

# Web ページの部分表示による PDA への対話的 Web 適応

\* 山本 浩司

東京工業大学大学院 総合理工学研究科 知能システム科学専攻

〒 226-8502 横浜市緑区長津田町 4259

Tel/Fax : (045)924-5218

yamamoto@ntt.dis.titech.ac.jp

Abstract: PDA 等で Web を閲覧する際には, PC 上の Web ブラウザで閲覧する場合に比べてスクロールの量が增大し, ユーザは目標とする情報に到達するまでに労力を要する. 本稿では, ユーザが興味を持っていると考えられる部分を検出し, スクロール量が增大しないようにページの上部に表示するシステムを開発した.

## 1 はじめに

携帯電話や PHS, 携帯情報端末 (PDA) 等の普及が進み, これらのモバイル装置を用いて Web ページを閲覧することも一般的になってきている. しかし, 画面が小さく通信速度やプロセッサの処理速度において PC に劣るモバイルの装置においては, 目標とする情報を得るまでに多大な労力を要する. この労力を軽減するために, 本稿では Web ページ中でユーザが興味のある部分に対話的に学習し, PDA に表示するシステムを開発する. ユーザが過去に頻繁に見た情報に関連する情報にユーザが興味を持っていると考え, そういった情報に重点をおいて表示し, ページをユーザに適応させる.

## 2 Web ページの部分表示

本研究では, 主に新聞社のページやニュースサイトを対象とする. その理由は, 第一に, これらのページは何度もユーザが繰り返し見るようなことが多く, 繰り返し見て, 更新があったとしてもほとんど構造が変化しないため, 適応に適していると考えられるからである. また第二に, 記事の見出し部分や, いろいろなカテゴリについての各記事へのハイパーリンクや, サイト内の他のコーナーへのナビゲーションメニュー等, いろいろなパーツから構成されている場合が多い. これらに対し, ユーザがそれぞれに興味を持っているか否かの区別が比較的つけやすいと考えられるからである.

本研究では, Web ページをいくつかの単位に分割し, それぞれに対してユーザがどのくらい興味を持っているかを判別する. 興味を持っているかどうかは, ユーザがその部分を閲覧するかどうかで判定する.

## 3 システムの概要

### 3.1 システムの機能

本システムは, ユーザがリクエストした URL を入力とし, そのユーザに適したように加工された HTML ファイルを出力し, ユーザの PDA に表示させる. ユーザが興味を持っている情報とは, 過去に何度も見たような情報であると考えられる. Web ページ中でユーザがどのようなトピックに興味を持っているかを知るために, ユーザの閲覧行動から, 加工した Web ページをユーザに出力しつつ, トピックの重要度を更新する. つまり, ユーザが閲覧した部分のカテゴリの重要度を高めるということを行い, ユーザに適応する.

### 3.2 処理の流れ

本システムはユーザの PDA からのリクエストを WWW へ中継するプロキシサーバとして動作する. 処理の流れを以下に順を追って示す.

1. 一般的なプロキシサーバと同様に, ユーザがリクエストした URL のファイルを取得する.
2. HTML ファイルを獲得した後, それを解析し, 記事部分を抽出しブロックを生成する.
3. 解析が終了した後, ユーザに対して提示するための HTML ファイルの加工を行う.

### 3.3 ユーザへの適応

ユーザの閲覧行動を基にしたフィードバックを用いて, カテゴリの重要度の修正を行う. ユーザがリンクを辿ると, その位置を記録しておくとともに, 以下のようにすべてのカテゴリの重要度  $I$  を更新する.

$$I = \begin{cases} I + 1 & (\text{クリックされたカテゴリ}) \\ I * \lambda & (\text{それ以外}) \end{cases}$$

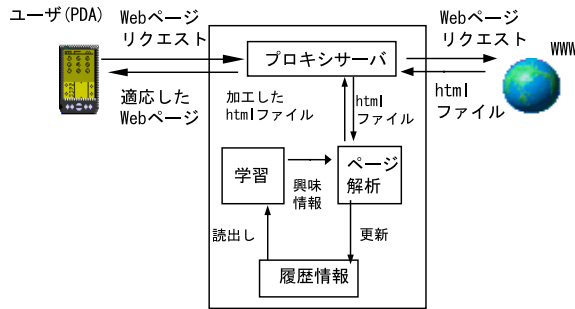


図 1: システム概要



図 2: 適応前の Web ページ



図 3: 適応後の Web ページ

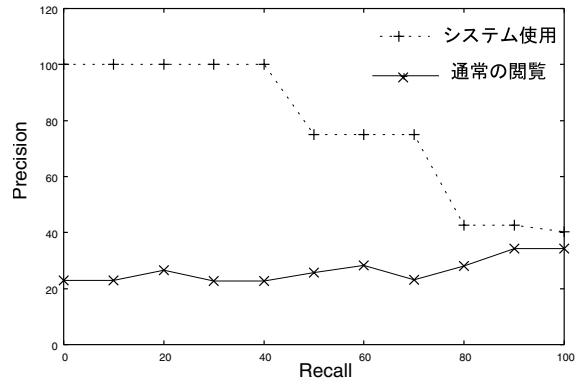


図 4: 実験結果

$\lambda$  は  $0 \leq \lambda \leq 1$  となる定数である．この更新を，ユーザがハイパーリンクをクリックし，新たな HTML ファイルをリクエストする度に行う．

## 4 実験

実験の対象として，読売新聞や朝日新聞などの大手新聞社 5 社のサイトを用いる．それぞれのサイトに対する評価は独立に行う．ユーザが必要な情報に到達するまでの効率の評価として，再現率と適合率の値を用いて評価する．Web ページ中で，上位  $N$  個に出現したブロック中で興味のあるブロック数を  $R_a$ ，ページ全体で興味のあるブロック数を  $R$  とし，適合率 (precision) と再現率 (recall) を以下のように定義した．

$$precision = \frac{R_a}{N} \quad recall = \frac{R_a}{R}$$

興味の有無については，それぞれのブロックについてユーザがあらかじめ決定しておくものとし，サイトのトップページの閲覧を 5 回繰り返し行った．また， $\lambda = 0.95$  を用いた．なお，図 2 はシステムを使わない通常の表示，図 3 はシステムを利用した動作例である．この例では，“最新ニュース”というカテゴリが重視され，上部に表示されている．

### 4.1 実験結果

実験結果を図 4 に示す．システムを利用している場合の線が，システムを未使用の場合のものに対して，recall が同等な場合には，precision が常に上回っていることが確認できる．

### 4.2 考察

一般的に，precision と recall のグラフは，傾向として右下がりの曲線になることが多い．しかし，本システムを使っていない場合の通常の表示の場合は，recall が低いときに precision が低く，両者ともそれほど高くない結果になっている．これは，通常の Web サイトは PC で閲覧することを前提にレイアウトが設計されており，そのレイアウトには table タグが使われている場合が多い．PDA 上では，画面の大きさの制約から，複数の列は表示できないため，テーブルの 1 列目を全て表示してからでない次の列を表示が行われない．そのため Web サイトの製作者の意図が反映されづらくなっているからであると考えられる．