

Web 閲覧履歴の並べ替えに基づく情報収集支援システムの試作 An Implementation of an Information Gathering System based on Arrangement of Browsing Histories

滝由貴[†] 大園忠親[†] 新谷虎松[†]

Yuki Taki, Tadachika Ozono, Toramatsu Shintani

1 はじめに

Web を用いて情報収集を行う場合、ユーザは通常、ブラウザのブックマーク機能を用いるかメモを残す。しかし、単純に記録を残し忘れたり、その時点では重要であると思わず、記録に残さなかった情報であっても、後々の情報収集を通してその必要性に気づくことがある。そのような場合、ユーザはそのとき取りこぼした情報の中から、ブラウザの閲覧履歴を用いて必要な情報を見つけ出す必要がある。

Web の閲覧履歴から、ユーザが必要とする情報を見つけ出すことを支援するシステムとしては、Memory-Retrievr[2] および俺デスク [3] がある。一般に、人間は時系列に沿った情報を利用したほうが、過去の情報を思い出しやすいと言われている。これらのシステムは、Web 閲覧時の行動履歴を収集する。例えば、Memory-Retrievr では、キーボードおよびマウスの入力履歴や Web ページのプリンタへの出力記録、俺デスクでは、Web ページの閲覧時間や同時に開いていたほかのアプリケーションの情報などである。2つのシステムは、収集した行動履歴をもとに、それぞれのシステムで設定された関数を用いてユーザの各 Web ページに対する着目度を算出し、時系列に沿って提示する。この着目度を利用して、ユーザは目的の Web ページを閲覧履歴の中から探し出す。

人が同時に複数の調べものを行う際、タブブラウザや複数のウィンドウを用いて情報を表示する。しかし、一度に調べられる内容はたかだか一つであり、通常複数の調べものは時分割で行われる。その結果、時系列に並べた閲覧履歴は同じ内容の Web ページがとびとびに表示されてしまい、ユーザが手動で履歴から同じ内容の Web ページを拾いださなくてはならない。ユーザにとって、過去に行ったそれぞれの情報収集ごとに、関連する閲覧履歴がまとめて表示されたほうが、過去の閲覧内容を想起したり、閲覧履歴から必要な情報を見つけやすいと考えられる。

本システムでは、Web ページの内容を分類し、その結果をもとに閲覧履歴を並べ替える。閲覧履歴を並べ替えるとは、画像化された Web ページの配置を変更することである。並べ替えを行った閲覧履歴は Web ページの内容毎にまとまっており、さらにその中で時系列に沿って並んでいる。ユーザはこの並べ替えられた閲覧履歴を用いて、過去に取りこぼした情報から必要な情報を見つけ出し、情報収集に役立てることができる。

2 閲覧履歴の並べ替えに基づく情報収集支援

本システムは、閲覧履歴の Web ページの内容をもとに分類を行う。その結果を用いて画像化された Web ページを並べ替える。これにより、閲覧した日付や時刻に関わらず、同じ分類に属する情報収集の履歴がまとめて並べられる。ユーザは収集したい情報に関して過去に収集したものを即座に把握することが可能となり、新たな情報収集に役立てることができる。

並べ替えの例を図 1 に示す。図中の番号はユーザが Web ページを閲覧した順を表す。「時系列」の図は、ユーザが 2 日間にわたって、合計 8 個の閲覧履歴を残した場合の閲覧履歴を表している。この履歴は、閲覧した時系列に沿って並べら

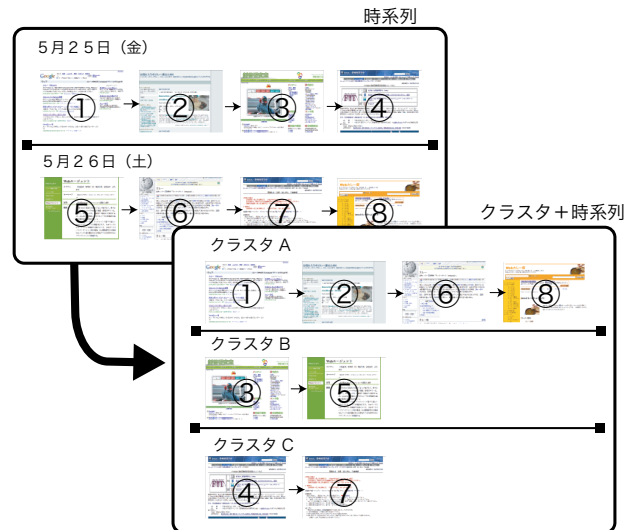


図 1: 並べ替え例

れており、この表示方法では、履歴のどれとどれが内容として関連を持っているのかがわからない。これを Web ページの内容に基づいてクラスタリングし、さらに時系列に沿ってならべる。結果は、「クラスタ+時系列」の図の通りとなる。8 個の履歴の中に合計 3 つのクラスタを持っていた閲覧履歴は、クラスタ毎に分けられ、さらにクラスタの中で時系列に沿って並べ替えられている。

