2007－HCI－123 (9)
2007／5／11

Editable Web Browser: 編集操作の伝播によるウェブ閲覧支援

中村 聡史† 山本 岳洋† 田中 克己†

†京都大学大学院

〒606-8501 京都市左京区吉田本町

E-mail: †{nakamura, tyamamot, tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本研究は，ユーザがウェブページを閲覧している際に，削除や強調といった編集操作を可能とするものである．ユーザは削除および強調操作を利用することより操作対象が必要であるか，不要であるかといったことを明示的にシステムに伝える．一方，システムは編集内容からユーザの意図を推定し，コンテンツ全体に対して伝播させることでコンテンツを修正する．本システムの実現により，検索結果のリランキングや掲示板からの不要な投稿の削除などが手軽に利用可能となる．

キーワード ウェブ閲覧，リランキング，コンテンツ適応，編集操作

Editable Web Browser: Edit-and-Propagate Operation for Web Browsing

Satoshi NAKAMURA† Takehiro YAMAMOTO† and Katsumi TANAKA†

†Graduate School of Kyoto University

Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

E-mail: †{nakamura, tyamamot, tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract Our research enables users to do edit operations such as delete operation or emphasis operation for any portion of a browsed Web page at any time. The user can indicate that the target content is unwanted by delete operation and important by emphasis operation. The system recognizes user's intention from user's edit operation. Then, the system modifies the Web page by propagating user's intention. In order to realize our system, the user can re-rank Web search results, filter unwanted messages from Web forums, and so on easily.

Keyword Web Browsing, Re-ranking, Content Adaptation, Edit Operation

1. はじめに

近年ウェブブラウザは欠かせないソフトウェアとして広く利用されている．人々はウェブブラウザを利用してWWWにアクセスし，多くの情報を取得している．一方で，従来のウェブブラウジングには下記のような問題がある．

- **不要なコンテンツ**：掲示板におけるオフトピックな投稿やSPAM，レビューサイトにおいて閲覧対象でないコンテンツ，ウェブページ上の広告などのことであり，コンテンツの可読性を低下させる要因となっている．
- **検索の精度**：検索結果がユーザの意図に沿わないことが多々ある．これはSEO（Search Engine Optimization）などが多用されていることにより検索クエリとの適合度の低いコンテンツが上位にランクされることが多くなったことや，入力された検索クエリだけからユーザの意図を推定することが困難であることなどに起因する．
- **オブジェクト認識**：検索結果ページやニュースサイトなどではリンク先にあるオリジナルコンテンツの一部のみをサマリとしてユーザに提示している．ユーザにとって，この省略されたサマリのみを閲覧して対象が求めるものであるかどうかを判断することは難しい．

一方，ユーザはコンピュータのアプリケーションを利用する際，削除や挿入，置換，強調などといった様々な編集操作を行って

いる．例えば，エディタでテキストファイルを編集する際に，不要な部分を選択して削除したり，重要な部分を下線または背景色の変更などにより強調したりすることは一般的である．また，必要に応じてテキストを挿入したり，部分的にテキストを置き換えたりもする．

一方，従来のウェブブラウザにはこうした編集操作を広く利用可能とする仕組みは存在しない．WYSIWIG などのサービスを利用することで，ローカルに保存したウェブページや，自身のウェブサイトを編集することは可能だが，その編集は対象としているその部分を編集することのみに主眼を置いており，利用によって操作性，可読性が上がるわけではない．

編集操作を可能としているウェブ上のサービスとしては，Wikipedia に代表される Wiki ベースシステムが広く知られており，ウェブブラウザを利用した多ユーザでの協調編集を可能としている．しかし，こうした編集操作は，そのサービス内でのみ許されているものであり，広く利用することはできない．

そこで本稿では，このように広く利用されており，ユーザが意図を伝達するのに適していると考えられる編集操作をウェブブラウジングに導入する．ユーザにウェブブラウジング中の編集操作を可能とし，ユーザの編集操作の種類や対象からユーザの意図を推定する．意図推定後，その意図をコンテンツ全体に伝播させることでコンテンツをユーザに適応化させるものである．ここでは特に「削除」と「強調」という操作を編集操作として導入し，その有用性を検証する．本稿で提案するウェブブラウザを Editable Web Browser と呼ぶ．

