

新たなサービスの追加が容易な Web ページフィルタリングシステムの提案

服部 佳隆 北 英彦 石田 伸吾 沖廣 充功 加藤 友治 林 照峯
三重大学工学部

Web 上では Proxy サーバと呼ばれるサーバ、クライアント間の要求、応答を中継するシステムが使われている。Proxy サーバでは HTML 文書の中継しているので中継の途中でその HTML 文書を加工することができる。本研究では Proxy 技術を利用して Web ページを加工し、様々なサービスを提供するシステムを Web ページフィルタリングシステムと呼ぶこととする。

本論文では、Web ページフィルタリングシステムについての従来の研究の問題点である、新規サービスの追加が困難、ユーザごとに細かなサービスの提供が出来ないといったことを解決するための手法を提案する。また、読込時間表示、推薦情報提供などの新たなサービスの追加が容易な Web ページフィルタリングシステムの実装例、利用例を示す。

Web Proxy System with a function of Inserting Additional Information into Web Pages

Yoshitaka Hattori, Hidehiko Kita, Shingo Ishida,
Michinori Okihira, Yuji Kato and Terumine Hayashi
Faculty of Engineering, Mie University

Web proxy servers are used for security, access control and traffic reduction. It relays requests from clients to servers and responses from servers to clients. Therefore, web proxy server can process HTML documents during it relaying them from servers to clients.

We propose a web proxy system with a function of inserting additional and valuable information into relaying web pages. The features of this system are as follows: (1) it is easy to add a new information providing service into this system, and (2) personalized information can be provided by user authentication function.

1 はじめに

インターネットのユーザ数は近年爆発的に増加している。特に、World Wide Web の利用が増えている。企業や学校などでは、Web を利用する際に、ローカルネットワークとインターネット間のアクセスを中継するために Proxy サーバを設置することがある。Proxy サーバを設置する目的としては、外部

からローカルネットワークのセキュリティを確保すること、内部からインターネットへのアクセスを制限すること、Web ページのキャッシングによるトラフィックを軽減することがあげられる。また、Proxy サーバは上記の目的のためのキャッシュおよびアクセスログを持つ。

Proxy サーバはサーバ、クライアント間で要求、応答の中継を行っている。そのため、中継途中で Web ページを加工することが可能である。このような Proxy サーバを利用した Web ページ加工システムを本研究では、「Web ページフィルタリングシステム」と呼ぶことにする。「Web ページフィルタリングシステム」を利用することにより、さまざまなサービスの提供を行うことができる。ここで、柔軟なサービスを提供するためには、ユーザごとに利用するサービスの選択が可能でなければならない。また、サービス作成者が新たなサービスを容易にシステムに追加可能であることが求められる。これらの機能を持つ、Web ページフィルタリングシステムを提案する。

2 Web ページフィルタリングシステム

Web ページフィルタリングシステムとは従来の Proxy の機能に加え、Web ページを中継する途中で加工し、さまざまなサービスを提供するシステムである。本研究での「フィルタリング」は、余分なものを取り除くだけでなく、サービスを提供するためのさまざまな情報を付加することを含むことにする。

以下に Web ページフィルタリングシステムの処理の流れを示す[図1]。このような処理を行うことにより、提供するサービスの目的に応じて中継途中の Web ページを加工することが可能となる。

3 Web ページフィルタリングシステムに関する従来の研究と問題点

Web ページフィルタリングシステムに関する研究として著者らは「Proxy を利用した新たなサービスの可能性について」[1]の研究を行ってきた。このシステムは Web ページの読み込み時間を Proxy サーバのキャッシュとログを元として計算し、Web ページ内のリンクの後ろに表示するシステムである。

このシステムの問題点として、以下の点があげられる。

- サービスの提供部分がプログラムに組み込まれているため、新規サービスの追加が困難
- 提供されるサービスが固定
- 動作が遅い

また、不要な情報の削除を目的とする個人向け Web ページフィルタリングシステムとして、オープンソースプロジェクト Muffin[2]が進められている。Muffin は与えられたサービスの中から自由に選択して利用することができる。

