

# Webgrep: 能動的 Web パーソナライゼーション

## Webgrep: User Centered Web Personalization

青木 裕伸<sup>†</sup>

Hironobu Aoki

田中 二郎<sup>††</sup>

Jiro Tanaka

<sup>†</sup>筑波大学 修士課程 理工学研究科

Master's Program in Science and Engineering, University of Tsukuba

<sup>††</sup>筑波大学 電子・情報工学系

Institute of Information Sciences and Electronics, University of Tsukuba

### 概要

本論文では、まず、WWWの利用形態のひとつである Web 巡回に関する問題点を整理し、次に、これらの問題点を解決し、能動的な Web パーソナライゼーションを支援するシステム Webgrep を提案した。ユーザは巡回したいページの指定から、最新情報の取得、閲覧、カスタマイズまでを Webgrep 上で行なうことができる。

## 1 はじめに

Web ページの利用は、ネットワーク環境の普及により、誰もが日常的に経験するものとなった。ユーザは明日の天気、チケット情報、買物の指針など、暮らしの情報を Web を通じて参照できるようになりつつある。

ユーザが Web の日常的な利用を繰り返すと、ごく頻繁に利用するページや、比較検討用に複数まとめて利用するページなどができる。ユーザは検索エンジンなどを用いて見つけて気に入ったページを、ブックマークに登録して再利用する。天気予報、価格情報、ニュースなどの情報は同じページで繰り返し更新されることが多いので、これらのページを順に見ていけば新しい情報を容易に得ることができる。ユーザはそれらのページを順に回って更新状況をチェックし、新しい情報がないか判断する。新しい情報を求めていくつもの特定のページを巡っていくこの作業は Web 巡回と呼ばれる。

Web の利点の一つは、多種多様な情報を大量に含んでいることであるが、反面、情報が膨大であり、どんな情報がどこにあるかが掴みにくい。サーチエンジンで得られる情報も、ユーザ自身がうま

く分類、整理する必要がある。多様な情報を管理するために、パーソナライゼーション、すなわち情報の個人化 [1] が試みられている。既存のパーソナライゼーションにおいては、サーバ側で何らかのユーザ情報を管理し、システムがユーザの興味を推定してユーザに渡す情報をフィルタリングする。My Yahoo! [2] や MyNetscape [3] に代表されるポータルサイトや、amazon.com [4] に代表されるこの種のサービスを利用するためには、ユーザはサーバに個人情報や渡す必要があり、システムに特化されたサービスしか受けられない。これは、ユーザ側から見ると、サービスのシステムに左右される部分が多く、受動的である。

一方、ユーザは、システムが整理を肩代りしてくれないサービスについてはブックマークの構成などを用いて自分の手で整理をする必要がある。我々は Web 巡回とユーザ自身による管理を、「ユーザ側からの能動的な Web パーソナライゼーション」と捉えた。

## 2 Web 巡回における問題点

日常的な Web 巡回における問題点を以下に示す。

1. 巡回対象のページが増えると、Web巡回に費やす時間も増える。
2. 対象ページの更新状況は、訪問して確認しなければならない。
3. ページ内の更新された箇所はユーザが目視で探し、自らの記憶と照らし合わせて判断する必要がある。
4. 類似のページを複数比較検討したい時、ブラウザのウィンドウをページの数だけ開かなければならない。また、比較したい部分はページの中の一部に過ぎない。

Webページの更新状況のチェックをしたり、サイトを自動的に保存してローカルディスクに再構築することで巡回作業を支援するツールの代表的なものとしては WWWC [5]、Web Auto [6] などが挙げられる。Web巡回ソフトやオートパイロットソフト、更新チェッカーと呼ばれるこれらのサポートツールで重視されているのは1、2の問題である。

上記のソフトは、保存した個々のページを別々に見ている点では、通常のWebの利用と変わるところがなく、利用方法についてはユーザが更に工夫をしなければならない。

### 3 Webgrep

1、2に加えて上記の特に3と4の問題を解決するシステムとしてWebgrepを提案する。

Webgrepは、以下の3点によってユーザ自身によるWebのパーソナライゼーションを補助する。

1. Web巡回の対象ページから最新情報を取得する。
2. ユーザの興味の対象になる部分を半自動的に切り出す。
3. 類似情報の比較検討、更新状況の把握のためのインターフェースを提供する。

ユーザはWebgrep上で、最新の情報を得るだけでなく、テーマによった各ページを複数並べて比較検討し、さらにテーマおよび個々のページを組み合わせることで日常的に欲しい情報を一覧できる。

