グループウェア 29−7
（1998．9．17）

読み込み時間を用いたWWWページのフィルタリング

村瀬 茂樹　北 英彦　林 照峯

三重大学工学部

インターネットの普及に伴い，WWW（World Wide Web）ページの数が急激に増加している．WWW ページはページごとに読み込み時間が異なり，中にはページが表示されるまでに非常に待たされるページも存在する．どうしても見たいページでなければ，このようになかなか表示されない WWW ページを訪問することを避けたい．

そこで，本研究では WWW ページの予想される読み込み時間をあらかじめ表示することで，ユーザが読み込みの遅い WWW ページを無駄に訪問することを避けさせるシステムを提案する．

The filtering system of WWW-pages by using time to finish downloading

Shigeki Murase　Hidehiko Kita　Terumine Hayashi

Faculty of Engineering, Mie University

The number of WWW(World Wide Web)-pages increase rapidly with the spread of the Internet. There are pages that take a long time to finish downloading. We want to avoid to visit such a page, unless we want to visit that page no matter what.

We suggest a system that let user avoid to visit WWW-pages which downloading times are too long by showing WWW-pages's forecasted downloading times.

1　はじめに

1960 年代後半に米国で誕生したインターネットはその後全世界に広がり，特にここ 2･3 年，WWW への関心は爆発的に高まっており，そのユーザ数を急激に伸ばしている．WWW ページには読み込みが早いページもあれば，遅いページも存在する．どうしても見たいページなら読み込みが遅くても我慢できるが，WWW ページが表示されるまでに時間がかかった上に，それが有用でなかったときにはユーザは苦痛を感じる．本研究は，プロキシサーバ（中継サーバ）のアクセスログファイル等を解析することで，WWW ページの予想される読み込み時間を求め，それをあらかじめ表示することによって，ユーザが読み込みの遅い WWW ページを無駄に訪問することを避けさせようとするものである．

WWW ページの読み込み時間があらかじめわかっていれば，調べたいものがあって探索を行っているユーザなら，読み込みの早いページから優先的に訪問でき，また，ネットサーフィンをしているユーザなら，遅いページを訪問するのを避けることができる．

本研究室には WWW ページへのアクセス回数等をもとに，人気 WWW ページを推薦するシステム「やじうまくんII」がある[1] [2] [3]．今回は，このシステムに各 WWW ページの予想される読み込み時間の表示を加えて，その有効性を評価することにした．

2　経済的情報フィルタリング

情報フィルタリング技術を大まかに分類すると，認知的情報フィルタリング(Cognitive filtering)，協調的情報フィルタリング(Collaborative filtering)，経済的情報フィルタリング(Economic filtering)，社会的情報フィルタリング(Social filtering)の 4 つに分類できる[4]．

ここで，経済的情報フィルタリングとは，情報を得ることによる利益と，情報を得るためのコストの比に基づくフィルタリングのことである．ここでいうコストとは，たとえば情報に対する課金や，情報を得るのにかかる時間，その他情報を見るときの疲れ等の心理的な要因なども含まれる．

WWW ページの読み込み時間を基に，フィルタリングを行うことは，経済的情報フィルタリングの一種と考えることが出来る．

3　WWW ページの読み込み時間の特性

WWW 上のページはサーバのディスク上にあり，そこからデータをダウンロードする（読み込む）ことによって表示される．WWW ページの読み込み速度は，主にそのサーバとクライアントとの間の回線の容量とその使用率に依存し，回線の使用率は主に 24 時間の周期で時間帯によって変化する．すなわち，WWW ページの読み込み速度は時間帯によって変化する．

また，WWW ページは基本的に HTML(Hyper Text Markup Language)ファイルと複数の画像，音声ファイル等で構成されているので，その読み込み時間は構成ファイルの容量にも依存する．すなわち，同じサーバ内にある WWW ページであっても，その構成ファイルの容量によって読み込み時間は異なるのである．

そこで，この WWW ページの読み込み時間の時間帯による変化を調査するために，本研究室からアクセスが多かった WWW ページを適当に選択して，1 時間ごとに 24 時間にわたってその読み込み時間を測定した(図 1)．その結果，WWW ページによって傾向は違うが，大学側が混んでいる昼間（11 時～19 時），及び NTT のサービスである「テレホーダイ」の時間帯でインターネット自体が混んでいる深夜（23 時

