

ブックマークエージェントによる URL の 協調的情報フィルタリング

Information Sharing of URL Using Bookmark-agent

森 幹彦
Mikihiko Mori

山田 誠二
Seiji Yamada

東京工業大学 大学院 総合理工学研究科
Tokyo Institute of Technology

An idea of bookmark-agent was studied on the basis of the file of bookmarks, which is registered user's preference URL addresses and is one of derivatives of the cooperative information sharing. The bookmark-agent we studied is more effective in searching information than existing search engines such as yahoo, altavista and so on. When an user tries to find for certain URLs (i.e. the home pages which is concerned with artificial intelligence) by browsing hyperlink, the agent begins to search for similar web pages to what is dealt with by user at present. At the same time, the agent requests for other agents to search for their own bookmarks, because one agent is stood up as one user in a small scale group, and can correspond with each other. As a result, the user can obtain similar pages by his agent as hyperlinks on his web browser. The information which he got is filtered beforehand and reduced.

1 はじめに

インターネットの急速な広がりとともに、ユーザは膨大な情報の渦中におかれるようになった。そこで、ユーザは欲しい情報を得るために多大な労力を必要とする。このため、検索エンジンなどの開発も進められているが、必要以上の情報を与えられてしまうのが現状である [2][3]。本研究では、比較的小規模のグループにおいて、ユーザが個々に持つ既にフィルタリングされた URL に関する情報であるブックマーク¹を必要に応じて共有することにより、ユーザの情報検索を支援するシステムを開発する。

我々はブックマークエージェントと呼ばれるシステムを提案する。すなわちブックマークエージェントと呼ばれるプログラムがユーザのブックマークを参照し、さらに他のエージェントとの通信により、他のユーザのブックマークの URL も参照して、現在ユーザのしている Web ページと類似したページを提示する。このように、ブックマークエージェントは、ユーザの明示的要求を必要とせずに、ユーザの欲する URL 情報を提示できる。また、ユーザは、ブック

マークエージェントに、明示的にキーワードを提示することも可能である。

ここで、留意して欲しいことは、ブックマークエージェントが類似ページを検索する範囲は、興味がある程度共通していると想定される小規模なグループのユーザのブックマークに限定されていることである。このことは、あるキーワードによる検索において、共通の興味という文脈を利用した、絞り込みが行われていることを意味し、一般の検索エンジンよりも効果的な情報フィルタリングが期待できる [1]。

また、この分野の先行研究 [4][5] では、ユーザが次にたどるであろうページの予測の学習に主眼が置かれているが、本研究では、グループ間の情報共有に主眼を置いている。

2 ブックマークエージェント

ブックマークエージェントのシステム概念図を図 1 に示す。

ブックマークエージェントは、1 ユーザあたりに 1 つ起動される。つまり、ブックマークエージェントは、一人のユーザに対し情報検索支援を行う。ブック

¹一度見た興味ある Web ページを記録しておくファイルのこと。

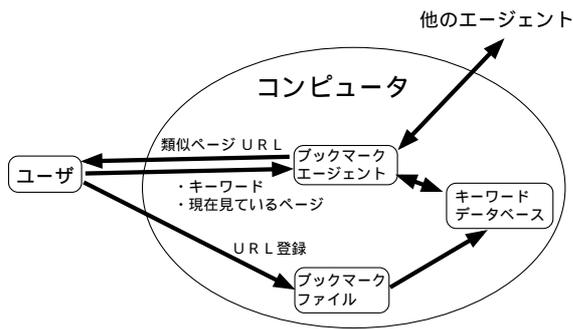


図 1 ブックマークエージェントのシステムの概念図

ブックマークエージェントの機能を以下に挙げる。

- 新規 URL のキーワード獲得：ユーザーのブックマークを監視し、新たな URL の追加があった場合、その URL の HTML ファイルからキーワードを抽出して URL データベースに追加する。
- 類似ページの提示：ユーザーが現在見ている Web ページからキーワードを抽出し、それを元にそのページの類似ページの URL をユーザーに提示する。さらに、ユーザーが明示的にキーワードをブックマークエージェントに与えて、類似ページを提示させることも可能である。ここで、類似ページとは、後述する類似度があるしきい値以上のページを意味する。類似ページの検索方法は、以下の 2 通りがある。
 - ブックマークエージェントが担当しているユーザーのキーワードデータベースを検索する。
 - あらかじめ設定されている範囲で、他のブックマークエージェントに URL 情報検索を依頼する。依頼されたブックマークエージェントは、担当ユーザーの URL データベースを検索して、見つかった類似ページを返す。
- 検索依頼の受領：他のブックマークエージェントから URL 情報検索の依頼があった場合、担当ユーザーの URL データベースで検索を行い、類似ページを返す。

