

検索のコンテキストによるブックマーク再検索の効率化

朱 文傑² 武内 雄太¹ 羽生 武史¹ 吉田 脩二¹ 石川 孝^{1,2}

1 日本工業大学, 2 日本工業大学大学院

概要

ウェブ検索から取得する情報量の増大に伴って、一度見つけたウェブページをもう一度見つけることが困難になる。ブックマークはウェブページ再検索のためによく利用されるが、ブックマークの数の増加につれて、ブックマーク自体の再検索が問題になる。本論文は、ウェブページの検索に使われたキーワードをコンテキストとしてブックマークを構造化することで、ブックマークの再検索を効率化するブックマークシステムの設計と評価について報告する。

Efficient Re-finding of Bookmarks by Search Context

Wenjie Zhu, Yuuta Takeuchi, Takeshi Hanyuu, Syuuji Yoshida and Takashi Ishikawa
Nippon Institute of Technology

Abstract

With the increase of information gained by web search, it becomes difficult to re-find the web pages found before. Though bookmarks are used frequently to help re-find web pages, re-finding bookmarks itself becomes difficult as the number of bookmarks increases. The paper reports on the design and evaluation of a bookmark system which makes the re-finding of bookmarks more efficient by organizing bookmarks by keywords as search context in web search.

1. はじめに

インターネットと WWW 関連技術の発展に伴って、ウェブ検索による情報取得が簡単になってきた。WWW 利用者は WWW から必要な情報を便利に取得できるようになった一方で、一度見つけた情報をもう一度見つけることの再検索の煩わしさに悩まされている。この論文で扱うウェブページの再検索とは、ウェブ検索で一度見つけたウェブページを後でもう一度見つけることである。ウェブページの再検索はユーザのアクセスした情報量の増加につれて次第に困難になるため、Personal Web の効率化を目的とした個人向けの再検索に関する研究が盛んに行われている[11,6]。ブックマークがウェブページの再検索にはよく

使われているが、ウェブブラウザのブックマークシステムは再検索のための機能がまだ不足している[7]。本研究は検索に使われたキーワードを検索のコンテキストとしてブックマークを構造化することでブックマーク再検索の効率化を試みる。

本論文は以下のように構成される。2 章では、検索のコンテキストについて述べ、3 章では本研究の再検索支援機能について説明する。4 章では設計した機能を検証する評価実験について説明し、提案するシステムの有効性について考察する。5 章では関連研究を紹介し、最後に 6 章で今後の課題について述べてから、本論文の結論をまとめる。

2. 検索のコンテキスト

ウェブ検索のコンテキストは、ユーザはウェブ検索から情報を取得する際に目的に合わせて情報を特定するなどの大きな役割を果たしている[12]。日常会話において人々が文脈の中で相手の話の意味を理解すると同じように、ウェブ検索する際にユーザは検索のコンテキストを意識しながら情報を探している。たとえば、熱帯島に旅行しようとしているユーザとプログラミングを学びたいユーザは同じく「Java」というキーワードで検索を行う場合、検索のコンテキストが明らかに違う。以下はウェブ検索におけるコンテキストとは何かを説明する。

ウェブ検索のコンテキストは、Steve Lawrence 氏の研究[1]によれば、注釈、閲覧履歴、キーワードリストの3種類に分類できる。

(1)注釈 ウェブページに対し追加するタグやコメントなどの付加情報で、ユーザの付与した明示的なコンテキストである[2]。注釈はウェブ情報の再検索に役立つが、注釈を追加すること自体は余計な手間がかかってしまう。

(2)閲覧履歴 ユーザのウェブナビゲーションのコンテキストとも考えられており、閲覧されたウェブページの集合により表現できる[3]。閲覧履歴は多くのブラウザの付加機能になっているが、ウェブページの再検索とブックマークの再検索に用いられたことはほとんどない。

(3)キーワードリスト ウェブ検索に使われた順序付きのキーワードのことで、ある程度ユーザの目的や昔の経験を反映できる[4]。また、キーワードリストはウェブ検索の結果リストにあるウェブページのトピックパス[5]と考えられており、ウェブ検索の情報氾濫問題(abundance problem)を大いに解決できる。

本研究で扱う検索のコンテキストは、主に上述のキーワードリストである。また、本研究では、キーワードリストの順序が上位概念から下位概念への順序と考えられ、キーワードリストの最初のキーワードから最後のキーワードまでの経路が当該検索のトピックパスと呼ばれる。