～2時）には，読み込みに時間がかかることがわかった．

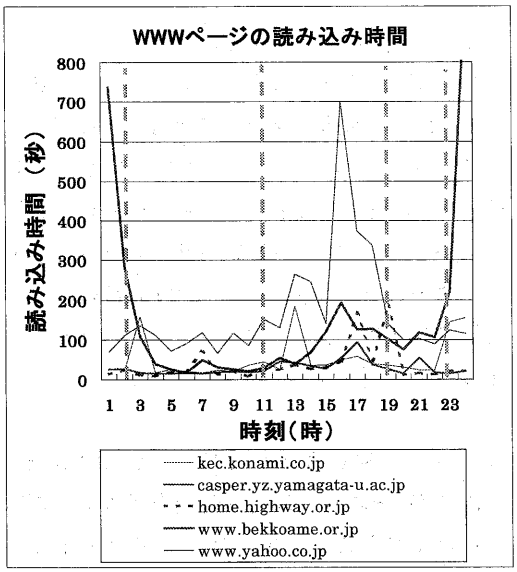


図1　各WWWページの読み込み時間の推移

4　WWW ページの読み込み時間を求めるシステム

4．1　用いる手段

proxy サーバである DeleGate のログ及び，キャッシュを解析することによって，過去に WWW ページを見に行ったときにかかった読み込み時間を時間帯ごとに求め，「やじうまくんⅡ」上でそれをユーザに提供する．

4．2　システム構成

図2に本システムのシステム構成図を示す．インターネットには莫大な数の WWW サーバがあり，それぞれの WWW サーバのディスクには複数の WWW ページが保存されている．ユーザはインターネットを介して WWW サーバにアクセスすることにより，WWW ページを見ることができる．

本システムはローカルエリアネットワークとインターネット間にプロキシサーバを設置することで得られるアクセスログファイルとキャッシュを利用を前提としている．ログファ

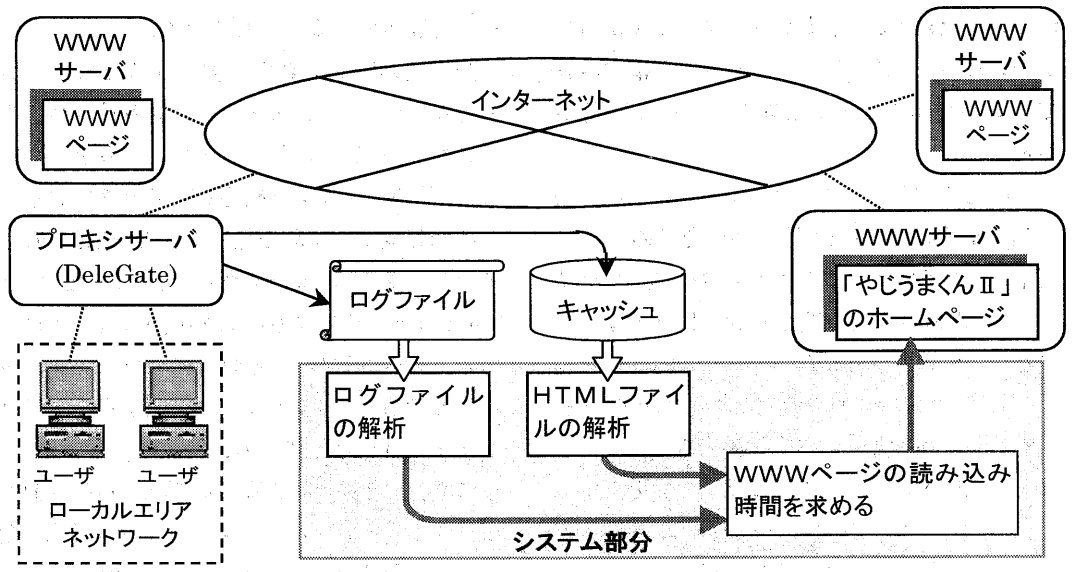


図2. システム構成図

イルとキャッシュ中の HTML ファイルの双方を解析することによって，WWW ページの読み込み時間を求めている．求めた時間は「やじうまくんII」の WWW ページ上で，各時間帯ごとに読み込み時間の長さを表す4種類のアイコンを用いて表示している．

4．3　WWW ページの読み込み時間の求め方について

本システムでは，プロキシサーバのログファイルとキャッシュ中の HTML ファイルの双方を解析することで，WWW ページの読み込み時間を求めている．これは，以下に述べる理由により，ログファイルのみの解析では読み込み時間を求めることができないからである．

WWW ページは主に HTML ファイルと複数の画像ファイル等から構成されている．ログファイルを解析することで，WWW ページの各構成ファイルの読み込み時間を求めることは可能だが，構成ファイルの読み込み時間の合計がその WWW ページの読み込み時間とはならない．これは複数のファイルが平行して読み込まれているためである．またログファイルからでは，WWW ページがどのファイルから構成されているかを知ることが出来ないという問題がある．

