

D-047

閲覧履歴を用いた個人用 Web アーカイブシステム

A Personal use System for Navigation of Web-Archive based on Web-Browsing History

吉川 晃生†
Akio Yoshikawa

阪口 哲男†
Tetsuo Sakaguchi

1. はじめに

近年、World Wide Web(Web)の普及によって Web を用いて情報を検索することが一般化しつつある。それに伴い、過去に閲覧した Web ページを改めて閲覧すること(再閲覧)が一般的になっている。利用者が行う検索の 30%は再閲覧を行うためのものであるとする報告[1]がある。しかし他方では再閲覧の 34.5%は失敗するという報告[2]がされている。この問題は Web で公開されている情報が変化しやすいこと、利用者がその Web ページを特定するための情報を覚えておらず改めて検索することが出来ないことが原因と考えられる。

本研究ではこの問題を改善するため、個人用 Web アーカイブシステムを構築した。利用者が再閲覧する際の支援として URL や Web ページを閲覧した日時、全文による検索の他に、利用者が Web ページを閲覧した順序(閲覧履歴)を用いる方法を提案した。

2. 個人用 Web アーカイブ

Web ページを再閲覧する方法としては Web ブラウザで提供されているキャッシュ機能を用いる方法がある。しかし、Web ブラウザで提供されている機能では Web ページの変化や消失が起こった際に対応することは出来ない。

そこで近年では、Web に公開されている Web ページを収集蓄積し、検索機能とともに提供する Web アーカイブサービスが注目されている。Web アーカイブは Web ページの収集方法によって大きく網羅的収集型と選択的収集型の二つに分けられる。

網羅的収集型は Internet Archive[3]に代表されるようにロボットなどと呼ばれる自動 Web 巡回アプリケーションソフトウェアを用いて Web ページの自動収集を行っている。そのため Web ページを収集する周期はシステムが決定し、Web ページの更新が行われるタイミングには同期していない。つまり網羅的収集型では Web ページのすべての更新された版を収集しているとは限らない。

一方選択的収集型は国立国会図書館における WARP[4]のように、収集する対象を限定する方法である。そのため、利用者の再閲覧要求を満たすのは Web アーカイブの収集対象と合致している場合に限られる。

本研究では個人が閲覧した Web ページを対象に収集する方法を用いる。これにより利用者個人が再閲覧したいと考え得る Web ページをもれなく収集することが出来る。この手法は先行研究でも採用されている手法である。

3. 先行研究

閲覧履歴を用いた再閲覧の支援は、一般的な Web ブラウザで提供されている。しかし、Web ブラウザに備わっている閲覧履歴機能は最後に閲覧した Web ページを再閲覧するための機能であり、長期間の利用に適さない。そこで個人用 Web アーカイブや閲覧履歴を利用した再閲覧の支援について、様々な研究が行われている[5,6,7]。これらの研究では、利用者が Web アーカイブを再閲覧した際に過去の閲覧順序を再現する目的で閲覧履歴を利用している。

WebScout[7]では、閲覧履歴の可視化によって再閲覧の支援を行っている。閲覧履歴の可視化とは閲覧順序を有向グラフとして表現したものである。しかし、WebScout の閲覧履歴では同一の Web ページでも閲覧した時点により別々の Web ページであると捉える。そのため、例えば「Java に関するある Web ページからリンクを辿って見つけた Web ページ」という要求には、基点となる Web ページのほかに、どの時点の閲覧履歴を検索するのか、という指定も必要となる。

4. アーカイブアクセスにおける閲覧履歴の利用

本研究ではこれに対して、Web の閲覧履歴を Web ページの遷移と捉える。遷移とは利用者がどの Web ページからどの Web ページへ辿ったかの流れであり、同じ URL ならば一つの Web ページとみなす。そのため、同一の Web ページを複数回閲覧しそれぞれ別のリンクを辿った場合、一つの Web ページからそれぞれのリンク先ページへ遷移があると考えられる。この方法は、基点となる Web ページの URL は分かるが、日時があいまいな場合を想定している。

また本研究では、Web ページの遷移情報と一般的な Web アーカイブの検索手法を組み合わせたアクセス手法を提案する。この方法により、日時の指定や URL、全文検索による手法で基点となる Web ページを絞り込み、その絞り込まれた複数の Web ページを基点とした遷移情報にアクセスすることが可能となる。

5. システムの実現

本システムはプログラミング言語として Java2SDK1.4、DBMS として Mckoi SQL Database[8]を利用した。本システムの動作確認は Windows2000 上で行った。

本システムは利用者の閲覧した Web ページを収集する部分、収集した Web ページを蓄積しアーカイブ化する部分、蓄積したアーカイブ内を検索する部分から構成される(図 1)。また、本システムでは、URL、Web ページの本文とヘッダ、閲覧履歴を蓄積する。

利用者の閲覧した Web ページを収集するために利用者 PC 上にプロキシサーバを構築する。このプロキシサーバを通して行われた HTTP による通信を監視することで、利

† 筑波大学大学院図書館情報メディア研究科

ユーザーが閲覧した Web ページを収集する。また、その Web ページに付与されているヘッダ (Last-Modified 等)、閲覧履歴を記録する。

アーカイブの閲覧は本システムのユーザインタフェース (図 2) から行う。検索可能な項目としては、URL の検索、アーカイブされている Web ページの全文検索、日時を指定しての検索などがある。遷移情報を利用した検索はユーザインタフェースの「履歴を前に N 個遡る」「履歴を後ろに N 個遡る」部分を用いて行う。遷移情報を利用した検索は他の検索項目との組み合わせによって挙動が変化する。例えば、URL と組み合わせた場合、ある URL を基点とした遷移情報が得られる。また、全文検索と組み合わせた場合、全文検索で得られた Web ページそれぞれを基点とした遷移情報が得られる。検索で得られた結果は別ウィンドウに表一覧の形式で提示される。利用者は提示された検索結果から閲覧したい Web ページを選択することで再閲覧することができる。Web ページの閲覧は利用者が通常利用している Web ブラウザを用いる。