3. 再検索支援機能の設計

本論文で提案する再検索支援機能は、ユーザが検索のコンテキストを選択すると、検索のコンテキストによって構造化されたブックマークを出力とする。検索のコンテキストは、検索に使われたキーワード(順序付き)で表現される。キーワードの優先順位からキーワードの重要度と検索の目標となる中心語を発見でき、複数のキーワードの関連から検索のトピックと目的を発見できる。例えば、図1のようにユーザは自分の目的に合わせて「java」の意味のある特定なトピックに限定したい場合は、往々にして「java」のほかに別の適切な単語をも見つけて、「java」と一緒にキーワードの組み合わせで検索する。

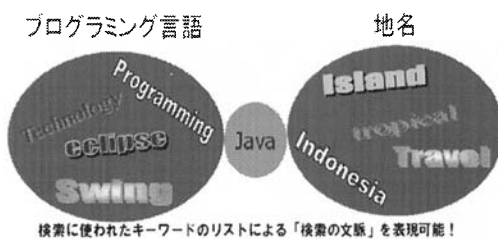


図1「Java」を中心語とする検索のコンテキスト

ブックマークは検索のコンテキストにより階層的に表示され、各ブックマークは特定なトピックパスを持つ。検索に使われたキーワードのリストから検索のコンテキストを抽出できるので、逆に考えれば、キーワードのリストによる検索のコンテキストの再現も実現できる。しかし、単なるウェブページの検索ごとに使われたキーワードからキーワードのリストを抽出し保存するだけでは、ウェブページの検索の回数が増加につれて、保存されるキーワードのリストの量もただならぬ数になってしまう。そこで、Yahoo!などのカテゴリ検索でコンテキストを示すためにパスを使うと同じように、キーワードのリストをパスに変換してから保存する方法が考えられる(図2)。

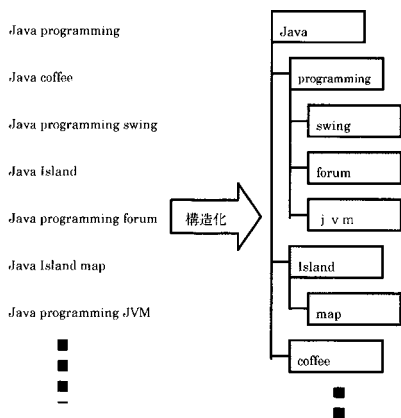


図2 検索のコンテキストを階層構造に変換

提案する機能の入力は検索に使われたキーワードであり、出力は検索のコンテキストによって構造化されたブックマークである。検索に使われたキーワードのリストから検索のコンテキストを自動的に取得しブックマークを構造化するために保存するアルゴリズムは以下である。

STEP1 入力されたキーワードから順番に単語を抽出し、ベクトル V に入れる

STEP2 V の先頭から単語を順番に取り出し、トピックによる構成されたツリー T のルートから各ノードにマッチングしてみる

STEP3 IF (マッチできなかった) この単語を新しいトピックとみなし T のノードとして追加する

ELSE マッチできた T のノードを現在の「検索のコンテキスト」とする

STEP4 V の先頭の単語を削除し、二番目の単語を先頭の単語にする

STEP5 V にまだ単語あれば、STEP2 へ、なければ終了

以上のアルゴリズムにより、検索に使われたキーワードのリストからブックマークのカテゴリが自動的に生成される。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。一番外側のカテゴリから内側のあるカテゴリまでの経路が、そのカテゴリに属するすべてのブックマークのトピックパスとする。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。「Forum」に属するすべてのブックマークのトピックパスは「Programming>Java>Forum」となる。従来の研究[13]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、提案する機能のアルゴリズムによるブックマークのカテゴリが複雑な構造になる心配は少ない。

4. 設計の評価

4.1 目的

評価の目的は、提案の検索のコンテキストによるブックマーク再検索を効率化する機能をウェブベースのブックマークシステムで検証することである。

評価実験のために、検索のコンテキストを利用したブックマークシステム ReSearchable(図3)のプロトタイプと従来のブックマーク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。

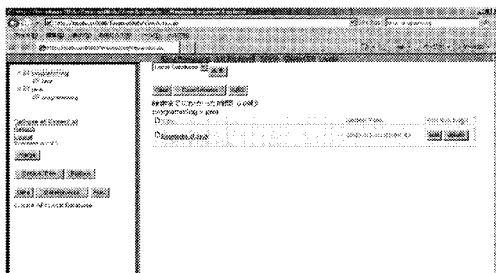


図3 ReSearchable のユーザインタフェース

4.2 方法

提案機能を評価する方法は、提案機能を実装した ReSearchable を使って、ブックマークの再検索にかかる時間を比較の基準とする評価実験を行うことである。評価実験は、ブックマークの再