ユーザの情報収集を支援するために、本システムでは、収集、分析および可視化の 3 つのフェーズを用いる。収集フェーズでは、ユーザがブラウザで行った操作を記録する。分析フェーズでは、閲覧履歴の画像を収集した情報に基づいて並べ替える。可視化フェーズでは、ユーザが閲覧した Web ページを画像化する。また、収集および分析フェーズで得られた情報と、画像化された Web ページをインターフェイスに表示する。

2.1 閲覧履歴の可視化

可視化フェーズは、ユーザが閲覧した Web ページを画像化し、閲覧履歴の可視化を行う。ユーザは、画像化された Web ページの履歴の中から、目的とするページを見つけることができる。可視化された閲覧履歴は、Web ページのタイトルのみからなる履歴よりも、ユーザにもう一度見たいと思っているページを判別しやすくさせる。また、以前に見た Web ページの内容に対するユーザの記憶が曖昧であっても、画像化された Web ページを見ることで、ユーザが過去の閲覧行動を想起することが可能となる。

2.2 閲覧履歴の収集

収集フェーズは、Web ページ閲覧中にユーザがブラウザに行った操作と閲覧日時および閲覧時間を記録し蓄積する。記録された情報を、可視化モジュールを用いて閲覧履歴に付加し、ユーザに提示する。収集された過去の閲覧行動をもとに、

[†]名古屋工業大学 情報工学科

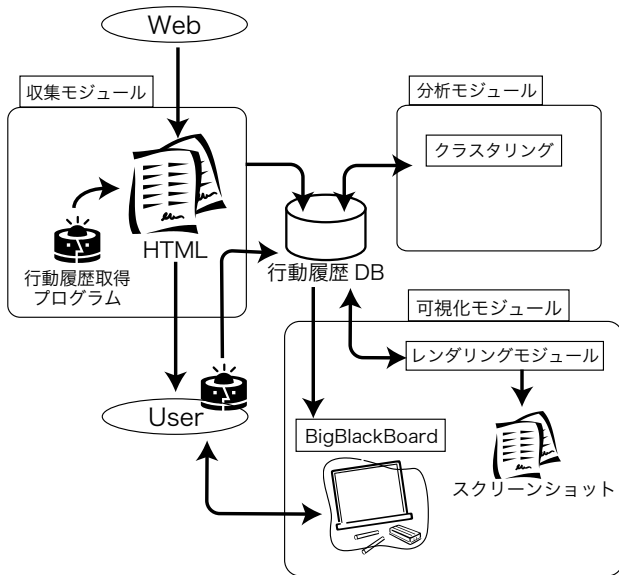


図 2: システム構成

ユーザは膨大な閲覧履歴の中から欲しい情報を絞り込むことができる。また、ユーザが、Web ページの画像だけでは、もう一度見たいと思うページを特定しきれない場合にも、収集された情報が、それを助ける。具体的には、マウスの移動履歴を記録し、ユーザがその Web ページのどこを見たのかという情報を提示する。また、スクロール情報を取得し、Web ページがブラウザ上で表示された範囲を示す。ユーザは 2 つの情報から、過去に自分がその Web ページで注目した箇所とその度合いがわかり、目的の Web ページを判別しやすくなる。閲覧履歴の絞り込みとしては、閲覧時間の短い Web ページは、過去の閲覧時に、ユーザがあまり興味を示さなかったか、もしくは必要な情報はないと判断した可能性が高い。そこで、Web ページの表示時間を用いて、ユーザが履歴を閲覧する際に履歴を絞り込むことができる。