3 システム構成

図 2 に、システム構成を示す。

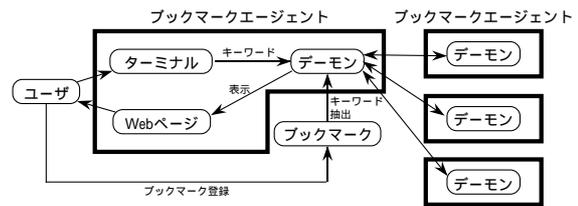


図 2 ブックマークエージェントのシステム構成

ブックマークエージェントは 3 つの部分に分かれる。

- ターミナル：ユーザーの明示的キーワード等のコマンドをエージェントに伝える役割を担う。エージェントからのコマンドの実行結果やエラーも表示される。また、ユーザーのエージェントの使用開始、終了の指示をする。
- デーモン：エージェントの本体。類似したページの検索、ブックマークからのキーワード抽出、エージェント間通信を行うプログラムである。各ユーザーに 1 個ずつ起動される。デーモンには以下の 2 種類がある。
 - 親デーモン：子デーモン同士の交信の中継をする。子デーモン同士を直接つなぐ煩わしさの解消のため。
 - 子デーモン：実際のデーモンとしての活動を行う。単にデーモンと呼ぶ場合はこれを指し、ユーザー毎に起動される。
 複数の子デーモンは、1 つの親デーモンに星状に接続される。
- Web ページ：類似ページのリストを表示する。

3.1 コマンド

コマンドは、ユーザーがターミナルからデーモンに命令を与える時に使用される。以下のようなコマンドがある。

- keyword キーワードを入力する。
keyword keyword1 keyword2
- stat ステータスを表示する。
- connect 子デーモンを親デーモンにつなぐ。
- disconnect 子デーモンを独立で動かす。

3.2 ターミナル

ターミナルには、以下のような機能がある。

- コマンド入力：UNIXのシェルコマンドのように入力を行う。しかし、ほとんどの入力がキーワード入力と考えられるので、キーワード指定コマンドは他のコマンドとは差別化しなければならない。
- コマンドの実行結果の表示：コマンドを実行した結果を表示する。keyword コマンドの場合は Web ページに直接表示されるが、他のコマンドの結果はターミナルに表示される。
- デーモンからのエラー表示：デーモンからのエラーを表示する。何らかの理由でデーモンが本来の実行できなかった時にその旨を表示する。

3.3 デーモン

デーモンには、以下のような機能がある。

- コマンドの実行：ターミナルから送られてきたコマンドを実行する。結果は、ターミナルまたは Web ページに表示される。
- キーワード抽出：ユーザの現在見ているページを監視し、そのページのキーワードを抽出する。ここで抽出されたキーワードは、次の類似ページの提示に用いられる。
- 類似ページの提示：キーワード抽出された Web ページと次に後述する方法で構築したキーワードデータベースに登録されたキーワードとを比較し、類似ページを検索する。検索結果はエージェントの Web ページに表示する。
- エージェント間通信：ユーザのコマンドは、場合によっては他のエージェントに伝えられる。例えば、明示的なキーワードや暗黙のキーワード抽出でのキーワードは、他のエージェントに参照される。もし、他のエージェントからキーワードの検索依頼があった場合は、検索しその結果を返信する。
- キーワードデータベースの構築：ブックマークエージェントは、ユーザのブックマークを参照し、その URL 情報を用いて検索を行う。データベースを構築するタイミングは、デフォルトの指定と、オプション指定で併用する。データベースには URL とそのキーワード、重みが登録されている。また、この URL に対する他のユーザのアクセス権も登録する。