本システムを用いて再閲覧を行っている際はナビゲーション用ユーザインタフェース部分 (図 3) に現在再閲覧している Web ページに関する情報が表示される。そこには再閲覧している Web ページを過去に閲覧した日時の一覧、その Web ページを閲覧していたときにその前後に閲覧した Web ページの情報などが表示され、選択することができる。これらを選択すること選択された Web ページを簡単に再閲覧することができる。

6. おわりに

本システムでは過去に閲覧した Web ページへのアクセス方法として遷移情報と通常の Web アーカイブ検索手法とを組み合わせたアクセス手法を提案した。これにより利用者が再閲覧したい Web ページを特定する情報ははっきりと覚えていない場合でも、その前後に閲覧した Web ページを検索できれば、そこを基点として再閲覧することが可能になった。また、基点となる Web ページがわからない場合においても、通常の検索手法と組み合わせることで再閲覧することが可能となった。

今後の課題としては本研究で提案した手法を長期にわたって使用した場合の問題やアクセス機能の改善のため、多人数の被験者を対象に調査することが挙げられる。

また、本システムでは複数の Web ページを基点とする遷移情報を利用者に提示する際の表現方法として、表による一覧という形式を用いた。しかし、この方法では遷移の状態を直観的に理解することは難しいと考えられる。そのため、複数の Web ページの遷移情報を利用者にわかりやすく提供するユーザインタフェースの提供が課題として挙げられる。

参考文献

- [1] 検索時の 30% は過去に訪問したページにアクセスしようとしている, SEM リサーチ, 2004.4, <http://www.sem-research.jp/sem/statistics/20040406014837.html>
- [2] 新垣紀子 ほか, 電子メディア社会における人の情報探索プロセス, CMCC 研究会第一回シンポジウム, 東京, 1999 <http://www.nozy.org/nojimaweb/nojima-old-WWW/Archive/shinnoz-1999.pdf>

- [3] Internet Archive, <http://www.archive.org/>
- [4] 国立国会図書館インターネット資源選択的蓄積実験事業 (WARP), <http://warp.ndl.go.jp/index.html>
- [5] 橋元 直 ほか, Packrat: 過去に閲覧したページの再閲覧を可能にする Proxy Server, 富士ゼロックス株式会社. 総合研究所システム実験研究所, 1997.12, http://www.fujixerox.co.jp/randd/12/7_hayasi/ttws/jwwwc97/packrat_paper.html
- [6] 矢野慎一郎. 個人の Web ブラウジング履歴再現システム, 2002 年度 システム情報工学セミナー発表, 2000 年度 筑波大卒業研究, <http://www.iplab.is.tsukuba.ac.jp/~yano/2002SIE/>
- [7] Natasa Milic-Frayling, Ralph Sommerer, Kerry Rodden. WebScout: Support for Revisitation of Web Pages within a Navigation Session. IEEE/WIC International Conference on Web Intelligence (WI'03). 2003. <http://csdl.computer.org/comp/proceedings/wi/2003/1932/00/19320689abs.htm>
- [8] Mckoi SQL Database Home Page, <http://mckoi.com/database/>

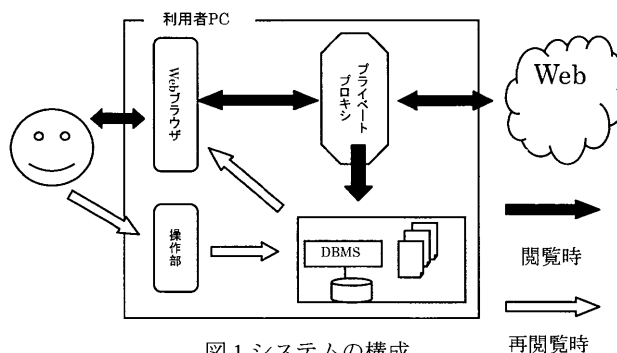


図 1. システムの構成

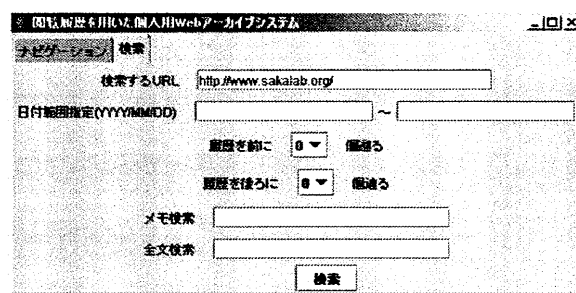


図 2. 検索用ユーザインタフェース

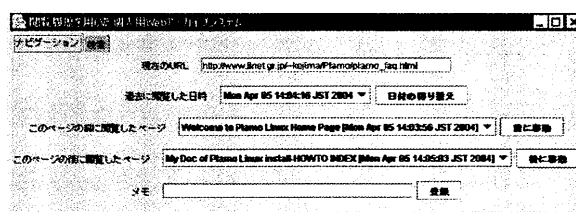


図 3. ナビゲーション用ユーザインタフェース