

グループでウェブの探索を効率化する検索共有インタフェース

Improving the efficiency of web search by reusing search history of group members

武田 達弥 五十嵐 健夫*

Summary. When searching a web using a keyword, the search engine returns many results and the user needs to pick desired one manually by visiting the returned many pages one by one. We try to reduce this burden by reusing the result of this task by other group members. Our basic assumption is that group members are expected to share general interest. For example, if one member picks page A as an appropriate search result for search query X after examining many search results, it is likely that another group member also wants to get page A when submitting search query X. This paper proposes search history sharing system that allows the users to get page A immediately when the second group member submit query X to save the user from checking many search results. Finally, we discuss our preliminary test.

1 はじめに

近年、ウェブ上の情報が爆発的に増え、それらを効率的に入手するためのシステムや研究が急速的に発展し、検索画面や表示される結果のパーソナライズなど各個人の情報活動効率を高めるユーザインタフェースに関する技術が注目を集めるようになってきている。しかし、検索という作業は一人で行うものばかりではない。検索活動には、他の人と協力する場面や、協力出来るような場面がある。例えば、グループで取り組むプロジェクトに関連した情報やグループ旅行の目的地などを探す場合がそれに当たる。また、そのように明示的に協力しなくても、グループでは興味・関心が共有されることが多いため、個々人の検索活動には重複があり、グループのメンバー同士で協力できる可能性がある。もし他人と協力出来るならば、検索が下手な人でも上手な人の力を利用することで有用な情報に簡単にたどり着くことができるようになるなど、検索効率の向上が期待出来るだろう。

本稿では、そのようなグループでの検索活動を支援する手法の一例として、グループのメンバーの過去の検索履歴を用いる手法を提案する。さらに、作成したプロトタイプシステムおよび予備的な調査の結果について報告する。

2 関連研究

2.1 グループを対象とした検索支援

グループを対象に検索を支援した研究がいくつかある。[1] と [2] は、グループのメンバーが同期的に協

力して探すのを支援した研究である。[1] は、グループのメンバーが使ったクエリを表示することと検索結果中のどのページが既に他のメンバーに見られているかを表示することによって、 Awareness を提供し、重複して同じページを探すという無駄を省く。また、見たウェブページに評価をつけたり、他のメンバーにウェブページを推薦したりすることにより、検索効率を上げる。[2] は、グループで何かを決めるために検索を行う場合に、発見したページへグループのメンバーがどのようなルートでたどり着いたかを視覚化することで、発見したページがどのような経緯で発見されたかわかるようにし、発見した情報を用いたグループの意思決定を支援する。[3] は、非同期な協力を促した研究である。見た web ページに対して Annotations を残せるようにして、次にそのページを見た人が効率的に情報を得ることが出来るようにした。これらのシステムに対して、われわれのシステムは、より暗黙的に協力できるシステムを目指した。ここで暗黙的な協力とは、協力するにあたって通常の検索作業以外の特別な作業が必要なく、同じ時間に検索する必要が無く、いつの間にか協力できていることを指す。

2.2 協調フィルタリング

[4] のような協調フィルタリングを用いたシステムは他人の情報を用いて、有用な情報を推薦するシステムである。協調フィルタリングも暗黙的に他人と協力して情報を見つけれられるシステムであるが、協調フィルタリングは、膨大なユーザコミュニティの中から趣向の異なる人たちの中から趣向の近い人を見つけてそれを推薦するという方式であり、本論文は小規模のグループを対象としているので、その点で異なる。

Copyright is held by the author(s).

