

KiriBariWeb : スクリプティングによる個人化ブラウジング

内山 匡 滝川 大介 宮本 勝

日本電信電話株式会社
NTTサイバーソリューション研究所

1. はじめに

近年のインターネットの普及により、利用環境や利用目的、そして利用者層は極めて多様化してきている。これら多様化した利用の形態に如何にきめ細かく対応していくかは、特に利用者とインターネットとの接点にあるポータルサービスの大きな課題である。

ここで紹介するKiriBariWebは、利用者が自分自身の嗜好に応じてWebページを閲覧するためのスクリプティング環境である。Webからの情報取得に対する利用者の多様なニーズに応えるツールとして様々なシーンでの利用が期待できる。

2. 従来技術とKiriBariWebの特徴

近年、利用者ごとにカスタマイズ可能な統合型ポータルソリューションが企業向けに数多く提供されている(例えば[1][2])が、多くはXMLコンテンツ等を新たに用意することが必要であり、HTMLで記述された膨大な既存リソースをそのまま活用することはできない。またHTMLをXMLへ変換するwrapper[3]は、DBから作成されたWebページのようにある程度固定的な構造をもったHTMLへの適用に限定される上に、on-the-flyでHTMLからXMLへ変換する場合にはサーバ負荷の問題があり大規模なサービスには適さない[4]。

これに対してKiriBariWebには以下のような特徴があり、HTMLで記述された既存のリソースを最大限に活用することができる。

- HTMLで記述された既存ページの指定部分を自在に合成して表示することができる。
- 当該ページの内容更新後も指定部分を高い精度で特定し最新の内容を抽出できる。
- HTMLの構造を可視化することで直観的な操作で部分指定できるほか、一般のドローソフト同様の各種編集を行うことができる。

また、ほぼ全ての処理はスクリプトによってブラウザ側

で行われるために、サーバ負荷が軽いという特徴をあわせ持っている。

3. システム構成と処理の流れ

システムはサーバとブラウザ (Windows/IE6.0以上) から成る。サーバから配信されるスクリプト (JavaScript) により、ほぼ全ての処理はブラウザ側で行われる。サーバ側の処理は、スクリプトの配信、利用者設定情報の保存/配信、スクリプトでニュースサイト等の外部ページを操作するために必要な外部ページのドメイン変換のみである。

