

時間的差分データの監視を目的とした携帯端末画面への表示システムに関する研究

小谷忠史[†], 中村正人[†], 岩沼宏治^{††}, 鍋島英知^{††}
山梨大学工学部コンピュータメディア工学科^{††}
山梨大学大学院工学研究科電子情報工学専攻[†]
e-mail[†]:kotani@iw.media.yamanashi.ac.jp

概要: 現在 Web へのアクセス法として PDA など携帯端末の利用が増えてきている。PDA は Web ページを何時でも何処でも表示できる利点があるが、幾つかの問題もある。小さな表示画面、遅い通信速度、操作性の悪さの問題などにより、PDA によって欲しい情報を即座に簡単に得るのは難しい。本論文では Web ページの時間的差分データの表示を念頭において、ユーザが Web ページの表示する部分を事前に指定して、その部分の最新の差分情報を即座に容易に表示する手法を提案・考察する。また以上を実現する時間的差分データ監視・表示システムを実装し評価したので報告する。

A PDA Display System for Watching Temporal Differences of HTML Documents

Tadashi kotani[†], Masato Nakamura[†], Koji Iwanuma^{††}, Hidetomo Nabeshima^{††}
Department of Computer Science and Media Engineering, Yamanashi University^{††}
Graduate School of Electrical Engineering and Computer Science, Yamanashi University[†]
e-mail[†]:kotani@iw.media.yamanashi.ac.jp

Abstract: Access to World-Wide-Web (Web) by Personal Digital Assistants (PDAs) has been increasing. PDAs can browse Web site With PDAs anywhere at any time, but there are some difficult problems. For example, PDA's screen size is rather small, the transmission speed is considerably low and PDA's usability needs to be still improved. Therefore it is difficult for a user to get his information easily and quickly. To solve those problems, we study how to display some parts of a Web page, which a user wants to watch continuously. In this paper we propose a system which asks a user preliminarily which parts of a Web page should be displayed on a PDA's display. This system continuously tries to detect the temporal difference of the target pages, and informs the user about the detected difference with PDAs in the appropriate form.

1 はじめに

PDA は Web ページを何時でも何処でもブラウジング可能であり、Web 上の情報を得るのに便利である。そのため、PDA から Web ページを閲覧したいという

要求が増えてきている。

PDA によるブラウジングは携帯性、即時性に優れているが、しかし画面の大きさ、通信速度の問題、操作性の悪さなどの問題がある。既存の Web ページは PC

の画面サイズにあわせてデザインされており、PDA 画面では表示しにくい。また画面が小さいので何度もスクロールしなければ見たい場所がなかなか見つからない。PDA のブラウザの種類によっては、HTML の表内部データの折り返しが増えて、可読性が低下することもある。また現在の無線通信では通信速度が低いので、Web ページが大きいと表示するのに時間がかかる。更に、PDA では小さなボタンやタッチペンでの入力のため操作性が悪い。これらの問題を解決するために、Web ページの文書要約や Web ページを携帯端末画面にあわせて自動生成する方法が必要になってくる。

本論文ではこれらの問題をユーザの視点にたつて実用的解決法を探る。Web ページは日々変化している。注目している最新の情報は、Web ページの時間的差分データを知ることで得られる。ユーザはその差分データから自分の知りたい情報のみを得られれば良い。そこで本論文では、Web ページの時間的差分データ表示を念頭において、ユーザが知りたい最新の情報のみを見やすく表示する手法を考えたい。以上を実現する、携帯端末上の時間的差分データを監視し、ユーザの見たい場所のみ表示するシステムを実装し評価した。

2 携帯端末の表示問題

PDA を用いた Web ページのブラウジングには、表示画面の問題と PDA のブラウザの問題 [2] がある。

2.1 表示画面の問題

PDA では既存の Web ページを閲覧しにくい。主な原因は、既存の Web ページが PC の画面サイズにデザインされているためである。既存の Web ページを PC で表示した例を図 1 に、また同じ Web ページを PDA で表示した例を図 2 に示す。

図 1, 2 で示すように、PC の場合は、画面サイズが大きく Web ページ全体を見渡せ、欲しい情報をすぐに発見できる。PDA の場合は、画面が小さいため一画面内にデータが収まらず可読性が低下する。また、Web ページ全体を見渡せないため画面のスクロールなどの操作量が増え、欲しい情報がすぐには見つからない。