* Tatsuya Takeda and Takeo Igarashi, 東京大学大学院情報理工学系研究科 コンピュータ科学専攻

3 検索履歴の利用

本システムは、グループの検索を支援するために、グループのメンバの検索履歴を利用する。グループでは、興味・関心の方向性が近い場合、ウェブで求める情報も似ていることが予想される。例えば、研究室で共同研究をしているときの関連論文探し、出張や旅行をするときの計画立てなどウェブで同じ対象を求めることがあるだろう。そのようなとき、通常はメンバ毎に求めている情報を探す、つまり検索エンジンにクエリを打ち、結果として提示されたウェブページを順番に見ていくことになる。この作業は、求める情報によっては、労力と時間のかかる作業になることがある。そこで、グループの誰かが見つけた有効そうな情報を共有することが出来れば、他のグループメンバはその情報を利用することが出来、作業の効率が上昇することが予想できる。

具体的には、検索エンジンに打たれたクエリと、価値のあるページを結び付け、次にそのクエリが使われたときに表示するという方法で支援する。ユーザが何か検索しようとして、検索エンジンに情報を打ち込んだときに、システムによって共有されている価値のあるページが提示される。こうすることで、ユーザは本システムを使っていないときと変わらず通常通り検索しながら支援を受けることが出来る。

また、システムは価値のあるページを見つけたときに直接利用したクエリと結びつけるだけでなく、そのクエリの前に使われたクエリとも結びつける。これは、ユーザは検索するとき、いくらかのクエリを試行錯誤しながら検索エンジンに打ち込んで、ページを調べていくので、それらのクエリも結びつけることで、その検索タスク中に使われていたクエリのどれかが次に用いられれば、価値のあるページが表示されるということになる。こうすることで、グループ内のユーザ Alice が、query1 を使ってページを探したが見つからなかったため、次に query2 を使って、page3 という求める情報を発見したという場合に、それ以後、ユーザ Bob が query1 を打つだけで、page3 を見つけることが出来ると期待される (図 1)。これは、もし Bob の検索能力が未熟で query1 を打ったあとに欲しい情報を見つけないことが出来ず、query2 というクエリを思いつかなかったような場合でも page3 を発見できるということになり、有効であると考えられる。

3.1 技術的課題

この手法を実現するための技術的課題は二つある。一つ目は、検索履歴のページの内、どれがその人の探していたページで価値のあるページであるのかを判断しなければならないということである。二つ目は、価値のあるページを判断出来たとして、それを見つめるために入力されたクエリはどれであるのかということである。

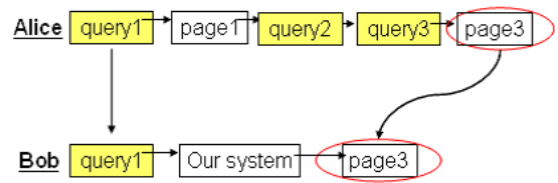


図 1. システムによる検索作業支援

3.1.1 価値のあるページの判別

グループのメンバの検索履歴から有用な情報を決める必要がある。検索履歴は、価値のある情報と価値のない情報からなる。我々は検索するとき、情報を発見するまでいくつかのウェブページを見て、それぞれが求めている情報かどうか判断している。求める情報のないページなら、その検索タスクでは価値がないページであろうし、求める情報があるページなら価値のあるページとなる。この二つのうち、グループのメンバに提示すべき情報は、価値のあるページである。今回、われわれは、価値のあるページの判定として、ユーザのブラウジング中の特定の操作を用いた。ユーザはブラウジングしているとき、見ているページをブックマークする、印刷する、ページやファイルを保存するなどの行動を取る。経験上、これらの行動は、価値のあるページに対してなされることが多いはずである。よって、ユーザのこのような行動が取られたページを価値のあるページであるとみなし、グループのメンバに提示することとする。

3.1.2 価値のあるページとクエリの結び付け

価値のあるページをクエリと結びつけて、ユーザに提示するためには、価値のあるページと結びつくクエリがどれなのかを判定しなければならない。図 3 で page3 が価値のあるページだとすると、直前に打たれた query2 は page3 を見つけるために打たれたと判定できそうだが、その前に打たれている query1 が page3 を見つけるために打たれたかどうかは容易には判定できない。