2. 関連研究

2.1 不要なコンテンツの削除

不要なコンテンツの削除に関しては多くの研究がなされている．e-mail の SPAM フィルタ[3]はその代表的なものであり，その技術が掲示板や Blog などでも利用されている．一方で，スパマーと管理者の間でのやりとりがいたちごっこの様相を呈しており，本質的な解決にはならないことが問題となっている．また，こうしたシステムはサーバ側で対処することに主眼を置いており，クライアント側で自由に削除して可読性を向上させることはできない．一方，SPAM 以外で，ユーザが必要としていないコンテンツを削除したい時などにもこのシステムを利用することができない．

クライアントサイドで広告を削除し可読性を向上させる仕組みとして AdBlock[1]などがある．AdBlock は FireFox の拡張の一種で，JavaScript やテキスト，画像，ビデオなどで作成された広告を自動的に削除するものである．一方，このシステムを利用することでユーザは有用な広告を見逃してしまうという問題もある．

商品検索サイトなどで動的なフィルタリングを可能とする仕組みとして[4]がある．ユーザがサイドバー上に表示されている「抽出」ボタンを押すと，システムは現在ブラウザ上に表示されているページを解析し，各要素を抽出する．ユーザはサイドバーに表示された抽出情報の ON/OFF を切り替えることでコンテンツをフィルタリングすることで，必要なコンテンツの表示/非表示を切り替えることができる．これによりユーザは必要なコンテンツのみを表示し，比較をすることができる．しかし，この仕組みはコンテンツの要素を利用してフィルタリングするものであり，コンテンツの内容を考慮したリランキングやフィルタリングは行えない．

2.2 ユーザ操作に基づくリランキング

Yahoo Mindset![2]はスライダー上のポインタの位置に基づきコンテンツのリランキングを行う仕組みである．システムはあらかじめコンテンツが物販系であるか調査系であるかによって重みづけを行っている．ユーザは，検索クエリを入力したのち，スライダー上のポインタを物販系または調査系に移動することで検索結果をリランキングすることができる．

One to one ranking system[3]はコンテンツを「コストパフォーマンス」「雰囲気」な

[1] http://adblock.mozdev.org/
[2] http://mindset.research.yahoo.com/
[3] http://www.121r.com/

ど複数の軸により重みづけをおこなっているものである．ユーザはレダーチャートの各軸の値を変更することにより，コンテンツをリランキングすることができる．

山家らのシステム[5]は，ウェブ検索結果の各ページがソーシャルブックマーク上でどのようにブックマークされているかを分析し，ウェブ検索結果をリランキングするものである．ユーザはシステムによって提供されるいくつかの軸を調整することでコンテンツをリランキングすることができる．

一方で，こうしたシステムはあらかじめ決められた軸以外での並び替えなどは行えない．本システムは，ユーザの自由な観点に基づきコンテンツをリランキング可能とするものである．

適合フィードバック[2]は検索されたページ群の中からユーザがページを選択し，選択されたページの特徴ベクトルを用いて，もとの質問ベクトルを修正するというものである．ユーザが適合または不適合といった評価を各ページに下すと，システムはその評価に基づき再検索または検索結果のリランキングを行う．一方，ユーザはページ単位でしか評価することができないという問題がある．なお，こうした研究により積み重ねられた各手法は我々の研究においても有用であると考えている．

3. Editable Web Browser

我々のシステムは，ユーザに対していつでもコンテンツを編集できるようにするものである．システムはユーザの編集操作と編集対象に基づきユーザの意図を推定し，その意図をコンテンツ全般にわたって伝播させることでコンテンツの適応化を行う．