図 1: 情報処理学会のページを PC で表示した例

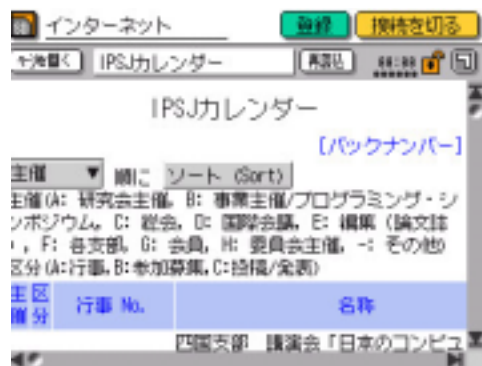


図 2: 情報処理学会のページを Zaurus[4] で表示した例

2.2 PDA のブラウザの問題

PDA のブラウザはベンダによって機能は異なる。例えば、カラー表示、HTML のタグの動作、スクロールバーの有無などは、ブラウザによって大きく異なる。

図 3 は、PDA ブラウザの Palmscape[5] で表を表示した例 [2] であるが、横スクロールバーがないので、データの折り返しが起こっている。横スクロールバーが付いていないと、表や文章が横に長ければ長いほどデータの折り返しが増加し、ユーザは何を示しているか理解しがたくなる。このような問題に加え、小さなボタンやタッチペンの操作性の悪さや通信速度の低さから、PDA では PC のように Web ページを辿りながらブラウジングするのに向いていない。

PDA のような小さな画面に既存の Web ページを表示させるとき、そのまま表示させては問題が発生する。そこで何らかの加工を施す方法を考えなくてはならな

番号	題名	著者	出版社	発行年月
オブジェクト指向プログラミング - 情報処理学会 00'98シンポジウム	ソフトウェア工学研究会	情報処理学会	9子案	1997.06
オブジェクト指向プログラミング - 情報処理学会 2000シンポジウム	ソフトウェア工学研究会	情報処理学会	9子案	1997.06

図 3: Palmscape で表示した例 [2]



図 4: レイアウトによるテーブルの使用例

い。これまでの研究として、インターネット端末にコンテンツを対応させる方法がさまざまな研究されてきた。例えば、Digester システム [1] は PDA/Laptop 用知的プロキシであり、大きな Web ページを複数の小さなページに分割し、その小ページをリンクでつなぎ、PDA 画面でも既存の Web ページを見やすくしている。ただし Digester システムはデスクトップ用の画面を細切れにするため、ユーザはしばしば混乱してしまうことがある。また、HTML 表データの折り返しを見やすく表示する研究 [2] もなされている。そこでは実際に Web 上に存在する表を調査・分類し、HTML の表を携帯端末に適した形に自動変換している。結果として、ある程度可読性の低下を解消できたが、多種多様な HTML の表を分類するのは大変であり、場合によっては変換によりデータ量が増加し可読性の低下につながることもある。

3 差分データの表示を目的としたユーザ指定部分の絞り込み表示

携帯端末の表示問題を一般的かつ実用のレベルで解決することは極めて難しい。本研究では時間差分データの表示問題等に限定して、もっと単純で実用性のある方法を提案する。先に述べたように PDA では Web ページを辿りながらブラウジングするのに向いていない。そこで、「ユーザは Web ページの何処を見たいか」に注目する。ユーザに Web ページ中のどの部分の情報が見たいかを選択させ、表示させる。情報を絞り込むことで、PDA の小さい画面でも必要な情報を見やすく表示することが可能となる。また、HTML ページが更新されたとき、更新前後の HTML ページの差分を取

ることにより、ユーザの選択した情報が更新されたか否かを検出し、情報が更新されたときのみユーザに告知する。これによりユーザは頻りにインターネットにアクセスする必要がなくなり、少ない手間で最新の情報を取得することが可能となる。このシステムを以下ページ監視表示システムと呼ぶことにする。

3.1 Web ページ中の HTML テーブル選択による絞り込み

本システムでは、ユーザが Web ページの何処を定期的に表示したいのか指定させ、その部分のみを絞り込み表示する。本システムでは、ユーザが HTML テーブルのみを選択するようにした。HTML テーブルとは、HTML の <TABLE> タグで記述されるものであり、以後テーブルと呼ぶ。テーブルの中で HTML の <TR>、<TR>、<TD> タグで記述され、それらによって区切られている部分を以後セルと呼ぶ。

HTML ページ中のテーブルを対象とする理由は、現在多くの Web サイトがテーブルを多用しているからである。テーブルは以下の 2 種類の用途がある。

- 本来のテーブルとしての利用:
本来のテーブルとはテーブルの枠がユーザが見て分かるテーブルのことであり、ランキングサイト、株価サイト、テレビ番組表などで利用されている。
- ページレイアウトのためのテーブルの利用:
ページレイアウトのためのテーブルとは、そのテーブルの多くが枠が見えないものである。例えば、ニュースサイトやポータルサイトなどで利用されている。

