

## 個人プロフィールと検索履歴を用いた WWW 情報検索支援モデルの提案

臼井 大介<sup>†</sup> 塚本 享治<sup>†</sup>

東京工科大学大学院メディア学研究科

### 1 はじめに

キーワードに基づく WWW サーチエンジンではユーザが得たい情報を取得するのが困難な場合が存在する。特に専門性の高い情報や、それほど有名ではない Web サイトを検索する場合に、2 つ以上の検索語を同時に用いないと、ユーザの得たい情報が検索結果の上位に上がってこないことが多い。ユーザがインターネットの利用に熟練していない場合や、検索対象の知識を十分に持っていない場合に、適切な検索語を思いつくのは困難である。このような問題を解決する手がかりとして、次の 4 つの情報が利用できると考えた。

入力した検索語と関連して、他のユーザが

- ・ 利用した検索語
- ・ 利用した Web ページの URL
- ・ ブックマークに追加した Web ページの URL
- ・ ブックマークに追加する際のフォルダ名

そこで、これらの情報の利用して、ユーザの情報検索を支援する、WWW 情報検索支援モデルを提案する。

### 2 関連研究

関連研究として、特定サーバ上で不特定多数のユーザの検索履歴を収集し、その情報を元に関連する検索語をユーザに提示する手法[1]や、ユーザ間でブックマークを共有し、現在訪れている Web ページを元に、そのページに類似したページを他のユーザのブックマークの中から探し提示する手法[2]、ユーザ間でブックマークを共有し、利用者の類似度とブックマークの階層構造の類似度を判定し、協調フィルタリングに基づいて Web ページを推薦する手法[3]等がある。

しかし、これらの関連研究では検索履歴がブックマークのどちらか一方しか用いておらず、またユーザの趣味や職業といった属性を考慮していない。

### 3 WWW 情報検索支援モデル

#### 3.1 提案モデル概要

提案モデルは、各ユーザが個人プロフィールを設定した P2P ネットワーク上で、検索履歴とブックマークを共有し、ユーザの入力した検索語と関連する情報を他のユーザの検索履歴、ブックマークの情報から収集して提示するものである。

#### 3.2 個人プロフィールの利用

インターネット利用調査[5]によれば、インターネットユーザは趣味に関連する Web ページを閲覧することが多いと報告されている。そのため、ユーザは自分の趣味に関する Web ページを頻繁に検索したり、ブックマークに登録しており、自分の趣味に関する専門的な知識をある程度有していることが多いと考えられる。また、趣味だけではなくユーザの職業や年齢も利用する Web ページに影響を与えると予測する。そこで、本研究ではユーザの属性情報(個人プロフィール)として以下の 3 つを用いる

- ・ 年齢
- ・ 趣味
- ・ 職業

これにより、ユーザは自分の入力した検索語がどのような年齢、趣味、職業のユーザでよく用いられているかという情報が得られると同時に、これらの属性情報を用いることでユーザ自身による検索の絞込みを効率的に行えるようになると考えられる。

#### 3.3 ユーザのマシン上での処理

提案システムでは検索履歴とブックマークをユーザごとにローカルマシン上の適当なデータベースに保存し、そのデータベースの情報を P2P ネットワーク上で共有する。また、データベース内の情報の関連を構造化するために[4]で提案された Query Network(略して QN)を用いて検索履歴とブックマーク情報を構造化し、この情報も他のユーザと共有する。

検索履歴を収集する際は、特別なブラウザを用意し、ユーザがブラウザを通してサーチエンジンによって検索を行う際の、検索語と検索結

A Proposal of Information Retrieval Assistance Model using Personal Profile and Retrieval History.