処理の流れを図1に示す。図中①～⑤が部分指定時のフロー、⑥～⑫がページ編集時のフローである。カスタマイズされたページの閲覧は⑥～⑩にしたがう。

4. HTMLの部分表現

従来の wrapper の多くはDOM (Document Object Model) [5]によってHTMLの木構造に対する操作を定

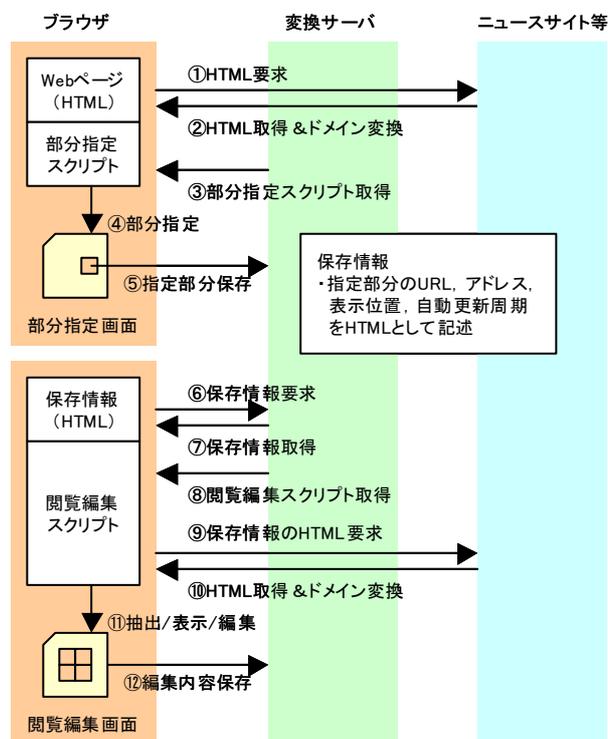


図1. 処理の流れ

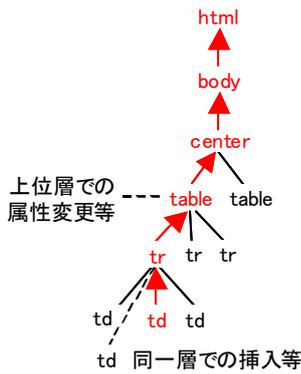


図2. 木構造中の部分指定

セル幅等タグの属性を考慮したアドレス表現を導入して類似タグを識別している. さらに上位層でのタグの挿入/削除/属性変更への対応として, 下位ノードから逐次部分木マッチング(図2中実線矢印)することにより挿入等の影響を最小限に抑えている.

5. Webページの部分指定と編集

HTMLの木構造における部分木をオブジェクトの単位としてこれを可視化し抽出するスクリプト, および抽出した部分木をオブジェクト化しこれを編集操作するスクリプトを開発した. 前者は部分指定画面(図3上)で, 後者は閲覧編集画面(図3下)でそれぞれ機能する. 閲覧編集画面はさらに閲覧モード(デフォルト)および編集モードのいずれかのモードで動作する.

図3上は既存のWebページ上で部分指定する場合の例である. ページ上に重ねて表示される選択メニューで[+][-]をクリックすることで, 部分木に対応する選択部分が拡大(縮小)する. 所望の部分が選択表示されたところで[select]をクリックすると選択部分が抽出され, オブジェクトとして閲覧編集画面へ挿入される(閲覧編集画面上に<iframe>タグを挿入し選択した部分が収まるように表示位置等を設定している).

図3下は閲覧編集画面(編集モード)の例である. 各オブジェクトを単位として, ドラッグ&ドロップによる配置やカット&ペーストによる編集等, 一般のドローソフトと同様の各種編集が可能になっている. さらに定期的に自動更新するよう, オブジェクトごとに設定することができる.

6. まとめ

利用者が自分自身の嗜好に応じてWebページを閲覧するためのスクリプティング環境であるKiriBariWebを紹介した. 今後は, システムのユーザビリティ評価,

義している. KiriBariWebでは, ニュースサイトにおける速報の挿入/削除等, HTMLの木構造の変化に対してもロバストに指定部分が抽出できるように次のような改良を施している.

まず同一層でのタグの挿入/削除への対応として, タグ名に加えて

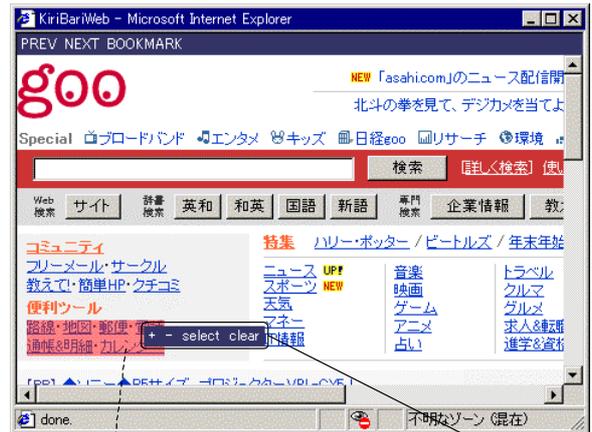


図3. 部分指定画面と閲覧編集画面(編集モード)

および実際の利用形態の調査を予定している. またカスタマイズの対象を, フォームで実現されるような機能レベルにまで拡張することを検討している.

参考文献

[1] WebShere Portal Server, <http://www.ibm.co.jp/>
 [2] Oracle 9i/Application Server Portal, <http://www.oracle.co.jp/>
 [3] Laender, A. H. F. and Ribeiro-Neto, B. A., "A Brief Survey of Web Data Extraction Tools," SIGMOD Record Vol.31 No.2 (2002)
 [4] 注目製品選択のポイント「WWWシステムを携帯電話対応にするソフト」, 日経オープンシステム No.90 pp.140-147 (2000)
 [5] Flanagan, D., "JavaScript: The Definitive Guide, Fourth Edition," pp.274-285 O'Reilly & Associates (2002)