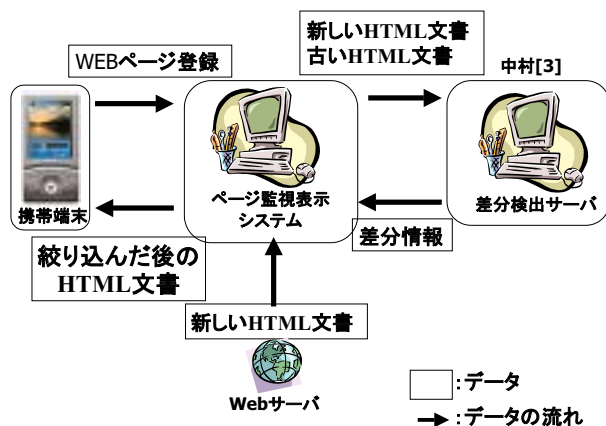


図 5: ページ監視表示システムの概要図

図4はポータルサイトであるYAHOO!JAPANの例である。左側は、PC上で普通に見たものである。右側はページ中の全てのレイアウトのためのテーブルの枠に色を付けて表示したものである。YAHOO!JAPANのトップページは、実際に目に見えるテーブルは3つしかないが、レイアウトとしてのテーブルが14ある。また、全ての項目がテーブルの中に記述されている。

最近のWebページでは見た目を重視する傾向が強く、頻繁にページレイアウト用のテーブルが使われている。多くの場合、テーブルやセル毎に目的の異なる情報が分けられ、長短のテキストや画像など種類は豊富である。例えば図6のYAHOO!JAPANのトップページでは、トピックス欄、お知らせ欄、広告欄などがそれぞれのテーブルやセルの中に入っている。このように一貫性のあるデータ集合は、ユーザにとって興味のあるデータ集合となる場合が多い。

テーブルの特徴は、多数サイトが利用しているという汎用性、テーブルデータの一貫性、豊富な種類のデータの取り扱いが出来ることである。この特徴から、ユーザはテーブルデータに着目する機会が多いであろうと考え、本システムではテーブルを対象にして情報を絞り込み表示することにした。

4 時間差分ページ監視表示システム構成

本システムの概要図を図5に示す。

本システムの重要なポイントは次の2点である。

- ユーザは登録したいWebページのテーブルまたは

セルをシステムに事前に登録する。

- システムはユーザの指定したWebページを定期的に監視し、差分検出サーバを使いWebページが更新されたかどうか調べる。ユーザが注目する部分が更新されていれば、その部分を絞り込み表示する。

これらのことについて、詳しく説明していく。

4.1 Webページの対話的登録方法

本システムは、ユーザとシステムが対話することにより、Webページを登録する。本システムとユーザとの対話はJavaServletで実装している。

本システムでは、ユーザがテーブルやセルを選択するのにチェックボックスを利用している。各テーブルや各セルにはチェックボックスが付いている。ユーザは、選択したいテーブルやセルのチェックボックスにチェック付けることで、テーブルを選択させている。ユーザは表全体でも複数のセルでも、また一つ一つのセルでも選択可能である。図6は全てのテーブルにチェックボックスが入っている本システムの表示画面の一例である。

本システムの対話的手法は次のようにした。ユーザはWebブラウザからシステムに表示したいテーブルやセルをチェックボックスにより指定すると、システムはユーザに選択されたテーブルやセルのみを絞り込み表示する。最終的にユーザが表示させたい部分的な表が絞り込み表示されるまでこのやり取りは繰り返される。これにより本システムはテーブルやセルを徐々に絞り込みユーザに提示することが出来る。ユーザによる絞り込みが終了した時、システムはユーザがどのテーブルのどのセルを選択したかを保存しておく。

図7の右側のPDA画面は、本システムを使いYAHOO!JAPANのトピックスを表示させた時の画面であり、左側のPDA画面は、普通にトピックスを表示させた時の画面である。図7の2つの表示画面を見てわかるように、トピックス欄は見やすく表示されていることがわかる。

4.2 時間差分データの表示

Webページは日々変化しているものであり、Webページに変化があればユーザに提示する情報を更新しなけ



図 6: チェックボックス入りの表

```
<HTML>
<BODY>
<TABLE>
<TR>
<TD> 牛肉 </TD>
<TD> 200 円 </TD>
<TD> 和牛 </TD>
<TD> 北海道 </TD>
</TR>
</TABLE>
</BODY>
</HTML>
```