<sup>†</sup> Tokyo University of Technology, Graduate School of Media Science

果の中からユーザがクリックした URL の組をデータベースに格納していく。またブックマークも同様に、ユーザが興味を持った Web サイトをブックマークに追加するたびに、その URL とブックマークのフォルダ名をデータベースに追加する。次に、データベースに格納された情報を元に QN を構築する。QN とは[4]で提案された手法で、ユーザ(行為者)、検索語、検索結果中から選択した Web ページの URL という 3 つの属性データの相互関係をグラフ構造(無向グラフ)によって表現する手法である。図 1 に QN の最小単位グラフを示す。

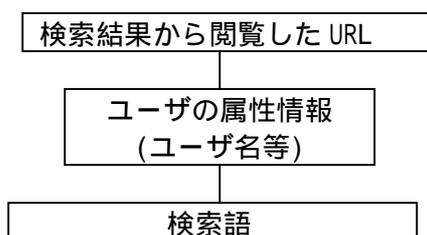


図 1: QN の最小単位グラフ

[4]では、上記 3 つの属性のみを用い、また検索語は一語のみ用いているレコードのみ処理対象としていたが、本研究では QN に 3.2 で述べた個人プロフィールを追加し、さらに検索語も 2 語以内のものを処理対象とする。例えば、「2005 年 1 月 1 日 2 時 15 分 0 秒に、職業がシステムエンジニア (SE) であり、趣味がコンピュータである John という 25 歳のユーザが「サーチエンジン」、「グーグル」という 2 つの検索語を用いた検索結果から <http://www.google.co.jp/> という Web ページを閲覧した」という事実がデータベースに「2005/01/01 02:15:00 SE コンピュータ John 25 サーチエンジン グーグル <http://www.google.co.jp/> 」というレコードとして記録される。このレコードを QN の最小単位グラフに対応させると図 2 のようになる。

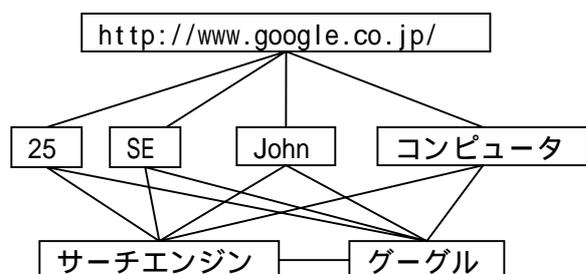


図 2: 本研究で用いる QN の例

QN により、検索語や URL を次々と接続していくことで、関連する検索語やブックマークのフォルダ名、URL がグラフ上で接続する。

### 3.4 P2P ネットワーク上での検索

P2P ネットワーク上の検索では、検索対象の絞り込みを行えるよう 2 段階の方法で行う。第 1 段階ではユーザが検索語を入力すると、その検索語を P2P ネットワーク上にマルチキャストし、他のユーザのデータベースから検索語を含むレコードをユーザに返す。第 2 段階は、第 1 段階で取得した結果の中からユーザが個人プロフィールによるフィルタリングをかけ(例えば職業は研究者であるユーザにのみ検索対象とするなど)、それらのユーザの QN から、目的の検索語と隣接する情報を返すというようにする。こうすることで、第 1 段階では得られなかった、検索語を一切含まないが、QN によって隣接している検索語や URL の情報を得ることが可能になる。

### 3.5 提示された情報の利用

ユーザは本モデルにより提示された検索語や URL を利用することで、より専門性の高い Web サイトを発見したり、得たい情報を取得するための適切な検索語を発見できるようになると考える。

## 4 課題

3.2 の個人プロフィールが検索を行う際に有用であるかどうか検証・評価する必要がある。また、P2P ネットワーク上での検索は検索時間とネットワーク負荷が大きい点も課題である。

### 参考文献

- [1]原田昌紀, 清水奨, "WWW 検索システムにおける不特定多数の操作履歴の活用", 情報処理学会研究報告, 97-OS-74/97-DPS-81, pp. 61-66, 1997
- [2]森幹彦, 山田誠二, "ブックマークエージェント; ブックマークの共有による情報検索の支援", 電気通信学会論文誌 J-83-D-1, pp. 487-494, 2000
- [3]佐保田圭介, 波多野賢治, 宮崎純, 古川正俊, 植村俊亮, "ブックマークの階層構造情報を組み込んだ強調フィルタリングによる Web ページの推奨手法", 電子情報通信学会, DEWS2004, 6-B-4, 2004
- [4]佐藤進也, 原田昌紀, 風間一洋, "検索手法可視化の一手法", 情報処理学会報告 DBS-130/FI-71, 2003
- [5]インターネット利用調査結果報告 1996 年 11 月 <http://www.ntv.co.jp/bekkoame/>