

Personal Web Archive: Web ページのバージョン閲覧システム

若松 亮太[†] 志築 文太郎[†] 田中 二郎[†]

Personal Web Archive: System for Browsing the Versions of Web Page

RYOTA WAKAMATSU,[†] BUNTAROU SHIZUKI[†] and JIRO TANAKA[†]

1. はじめに

Web ページは更新を繰り返している。このため、過去に閲覧したページで見た情報を再び見ようとしても、ページから見た情報が削除されていること、あるいはページそのものが削除されていることがしばしばある。ページを保存しておくことで削除後もその情報を見ることができるが、閲覧中必要な情報が後に必要になることがあるため、保存しておくべきページを見極めるのは困難である。

Internet Archive¹⁾ のような Web アーカイブを利用することで、削除された情報を探すことはできる。しかし、利用者が閲覧したバージョンが保存されているとは限らず、逆に閲覧していないバージョンが存在するため、それらのバージョンを比較しながら目的の情報を持つバージョンを探しだすのは手間がかかる。

本論文では、利用者の Web 閲覧によって Web ページのバージョンを蓄積し、時系列と更新差分の提示によるバージョン比較・閲覧の支援を行うシステム Personal Web Archive について述べる。

2. Personal Web Archive

本システムでは、前述の削除された情報を閲覧する際の問題点を踏まえ、利用者の閲覧に即した個人的な Web アーカイブの作成を行う。まず、利用者が閲覧した Web ページに対し、スタイルシートや画像などを含めてその時点でのバージョンとしてローカルディスク上に保存する。また、同時に URL や閲覧時刻などの情報も閲覧履歴として記録しておく。このようにし

て、明示的な保存を必要とせず、閲覧したページのみを保存した個人的な Web アーカイブを作成する。

2.1 アーカイブの視覚化

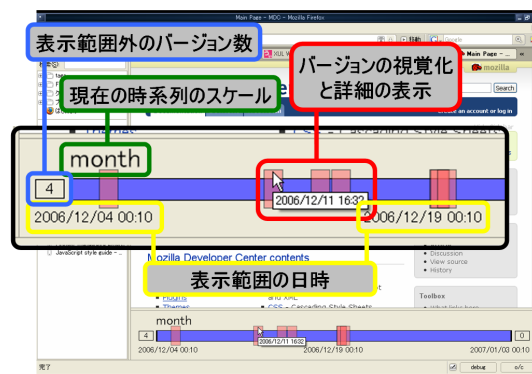


図 1 アーカイブの時系列表示

作成したアーカイブは、閲覧時刻を基にした時系列データとして視覚化を行う。図 1 は、Personal Web Archive を組み込んだ Web ブラウザである。ブラウザの下部に時系列を表す矩形を表示し、その上に閲覧中のページと同じ URL を持つアーカイブのバージョンを表すアイテムを提示している。時系列は表示範囲のスケール (10year, year, month, week, day, hour) を持ち、マウス右ボタン + 移動のジェスチャでそれを切り替える (上下でスケールアップ・ダウン, 左右で同一スケール内でのバック・フォワード)。表示範囲の外に存在するバージョンは矩形の両端にその数を提示する。アイテムは透明度を持ち、周辺の透明度が低いほどその周辺の閲覧間隔が短いことを表す。アイテムを選択することにより、そのアイテムが示すバージョンを更新を強調して提示する。

[†] 筑波大学コンピュータサイエンス専攻
Department of Computer Science, University of Tsukuba

2.2 更新差分の提示

単一バージョンの閲覧

時系列上においてクリックによりアイテムの選択を行うと、そのアイテムが示すバージョンとその直前のバージョンの差分を強調表示する(図2参照)。追加された情報は背景を赤く強調表示、削除された情報は元の位置に復元し、背景を青く強調表示をする。強調部分を確認しながらバージョンを順に閲覧していくことにより、目的の情報を見つけることができる。

