

# 外国人等のための仮名訳機能付き HTTP プロキシおよび POP3 プロキシの開発

石川 准  
静岡県立大学国際関係学部

## 1. 開発の目的

外国人、知的障害者、学習障害者、子供等、漢字が読めない人々のインターネット・アクセシビリティを保障するための各種プロキシサーバの構築とその普及を目的とする。

## 2. 仮名訳機能付き HTTP プロキシ

### 2.1 仮名訳機能付き HTTP プロキシの概要

仮名訳機能付き HTTP プロキシの仕組みの概要は次のとおりである。

A は任意の日本語 Web サイト、B は仮名訳機能付き HTTP プロキシ、C はユーザのクライアントとする。C が B を HTTP プロキシとして設定することで A にアクセスした際の表示が分かち書きの仮名になる。B を実現するために、自動点訳・仮名訳エンジン EXTRA (UNIX 版) の開発と HTTP プロキシのフィルタ機能を利用するための、HTTP 仮名訳プログラム `htmllexw` の開発を行った。

### 2.2 `htmllexw` について

`htmllexw` は EXTRA for UNIX などを使用して HTML ファイルを仮名に変換するプログラムである。HTTP の proxy、POP3 proxy などの用途が想定されている。`htmllexw` は標準入力から EUC の漢字コードの HTML ファイルを読み込む。次に、仮名訳に変換する対象を切り出し、EXTRA にて仮名訳を行う。翻訳後の内容は、元の場所に埋め戻され、最終的に仮名訳化された HTML ファイルを標準出力に書き出す。翻訳対象の文字列は、タグに挟まれた文字列と、`img` タグ内の `alt=""` で指定された文字列である。例外として、`<script>` と `</script>` で挟まれた箇所、あるいは、`<!--` と `-->` で挟まれた箇所は翻訳対象外としている。`htmllexw` を DeleGate のフィルタとして設定することで HTTP の仮名訳 proxy を実現することができる。

## 3. 仮名訳機能付き POP3 プロキシ

### 3.1 仮名訳機能付き POP3 プロキシの仕組み

A はプロバイダのユーザのメールボックス、B は仮名訳機能付き POP3 プロキシ、C はユーザのクライアントとする。

ユーザは、POP3 のサービスを使用するために、B の Web ページから A のアカウント情報を登録する。アカウントは B 内の PostgreSQL のデータベースに登録される。CGI の開発言語としては PHP を使用している。セキュリティの維持のために、この Web ページは HTTPS を使用する。また、不正なセッションを防ぐため、MD5 を利用したセッション ID と、セッションの有効期限の仕組みをインプリメントしている。アカウントが B に登録されると、一定時間 (デフォルトで 1 時間) 毎に、B のサーバは `fetchmail` を使って、A のメールボックスの内容を POP3 もしくは IMAP のプロトコルを使って吸い上げる。`fetchmail` で吸い上げたメールの内容は、`fetchmail` のメール転送エージェントで指定された、今回開発した `m1prx_trans` に渡される。

### 3.2 `m1prx_trans` の仕組み

`m1prx_trans` は次の手順でメールを仮名訳/カタカナ訳/ローマ字訳/点訳 (NABCC) する。(なお、`m1prx_trans` は Perl で記述)

MIME の解析、および文字コードの自動判定 各パートの翻訳、テキスト部分の翻訳には、点訳・仮名訳エンジン EXTRA を Perl から呼び出すライブラリを使用する。また、HTML のパートの翻訳には、`htmllexw` という HTTP プロキシのために開発したプログラムを使う。なお点訳の場合には、点字コードの一部が別の意味で解釈される可能性があるため、メールのヘッダと HTML のコンテンツは、翻訳対象外としている。 翻訳結果を MIME でパッケージ

翻訳済の内容を各ユーザの Maildir 形式のメールボックスに保管する。B のメールボックスに保管されたメールの内容は、C から POP3 のプロトコルを使って、読み出す。

## 4. 今後の課題

これらのサーバを実際に可動させ、システムの評価と改良を行う。