このシステムの問題点として、以下の点があげられる。

- 新規サービスの追加には Java で書かれたプログラム内部に手を加える必要がある
- ユーザが直接利用するパソコン上で動作するシステムであるので、グループで情報を共有できない

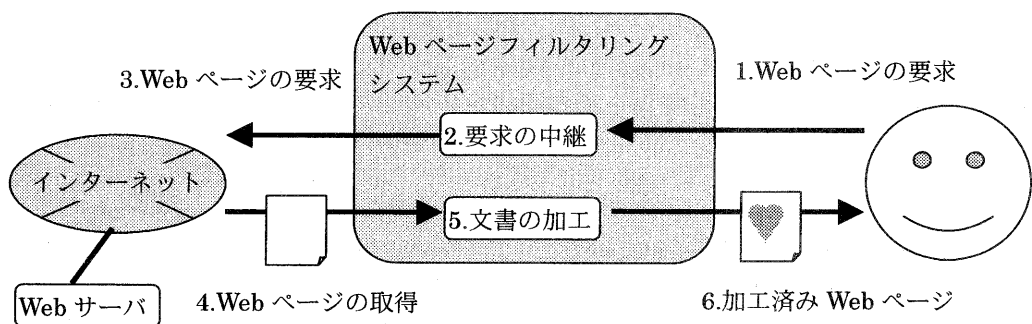


図1 Web ページフィルタリングシステム

本研究では、これら従来の研究の問題点を解決し、特に新規サービスの追加を簡単にするシステムを提案する。

4 Web ページフィルタリングシステムを利用することにより提供可能なサービス

Web ページフィルタリングシステムを利用することによりさまざまなサービスが提供可能である。

提供可能なサービスは次の二種類に分けることができる。

●個人向けサービス

Web ページを閲覧する上でユーザの助けとなるサービス。

- Web ページの読み込み時間表示
- Web ページの更新チェック

●グループ向けサービス

企業の部課や大学の研究室などのグループ単位で提供するサービス。

- Web ページ推薦
- スケジュール、お知らせ情報の表示

どちらの種類のサービスを実現する上でもユーザ管理とユーザごとに利用するサービスの選択が可能であることが求められる。また、さまざまなサービスを実装し検証するためには容易にサービス提供のためのプログラムを作成可能なことが求められる。

5 本研究で提案するシステム

5.1 従来の研究の問題点を解決する手法

Web ページフィルタリングシステムの利用者は、提供されるサービスを利用する「ユーザ」、具体的なサービスを提供する「サービス提供者」に分けることができる。

ユーザ側から見た従来の研究の問題点として、

- サービスの組み合わせが選べない
- 自分専用のカスタマイズできない

があげられる。

また、サービス提供者から見た問題点として、

- 新規サービスの追加が困難
- 複数のユーザを扱えない

があげられる。

これらの問題点を解決するために本システムでは以下の手法を用いた。

(1)Proxy 認証を利用したユーザ管理

Proxy 認証を用いることでサービスを提供するユーザの特定が可能である。また、正規のユーザ以外のアクセスを制限することが可能となる。

(2)Web インタフェースによるサービス選択

Web インタフェースを利用することにより、ユーザは自分の利用したいサービスをWebブラウザ上で自由に組み合わせて選択することが可能である。また、設定ファイルを直接編集する必要がなくなる。

(3)CGI ライクのフィルタプログラムインタフェース

フィルタプログラム作成のためのインタフェースをWeb上で利用されるCGIに似た形式とした。CGIとは、掲示板などの動的なWebページを提供するためにWebサーバから呼び出されて動作するプログラムである。

このことにより、CGIプログラム作成経験者であれば、サービスを提供するフィルタプログラム作成のための学習量を減らすことが可能であると考えられる。

これらを実現するために、システムをProxy機能とユーザ管理機能を持つProxyエージェントおよび、具体的なサービスを提供する複数のフィルタリングエージェントおよび、フィルタリングエージェント下で動作するフィルタプログラムに分離した。それぞれの詳細を以下に示す。