編集操作については削除，挿入，置換，強調など様々なものが考えられるが，本稿では削除操作と強調操作に注目し，システムの提案を行う．

ここでは，どのようにして問題を解決できるかということを明確にする．

3.1 掲示板やレビューサイトの閲覧性向上

人々はウェブ掲示板を利用することで各種の情報交換を行っている．ウェブ掲示板には政治や経済，趣味や料理など多岐にわたる情報がやり取りされており，有用であるといえる．一方で，ウェブ掲示板上では多くの内容についての書き込みが集まるため，ユーザはしばしば関連のない情報をフィルタリングしたいと考える．

一方，レビューサイトも同様に広く利用されているウェブサービスである．ユーザは商品を購入しようとする際に，レビューサイトをチェックすることで，商品の比較をすることができる．れビューサイトは物品系のみならず，ホテルやレストラン，映画など多岐にわたる．こうしたレビューサイトをチェックする際，ユーザはその対象を精査するため，しばしば偏ったレビューのみを閲覧しようとする．

そこで，我々はこのウェブ掲示板やレビューサイトにおけるコンテンツの閲覧において，削除という編集操作を導入する．ユーザが掲示板やレビューサイトにおいて削除操作を行った場合は，それに関するメッセージを削除することがユーザの目的であると判断する．これにより，不要なメッセージやレビューをまとめて削除することが可能となる．レビューサイトでは「高い」「安い」「悪い」などの評価に関するキーワードでフィルタリングできるようになる．なお，この仕組みは Blog のような他のサービスにも利用可能であると考えられる．

3.2 ウェブページからの広告の削除

多くのウェブサイトの管理者は，収入を得るために運営しているウェブサイトに広告を張り付けている．広告はしばしばコンテンツに溶け込んでいるため，コンテンツの閲覧を阻害する．我々のシステムはそうした不要な広告をユーザの意図により削除する仕組みを提供するものである．

ここでは，ユーザがコンテンツ閲覧中に広告の一部を削除すると，それに合わせてコンテンツに含まれる似た広告を削除するというものである．これにより，コンテンツの可読性を向上させることができると考えられる．

3.3 ウェブ検索結果のリランキング

情報を検索する際において，ウェブ検索の利用が一般化している．ウェブサーチエンジンのカバー率や速度も向上しているため有用であるといえるが，近年 SEO などが

広く利用されることになったことにより，上位に必ずしも適合ページが表示されなくなってきている．また，検索において検索クエリのみからユーザの意図を推定することは困難であるという問題もある．

そこで，本稿ではユーザのウェブ検索結果に対する編集操作を検索結果のリランキングに利用する．ユーザが検索結果上でキーワードを削除した時は，削除されたキーワードを含んでいる検索結果を下位にするようリランキングする．一方，キーワードを強調した場合は，強調されたキーワードを含んでいる検索結果を上位にするようリランキングするといったようにである．この手法は，検索結果の各アイテムのみならず，URLなどにも応用可能である．例えば，URLからblogというキーワードが削除された場合は，Blogに関するページを下位にリランキングしたり，JPが強調された場合はJPドメインのコンテンツを上位にするようリランキングするといったようにである．

システムは検索結果のすべてについて下記の式に基づきスコアを決定する．ここではNは初期状態での順位を意味している．

$$Score(N) = number\_of\_results - N \qquad (1)$$

削除の後，リランキングモジュールは検索結果アイテムのスコアを下記の式に基づき再計算する．

$$Score_{new} = Score_{last} - number\_of\_results \qquad (2)$$

再計算の結果をもとに検索結果リストの並び替えを行い，ユーザに提示する．図1はサーチ結果のリランキングを行っている様子である．

3.4 ウェブ検索結果のスニペット修正

スニペットはサーチエンジンが自動的に生成するクエリ依存のコンテンツのサマリである．スニペットは通常サーチクエリとその周辺テキストを含んでおり，その情報は検索結果ページ上に表示されている各ページが有用であるかどうかを判断するのに役に立っている．しかし，スニペットは情報量が少ないため，情報の深さや広さなどを調べるのには十分であるとは言い難い．