古い HTML 文書

```
<HTML>
<BODY>
<TABLE>
<TR>
<TD> 牛肉 </TD>
<TD> 100 円 </TD>
<TD> 北海道 </TD>
</TR>
</TABLE>
</BODY>
</HTML>
```

新しい HTML 文書

```
<!DOCTYPE DOC>
<!ELEMENT DOC ((NOCHANGE|CHANGE)*)>
<!ELEMENT NOCHANGE PCDATA>
<!ELEMENT CHANGE(OLD?,NEW?)>
<!ELEMENT OLD PCDATA>
<!ELEMENT NEW PCDATA>
```

XML 文書の DTD

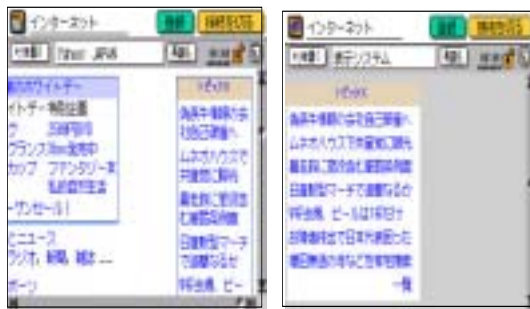


図 7: PDA での表示画面の違い : YAHOO!JAPAN のトップページの "トピックス" を表示

ればならない。そこで本システムでは、中村らの開発した差分検出サーバ [3] を使い差分データを得る。差分検出サーバは、文書構造の類似した HTML 文書間の対応関係を検出する。更新前の HTML 文書と更新後の HTML 文書の対応関係を検出し、差分データは XML 形式で返す。

HTML 文書間の時間差分データの変化は以下の 4 通りである。

1. データに変化なし
2. データに変化あり
 - (a) 新しいデータが挿入された
 - (b) 古いデータが消えた
 - (c) 古いデータが新しいデータに変化した

これを考慮し本研究で用いた XML 文書の DTD とそれを用いた XML 文書の簡単な例を図 8 に示す。

ここで、HTML 文書に変化がある場合、新規データが挿入された場合は古いデータはないので<OLD>タグ

```
<NOCHANGE><HTML></NOCHANGE>
<NOCHANGE><BODY></NOCHANGE>
<NOCHANGE><TABLE></NOCHANGE>
<NOCHANGE><TR></NOCHANGE>
<NOCHANGE><TD></NOCHANGE>
<NOCHANGE>牛肉</NOCHANGE>
<NOCHANGE></TD></NOCHANGE>
<NOCHANGE><TD></NOCHANGE>
<CHANGE>
<OLD>200 円</OLD>
<NEW>100 円</NEW>
</CHANGE>
<NOCHANGE></TD></NOCHANGE>
<NOCHANGE><TD></NOCHANGE>
<CHANGE>
<OLD>和牛</OLD>
<NEW> </NEW>
</CHANGE>
<NOCHANGE></TD></NOCHANGE>
<NOCHANGE><TD></NOCHANGE>
<NOCHANGE>北海道</NOCHANGE>
<NOCHANGE></TD></NOCHANGE>
<NOCHANGE></TR></NOCHANGE>
<NOCHANGE></TABLE></NOCHANGE>
<NOCHANGE></BODY></NOCHANGE>
<NOCHANGE></HTML></NOCHANGE>
```

XML 文書

図 8: DTD と XML 文書の例

の欄には“ ”というギャップを入れる。古いデータが消えた場合にも同様に<NEW>タグの要素にギャップを入れる。

例の XML 文書は“肉の種類”，“肉の値段”，“ブランド名”，“肉の産地”が書かれた HTML 文書が変化した時にどのような XML 文書ができるかを示す例である。例では、変化した箇所は，“肉の値段”と“ブランド名”である。変化していない HTML タグやテキストデータは<NOCHANGE>で囲んであり、変化した肉の値段は<CHANGE>タグで囲み、古いデータ 200 円は<OLD>タグ、続いて新しいデータ 100 円は<NEW> タグで囲んでいる。ブランド名も同様である。ただし、<NEW> タグのデータは“ ”である。本システムでは先ほどの