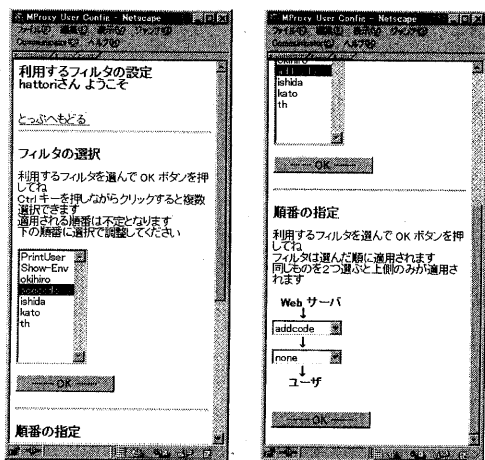
5.2 Proxy エージェント

Proxy エージェントはHTTP要求、応答の中継といったProxyサーバとしての機能を提供する。

加えて、ユーザ認証、ユーザプロファイルの管理、ユーザごとのフィルタリングエージェントとの接続の管理を行い、フィルタリング対象のデータ

(HTML 文書)のやり取りを行う。

また、簡易 Web サーバ機能を実装することにより、Web インタフェースを持つ。これを利用して、各ユーザは利用するサービスを自由に選択することが可能である。ユーザ追加、削除といったシステム管理にも利用することが出来る[画面1]。



画面1 Web インタフェース

5.3 フィルタリングエージェント

フィルタリングエージェントは Proxy エージェントとの接続部およびフィルタプログラムのためのインタフェースにより構成される[図2]。

Proxy エージェントから HTTP 要求、応答ヘッダと対象の HTML ファイルを受け取る。HTTP 要求、応答ヘッダのうち必要なものを環境変数に格納し、フィルタプログラムを子プロセスとして起動する。その後、フィルタプログラムに対し、標準入出力を通じて対象の HTML ファイルの受け渡しを行う。また、フィルタプログラムが何らかの原因により暴走した場合の復旧も行う。

フィルタリングエージェントを用いることでサービス提供者は HTTP の詳細に触れることなくサービス提供のためのフィルタプログラムを容易に作成することが可能となる。

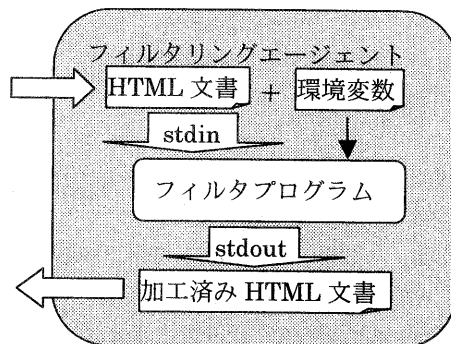


図2 フィルタリングエージェント

5.4 フィルタプログラム

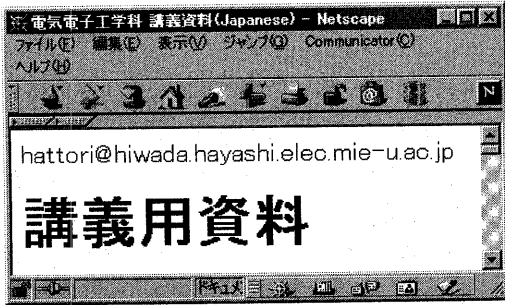
フィルタプログラムはサービス提供のために必要な Web ページの加工、データベース操作などを行う。

フィルタプログラムはフィルタリングエージェントの子プロセスとして実行され、標準入出力を通じて対象の HTML ファイルの読み込み、書き出しを行う。また、環境変数に HTTP 要求、応答ヘッダのうちの重要なもの、現在接続しているユーザ名などが格納されている。これらを利用して提供するサービスの内容をユーザごと、対象 URL ごとに細かく変更することもできる。