検索の効率において ReSearchable とベースシステム（検索のコンテキスト関係の部分を除いて、ユーザインタフェースは ReSearchable と同じ）の比較と、ReSearchable と IE の比較によって構成される。また、いずれの評価実験も検索と再検索との二つの部分によって構成され、20 名の被験者を無作為に 10 名ずつの A グループ B グループに分けてから実施された。以下は二つの評価実験の内容を順番に説明する。

(1) ベースシステムと比較する実験 大学院生 20 人を被験者とし 3 週間に渡って行われた。検索の部分では、被験者に 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もらい、キーワード検索で WWW から見つけた価値あるウェブページをブックマークとして約 100 件を登録してもらい。集められたブックマー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。それに、順序効果を考慮して、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ことにする。検索の実験で約 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する。検索の実験の段階ですでに登録された約 100 件のブックマークから無差別に選んだ 10 件のブックマークのタイトルを文字で被験者に見せて、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、ウェブページの再発見までにかかった時間をデータとして一件ずつに測定する。

(2) IE と比較する実験 卒研生を主にした 20 人を被験者とし 2 週間にわたって行われた。A グループは、登録と再検索の際、先に IE を使ってから ReSearchable を使う。B グループは、登録と再検索の際、先に ReSearchable を使ってから IE を使う。登録の実験は、被験者に論文を書く際に生じた疑問点を 2 つあげてもらい。被験者は、ReSearchable と IE それぞれに WWW から約 30 件（誤差 3 件以内）のブックマークを行う。登録

を行う基準は、疑問点に対する回答が存在するかである。ブックマーク後は、そのウェブページから回答の部分の最大 1 段落で抜き出してもらう。被験者が、登録開始から登録終了までにかかった時間を登録時間として測定した。約 30 件のブックマークを登録後の 1 週間（前後 3 日間の誤差）後に、再検索の実験を実施する。再検索の実験は、約 30 件の中から 10 件のブックマーク登録時に抜き出してもらった文章を抽出し、それらを約 30 件のブックマークの中から再度見つけてもらう。被験者が、再検索するときブックマークの再発見までにかかった時間を 1 件ずつ測定する。

4.3 結果

比較対象が違う二つの評価実験のいずれも、ブックマークの再検索において ReSearchable が比較対象と同じかより効率的という結果を得た。具体的なデータと分析は、以下に実験ごとに説明する。

ベースシステムと比較する結果は図 4 で示す。青色の縦棒は ReSearchable での被験者ごとの再検索時間の平均で、紫色の縦棒はベースシステムでの被験者ごとの再検索時間の平均である。ここでいう再検索時間の平均とは、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。また、星印の付いたところは有意水準 10% で有意差があるデータである。1 番から 10 番までの被験者は A グループであり、11 番から 20 番までの被験者は B グループである。

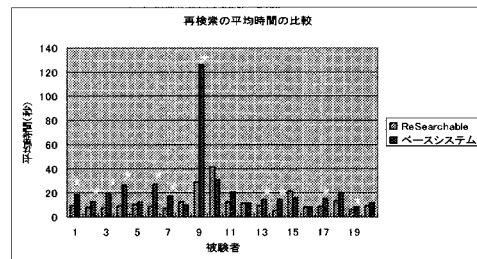


図 4 ベースシステムとの比較結果

IE と比較する結果は図 4 で示す。IE と ReSearchable の再検索時間の差で t 検定を行った

場合、有意水準 5%より小さく、再検索時間の平均が IE より ReSearchable が早い被験者が 5 人いた (図 5 中に星印で示す)。有意水準 5%より小さく、ReSearchable より IE が早い被験者がいなかったため、IE と ReSearchable で再検索の時間が変わらない、または ReSearchable が早いという結果となり、再検索において ReSearchable が有効的である。また IE と ReSearchable のウェブページの登録時間の差で t 検定を行った場合、A グループの確率 p の 2 倍値は 0.561、B グループの確率 p の 2 倍値は 0.859 であり有意水準 5%より両グループとも大きいので、ブックマークレットによる登録と IE のブックマーク機能による登録に差がないといえ、ブックマークレットによる登録が実用に供する。

