

検索の文脈を利用して再検索を容易にするブックマークシステムの提案

朱 文傑 石川 孝
日本工業大学大学院

要旨

本論文は、自動に取得された検索文脈を利用して再検索を容易にするブックマークシステムを提案する。提案システムは、ウェブを閲覧する人々が検索で一度見つけた情報を「再検索」する効率を高めることを目標としている。この目標を実現するため、我々は検索の文脈を利用して検索結果を自動的に分類する機能と検索結果を文脈によって階層的に表示する機能をもつ再検索支援システムのプロトタイプシステムを開発した。本論文は開発したプロトタイプシステムを紹介してから、再検索機能の有効性に対する評価実験の結果について述べる。

A Bookmark System Facilitating Re-finding Bookmarks Using Search Context

Wenjie Zhu and Takashi Ishikawa
Nippon Institute of Technology

Abstract

This paper proposes a bookmark system which facilitates re-finding bookmarks by using automatically gathered search context. The goal is to increase the rate at which those browsing the web can re-find information they have previously bookmarked for later use. To achieve this goal, we have developed a prototype of a bookmark system which takes advantage of using search context to incorporate new features to automatically categorize bookmarks and provide a user-friendly bookmarks' view which presents search context hierarchically. The paper presents the new features of this system, along with an analysis of an evaluation experiment measuring the effectiveness of re-finding bookmarks.

1. はじめに

WWW はインターネットの普及とウェブ関係技術の発展とともに、人類社会においても重要な「情報源」の一つに急速に成長してきた。この常に更新されている「情報源」から情報を取得するには、さまざまなウェブページを何回も閲覧するしかない。そのため、お気に入りのウェブページを再訪することはすでにWWWの利用者の日課になっている[1]。

ブックマークはページを再訪するためによく使われる手段である。ブラウザに付属するブックマークシステムが広く使われているが、Spurl[5]や iKeepBookmarks[6]などのウェブベースのブックマークシステムもある。WWWの利用者は効率的にウェブページを再訪するために、お気に入りのウェブページをブックマークとしてブックマークシステムに追加し、タグによって分類する。多くの人が大量のブック

マークを集めている[7]。

この論文で扱うブックマークの「再検索」とは、すでに追加されたブックマークの中から必要なブックマークをもう一度見つける作業である。ブックマークの量が増えるにつれて、ブックマークの再検索がますます困難になる。従来のブックマークシステムは、ウェブページの URL とタイトルをリスト構造に保存するなどの単純な機能にとどまっている[1]ので、将来的に情報を再利用する際に必要なブックマークを効率的に見つける「再検索」というニーズには応えにくい[2]。ブックマークの追加と整理に存在する困難は、ブックマークを十分に利用することを思いとどまらせる。ある研究者の言う通り、「情報はどんどんコンピュータに入るが、出てこない」[8]。

ブックマークの「再検索」の効率向上のために、検索の文脈を利用することが考えられる。検索の文脈は、ユーザが検索を行った状況を表す情報であって、検索の目的やトピックによって表現されうる。検索の文脈を明示的に残せば再検索に役立つ[9]。しかし、検索の文脈は通常、検索をする人の頭の中にしか存在しないので、検索の文脈を取得し利用することが難しい。そこで、本論文では、検索に使われたキーワードから検索の文脈を取得し利用するアプローチによって、従来のブックマークシステムより再検索を容易にするブックマークシステムを提案する。

本論文は以下のように構成される。まず、従来のブックマークシステムの関連研究について概観して(2章)、次に本研究の提案システムについて説明する(3章)。そして、提案システムの有効性を検証する評価実験を紹介し、提案システムの有効性について考察する(4章)。最後に、今後の課題について述べて(5章)から、本論文の結論をまとめる(6章)。

2. 関連研究

情報検索と知識管理の分野において、多くの研究者たちは膨大な情報世界から自分の関心を持つ情報を効率に見つけ出し、組織化することに悩まされている。そのため、効率な個人向けの情報管理システムに関する研究がコンピュータの誕生当初に提案された Memex[10]から絶えることなく続けてきた。90年代に入ると、インターネットと WWW の発展とともに、Haystack[11] や WebTop[3] を代表とする Personal Web と呼ばれる WWW を対象とした個人向けの情報管理システムに関する研究が行われた。Personal Web は WWW を個人の視点から整理した Personal View を提供することを目的とした個人情報空間の構築と組織化に関する研究である[3]。いかにして WWW の膨大な情報からユーザの関心を持つ情報を抽出し、さらに、いかにして抽出した情報を個人向けに組織化するかが Personal Web 研究の鍵となっている。ブックマークがその馴染みやすさと広範な使用のために、幾つかの Personal Web システムで利用されている。本論文で提案する再検索支援のためのブックマークシステムは Personal Web システムの一種と考えられる。