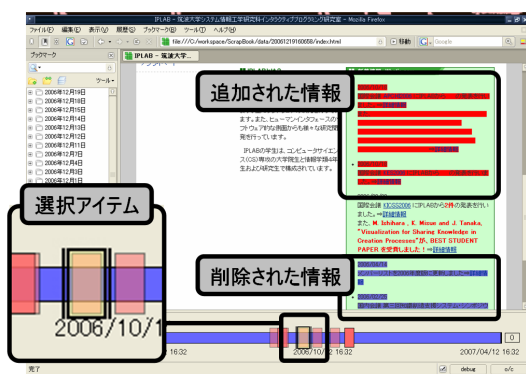


図2 更新差分の提示

複数バージョンの閲覧

時系列上においてマウスをドラッグすることで複数アイテムの選択を行う。このとき、選択範囲内のアイテムが示すバージョンとそれらの差分をマージして強調表示する(図3参照)。提示方法は単一のバージョンを閲覧する場合と基本的に同様であるが、追加や削除の鮮度によって背景色の濃度を調節し、新しい追加(古い削除)であるほど濃度を高くする。利用者は、記憶から想起するなどによってある程度の時期を推定して範囲を選択し、その結果から目的の情報を探していく。現在提示されている最も新しい追加(古い削除)より未来(過去)にその情報があつたと感じた場合、選択範囲を未来方向(過去方向)に拡げて探すことで見つけることができる。また、選択範囲のバージョン内に探している情報がある場合、選択範囲を狭めていけばよい。

3. 実装

本システムはWebブラウザFirefoxの拡張機能として開発を行った。アーカイブの作成には、本システムと同様にMozilla Firefoxの拡張機能として公開されているScrapBookとScrapBook AutoSave²⁾を用いた。また、URLや閲覧時刻などのデータは、ブラウ

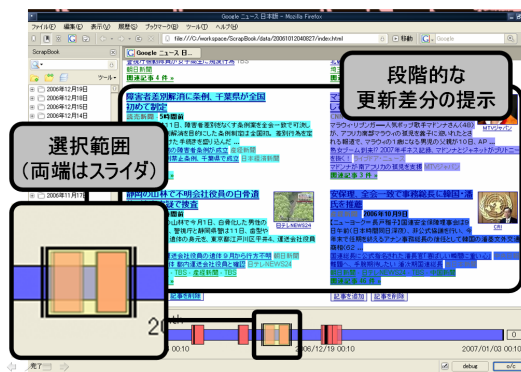


図3 複数バージョン間の差分

ザ付属の閲覧履歴とは別にRDF形式での保存を行う。

バージョン間の更新差分については、ページを提示する際に閲覧中のバージョンと直前のバージョンそれぞれのDOM(Document Object Model)ツリーを走査し、追加されたノード、削除されたノードをそれぞれ検出し、削除されたノードを閲覧中のバージョン内に復元した後、ノードのstyle属性のbackground-colorの変更による強調を行う。

4. 関連研究

白井らは閲覧したWebページを保存し、閲覧中のページと類似、関連のあるページを提示するシステムの構築を行った³⁾。また、A. JatowtらのPast Web Browser⁴⁾では、時系列上にアーカイブを視覚化し、2つのバージョン間の差分の提示を行った。我々は、蓄積したページの複数バージョン間の更新差分をWebページと同一のビューで閲覧、比較することに重点を置いて研究を進めている。

参考文献

- 1) Internet Archive: <http://www.archive.org/>.
- 2) ScrapBook:
<http://amb.vis.ne.jp/mozilla/scrapbook/>.
- 3) 白井良成, 中小路久美代, 山本恭裕. インタラクシオンヒストリによるwebブラウジング拡張. インタラクシオン 2006, pp. 223-224, 2006.
- 4) A. Jatowt, Y. Kawai, S. Nakamura, Y. Kidawara, and K. Tanaka. A Browser for Browsing the Past Web. In WWW '06: Proceedings of the 15th international conference on World Wide Web, pp. 877-878, New York, NY, USA, 2006. ACM Press.