Webgrepのシステム構成を図1に示す。、Webgrepは各ページをチェックして更新情報を取得し、パターンと照合する解析部、更新情報とユーザの操作からレイアウトを決定し、表示を行なう表示部の2つの部分から構成される。

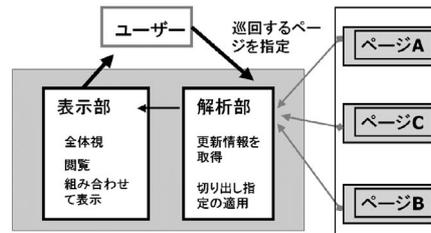


図1 Webgrepの構成

#### 対象ページの取得

解析部はユーザが選択したページ群に共通するキーワードを推定し、そのキーワードを含むブロックを各ページの切り出し候補としてユーザに提示する。

ユーザがWebgrepを利用するには、まず巡回対象のページをWebgrepに登録する必要があります。

巡回したいページのリストの指定には、既存のブラウザであるNetscape Navigatorのブックマークファイルを利用する。NetscapeのブックマークファイルはHTML形式で記述されており、ユーザが気に入ったページを登録する。ブックマークファイルはユーザが使い込むにつれ、整理・分類され、ユーザ自身が使い易いようなカスタマイズがなされている。あるWebを程度使い慣れたユーザの場合、ブックマークは階層状に管理されており、分類やテーマ毎に分類がされていると考えられる。

Webgrepは、ブックマークファイルを読み込み、ブックマークの階層構造をユーザに提示する。ユーザは、提示された情報を元に対話的にある1つのテーマに関するページ群を巡回ページとして選択する。

#### ページの切り出し

次に、選択した巡回ページの各々に切り出しの情報を与える必要がある。各ページの重要部分を切り出す方法としてまず考えられるのは、ユーザが切り出したい部分を直接指定することである。Internet Scrapbook [7]では、ユーザがHTML文書をコピー&ペーストによる例示入力を与える。この例示入力を、タグのマッチングや記事のタイトルなどによって補完し、切り出し部分が更新された文書のどこに相当するかをチェックする。

我々の手法では、まずユーザはテーマ名を入力し

た後、テーマに基づいたページ群を指定する。巡回ページには単語などに共通点が見られる。これらのページのキーワードとHTMLタグの構成から類似点を見つけ、各ページにおいて他のページと多くの共通点を持つ部分を、切り出し候補としてユーザに提示する。

Webgrepが提示した切り出し候補に対し、ユーザはこれで良いか答える。気に入らない場合は次候補が示される。すべての巡回ページの切り出し部分を決定すると、一つのテーマに関する登録が完了する。

同様にして、ユーザは複数のテーマに関する登録を行なうことができる。

### 表示とインタフェース

表示部は、テーマを管理し、登録したページ全体の更新状況を確認するための全体視表示と、実際に各ページを閲覧するための個別閲覧表示のふたつの表示モードを持つ。

一つのグラフは図2に示すように、テーマ名を表すノードを中心に、テーマに関する巡回ページを放射状に表示する。放射状に表示するノードの中にはページのタイトル、差分情報や更新日時を表示するので、ユーザはこの表示から最低限の関連情報を知ることができる。放射状に表示したノードの大きさは更新の新しさを示す。グラフの大きさ全体は構成要素である各ノードの大きさから決まるので、新しいページを多く持つグラフはより面積が広がる。

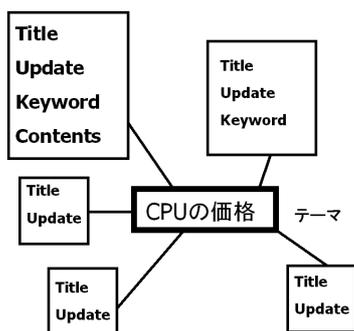


図2 グラフ表示

ユーザは既に登録した複数のグラフのノードを組み合わせ、新しいテーマについてのグラフをドラッグ&ドロップで作ることができる。ドラッグ&ド

ロップされるノードは、テーマを示すノードとページを示すノードのどちらでもよい。

あるテーマに関して各ページの詳細な閲覧を行ないたい時はページを示すノードまたはテーマを示すノードをクリックすると、個別閲覧表示に切り替わる。

ページを示すノードをクリックした場合には、ページ全体を表示し、切り出し部分を強調表示する。テーマを示すノードをクリックした場合には、登録されているページを切り出したものを一覧表示する(図3)。一覧表示ではテーマに登録されているページを切り出して複数表示する。各ページはグラフ表示に対応してレイアウトされる。優先度の高いページから、切り出した部分だけを表示する。1画面では表示しきれなくなったページについては、一つのノードに情報をまとめる。ユーザはこのノードをクリックすると表示できなかったページを表示する。一覧表示は、テーマに登録されたページの類似情報の比較検討などに用いる。