そのため本システムでは，以下のような方法で WWW ページの読み込み時間を求めた．

（1）　ログファイルの中から HTML ファイルの読み込み開始時刻を抜き出す．

（2）　キャッシュに残っているその HTML ファイルを解析して，その WWW ページの構成ファイルを調べる．

（3）　再びログファイルに戻って，その構成ファイルの中で一番最後に読み込みが完了したファイルを探す．

（4）　その読み込みが完了した時刻と HTML ファイルの読み込みを開始した時刻の差をとって，WWW ページの読み込み時間を求める．

4．4　時間帯の分け方について

先に延べた通り，同じ WWW ページであっても，その読み込み時間は時間帯によって変化する．従って本システムでは 1 日を 2 時～11 時，11 時～19 時(大学側が混んでいる時間帯)，19 時～23 時，23 時～2 時(「テレホーダイ」でインターネットが混んでいる時間帯)の 4 つの時間帯に分けて，それぞれの WWW ページの読み込み時間を求めることにした．

4．5　「やじうまくんII」への組み込み

図 3 が，求めた読み込み時間を組み込んだ新しい「やじうまくんII」の WWW ページである．図 3 は 11 時～19 時の読み込み時間を表示したページであり，同様に他の時間帯の読み込み時間を表示したページも存在する．「やじうまくんII」の WWW ページを見に来た時刻によって，自動でその時刻に対応した時間帯のページを表示するようにしている．

読み込み時間を数字で表示するとわかりづらくなるので，本システムでは読み込み時間の早さを 5 段階の時計のアイコンで表示している．時計のアイコンは動画 GIF でできており，針の回転速度と時計の色によって，速さを表している．また Windows 版の Internet Explorer Ver.3 以上，及び Netscape Navigator Ver.4 以上では，図 3 の時計のアイコンにマウスカーソルをあわせると，図 4 のように説明がポップアップするようになっている．

また，図 3 の「他の時間帯の読み込み時間も表示する」というボタンをクリックすると，全時間帯の読み込み時間を表示した WWW ページを表示するようにしてある(図 5)．

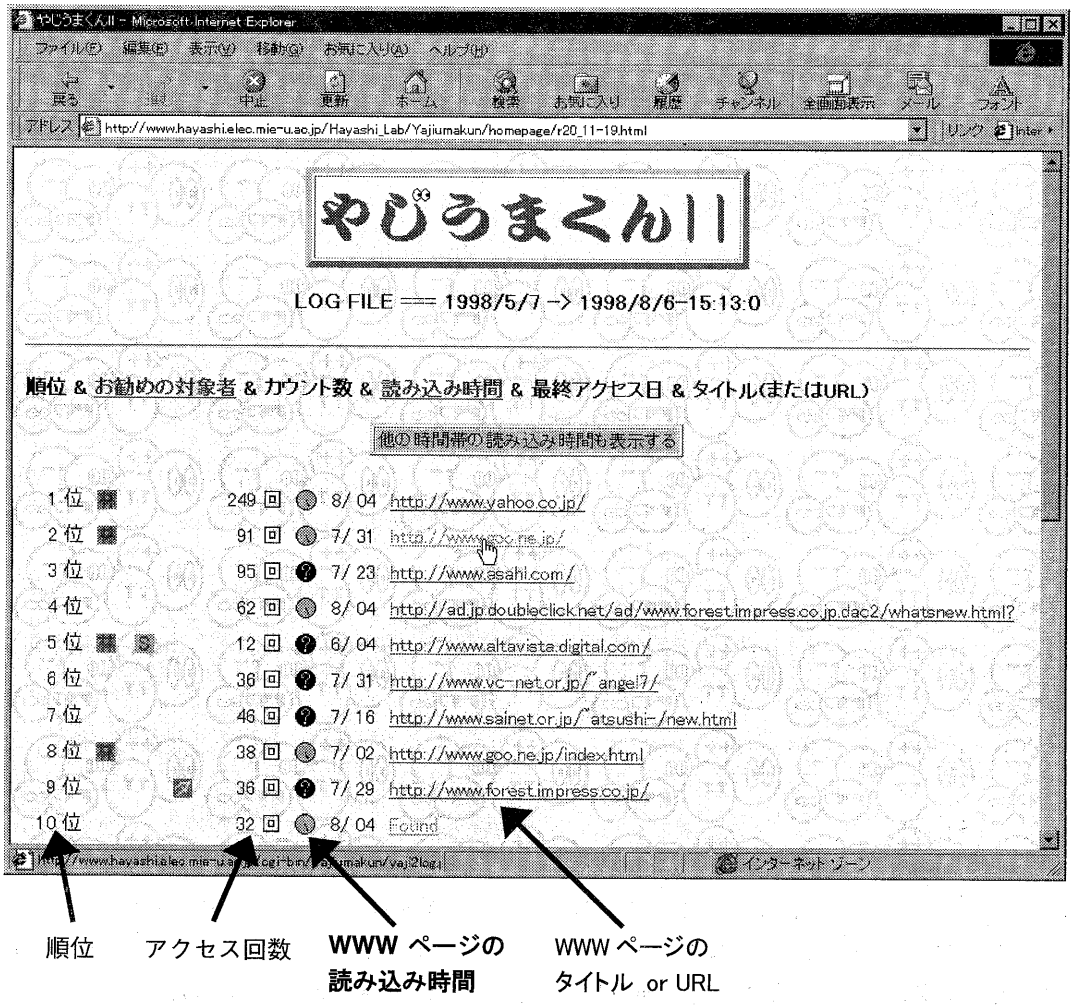
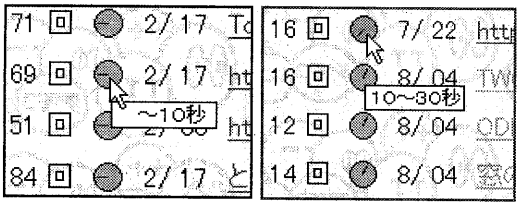


