録の状況、図 4 には、実際の article の表示を示す。

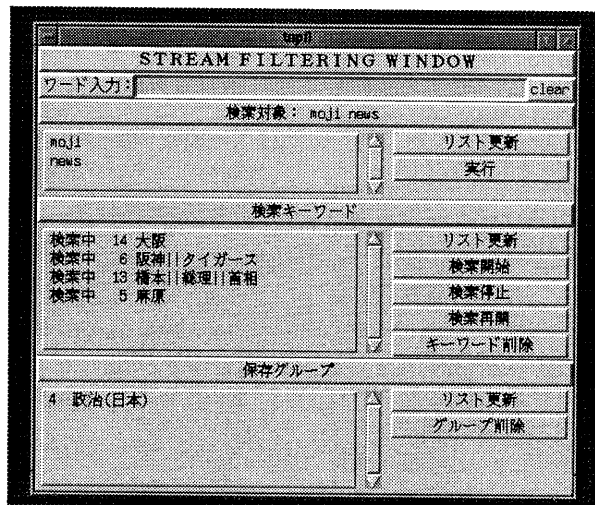


図 3: キーワードの登録

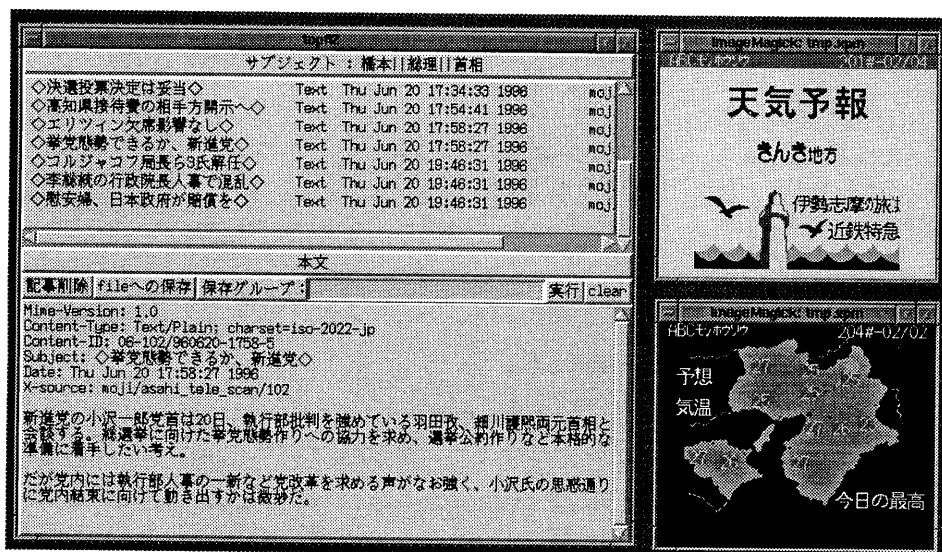


図 4: article の表示

## 5 評価および今後の課題

CMFS 情報源でデータを統一フォーマットにすることによって、CMFS サーバから提供されるデータの取り扱いが容易になり、一つの filter プログラムで種類の異なる複数の情報源からの情報をフィルタリングすることができた。

IP マルチキャスト通信を行なうことによりユーザの数が増えても通信路にかかる負荷が変わらず、スケーラビリティがある通信ができる。同時に、再送用のバッファをもうけることにより通信の信頼性も確保できた。

今回の実装では、キーワードによるパターンマッチの and/or によるフィルタリングを採用した。しかし、パターンマッチだけでは、本当にそのキーワードを含む article がユーザにとって必要な情報かどうかはわからないので、別の手法も組み合わせる必要がある。別の手法としては、同じキーワードにマッチした情報の比較などを用いて、不必要な情報の排除を行なうことを考えている。

今回の実装では、構造化されたデータのストリームを発信しているメディアを情報源とすることによって、MIME のストリームにして、各種のデータを統一的に扱えた。しかし、テレビなどの動画の情報を CMFS モデルで適用可能かどうか検討する必要がある。

## 6 まとめ

本研究では、大量に提供されている連続メディア型情報の中から必要とする情報だけを取り出す手法として、CMFS モデルを提案した。CMFS モデルでは、以下の機能を提供する。

- i) 様々な種類の連続メディア型情報の統一フォーマットへの変換。
- ii) フィルタリングによる目的の情報だけの入手。
- iii) スケーラビリティと信頼性を考慮した情報の配送。

また、提案したモデルの有効性を検証するために文字放送と NetNews を情報源とするシステムを実装した。実装したシステムを実際に運用することによって、提案したモデルの有効性を示した。

## 参考文献

- [1] 森田昌宏 篠田陽一. 情報洪水緩和のための情報フィルタリングの実現. *JAIN Symposium 1994*, 1994.
- [2] 中村友昭. 連続メディア型情報提供に対するフィルタリングサービスシステムの構築, *naist-is-mt9451082*, 1996.
- [3] N. Borenstein and N. Freed. MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies, *RFC1521*. September 1993.
- [4] K. Moore. MIME (Multipurpose Internet Mail Extensions) Part Two: Message Header Extensions for Non-ASCII Text, *RFC1522*. September 1993.
- [5] 松本裕治 黒橋禎夫 宇津呂武仁 妙木裕 長尾真. *Naist technical report, naist-is-tr94024*. Technical report, NAIST, 1994.