これらにより、サービス提供者は標準入出力と環境変数の利用方法といった基礎的なプログラミング能力、そして HTML の知識があれば、好みのコンピュータ言語を利用して、容易にフィルタを記述することが可能となる。

フィルタプログラムの例(ユーザ情報表示)

```
#!/usr/bin/perl
$user=$ENV{'REMOTE_USER'};
$host=$ENV{'REMOTE_HOST'};
while(<STDIN>) {
    s/(<body.*?>)/$1$user%@$host/is;
    print;
}
```



画面2 表示結果

5.5 システムの動作

上記の Proxy エージェント、フィルタリングエージェントおよび、フィルタプログラムを組み合わせるとどのように動作し、ユーザごとに異なったサービスを提供するか説明する。

本システムは、Proxy エージェントに対し、複数のフィルタリングエージェントがぶら下がっている構成となっている[図3]。前述の通り、各フィルタリングエージェントはサービス提供のためのフィルタプログラムを持つ。

Proxy エージェントはユーザごとのユーザプロフィールに基づき、ユーザの利用するサービスを

提供するフィルタリングエージェントに対してフィルタリング対象の HTML ファイルを送り、フィルタリングされた結果を受け取る。

ユーザが複数のサービスの利用を希望した場合は、フィルタリングされた結果の HTML ファイルを次のサービスを提供するフィルタリングエージェントに対して送り、結果を受け取ることを繰り返し、すべてのサービスを適用した HTML ファイルをユーザに送る。

6 本システムの適用例

本システムの適用例として以下のサービスを作成中である。

6.1 リンク先 Web ページの読み込み時間

表示サービス

本システムの Proxy エージェントのログを解析することにより、リンク先の読み込み時間を予測し、そのリンクの後ろに読み込み時間に応じたアイコンを表示する。また、リンク先が何らかの理由でアクセス不可能となっているときにはリンク切れアイコンを表示する。

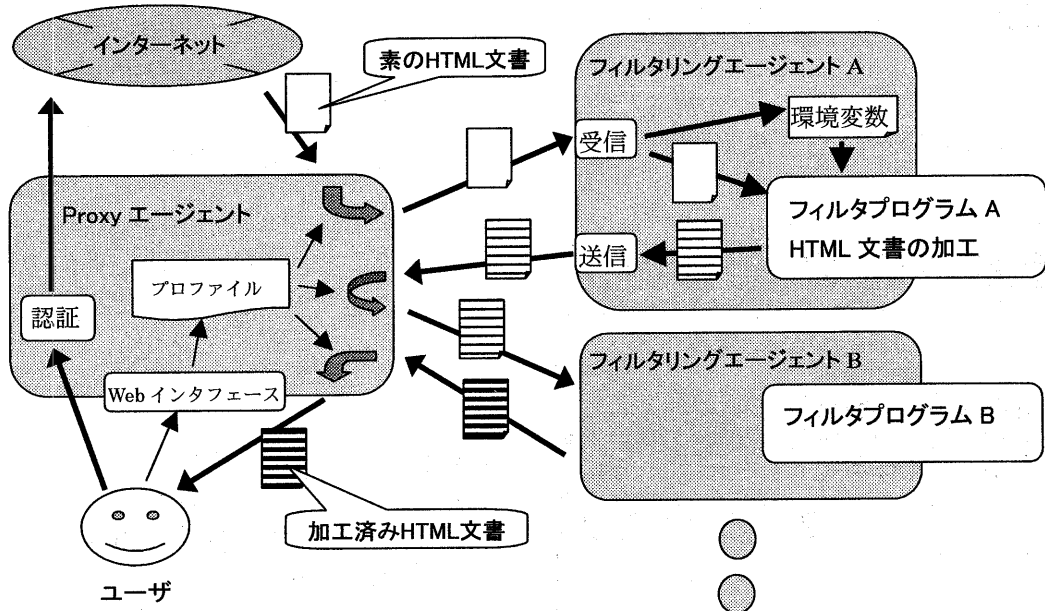
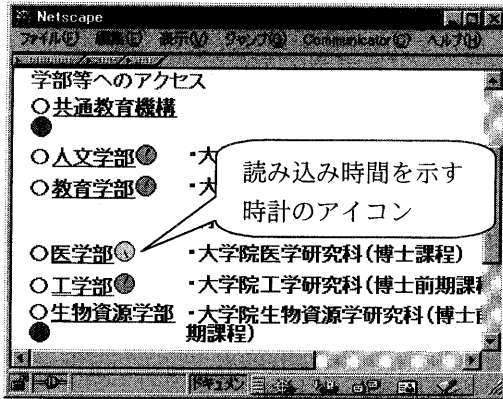
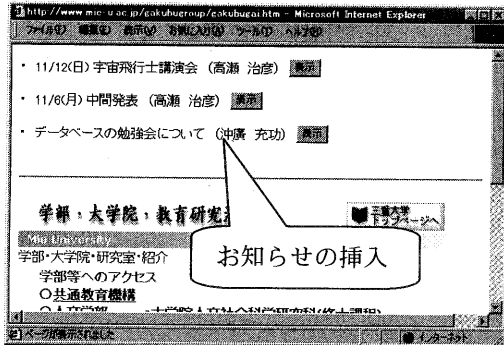