図3　読み込み時間の表示を加えた「やじうまくんII」のホームページ



(a) 青色のアイコン　(b) 緑色のアイコン

図4　ポップアップするアイコンの説明

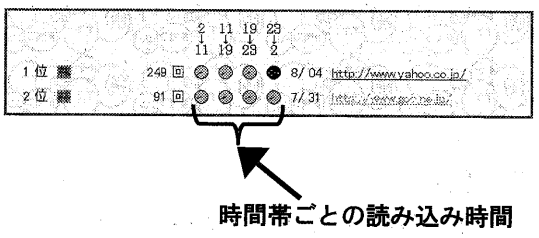


図5　全時間帯の読み込み時間を表示した「やじうまくんII」のWWWページ

5　試行実験

5．1　実験方法

本システムを評価するために，1997年11月～1998年3月の約4ヶ月間の試行実験を行った．対象は主に本研究室の学生である．

表1　システム導入前後のアクセス傾向の変化

	システム導入前 (11/4～1/16)		システム導入後 (1/17～3/30)	
読み込み時間	アクセス回数	割合	アクセス回数	割合
～10秒	228	82.0%	195	92.0%
10～30秒	20	7.2%	11	5.2%
30～60秒	27	9.7%	3	1.4%
60秒～	3	1.1%	3	1.4%
計	278	100%	212	100%

5．2　実験結果

本システムを「やじうまくんII」に導入する前と，導入した後での「やじうまくんII」からのアクセスの傾向について調べた結果が表1である．

本システムを導入したことによって，ユーザは読み込み時間が長くかかるWWWページを避け，読み込み時間が短いWWWページを多く見るようになったことがわかる．

6．今後の課題

最近ではキャッシュを取ることが出来ないWWWページ（サーバの設定により可能）が増加してきている．そのため，プロキシサーバのアクセスログとキャッシュの双方を使用してWWWページの読み込み時間を求める方式では，キャッシュが残らないWWWページの読み込み時間を求めることが出来ない．今後はこれらのWWWページの読み込み時間も求められるようにする必要がある．

7．終わりに

本研究では，有用なWWWページを探すときに，読み込み時間が長く，かつ，それが有用でなかったときには苦痛を感じるということから，WWWページの予想される読み込み時間を表示することを提案した．

そしてこのシステムの有効性を確かめるために「やじうまくんII」に組み込んでみた結果，読み込み時間が短いWWWページを見る比率は増加し，反対に読み込み時間が長いWWWページを見る比率は減少した．すなわち，本システムを導入したことによって，ユーザは読み込みの早いWWWページを多く訪問するするようになった．このことより，WWWページの読み込み時間は，ページを訪問するかしないかの1つの判断基準になっていることがわかる．

参考文献

[1] 杉井俊彦，北英彦，林照峯："アクセス回数を利用したWWWの人気ホームページ道案内システム"，グループウェア研究会（1997年），p.235-240

[2] 杉井俊彦，北英彦，林照峯："WWWのお勧めホームページについての情報共有システム"，グループウェア研究会（1998年），p.103-108

[3] 杉井俊彦，北英彦，林照峯："人気・お勧めホームページの情報共有システムに関する研究"，三重大学大学院修士論文 1998年

[4] Thomas W. Malone, Kenneth R. Grant, Franklyn A. Turbak, Stephen A. Brobst and Michael D. Cohen: Intelligent Information-Sharing Systems. Communications of the ACM (1987) Vol. 30 No. 5, pp. 390-402