利便性と人気とがあるにも関わらず、従来のブックマークシステムには以下のような不足がある。

- ① 文脈の遺失 タイトルと URL のみが記録される
- ② 編集の困難 編集の方法は明示的にユーザに示されていない
- ③ 構造の制限 ブックマークの構造は単純なリストにと止まっている
- ④ 検索の不便 専門な検索機能が備えられていない

ブックマークシステムの不足を克服するた

めの、ブックマークに新要素を加える以下の三つのアプローチと関連システムを以下で紹介する。

CoWing[12]は、ブックマークシステムにおいて複数のユーザによって形成されるコミュニティに対して動作する協同エージェントを利用し、ユーザのブックマークの分類を支援するとともに、同じコミュニティにいる他のユーザの追加したブックマークを推薦する。

Dynamic Bookmarks System[13]は、閲覧されたウェブページから情報を集めるエージェントを利用する。このエージェントは、各ブックマークのリンク先のウェブページの相互関係とユーザの閲覧パターンとを分析することによって、ブックマークの構造化と表示を支援する。また、同じくブックマークの構造化のためにウェブページの内容を調べる手法を使ったブックマークシステムとして、PowerBookmarks[14]がある。

MyLifeBits project[15]は、ユーザが閲覧したすべてのウェブページを注釈付きでアーカイブして保存する。Yahoo!とGoogleなどの検索エンジンも類似の機能を提供している。たくさんの複雑な機能付きのブックマークシステムもあるが、多くの研究はユーザがそれらのブックマークシステムを使う意欲が高くないということを示した[4]。

これらのアプローチに対して、本論文は検索の文脈を自動的に取得して利用するアプローチで「再検索」を効率的に支援することを提案する。

3. 提案システム

ブックマークの再検索支援に対する本論文のアプローチを評価するために、提案システムのプロトタイプを ReSearchabl と名づけ、開発した。提案システムはブックマークの表示に

において、PWD(Personal Web Directory)という検索の文脈による階層表示を提供する。PWD は、ブックマークをカテゴリごとに内容を一覧に表示するだけでなく、サブカテゴリの内容も同時に一覧を表示し、カテゴリの名前によって文脈情報を表示する。本章は、まず ReSearchable の新要素を最初に紹介し、次に ReSearchable の設計について述べ、その後に ReSearchable の実装について概説する。

3.1 新要素

従来のブックマークシステムと比べて、ReSearchable には以下の三つの新要素がある。

(1) 検索に使われたキーワードから検索の文脈を取得する。検索に使われたキーワードから検索の文脈を取得することによって、前述の従来のブックマークシステムの文脈の遺失という不足を改善できる。従来のブックマークシステムはタイトルと URL などのごく限られた情報しかを保存していないため、ユーザはブックマークを追加した当時の目的やトピックなどを思い出せなくなり、「再検索」が難しくなる。Web 検索では、ユーザは興味を持つトピックに対するキーワード検索によってウェブページを発見することが普通なので、検索の文脈の情報の一部が検索のキーワードに反映されることが考えられる。本論文は、検索に使われたキーワードから検索の文脈を抽出し再現する方法を提案する。

(2) 検索の文脈によってブックマークを自動的に分類する。ブックマークを自動的に分類することによって、従来のブックマークシステムの編集の困難という不足を改善できる。従来のブックマークシステムの多くはユーザがブックマークを追加する時に手でフォルダを作って指定しなければならないので、ユーザに煩わしさを感じさせるほかにブックマークを複雑な構造に細かく分類することを妨げる。本論