図1．リランキングの様子．東西線というクエリでの検索結果から東京を削除することで京都の東西線に関する情報が6位から2位に上昇している

図2．削除操作によるスニペットの編集例

この問題を解決するため，我々はスニペット修正の仕組みを導入する．ユーザはスニペットを閲覧し，不要であると感じた部分を選択し削除すると，それとは異なる部分のコンテンツをオリジナルページから取得し，提示するというものである．

Google や Yahoo!などの検索エンジンでは，スニペットを生成する際，部分的に省略すると「...」という文字列を挿入する．そこで，例えばユーザが「...」の部分を削除するとその部分を率先して展開したりすることが考えられる．

なお，このスニペットの再生成についてはオリジナルコンテンツへのアクセスが必要になるため，処理が重くなるという問題がある．スニペットの再生成機能は検索エンジン側での対応が必須であることは明白であり，その有用性を示す必要がある．

4. 実装

提案システムを実現するため，図 3 のようにシステム設計を行い，各モジュールの実装を行った．

ブラウザインタフェースモジュールはユーザに対し，通常のウェブブラウジング操作環境を提供する．それに加え，ウェブページ上での編集操作を可能とする．なお，編集操作に伴うコンテンツ適応化においては，メモリ上にコンテンツを保持し操作に応じてページのリロードをすることなく適応化されたコンテンツの提示を行う．

コンテンツ分割モジュールは，取得したウェブページを複数のパーツに分解する．まず，ページをヘッダやフッタ，メニュー，広告，メインコンテンツといった各ブロックに分割する．次に各ブロックの内部を解析し，掲示板におけるメッセージリストや，ウェブ検索における検索結果のリストのような形に分割する．さらに，各アイテムを要素に分割する．例えば，掲示板の場合であれば，ユーザ名やユーザ ID，メッセージの内容，日時など．ウェブ検索の場合であれば，タイトルや URL，スニペットなどである．図 4 は分割の例である．

モニタリングモジュールはユーザのマウス操作を監視し，コンテンツ中のどの部分が選択され編集されたかということを検知する．ユーザがコンテンツの一部を選択するとマウスカーソルの近くに編集操作ボタンを表示し，コンテンツを操作可能とする．なお，ユーザはマウス操作のみならずショートカットキーを利用して編集操作を行うこともできる．本モジュールは検知したユーザの編集操作を認識部に送信する．

認識モジュールはユーザの編集に関する情報を受信すると，編集のタイプや編集の対象，サイトの情報などを分析する．掲示板やレビューサイトで編集操作が行われた

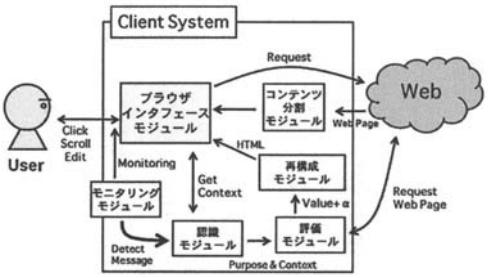

図 3. システムデザイン

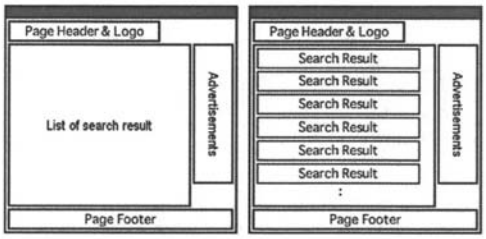

図 4. 検索結果ページの分割例

場合は，投稿者や投稿者 ID，メッセージ部分など，どの部分が編集されたかを判断し，それに基づきユーザの意図を推定する．例えば，投稿者 ID が削除された場合は，同一の投稿者 ID により投稿されているメッセージを削除する．メッセージの一部が削除されている場合は，その削除されたテキストを含むメッセージを削除するといったような形である．一般的なウェブページにおける広告についても同様に処理を行う．