3 実装

3.1 システム構成

図 2 に本システムの構成を示す。本システムは、収集モジュール、分析モジュール、可視化モジュールおよび行動履歴データベースから構成される。ユーザは収集モジュールを用いて Web を閲覧する。収集モジュールはユーザの Web 閲覧行動と閲覧履歴を収集し、行動履歴データベースに格納する。分析モジュールによって、閲覧履歴のドメインと内容をもとにクラスタリングが行われる。さらにクラスタ毎に閲覧履歴を時系列に並べ、その順番を行動履歴データベースに格納する。可視化モジュールはレンダリングモジュールと BigBlackBoard[1] から構成される。BigBlackBoard とは、1 枚の広大な Web ページ上の好きな場所に画像や Web ページを配置できるというシステムである。レンダリングモジュールは閲覧履歴にある Web ページを画像化し、BigBlackBoard によって、表示する。各モジュールの詳細は後述する。

本システムは、Web アプリケーションとしてユーザに機能を提供する。これは、異なる計算機環境から、ユーザが自分の閲覧履歴を利用することを可能にする。

3.2 収集

ユーザからサーバに URL がリクエストされると、サーバは Web から HTML を取得する。このとき、収集モジュール

は取得した HTML の URL とリファラー、取得日、取得時刻を行動履歴データベースに格納する。さらに、行動履歴取得プログラムである Script タグを HTML に埋め込みユーザに返す。行動履歴取得プログラムはユーザがその Web ページ閲覧中に行ったスクロールとマウス操作の情報、ブラウザの表示サイズ、Web ページの閲覧時間をサーバに送信する。閲覧時間とは、その Web ページがブラウザ上で表示されていた合計の時間を指す。収集モジュールは、行動履歴取得プログラムから送信された URL および閲覧日の情報と行動履歴データベースに格納されたそれら 2 つの情報をマッチングさせ、行動履歴取得プログラムから送信されたほかの情報を、行動履歴データベースに追加する。

3.3 分析

分析モジュールは、行動履歴データベースに蓄積された情報を取得し、まずリファラーから得られる閲覧の流れと閲覧時刻をもとに並べる。行動履歴データベースには分析モジュールによって時系列順の番号が記録される。次に、Web ページの内容とドメイン名によりクラスタリングを行い、クラスタ毎に閲覧日と閲覧時刻をもとに並べ替えを行う。行動履歴データベースにはクラスタ毎に割り振られた番号とクラスタ内の並びを示す番号を追加する。さらに、スクロール情報から Web ページがブラウザに表示された範囲を求め記録する。最後に、マウスの移動履歴から、マウス座標の y 座標を集計し、行動履歴データベースに記録する。

3.4 可視化

レンダリングモジュールは行動履歴データベースから URL を取得し、レンダリングを行い、Web ブラウザで閲覧した状態でスクリーンショットを作成する。作成されたスクリーンショットは Web サーバ上に保存され、画像への参照 URL が行動履歴データベースに格納される。本システムはインターフェイス起動時に、行動履歴データベースに格納された情報を BigBlackBoard 上に取得する。さらに、分析モジュールが生成した、履歴の並び順のデータをもとに、BigBlackBoard 上の Web ページの画像を配置し、閲覧日、閲覧時刻、閲覧時間を表示する。また、画像化された Web ページにマウスカーソルを当てた場合に、拡大画像を表示する。拡大画像上にマウスの移動履歴やスクロール情報の分析結果を表示する。

4 まとめ

Web 閲覧履歴に基づく情報収集支援システムの試作を行った。画像化された Web ページを時系列に配置し、閲覧履歴を可視化した。さらに、ユーザの収集したい情報に関連する閲覧履歴を提供するために、クラスタリングを行い、可視化された閲覧履歴を並べ替えた。また、ユーザによる Web ブラウジングの行動履歴を分析し、閲覧履歴の絞り込みを可能にし、ユーザが必要とする履歴の判別を支援した。本システムにより、ユーザは過去に取りこぼした情報から必要な情報を見つけ出し、情報収集に役立てることができる。

参考文献

- [1] 小長谷武志, 大園忠親, 伊藤孝行, 新谷虎松: “情報発信と情報交換の統合に基づく Web 情報共有システム Big Black Board の実装,” 第 19 回人工知能学会全国大会論文集 (CD-ROM), 人工知能学会, 2005.
- [2] 森田哲之, 倉恒子, 日高哲雄, 大浦啓一郎, 田中明通, 加藤泰久, 奥雅博: “Memory-Retriever: 体験獲得情報を想起させる行動検索手法,” 情報処理学会論文誌, Vol. 48, No. 3, pp. 1197-1208, 2007.
- [3] 大澤亮, 高汐一紀, 徳田英幸: “俺デスク ユーザ操作履歴に基づく情報想起支援ツール,” 情報処理学会第 47 回プログラミング・シンポジウム, 2006.