文は、取得された検索の文脈によって、ブックマークを追加する時に自動的にカテゴリ化する方法を提案する。

(3) 検索の文脈によってブックマーク表示を PWD で提供する。ブックマーク表示を PWD で提供することによって、従来のブックマークシステムの構造の制限という不足を改善できる。従来のブックマークシステムはフォルダのリストとフォルダに含まれるブックマークのリストを表示することしかできなく、ブックマークのタイトルをユーザに見せるだけでは再検索の不効率がもたらされる。本論文は、カテゴリ名の代わりに文脈情報を、リスト構造の代わりに複数のツリーによる森構造を PWD でユーザに提供する方法を提案する。

これらの三つの新要素のほかに、ReSearchable は Connotea[16]などのソーシャルブックマークシステムでよく使われているブックマークレットによるブックマーク追加機能を取り入れたので、ブックマークの追加はスムーズになっている。

3.2 機能設計

ReSearchable は検索に使われたキーワードから検索の文脈を自動的に取得する。同じキーワードは違う文脈の中では意味がまったく違うことがある。たとえば、「Java」は地名と、プログラミング言語の一種と、コーヒーの一種のように三つ意味があり、それぞれは違う文脈の中で使われる。ユーザは検索するとき、キーワードを自分の目的に合わせて特定な文脈の中で限定したい場合は、文脈を反映できそうなキーワードを元のキーワードに追加してから検索をする。「Java」の例をとると、プログラミング言語の Java に限定したい場合は「Programming Java」、「Technology Java」などのキーワードを使い、地名の Java に限定

したい場合は「Island Java」、「Travel Java」などのキーワードを使うのが普通である。そこで、ReSearchable では「Programming\Java」と「Island\Java」のように検索に使われたキーワードに応じて検索の文脈情報を残し、検索の結果から追加されたウェブページを相応の文脈に関連付ける。

ReSearchable は検索に使われたキーワードからブックマークのカテゴリを自動的に生成する(図 1、2)。生成されるカテゴリの名前は暗黙的に検索キーワードの名前を使う。複数のキーワードが検索に使われた場合は、キーワードの順番に従って、カテゴリからサブカテゴリへという形で階層構造のカテゴリを生成する。たとえば、「Programming Java Forum」のようなキーワードで検索された場合は、親カテゴリは「Programming」、子カテゴリは「Java」、孫カテゴリは「Forum」のようにシステムが自動的にカテゴリを生成する。従来の研究[17]により、一度検索のクエリーに含まれたキーワードの数は平均的に三つ以下なので、自動生成によるカテゴリが複雑な構造になる心配は少ない。

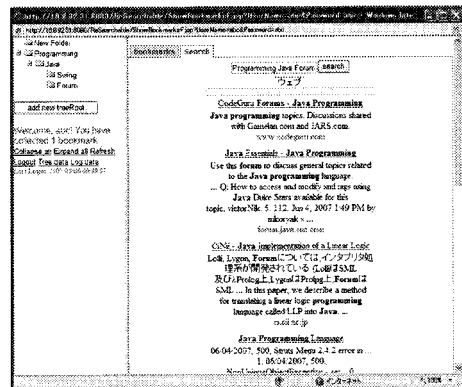


図 1 ReSearchable の検索画面

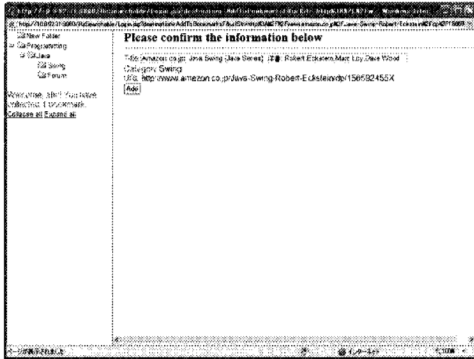


図2 ブックマークの追加画面

ReSearchable は PWD による各カテゴリのブックマークを表示する (図 3)。ユーザがあるカテゴリの内容を確認したい時は、単純なリストでブックマークを一覧表示するのではなく、サブカテゴリの内容も Yahoo! のカテゴリ検索のように階層構で表示する。PWD を利用して、ユーザは文脈情報による自動的に構造化されたブックマークをカテゴリ検索のように再検索できる。

