

自動辞書引き機能つきWWWブラウザ\*

遠藤進

(株)富士通研究所 情報社会科学研究所

E-mail: susumu@iias.flab.fujitsu.co.jp

5S-8

1 はじめに

WWWは、Internet上でサーバー上に構築された情報を、Mosaicなどのブラウザでハイパーテキスト的に閲覧するハイパーメディアであり、個人ベースでも簡単に情報発信ができるようになっている。

しかし、現在のWWWにおいては、基本的に処理はサーバー側で完結しており、ユーザ側の事情に応じた処理を実行する事は難しい。サーバー側にクライアント側の情報を送って処理させる事はできるが、情報を送るのに手間がかかる上、他人に知られたくない情報の場合はセキュリティ上問題がある。

本研究では、上記の問題を解決するために、ブラウザが取得した情報を、クライアント側でユーザの状況に応じて適切に編集し表示する仕組みの導入によって、ユーザの関心に密着した情報の提供を可能にする。応用例として、自動辞書引きツールと折り込み広告システムの二つのシステムを作成した。

2 Client-Side Stream Filter

2.1 概要

現在のWWWにおいては、サーバーが提供した情報をそのままブラウザが表示している。基本的に、提供された情報はサーバー側の処理等によって決められてしまい、クライアント側で変更する方法が無い。

よって、現在のWWWの枠組では、処理内容をクライアント側の情報にあわせて少しずつ変えるといったカスタマイズは難しい。ここでいうクライアント側の情報とは、例えば、ユーザの英語力や興味分野などの個人情報指す。

本研究では、クライアント側で個人情報を参照した処理を行うための仕組みとして、Client-Side Stream Filter(CSSF)を提案する。

ユーザがブラウザからアンカーをたどるなどして、次の情報にアクセスすると、ブラウザが情報を受け取る直前にCSSFが起動される。CSSFは、起動されるたび、設定しておいたプログラムを実行する。このプログラムは、標準入力から、ブラウザが取得したページの内容とそのURLを受け取り、結果を標準出力に出力する形式の物で、個人情報を参照しながらページの内容をカスタマイズする。CSSFはその出力をブラウザの入力として返し、ブラウザはこれを表示する。(図1)

CSSFで実行されるプログラムからは、ユーザのファイルへのアクセスや他のプログラムの実行が可能なので、個人情報の読み書きを行ったり、ウィンドウを開いたり

する事ができる。これによって、よりユーザに親しいサービスの提供が行える。

2.2 他の方法との比較

従来、CSSFのようにユーザの個人情報を扱う方法としては、入力フォームに入力させて、サーバー側で処理をするのが一般的だったが、この方法では、入力は手動で行わなければならないので手間がかかる上に、Internetを経由するためユーザが知られたくない情報を扱う事ができない。

また、CSSFと同じような仕組みで、クライアント側で処理を行う物として、DeleGate[1]が存在する。DeleGateは、漢字コードの変換や防火壁(firewall)を越えるための汎用の中継サーバーだが、主にサイト向けのサービスであり、クライアント側の個人情報を利用するサービス利用には向かない。

3 自動辞書引きツール

自動辞書引きツールとは、WWWで英語の文章を読んでいる時にわからない単語が出てきた場合、自動的にその意味を表示したり、詳細を表示するページへのリンクを作成したりするツールである。図2はdogにその意味を表示するページへのリンクが張られた場合を示している。

WWWで英語の文章を読んでいる場合、知らない単語の意味を調べたいと思う事がよくある。この場合、辞書ツールが独立しているシステムでは、単語を再入力しなければならない。

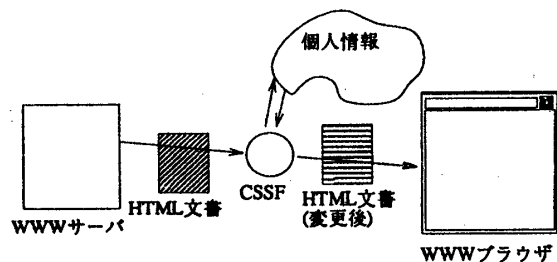


図1: CSSFの仕組み

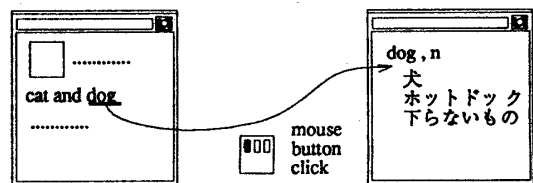


図2: 自動辞書引きツールの動作画面

\*The WWW Browser with automatic dictionary search; Endo Susumu; Institute for Social Information Science, FUJITSU LABORATORIES LTD.