ウェブ検索結果においては，タイトルおよびスニペットの一部を削除された場合は，ユーザの意図は削除されたテキストと同一のテキストを含む検索結果を下位にするようリランキングすることであると推定する．強調された場合は，強調されたテキストと同一のテキストを含む検索結果を上位にするようリランキングすることがユーザの目的であると推定する．URL についても同様で，削除または強調された URL の一部を含むものを下位または上位にリランキングすることがユーザの目的であると推定する．

評価モジュールは認識モジュールからの情報に基づき，処理の内容を決定する．処理の目的が削除の場合は，対象となる部分のスコアを閾値以下に設定する．一方，処

理の目的がリランキングの場合は，式(2)に基づき検索結果のスコアを再計算する．スニペットの修正の場合は，オリジナルコンテンツにアクセスし，どの部分が重要視されているかということを判断し，オリジナルコンテンツ自体に重みづけをおこなう．すべての処理について，スコアリングの後，情報を再構成モジュールに送信する．

再構成モジュールでは評価モジュールから取得したスコアに関する情報をもとに，ユーザに提示すべきウェブページを再構成する．なお，この処理においてはサイトの種類を考慮し，掲示板やレビューサイトの場合は ON/OFF を切り替えるものとして働き，検索エンジンの場合は並び替えを行うものとして働く．その後，HTML を再生成し，ブラウザモジュールに送信する．なお，スニペットの修正については，評価モジュールからの情報に基づきオリジナルページにアクセスし，修正を行う．

5. 評価実験

削除操作によるウェブ掲示板上の不要なメッセージの削除に関する有用性と，編集操作によるウェブ検索結果のリランキングの有用性を調査するため 2 つの評価実験を実施した．

5.1 不要なメッセージの削除

実験では 2 ちゃんねる[4]の掲示板より 10 個のスレッドをピックアップした．各スレッド毎に不要なメッセージをピックアップし，何回の削除操作ですべての不要なメッセージを削除できるか調査した．

このタスクでは，ユーザは掲示板を上位から順に閲覧していくことで不要メッセージをチェックし，不要なメッセージにたどり着くと不要メッセージの部分テキストを選択し，削除操作を行う．なお，削除されたメッセージが他のメッセージに対する返信である場合，そのメッセージも合わせて削除していく．評価実験において，削除操作回数と不要なメッセージの削除の関連性を調査するため，ユーザの操作と各操作におけるメッセージの削除を記録した．

[4] http://2ch.net/

図 5 は削除操作を繰り返すたびにどの程度不要なメッセージが削除できているかを表したものである．横軸は削除回数，縦軸はすべての不要なメッセージに占める現在残っている不要なメッセージの割合を表している．すべての掲示板のスレッドにおいて，1 回の削除操作により半数の不要なメッセージを削除できていることが分かる．また半数のウェブ掲示板において 90%の不要なメッセージが 1 回の操作により削除できていることが分かる．つまり，削除操作がウェブ掲示板において有用であることが分かる．

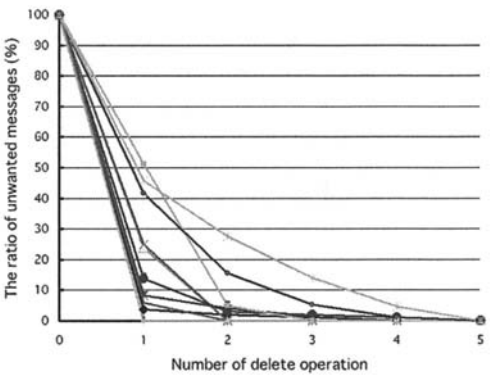

図 5. ウェブ掲示板における削除の効果

5.2 サーチ結果のリランキング

検索結果のリランキングにおいては 15 のクエリとその正解セットをあらかじめ用意した．ここでは Google を検索エンジンとして利用した．システムは 500 件の結果を提示する．

削除操作によるリランキングでは，ユーザは検索結果を上位から順に閲覧する．不適合の検索結果を発見すると，不要と思われるキーワードを選択し削除操作を行う．この操作を 3 回繰り返し，その度に上位 K 件の平均適合率を記録する．