XML 文書を解析し、ユーザ指定の場所を同定し、その場所に変化があるか否かを検査し、変化があれば必要部分だけを絞り込み表示する。

5 本システムによる表示例と考察

ここでは、PDA にシャープのザウルス MI-E1[4] を用いた実際の表示画面の例を図 9 に示す。

図 9 の (a) は PC 上で asahi.com の株価ページを表示させ、(b) は本システムを使わず PDA で同じページ表示させ、(c) は本システムにより絞り込み表示させた表示画面させたものである。PC 上の表示画面の四角の枠は、ユーザが見たい場所を指す。本システムを用いず PDA で表示した場合、ユーザは知りたい情報を得るのに画面をスクロールする必要がある。しかし、本システムを使用した画面では一画面内に表示されることですぐに知りたい情報を得ることが出来る。変更があった場合、本システムでは E-mail で知らせることになっている。これによりユーザは Web ページを頻繁に確認する必要がなくなる。また、変更後の表示例は (d) になる、変更された部分のセルには色付けがされ表示される。

このように、本システムの利用することで、既存の Web ページの表示データを少なく出来る。そしてユーザは必要な情報のみを閲覧でき、Web ページに変更があれば即座に知ることが出来る。

6 今後の課題

Web ページのレイアウトとして使われるテーブルは、Web ページの構造がほとんど変化しない場合には、ユーザの見たい場所を特定するのに有効である。しかし、ユーザが見たい Web ページにテーブルが使っていないときは、Web ページの場所の指定が出来ない。また、テーブルが使ったとしても、見たい場所がテーブルの外にある場合には、場所を指定することが出来ない。Web ページのテーブル以外のところも指定出来るようにする必要がある。

本システムでは、ユーザが表示させたい Web ページの場所（テーブルやセル）の時間差分データを表示できる。しかし本システムの場合、表示画面では 1 つのページを出力しているに過ぎない。ユーザの多くは複数ページを登録し、監視してもらおうほうがありがたい。現在、複数ページの同時表示について実装を行ってい

コード	銘柄名	始値	高値	安値	現在値
3001	片倉工業	825	838	818	818
3002	グンゼ	565	518	492	493
3007	神戸生絲	42	44	42	42
3009	川島産物	118	128	118	119
3181	東洋紡績	284	287	199	208

(a)PC：普通に表示

コード	銘柄名	始値	高値	安値	現在値
3001	片倉工業	825	838	818	818
3002	グンゼ	565	518	492	493
3007	神戸生絲	42	44	42	42
3009	川島産物	118	128	118	119
3181	東洋紡績	284	287	199	208

(b)PDA：普通に表示

コード	銘柄名	現在値	前日比	出単高
3001	片倉工業	818	-29	49
3002	東シ	348	+1	3878

(c)PDA：絞り込み表示:更新前

コード	銘柄名	現在値	前日比	出単高
3001	片倉工業	794	-29	117
3002	東シ	347	+1	3833

(d)PDA：絞り込み表示:更新後

図 9: asahi.com: 金融経済：東証一部：繊維製品の表示例

る最中である。

複数の Web ページを同時表示する場合、どのように Web ページの集合を再構成表示するか考える必要が出てくる。Web ページを再構成する例として、My Yahoo![6] がある。これは、ヤフー内の記事からユーザーの見た記事のみ表示させることが出来るものである。My Yahoo! の場合は好きな様に記事を入れ換えることが可能だが、操作性が悪く、初めて使う人には使いづらいものになっている。Web ページのより良い再構成表示の方法は今後の課題である。

[6] YAHOO!JAPAN : <http://my.yahoo.co.jp/>

7 まとめ

本研究では、PDA は画面が小さいことから既存の Web ページが閲覧しにくい問題を、ユーザーに表示してほしい Web ページのテーブルやセルをあらかじめ指定させ、それを監視することで、PDA にすぐに見やすく表示するシステムを作製した。このシステムは HTML 文書の時間的差分データを常時監視表示するシステムへ自然に応用出来る。実装システムはデータの変更をすぐに発見でき、リアルタイムで表示できる。

参考文献

- [1] Tomothy W.Bickmore and Bill N.Schilit ; Digestor@:Device-independent Access to the World Wide ; In Proceedings for the Sixth International World Wide Web Conference, 1997, pp. 655-663.
- [2] 安富 大輔, 増田 英孝, 中川 裕志 ; “ 携帯端末画面への HTML 表データの表示方法 ”, 情報処理学会, デジタルドキュメント研究会 DD-29, pp. 17-24, 2001.7
- [3] 中村 正人, 梅原 雅之, 岩沼 宏治, 鍋島 英知 ; “ HTML 文書の時間的差分の自動検出 ”, 第 16 回人工知能学会全国大会 3C1-04, 2002.5, <http://www.ymd.dis.titech.ac.jp/jsai2002/programs/list/3C1-04.html>
- [4] 株式会社 シャー プ : Zaurus,MI-E1 <http://www.mobile.ezaurus.com>
- [5] 株式会社 イリンクス : Palmscape3.1, <http://www.ilinx.co.jp>