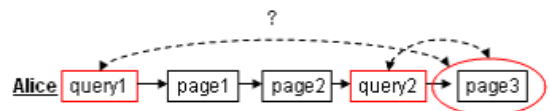


図 2. 価値のあるページとクエリの結び付け

われわれは、ユーザのブラウジングログから作成した決定木を用いて価値のあるページとクエリの結び付けの判定を行う。決定木の作成には、ユーザのブラウジング中の特徴的な活動を利用する。ユーザはブラウジング中にウィンドウを開いたり、ウェブ

メールを見たりといった活動を行っている。これらの活動がクエリと見つけられたページが結び付くかどうかを判定する手がかりとなりそうである。例えば、ウェブメールを見るということはひとつの作業であり、検索という作業と並列には起きることが少ないと予想され、検索作業の切れ目に起き易いかもしれない。価値のあるページとクエリとが結び付くかどうかを判定するために、検索作業の切れ目となりそうな以下の活動に注目した。

時間 あるクエリが打たれてから次のクエリが打たれるまでの時間である。間隔が長いと関連性が薄くなると予想される。

特定のページへの訪問 日常のブラウジング中によく訪問されるページがある。そのようなページが訪問されるということは検索活動がそこで途切れている可能性が高く、クエリ同士の関連性も薄くなりそうである。今回、そのような特定のページとしてウェブメールのページを採用している。

ウィンドウの開閉 ブラウザのウィンドウを開いたり閉じたりといった活動は、ブラウジングの終わりや始まりに起き易く、そこが検索活動の切れ目になりやすいと予想した。

これらの活動を独立変数とし、あるクエリと価値のあるページが結び付くかどうかを従属変数とする決定木を作成した。決定木作成の手順は以下の通りである。まずユーザのブラウジングログを集め、その中から価値のあるページを抽出する。そして、ひとつの価値のあるページに対して、直前に打たれたクエリ 10 個までをそのページと結び付くか結び付かないかを手動で判別し、正解集合のログを作成する。正解集合のログの各データは、あるクエリと価値のあるページとの関係を表す次のような形になる。

```
<時間>X(s)</時間>
<特定 URL 閲覧>true</特定 URL 閲覧>
<Window>true</Window>
<connect>true</connect>
```

〈時間〉は、クエリが打たれた時間とページが閲覧された時間との差を表し、〈特定 URL 閲覧〉はクエリが打たれた時間とページが閲覧された時間の間に、特定 URL が閲覧されていれば true となる。〈Window〉はクエリが打たれた時間とページが閲覧された時間の間に、ウィンドウの開閉が行われていれば true となる。〈connect〉はクエリがそのページを見つけるために打たれたものであれば true となる。そして、出来上がった正解集合を用いて、結び付く確率を返す決定木を作成する。決定木は、図 3 のようになる。