強調操作によるリランキングでも同様にユーザは検索結果を上位から順に閲覧し，適合する検索結果を発見すると必要であると思われるキーワードを選択して強調する．この操作を 3 回繰り返し，その度に上位 K 件の平均適合率を記録する．

図 6 は削除操作とそのリランキングについて，上位 K 件の平均適合率の変化を取っ

たものである．横軸は件数を，縦軸は適合率を表している．図 7 は強調操作について同様のグラフを生成したものである．

削除操作によるリランキングでは，2 回の削除操作だけで上位 5 件の平均適合率が 44%から 74%に上昇していることが分かる．一方で，2 回目と 3 回目の削除操作では大差がないことが分かる．

強調操作によるリランキングでは，1 回の強調操作だけで適合率が 90%を超えていることが分かる．また，2 回目の強調操作で上位 20 件のほとんどの検索結果が正解となっていることが分かる．なお，3 回目の強調操作には意味がないことが分かる．

6. 議論

実現したプロトタイプシステムを 1 ヶ月にわたり利用した．その利用の中で，ウェブ掲示板の閲覧に非常に有用であることがわかった．ウェブ掲示板の閲覧においては，2 回または 3 回の操作によりほとんどの不要なメッセージをフィルタリングできた．レビューサイトではあまり定型の文章が存在しないため有用性を発揮することはなかった．これについては，評価に関する辞書をあらかじめ用意しておくことで解決できると考えられる．また，評価に関するキーワードのみならず，日時に関する情報やスコアなどの情報を利用してフィルタリングすることが可能になると有用性が向上すると考えられる．

検索結果に対する削除操作によるリランキングについては悪い結果ではないものの，素晴らしい結果というには遠いものであった．これは今回の手法が削除されたキーワードだけを対象としていることが要因である．あるキーワードが削除されたときに，それにあわせて関連語句を抽出し，そうした関連語句を含むものもあわせて削除対象とすることで有用性を向上させることができると考えられる．なお，効果的に関連語を抽出するには，クラスタリングなどに関する処理が不可欠であるといえる．今後はこうした点についても取り組む予定である．

強調操作による検索結果のリランキングは，手軽に適合結果を上位にすることができるため，非常に有用であった．一方で，どの語句を強調すると効果的にリランキングできるかということが分かりにくいという問題がある．こうした語句を提示する手法について今後取り組む予定である．なお，強調操作においても，関連語句の抽出は，リランキングの精度を向上につながると考えられる．その点についても同様に取り組む予定である．

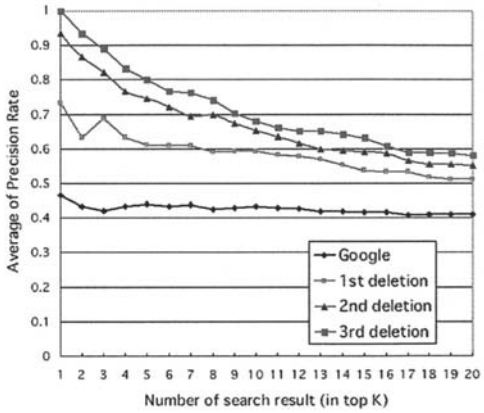

図 6. 削除操作による適合率の変化

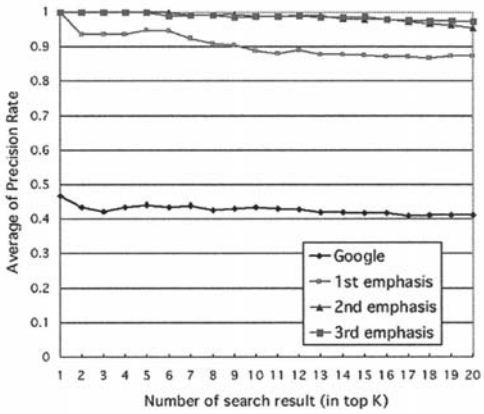

図 7. 強調操作による適合率の変化

ウェブ検索における削除と強調操作およびその伝播は，複数の対象を比較する際に有用であると考えられる．複数の検索対象に対する並列検索において，一方の編集操作を他方に反映することで，検索効率が向上すると考えられる．