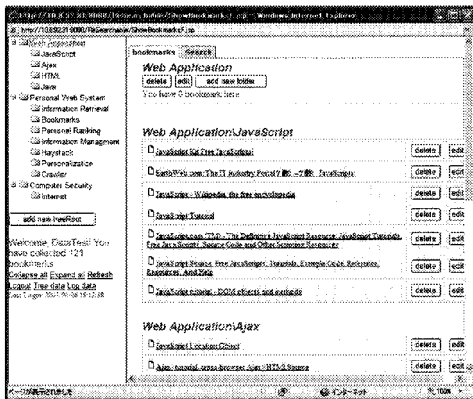


図3 文脈による階層表示

3.3 実装

ReSearchable は上述の三つの機能を評価するために開発されたプロトタイプシステムであるため、より多くのユーザが評価実験に参加できるように、GoogleAPI、Ajax 技術および

JSP を使ったウェブアプリケーションとして開発された。

ReSearchable はブックマークレットを使って、すべての操作の始まりは「Add to ReSearchable」と「Show ReSearchable」との二つのブックマークレットボタンからである。ブックマークレットボタンはブラウザに追加されると、ブラウザのリンクバーに常に表示されている。ユーザはブックマークを追加する時は「Add to ReSearchable」を、ブックマークを閲覧する時は「Show ReSearchable」をワンクリックすることによってシステムにアクセスできる。また、「Show ReSearchable」による呼び出された ReSearchable のブックマークの表示と編集の画面において、ボタンによる操作とドラッグ&ドロップ操作がともにサポートされている。

4. 評価実験

ReSearchable の再検索支援機能の有効性を評価するため、20 名の大学院生を被験者とする評価実験とアンケート調査を実施した。ReSearchable と比較するために、従来のブックマークシステムの一般的な機能だけを実装したベースシステムを開発した。ベースシステムの操作は ReSearchable の三つの新要素を除いて、ReSearchable の操作と同じにした。

評価実験は検索と再検索との二つの部分によって構成される。検索の実験では、20 人の被験者は無差別に A と B との二つのグループに分けて、ReSearchable とベースシステムをそれぞれ一週間にわたって使って WWW から 100 件 (件数の誤差は 10 以内) のブックマークを集めてもらった。A グループは先にベースシステムを使ってから ReSearchable を使い、B グループは先に ReSearchable を使ってからベースシステムを使う。集められたブックマ

クは全部被験者の研究テーマに関連しているが、ReSearchableで集められた約100件のブックマークとベースシステムで集められた約100件のブックマークのトピックは違う。

約100件のブックマークの収集を完了した時から一週間（前後二日間の誤差）の後に、再検索の実験を実施した。再検索の実験では、集められた約100件のブックマークから事前に無差別に抽出した10件のブックマークのタイトルを被験者に見せて、それらを約100件のブックマークからもう一度見つけてもらう。被験者が再検索する時に、ブックマークの再発見までにかかった時間を一件ずつに測定した。

20人の各被験者についての平均再検索時間を図4に示す。ここでいう再検索の平均時間は、10件のブックマークの一件ごとの再発見までにかかった時間の算術平均である。

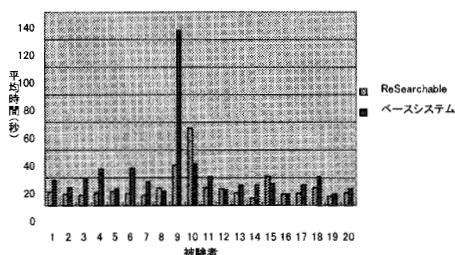


図4 再検索の平均時間

図4に示されたように、全体的に見ると、ReSearchableでの再検索時間はベースシステムでの再検索時間より少ない。特にAグループ（ReSearchableを後に使ったグループ）のほうがReSearchableでの再検索時間が小さい。例外としては、8番、10番、12番と15番の被験者は逆にベースシステムでの再検索時間は小さい。また、AグループとBグループの再検索時間の平均値の比較を図5に示す。

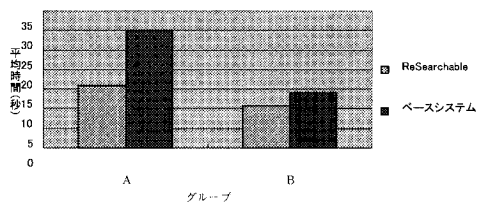


図5 グループ再検索時間の比較

図5に示されたように、AグループのほうがReSearchableでのブックマーク再検索を支援する効果はBグループよりよい。この差の原因は、Aグループは先にベースシステムで約100件のブックマークを取録したため、ブックマークの追加や編集などの共通作業の熟練度は高められたことが考えられる。