システムは、価値のあるページが見つかったと、こ

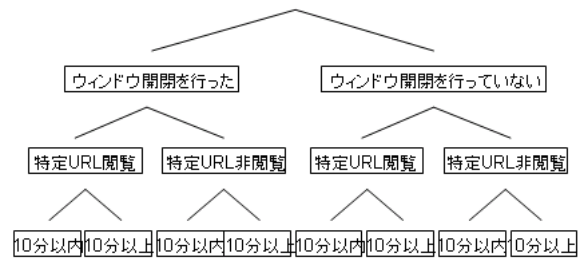


図 3. 決定木の例

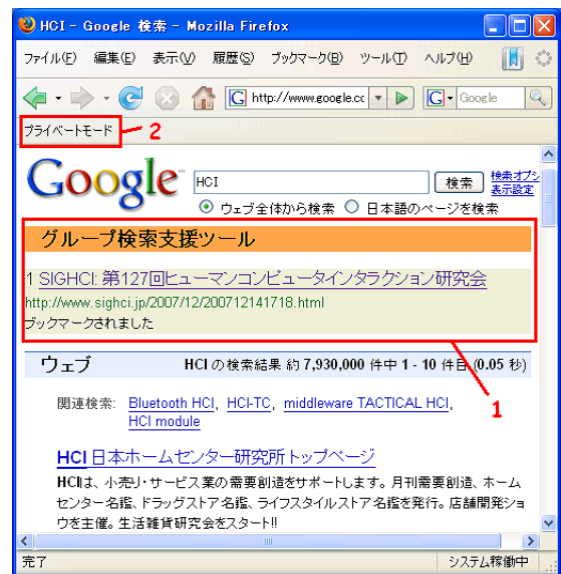


図 4. システム概観

の決定木を用いて、それ以前のクエリが結び付くかどうかを判定する。実際の運用では、予め複数人のログから作成された決定木を利用するので、ユーザが手動で特別な作業を行うことはない。また、今回の範囲では、個人間の検索活動の違いまで考慮していないが、個人間で検索の仕方が大きく異なり、他人のログから作られた決定木が利用できないとなれば、将来的にはいくつかの決定木を用意しておき、利用者にあう決定木を利用するなどということを考えていかなければならないだろう。

4 インタフェース

本システムの概観を図 4 に示す。本システムは、Firefox ブラウザの拡張として実装されている。ユーザが検索エンジン (現在は Google にのみ対応している) のページにクエリを打ち込むと、通常の結果の上に、システムが提示する価値のありそうなページが 10 件結び付く確率の高い順に表示される (図 4 の 1)。表示されるのは一つのページにつき、ページのタイトル、URL、それが選ばれた行動要因である。また、ステータスバーのボタン (図 4 の

2) を押すことにより、プライベートモードになり、システムへ履歴が送られないようにすることが出来る。もう一度押せば解除される。

本システムは、クライアントサーバモデルで、クライアントはFirefoxの拡張であり、サーバはRubyとMySQLで実装されている。クライアントであるFirefoxはウェブページへのアクセスとは別に本システムのサーバとも通信を行う。クライアントは、ブラウザ上の行動を監視し、クエリが打たれたり、印刷や、ブックマーク、ファイルの保存が行われれば、URL、どの行動かなどをサーバに送る。サーバ側は、それらを受け取り、クエリと価値のあるページの結びつけを行い、データベースに保存する。また、クライアントは、検索エンジンのページでクエリが打たれれば、そのクエリをサーバに送り、そのクエリと結び付けられているサーバが保持している価値のあるページのデータを受け取る。

5 評価

本手法がユーザの検索活動を適切に支援するかを調べるために2種類の予備的な調査を行った。一つ目は、他人の検索履歴は、検索の支援に有用なのかどうかを調べる目的で行い、二つ目は、価値のあるページとクエリを結び付けるアルゴリズムがどの程度正確に結び付けるのかを調べる目的で行った。

5.1 検索履歴の有用性の検証

他人の検索活動がユーザの検索活動の支援に役立つのかを調べるために、検索課題を用意して、特定の時間中検索課題に取り組むという短期集中型のユーザスタディを行った。ユーザスタディの場所は、筆者が所属する大学の研究室で行った。参加者は6人で、すべてコンピュータサイエンスを研究している学生である。また、この評価の目的は他人の見つけた情報が表示されたときに検索の支援になるのかどうかを調べることに絞り、価値のあるページの判定と価値のあるページとクエリの結び付けのアルゴリズムは評価の対象外とした。そのため、それらのアルゴリズムが巧く動いた場合を想定し、理想的な環境で実験を行った。具体的には、課題中に良いページや目的となる情報を見つけたときは、意図的にダウンロードやブックマークを取ってもらうように指示し、それらのページはその時間中に打たれたクエリすべてと結び付くようにするという事を行った。