現在，ウェブ検索における削除操作は検索結果のリランキングとスニペットの編集

を可能としている．2 つの削除操作の存在は混乱を招く可能性がある一方，スニペットの編集とリランキングを組み合わせると，削除操作によりスニペットを編集し，その編集を他の検索結果のスニペットにまで伝播させ，リランキングするということが可能になる．これはユーザにとって有用なものになると考えている．

本システムは PDA などのように表示スペースに制限がある一方で，ペン入力可能なシステムに有用であると考えられる．例えば不要なコンテンツの場合は，ペンを利用して打ち消し線を描くことで手軽に意思表示をすることができるだろう．強調の場合も，対象を○で囲むなどすることで操作できるかもしれない．今後は，こうしたシステムの可能性について調査するため，PDA でシステムを実装し評価実験を行う予定である．

今回は，ユーザの編集操作の履歴などは一切利用していなかった．ここで，ユーザの編集操作を蓄積し，その情報を利用するとコンテンツ適応の精度を向上させることができるようになるかもしれない．また，多数のユーザの編集操作をサーバで集積すると，編集操作の推薦などが可能となると考えられる．履歴を利用した精度改善については今後取り組んでいく予定である．一方で，サーバへの情報の送信はプライバシの問題を引き起こす可能性がある．

プロトタイプシステムではあらかじめテンプレートを用意してウェブページを分割していた．これは Gibson らの調査[6]で明らかになっている 40～50%のウェブページがテンプレートより自動生成されているという分析からも有用性が明らかである．一方，他の 50%に対応するため，ウェブコンテンツのセグメンテーションに基づくコンテンツ分割[1]についても取り組む予定である．

7. まとめと今後の課題

本研究では編集操作を可能とするウェブブラウザとして Editable Web Browser を提案し，プロトタイプシステムを実装した．またプロトタイプシステムを用いた評価実験により，本手法の有用性を示した．本提案はこれまでのクリックとスクロールというインタラクションに編集という新たな軸を導入したものであり，大きな可能性を秘めているといえる．

今後は，削除や強調操作だけでなく，ドラッグアンドドロップなどの操作に基づく移動や，コンテンツの挿入，置換といった各種の編集操作を導入し，新たなるインタラクションを提案していく予定である．また，長期的な利用に基づく実験や，削除・強調におけるキーワード推薦等にも取り組んでいく予定である．

実際にユーザに利用してもらうためには，ウェブブラウザに簡単に導入できるような仕組みを用意する必要がある．これには，GreaseMonkey[5]のような，各種のインタラクションを可能とする FireFox の拡張を利用し，システムを実装することが考えられる．

文　献

[1] Deng, C., Shipeng, Y., Ji-Rong, W., and Wei-Ying, M., VIPS: a Vision-based Page Segmentation Algorithm, Microsoft Technical Report (MSR-TR-2003-79).

[2] Ricardo. Baeza-Yates, Berthier Ribeiro-Neto, Modern Information Retrieval, Addison Wesley (1999).

[3] Sahami, M., Dumais, S., Heckerman, D., and Horvitz EM., A Bayesian Approach to Filtering Junk Email. AAAI Technical Re-port WS-98-05 (July 1998).

[4] Huynh, D., Miller, R. and Karger, D.: Enabling web browsers to augment web sites' filtering and sorting functionalities, Proceedings of the 19th annual ACM symposium on User interface software and technology, pp.125-134 (2006).

[5] Yanbe, Y, Jatowt. A., Nakamura, S. and Tanaka, K., Can Social Bookmarking Enhance Search in the Web?, ACM IEEE Joint Conference on Digital Libraries, 2007 to appear.

[6] D. Gibson, K. Punera, and A. Tomkins. The volume and evolution of web page templates. In Special interest tracks and posters of WWW'05, pp. 830-839.

[5] http://greasemonkey.mozdev.org/