図3 システムの動作



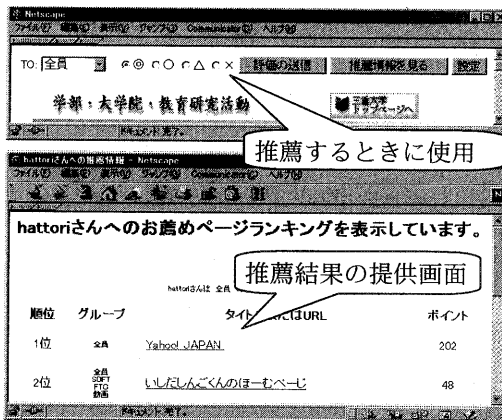
6.2 スケジュール、お知らせ情報表示サービス

Web ブラウザの上部にスケジュールとお知らせ情報を表示するサービス。



6.3 Web ページ推薦システムの評価入力

現在閲覧中の Web ページに対する評価をユーザにしてもらい、その結果を専用の Web ページで提供するサービス。



7 考察

本システムでは Web ページにサービスのための情報を追加している。このため、Web ページを加工することによる著作権の問題がある。本システムでは Web ページにサービスのための情報を追加しており、ユーザが自分の意思で最終的なビューを変更していることとなるので問題ないと考えられる。また、すべてのフィルタを容易に解除可能であるため、すぐにオリジナルのページを見ることが出来る。

8 まとめ

本システムでは、新規サービスの追加を容易にすること、ユーザごとに異なったサービスを提供しサービスを自由に選択可能とすることを実現するために以下の手法を用いた。

- (1) Proxy 認証を利用したユーザ管理
- (2) Web インタフェースによるサービス選択
- (3) CGI ライクのフィルタプログラムインタフェース

また、これらを実現するために、システムを機能ごとに Proxy エージェント、フィルタリングエージェント、フィルタプログラムに分離した。

以上より、提供されているサービスが固定されている、新規サービスの追加が困難という従来のシステムの問題点を解決することが出来たと考えられる。

今後の課題として、アクセス制限等のセキュリティの強化、ユーザの細かなグループ化を実現したい。また、新規サービスの追加をさらに容易にするために、フィルタプログラム作成ウィザードを作成しさらに容易にサービスを追加可能としたい。

参考文献

- [1] 村瀬茂樹, 北英彦, 林照峯: "Proxy サーバを用いた新たなサービスの可能性について", 情報処理学会研究報告 2000-GW-34, pp.19-24, (2000)
- [2] Muffin -- World Wide Web Filtering System
http://muffin.doit.org/