5.1.1 実験課題

実験課題として、2種類の課題を用意した。これらの課題には、われわれが日常行っている検索タスクを模倣したものを2種類用意した。それぞれ課題A、課題Bとする。

課題 A

課題Aは、波の音が入っている音声ファイルをGoogleで検索して集めるという課題である。本課題は、ウェブ上で探したいものが明確に決まっている場合に、本システムがユーザの検索活動を支援するかを調べるのが目的で行った。また、検索中は、音声ファイルを発見したら、PCに保存を行ってもらった。

課題 B

課題Bは、研究室で行うクリスマスパーティでのプレゼント交換用に、千円から2千円のプレゼントの候補を、探すという課題である。本課題は、目的はあるが、探す対象が曖昧で、ウェブを見ながら考えをまとめるというような場合に、本システムがユーザの検索活動を支援し、考えをまとめやすくなるかどうかを調べるのが目的で行った。また、見つけたプレゼントの候補や、プレゼントを考える上で役に立ちそうなページを見つけた場合に、ブックマークを取ってもらった。

5.1.2 手順

まず一人一台のPCを割り当て、Firefoxのインストールとクライアントであるブラウザ拡張をインストールしてもらい、システムと実験課題を簡単に説明した。次に6人を3人一組の2つのグループに分け、これをグループ1とグループ2とした。その2つのグループに20分間通常システム支援なしの検索、10分間休憩、20分間システム支援有りの検索という順番で課題を行って貰った。その際、システム支援有りの検索では、通常検索で作成された履歴をシステムが利用して支援を行った。課題の割り当ては、通常検索とシステム支援有りの検索で違う課題になるように表1のように割り当てた。これは学習効果などの影響を受けないようにするためであり、グループ2が課題Aを行うときには、グループ1が課題Aをシステムの支援なしに通常通り検索したときの履歴によって支援を受けることになる。グループ1はその逆である。

表 1. 課題の割り当て

| | 課題 A | 課題 B |
|-------------|--------|--------|
| 通常検索 | グループ 1 | グループ 2 |
| システム支援有りの検索 | グループ 2 | グループ 1 |

5.1.3 結果

実験中に取られたログとユーザの行動と作業後のアンケートの内容を議論する。表2は、システムの

提示したページは参考になったかという質問に対する回答である

まず、課題 A の結果を分析する。課題 A では被験者によって 92 種類のクエリが打たれ、そのうち 12 種類のクエリで、システムは他人の発見した価値のありそうなページを提示した。システムの提示したページは参考になったかという質問に対し、課題 A に参加した 3 人ともはいと答えた。理由として、「提示されたページから音声ファイルをダウンロードすることが出来、楽であった。」ということが上げられた。実際、被験者は音声ファイルをダウンロード出来るウェブサイトを 15 ページ発見したが、そのうち 4 ページはシステムからの支援で辿り着かれたものであり、システムは被験者の情報の発見を手助けしていた。次に、システムによって提示された結果は、通常の検索結果に比べてどうかという質問には、「洗練されていたと感じた。」という意見があった一方で、提示された結果には「良いときはとても良いが、タスクと関係ない結果が混じっていたと感じた。」という意見もあった。課題と関係ない結果が混じっていた原因として、課題 B を行っている被験者が間違いで課題 A 向けのクエリを打ってしまったということが上げられる。それが原因で一部課題 A と課題 B 向けの結果が混じた。今回は間違いのせいで混じたとはいえ、システムを長期的に利用した場合、関係のない結果が混じってしまうことは多いに有り得る。今後、関係のない結果が混じってしまった場合のことも考えて多少ノイズが混じっていても気にならないインタフェースなどを考えていった方が良いかもしれない。