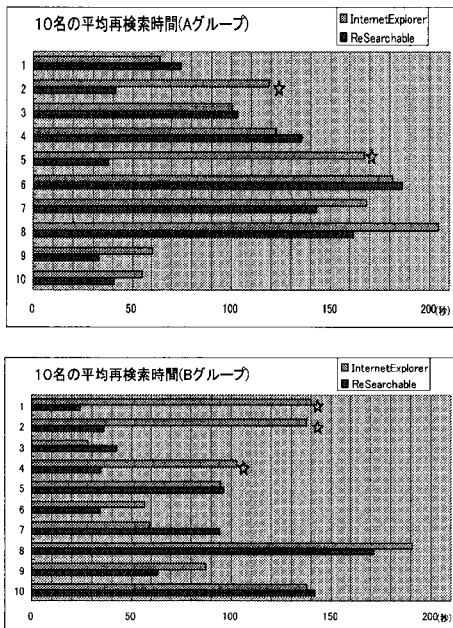


図 5 IE との比較結果

4.4 考察

ベースシステムとの比較では、平均値の差の検定結果により、有意水準 10%で有意差が出た被験者は過半数の 11 人もいることと、グループごとの平均値の差の検定結果により、A グループと B グループのいずれも有意差が出たことから、ReSearchable はベースシステムよりブックマ

クの再検索においてより効率的であることが分かった。

IE との比較では、ReSearchable のブックマークレットによる登録と IE の機能によるブックマーク登録に差がない事と、再検索時間の平均が IE より ReSearchable が早い事から、ReSearchable は IE よりブックマークの再検索においてより効率的であることが分かった。

IE との比較実験の再検索段階で ReSearchable を使う際の被験者の行動を観察する記録により、以下の傾向が見られた。

- 有意差のある被験者は往々にしてコンテキストのツリーを全部展開して、しばらく各トピックパスを見てからある特定のカテゴリ (フォルダ) を選ぶ。
- 有意差のない被験者は往々にして最初はコンテキストのツリーを展開せずにある特定のカテゴリ (フォルダ) を選ぶか、コンテキストのツリーを展開してもあまり考えずにある特定のカテゴリ (フォルダ) を選ぶ。

以上のことから判断すると、ReSearchable を使ってブックマークを再検索する際に、一旦展開されたコンテキストのツリーを見てしばらく考えることがブックマーク再検索の効率向上につながると考えられる。

5. 関連研究

ブックマークの再検索の効率の向上を目的として従来のブックマークシステムに新要素を加えるアプローチによる研究について、体系的な三つの関連研究を紹介する。

CoWing[8]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[9]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

ブックマークの構造化と表示を支援する。

MyLifeBits project[10]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。

6. おわりに

本研究は、検索のコンテキストを利用してブックマークの再検索の効率を向上させる機能を設計し、設計に基づいたプロトタイプシステム ReSearchable を開発して、評価実験を行い、ブックマーク再検索の効率が向上することが確認された。

今後の課題は、ReSearchable とソーシャルブックマークシステムの連携と ReSearchable のインタフェースの改善が考えられる。また、現在の ReSearchable では、ブラウザによっては WWW 上存在する PDF などのファイルを正常にブックマークとして登録できない場合はある。この問題点を改善して、ブックマークの追加をさらにスムーズになるようにすることも課題である。

参考文献

- [1] Context in Web Search, Steve Lawrence, IEEE Data Engineering Bulletin (2000).
- [2] Re-Finding Found Things: An Exploratory Study of How Users Re-Find Information, Robert Capra, Manuel A. Pérez-Quinones, Journal Title: CoRR: Human-Computer Interaction, Date: 2003.
- [3] Context-dependent Web bookmarks and their usage as queries, Nakajima, S. Oyama, S. Sumiya, K. Tanaka, K., WISE2002.
- [4] The Re:Search Engine: Helping People Return to Information on the Web, NSF PIM Workshop. Jaime Teevan. 2005.
- [5] A Survey On Web Information Retrieval Technologies, Lan Huang, Tzi-cker Chiueh, SUNY Stony Brook ECSL technical report ECSL-TR-120, 2000.
- [6] Abrams, D., Baecker, R. and Chignell, M., Information Archiving with Bookmarks : Personal Web Space Construction and Organization. In Proceedings of CHI'98. Los Angeles, CA, 1998
- [7] Robert, J., Ruiz, J. and Lank, E. Making Favorites Useful, 2005.
- [8] Kanawati, R. and Malek, M., A Multi-Agent System for Collaborative Bookmarks, Proc. Conference on Autonomous Agents and Multi-Agent Systems 2002.
- [9] 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.
- [10] Bell, G., Gemmell, J., and Lueder, R. The MyLifeBits Lifetime Store, ACM SIGMM 2003 Workshop on Experiential Telepresence.
- [11] Wolber, D., and Brooks, C. Navigating the Personal Web, JCDL Conference'04, 2004.
- [12] Using Web Search Engines to Find and Refind Information. IEEE Computer 38(10): 36-42. iee-dl. Capra, R., and Pérez Quinones, M.A. (2005).
- [13] Lawrence, S. and Giles, C. L. Accessibility of information on the web, Nature, Vol. 400, No. 6740, pp 107-109, 1999.