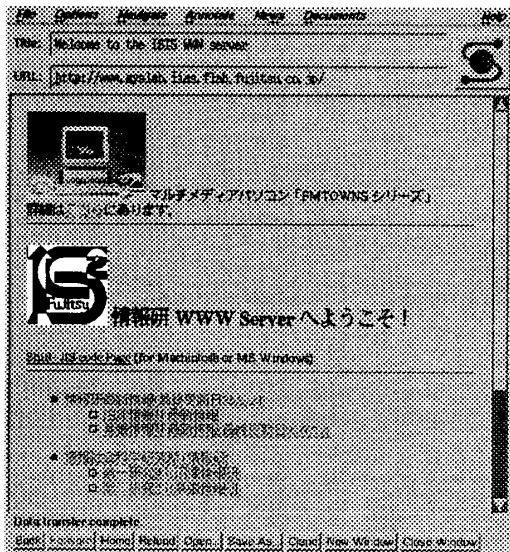


図 3: 折り込み広告システムの画面例

本ツールでは、文章は元の英文のまま表示し、難しいと思われる単語に、意味を表示するページへのリンクを張るか、あらかじめウインドウを開いて意味を表示するようにしている。

ある単語を知っているか知らないかはユーザごとに違うので、意味を表示する単語や意味へのリンクを張る単語は、ユーザに応じて、動的に変化する。システムは、単語ごとの難度を推測し、それにに応じて、リンクを張ったり、直接意味を表示したりする。

ここで、ユーザごとの単語の難度は、ユーザの行動からヒューリスティックに推測する。基本的な方針は以下の通りである。

- (1) 初出の単語は単語の形状等から難度を推測する。(長い単語は難しいなど)
- (2) 表示されたにもかかわらず、ユーザがその意味を調べない単語は、ユーザが既に知っている物として、難度を下げる。頻出する単語の場合も難度を下げる。
- (3) 一度調べた単語は、一時的に覚えるがすぐ忘れる事が多いので、しばらくの間、難度を上げ、元に戻す。
- (4) 難しい単語でなかなか出て来ない単語は、忘れやすいので、難度を上げる。
- (5) 難度が低くても不都合が無いような単語は、知っているものとみなし、扱わない。

CSSFとして組み込まれた自動辞書引きツールは、上記ヒューリスティックに従い、原文にリンクを追加したり、意味を表示したり、単語の難度を個人情報としてファイルに保存したりする。

本ツールと同様のシステムとして、PANSEE[2] や meshPLUS[3] がある。PANSEEは、DeleGateの仕組みを応用した全文翻訳システムだが、自動翻訳という性格上、誤訳が多く、正確な意味をとるには原文と翻訳文を何度も見比べる必要がある。また、meshPLUSは、ブラウザに辞書引き機能を組み込んだシステムであり、ユーザの知識に応じて、わからない単語をはじめから引いておくなどの処理はしない。

## 4 折り込み広告システム

折り込み広告システムとは、WWW上で元の情報とともに広告を表示するシステムである。ブラウザが情報を表示する際に、別のサーバー上に用意してある広告情報を並べて表示する。(図3)

従来の広告システムでは、基本的に無差別にユーザに広告を提供して来た。ユーザは基本的に見たい広告を選ぶ事はできず、提供者側が判断した方法で広告が提供されていた。しかしこれでは、ユーザの意識にマッチした広告を提供できるとは限らず、繰り返す数多くの宣伝をする必要があった。

WWW上においても、広告を表示する試みは行われているが、サーバー上に広告を置いた場合、ユーザはその記事を見に来てくれるとは限らないし、また、ページに広告を載せた場合、つねに誰に対しても同じ広告が表示され、ユーザ側の興味に対応した広告を表示する事ができない。

このような問題点を解消するために、CSSFを用いてブラウザ側に元ページと一緒に広告を表示する機能を持たせることを考案した。これにより、よりユーザの興味に合った効果的な広告を実現できる。本システムでは以下のような使い方を想定している。

このシステムには広告代理店が存在する。この広告代理店は各企業と連絡して各社の広告のページを管理する。ユーザは広告代理店と契約を結び、契約内容にもとづいて広告が表示される。契約の内容は自由であり、ユーザが興味を持つであろう範囲にしぼって、提供を受ける広告の種類等を選択する事ができる。また、広告と一緒に時事のニュースを流すなどのサービスを行う事も出来る。

ユーザがブラウザを使ったとき、CSSFによる広告表示プログラムが、広告代理店のサーバーと接続し、契約に基づいた広告を広告代理店のサーバーから受け取り、それを元ページと同時に表示する。その恩恵として、ユーザは広告の提供企業から、インターネットプロバイダの接続料などを一部負担してもらう事ができる。

広告代理店は、ユーザが欲する広告をうまく表示できれば、契約企業から広告料をより多く徴収できるので、ユーザの広告の参照履歴などを参考にしたり、その広告に対しての要・不要のボタンを表示したりするなどして、ユーザが望む情報を出来る限り正確に知ろうとし、それに成功する広告代理店がより利益を得る事が出来る。

## 5 まとめ

本研究では、ブラウザにフィルタリングの機能を追加し、よりユーザの意向を反映したサービスを提供するためのシステムとして、自動辞書引きツールと折り込み広告システムを作成した。

## 参考文献

- [1] DeleGate, <http://www.etl.go.jp:8080/etl/People/ysato@etl.go.jp/DeleGate/>
- [2] PENSEE for Internet, <http://www.oki.co.jp/OKI/JIS/PENSEE/index.html>
- [3] meshPLUS, <http://www.nec.co.jp/japanese/today/newsrel/9507/0503-01.html>