

Web ページの部分表示による PDA への Web 適応

Web Adaptation by The Partial Display of a Web Page

山本 浩司*1 山田 誠二*2 新田 克己*1
Koji Yamamoto Seiji Yamada Katsumi Nitta

*1 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
CISS, IGSSE, Tokyo Institute of Technology

*2 国立情報学研究所
National Institute of Informatics

In this report, we describe a system which learns the part user gets interested in a Web page interactively, and arranges and displays that the amount of scrollings does not increase on the upper part of a page. to each user. The system judges and detects a category with a user's interest by browsing action which clicks the anchor of the hyperlink, and importance of the category is corrected by user's feedback. By experiment by our system, we found that our system enabled a user to get information which the user saw frequently in the past and suitable for the user's interest.

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, 現状の大部分の Web ページは主に PC 向けに作られており, 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を対話的に学習し, 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.

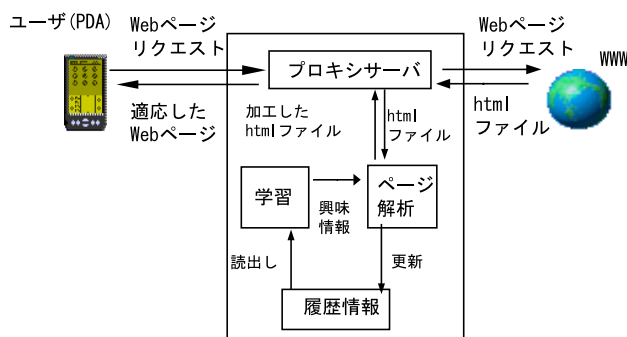


図 1: システム概要

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. 取得したページを HTML パーザにかけ, 木構造にする.
3. 木構造の根から順にノードを走査し, 文字列の長さの合計が閾値 θ を超えない範囲で最大のノードを, ニュースカテゴリを示す文字を含むノードとして検出する.
4. ニュースカテゴリの記述を検出したノードの弟のノードから, ニュースアンカーテキストに閾値 θ 以上の文字列を持つハイパーリンクが閾値 μ 以上含まれるノードを, ニュースの記事へのリンクとして検出する.
5. ニュースカテゴリの記述を検出したノードと, ニュースへのリンクを検出したノードを組み合わせると 1 つのブロックとする. また, カテゴリ名の単語をキーとするハッシュ



図 2: 適応前の Web ページ



図 3: 適応後の Web ページ

を生成する．単語をキーとし，値をそのカテゴリの単語の重要度とする．

6. 子に table タグを含まない table タグ (つまり一番内側の table) を検出し，1 つのブロックとする．
7. ブロックのニュースカテゴリの記述の重要度順にブロックをソートする．このとき，ソートされたブロックの N 番目は，ソート前の N 番目のブロックがあった位置に配置される．また，6. で作成されたブロックがあったノードの弟のノードを，木構造から削除する．
8. 結果を HTML ファイルに変換し，クライアントに出力する．

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．

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

λ 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

4.1 実験方法

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．そして，recall が 0%, 10%, 20%, ..., 100% までの 11 箇所の点における precision の値を求めた．

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

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 回繰り返して行った．また， $\lambda = 0.95, \theta = 10, \mu = 2$ 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．

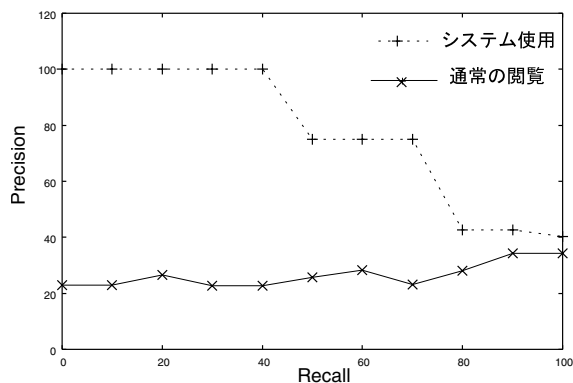


図 4: 実験結果

4.2 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．

4.3 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には，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．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で次の列を表示が行われにくい．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．こういった理由で，通常の Web サイトの HTML ファイルをそのまま PDA 上で表示しても，上部に不要な情報が集まるような事態が起こる可能性が高くなっており，それがシステム未使用の場合における precision と recall があまり高くないことにつながっていると考えられる．

5. まとめ

本稿では，PDA からの Web ページ閲覧を円滑にすることを支援するための対話的な適応を行うシステムを提案した．評価実験により，ユーザの興味を持つ部分を抽出して，その情報に適合する部分を重要視し，ページの上から表示することによって，ユーザが興味を持つ情報にすばやく接することができることを確認した．

参考文献

- [1] Corin R. Anderson, Pedro Domingos, Daniel S. Weld: Personalizing Web Sites for Mobile Users, In Proceedings of the Tenth International World Wide Web Conference, May 1-5, 2001, Hong Kong.
- [2] Mike Perkowitz, Oren Etzioni: Towards adaptive Web sites: Conceptual framework and case study, Artificial Intelligence, 118(2000) pp. 245-275.
- [3] Richard Baeza-Yates, Berthier Ribeiro-Neto: Modern Information Retrieval, 1999.