課題 B の結果について分析する。課題 B では被験者によって 64 種類のクエリが打たれ、そのうち 9 種類のクエリで、システムは他人の発見した価値のありそうなページを提示した。システムの提示したページは参考になったかという質問に対し、課題 B に参加した 3 人中 2 人がはいと答えた。はいと答えた被験者の意見として、「自分の思いつかなかったプレゼントを発見することが出来、考えが広まった。」「他の人と同じプレゼントを選ばないように気をつけることが出来た。」という意見があった。いいえと答えた被験者の意見は、「システムから他の人の見つけたページが支援されなかった。」ということであった。これは、その被験者の用いたクエリが他のメンバの用いたクエリと重複することがなく、支援を受けることが出来なかったということが原因であり、他のメンバが発見したページをクエリから提示する部分のアルゴリズムの問題点が示唆された。クエリが完全一致しないと他のメンバの発見したページが提示されないというのは、今後の改善点のひとつであることが分かった。次に、システムにより提示されたページは通常の検索エンジンの結果に比べてどうだったかという質問に対する意見として、「他

の人の選んでいる内容を見ることが出来たのは面白く、そればかり見てしまった。」という意見があった。本システムだと、通常の検索結果と違い、人間の作為で選ばれているので面白いと感じるのかもしれない。

結果をまとめると、他人の発見したページが見られることは、情報検索と思考の支援に繋がる可能性があることが分かった。一方で、インタフェースに対する改善点も示唆された。

表 2. システムの提示したページは参考になったか

| | はい | いいえ |
|------|----|-----|
| 課題 A | 3 | 0 |
| 課題 B | 2 | 1 |

5.2 クエリとページの結び付け手法の評価

クエリと価値のあるページ結び付けの手法の正答率を調べるために、コンピュータ科学を専攻するユーザ 1 人の 3 ヶ月分のブラウジングログを用いて評価を行った。手順はまず、ログ中に存在した 40 個の価値のあるページと、そのページそれぞれの直前に打たれたクエリ 10 個、合計 400 個のクエリに対して、結び付くか付かないかを手動で判定し、正解集合を作成する。その後、それらを半分に分け、半分を決定木の作成に用い、残り半分でその決定木の評価を行うという方法を取った。さらに、より単純な手法である価値のあるページが閲覧された時間から一定時間内に打たれたクエリを結び付ける手法での正答率との比較も行った。

5.2.1 決定木

3.1.2 の手法で、時間、特定 URL 訪問の有無、Window のクローズとオープンの有無を独立変数とし、ページとクエリが結び付くかどうかを従属変数とした決定木を作成した。なお、連続変数である時間は、1 分以内、1 分以上 5 分以内、5 分以上 10 分以内、10 分以上 30 分以内、30 分以上 1 時間以内、それ以上と分けることで、離散化している。また、作成した決定木の出力は、結び付いた件数 x 件、結び付かなかった件数 y 件となり、入力に対してどの程度結び付きそうかという確率を得ることが出来る。

5.2.2 結果

決定木の出力として、ある閾値以上の確率が得られれば結び付き、それ以下なら結び付かないと判断する。閾値による決定木の正答率は図 5 のようになった。

図 5 の「結び付く場合」の正答率は、価値のあるページとクエリが結び付くことが正解な場合に、結び付くと判定できたかどうかを表し、「結び付かな

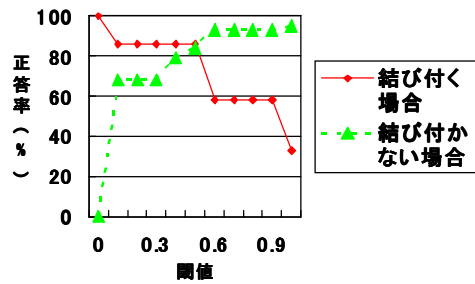


図 5. 本手法の正答率

い場合」の正答率は、価値のあるページとクエリが結び付かないことが正解な場合に、結び付かないと判定できたかどうかを表す。閾値を 0.5 にした場合、結び付く場合の正答率 86% となり、結び付かない場合の正答率 84% となった。