表示のための優先度は、更新日時、キーワード、含まれる数値の比較、ユーザがWebgrepを通して文書を参照した回数などで決定する。

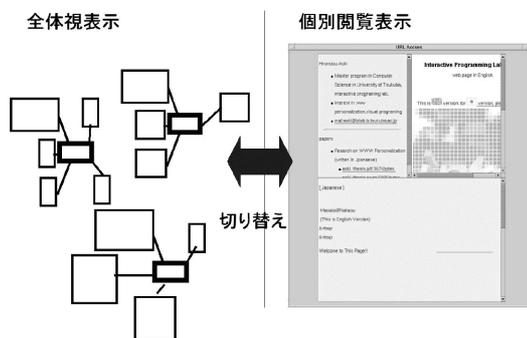


図3 2種類の表示

Webgrepは定期的に登録ページを巡回し、最新の情報を保存していく。ユーザはWebgrepが蓄えた情報について、全体視表示で更新状況を把握し、個別閲覧表示で個別ページの閲覧や、テーマに関する類似情報の比較閲覧をすることができる。

Webgrepにおける類似情報の比較の具体例としては、オンラインショップ毎の価格の比較、製品のスペックの比較などが挙げられる。さらに、天気、ニュースヘッドライン、頻繁に参照する掲示版の状



図4 Webgrep:個別閲覧表示:カスタマイズした利用

況などの組み合わせ表示も作成可能である。

図4は、「掲示板」、「天気」、「部品の価格」の3つのテーマを組み合わせで作成した新しいテーマ「関心事」を表示したときの例である。表示されている切り出し部分はそれぞれのテーマにおいて、もっとも優先度の高いものを表示する。

#### 4 関連研究

杉浦らによるInternet Scrapbook [7]は、部分切り出し機能を備えた巡回ソフトである。切り出し部分の決定においてはページのどこを切り出し部分はユーザが直接指定しなければならない。これに対し、Webgrepではテーマ毎にリストを与えることによって、半自動的に切り出し部分を決定することができる。また、ページの閲覧に関しては切り取ったページの一部を個々に閲覧する機能と、単純に縦に繋いで1枚のHTMLに合成して表示する機能があるが、Webgrepの全体視表示のような差分の表示や、個別閲覧表示のような優先度に基づいた表示機能はない。

坂上らによるANATAGONOMY [8]はサーバから配信されるニュースをユーザの嗜好に併せて閲

覧できるシステムである。しかし、対象は特定のサーバから配信されるニュースに限定されている[9]。Webgrepはサーバからの付加情報に頼らずに、Webページの優先度を決め、レイアウトする。また、ANATAGONOMYにはWebgrepの全体視表示のような管理をするための機能はない。

#### 5 今後の課題

ページリストの指定方法として、Web上にあるリンクリストも使用できる。リンクリストはブックマークと同じように人手で整理されており、大量に関連するページを得ることができる。また、リンク先の更新をチェックするような使用方法はWebgrepの応用範囲として有効であろう。

巡回を繰り返すにつれ、ページの更新頻度がわかるので、これを全体視のグラフ表示に反映させることも検討中である。

Webgrepの利用形態としては、ユーザが普段使っているブラウザと併用する形を想定しているが、ブラウザとの連動の形態は考慮する必要がある。

#### 6 まとめ

Web巡回の問題点を解決し、能動的なWebパーソナライゼーションを支援するシステムWebgrepを提案した。ユーザは巡回したいページの指定から、最新情報の取得、テーマごとの比較閲覧やテーマの組み合わせでの閲覧をWebgrep上で行なうことができる。

#### 参考文献

- [1] 神場 知成, 小池 雄一, 古関 義幸, 情報のパーソナライゼーションとその記述方式, 人工知能学会誌 Vol.14, NO.6, 1999.
- [2] My Yahoo!: <http://my.yahoo.co.jp/>
- [3] My Netscape: <http://my.netscape.com/>
- [4] amazon.com: <http://www.amazon.com/>
- [5] WWWC: <http://www.nakka.com/wwwc/>
- [6] Web Auto: <http://www.yanasoft.co.jp/>
- [7] Sugiura, A. and Koseki, Y., Internet Scrapbook: Creating Personalized World Wide Web Pages, Extended Abstracts of CHI '97, pp.343-344,1997.
- [8] 坂上 秀和, 神場 知成, 古関 義幸, パーソナル電子新聞ANATAGONOMYの開発と評価, インタラクティブシステムとソフトウェア IV, 近代科学社, 1996.
- [9] 読売COLiNS: <http://pnews.cplaza.ne.jp/>