ReSearchableの新機能に対するユーザのフィードバックを調べるためのアンケート調査を実施した。アンケートでは、被験者にReSearchableの1)ブックマーク自動分類、2)PWD、3)ブックマークの追加、4)ボタンの操作性、5)カテゴリの自動生成の五つの機能について五段階で採点してもらった。17人分の調査結果による、1)ブックマーク自動分類と2)PWDが一番高く評価された。具体的な評価結果は図6に示す。

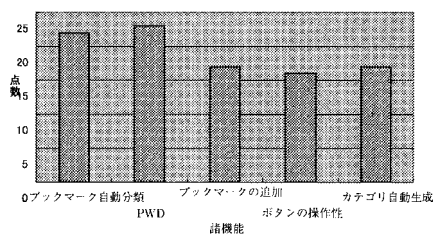


図6 各機能の評価

5. 今後の課題

評価実験とアンケート調査からReSearchableの新要素は再検索の効率向上に

役立つという有意義な結果を得たので、著者らは現段階の ReSearchable の不足と被験者から得たフィードバックに基づき、今後の研究課題を以下のように考えた。

- ① キーワード検索結果リストと PWD に個人向けのランキングを提供し、検索と再検索の効率向上を図る。
- ② キーワードの入力ミスを自動的に検出する。
- ③ ウェブページの metatag で提供されたキーワードが検索に使われたキーワードと違う場合は、ユーザに提示する。
- ④ PWD において複数の整列の基準を用意しておく。

6. まとめ

本論文は、ブックマークの再検索において検索の文脈を利用した再検索を容易にする新しいアプローチを提案し、この提案に基づき、プロトタイプシステム——ReSearchable を開発し、提案の有効性に対する評価実験を実施した。評価実験の結果により、提案したアプローチを導入した ReSearchable は従来のブックマークシステムと比べて、再検索においてはより効率的であることが確認された。

参考文献

[1] Robert, J., Ruiz, J. and Lank, E. Making Favorites Useful, 2005.
[2] Jones, W., Bruce, H., Dumais, S. Keeping Found Things Found on The Web. In Proceedings of CIKM 2001.
[3] Wolber, D., and Brooks, C. Navigating the Personal Web, JCDL Conference'04, 2004.
[4] Henderson, R., How do People Organize Their Desktops? Proc. CHI 2004.

[5] Spurl. Spurl.net 1.0 free online bookmark available at <http://www.spurl.net/>.
[6] ikeepbookmarks.com. A web-based bookmark manager available at <http://ww2.ikeepbookmarks.com/>.
[7] Cockburn, A. and McKenzie, B. What do web users do? An empirical analysis of web use. International Journal of Human-Computer studies, 4, June 2001.
[8] Boardman, R. and Sasse, M. A. "stuff goes into the computer and doesn't come out" a cross-tool study of personal information management. In Conference on Human Factors in Computing Systems – Proceedings, pages 583-590, 2004.
[9] Bharat, K., SearchPad: Explicit capture of search context to support web search. WWW2000.
[10] Vannevar, B. As We May Think, Atlantic Monthly, July, 1945.
[11] Adar, E., Karger, D.R., Stein, L.A. Haystack: Per-User Information Environments, 8th Conference on Information and Knowledge Management, 1999.
[12] Kanawati, R. and Malek, M. A Multi-Agent System for Collaborative Bookmarks, Proc. Conference on Autonomous Agents and Multi-Agent Systems 2002.
[13] Takano, H. and Winograd, T. Dynamic Bookmarks for the WWW: Managing Personal Navigation Space by Analysis of Link Structure and User Behavior, Proc. HyperText 1998.
[14] Agrawal, D., Bufl, C., Chang, C. Chang, E., Hara, Y., Hirata, K., Ito, R., Kimura, Y.,

Li, W., Mukherjea, S., Saito, Y., Shimazu, K., Vu, Q. and Wu., Y. PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management, Proc. SIGMOD 1999.

[15] Bell, G., Gemmell, J., and Lueder, R. The MyLifeBits Lifetime Store, ACM SIGMM 2003 Workshop on Experiential Telepresence.

[16] Connotea. Free online reference management for clinicians and scientists. <http://www.connotea.org/>.

[17] Lawrence, S. and Giles, C. L. Accessibility of information on the web, Nature, Vol. 400, No. 6740, pp 107-109, 1999.