一方、より単純な手法である価値のあるページの閲覧から一定時間以内に打たれたクエリを結び付けるという手法の結果は、図 6 のようになった。一定時間を 5 分以内、10 分以内、30 分以内、1 時間以内と変えた結果を図にしている。30 分以内に打たれたクエリを結び付けるとすると、結び付く場合の正答率 83% となり、結び付かない場合の正答率 91% となった。

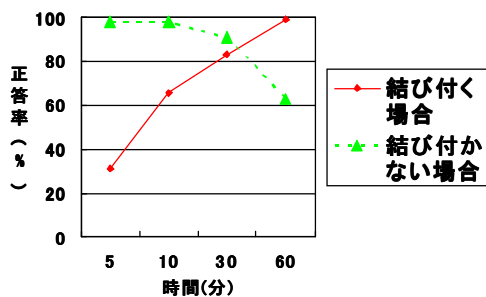


図 6. 単純な手法の正答率

結果、本手法とより単純な一定時間内のクエリを結び付ける手法の正答率は、さほど変わらないことが分かった。この理由としては、クエリの結び付くか否かの判定では時間の影響が大きいと、時間、ウィンドウの開閉、ウェブメールの閲覧だけでは時間だけを見る単純なアルゴリズムと大して変わらない結果になってしまうということが考えられる。今後、検索活動の切れ目を探す鍵となるユーザのブラウジング中の操作を探し出すなどしてより良いアルゴリズムを考えていく必要があるだろう。

6 まとめと今後の課題

本研究では、グループの検索履歴を用いて、グループでのウェブの探索を効率化する手法を提案した。また予備的な調査を行い、他人の検索履歴は、グループの検索支援に有用であるのかとクエリとページの結び付けのアルゴリズムの正確さについて調査した。結果、通常通りに検索するのに比べて、本手法による支援を受けると、情報を楽に発見できる可能性とユーザの思考を広げる可能性があることが示唆された。一方で、インターフェースに対するいくつかの改善点が明らかになった。また、クエリとページの結び付けアルゴリズムの評価では、現段階では、単純なアルゴリズムを用いた場合と同じ程度の結果であることが分かった。今後、検索の切れ目の判断になるユーザの活動を探しページとクエリの結び付きの精度を上げることが課題である。

次の課題として検索活動の個人差が挙げられる。本稿では考慮していなかったが、個人差を考慮してアルゴリズムを改善していく必要があるだろう。

また、一番の課題は、長期的な評価を行うことである。本稿では、数名の被験者が同じ課題に同時に取り組むという現実的にはほぼあり得ないような状況下での調査であった。長期的な評価を行うことで、現実の環境でシステムが有効であるのか調べなければならぬだろう。さらに、長期的なユーザスタディを行うことで、価値のあるページの発見アルゴリズムの評価や、クエリとの結び付けを行うアルゴリズムによって混ざる関係の無いページの煩わしさがどの程度になるかなどの評価が行えるはずである。今回の結果を踏まえ、今後長期的な評価を行っていくつもりである。

参考文献

- [1] Meredith Ringel Morris and Eric Horvitz. Search-Together: An Interface for Collaborative Web Search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, 2007
- [2] 伊豆 陸, 中島伸介, 田中克己. グループ支援型 Web 閲覧における閲覧履歴の視覚化と共有. *日本データベース学会論文誌 (DBSJ Letters)*, Vol.3, No.1, pp.121-124, 2004
- [3] Romano, Nicholas C. and Roussinov, Dmitri G. and Nunamaker Jr., Jay F. and Chen, Hsinchun. Collaborative information retrieval environment: integration of information retrieval with group support systems. In *Proceedings of the 32nd Annual Hawaii International Conference*, 1999
- [4] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom and John Riedl. and Chen, Hsinchun. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of the 1994 Computer Supported Collaborative Work Conference*, pp.175-186, 1994