3.4 Web ページ

Web ページは、キーワード検索の結果やユーザの現在見ているページの類似ページを表示する。また、ユーザ自身のブックマークを容易に見ることができるように、ブックマークファイルにリンクを張っている HTML で書かれている。

3.5 プロトコル

プロトコルのメッセージには、以下のようなものがある。

- keyword：キーワードを与える。
- result：コマンドの結果を返送する。
- stat：OK または NG。メッセージに対してその返事をする。
- cancel：先に送ったメッセージをキャンセルする。時間がかかりすぎる場合に用いられる。
- message：後に続く文字列をメッセージとして送る。デーモンがターミナルに対して実行結果を表示させるために送るときなどに用いられる。
- depend：要求を行う。接続要求はソケットで行うためこのプロトコルには関係ない。このメッセージは主にユーザ ID やユーザ名の問い合わせに使う。

3.6 キーワード抽出と類似度

テキストからのキーワード抽出は、出現頻度を用いるものを始めとして、さまざまな方法が提案されている [6] が、HTML という構造化された文書を扱うことから、本研究では、以下に述べるようなタグ構造に注目したキーワード抽出を用いている。

HTML ファイルにおいて、下記のタグ中に含まれる単語（名詞）に対し、括弧内の重みを加算しているとき、その重みの上位 5 つをその URL のキーワードとする。

• <META>(10) • <TITLE>(10)
 • <Hn>(6, 4) • (1) • <U>(1)

また、ページ間の類似度は、それぞれのキーワードの積集合の要素数とする。よって、類似度は、1 から 5 の値をとる。

4 実装

現在のところ、ブックマークエージェントは、UNIX マシン上に Perl を使って実装されており、WWW ブラウザとしては、Netscape Navigator™ を用いている。

4.1 ターミナル

ユーザからのコマンド入力をプロトコルを用いてデーモンに送る。デーモンからのメッセージを表示する。

4.2 デーモン

デーモンは、親デーモンとそれに星状に接続する子デーモンとに分かれる。親デーモンと子デーモンの大きな違いは、子デーモンが1つのターミナルと1つの親デーモンに接続するのに対して、親デーモンは複数の子デーモンと接続することである。

子デーモンは、そのデーモンを利用するユーザによって起動される。親デーモンは常時起動され、子デーモンの接続を待つ。

親デーモン

親デーモンは、子デーモン同士のコミュニケーションの仲介を行う。子デーモン同士は直接通信を行わない。子デーモンは、検索依頼などを親デーモンに送る。親デーモンは、送ってきた子デーモン以外の子デーモンにメッセージを送る。子デーモンからのコマンド実行結果は、適当な子デーモンのために中継する。

子デーモン

子デーモンの動作は以下の通りである。

- 子デーモンは、ターミナルからのメッセージと親デーモンからのメッセージを処理する。
 1. 子デーモンは、何らかの理由で親デーモンが接続を拒否したとき、あるいは、親デーモンに対する接続の失敗または親デーモンの終了により親デーモンが存在しないときは、独立して起動される。この場合はデーモンの参照領域は、その子デーモンの起動されたマシン内のみ限定される。また、起動後に親デーモンとの接続

が切れた場合は独立型で動作を続け、親デーモンの起動を待って接続する。

2. ターミナルからの検索依頼は、親デーモンに伝える。また、親デーモンからの依頼を受ける。
 3. ユーザが Netscape Navigator を起動していない場合は、それを起動する。ブラウジング画面にブックマークエージェントの Web ページが表示されていない場合は新しいブラウザを開いて、表示する。
 4. ユーザが新しい Web ページを閲覧するごとに、そのページの類似ページをエージェントの Web ページに表示する。類似ページを提示するために、キーワードデータベースを用いてキーワード検索を行う。
 5. ユーザからのコマンドに対しては、その結果を送る。
- ブックマークファイルの更新にしたがって、キーワードデータベースを構築する。キーワードデータベースを構築するために、以下の作業をする。
 1. ブックマークファイルのどの部分が追加、削除または変更されたのかを知るためのブックマークデータベースを作成する。このデータベースは、ユーザのプライバシー保護のために他のデーモン（ユーザ）には公開しない URL の指定ができる。フォルダ単位でこの指定を行うこともできる。
 2. ブックマークデータベースを参考にしながらキーワード抽出を行い、キーワードデータベースを構築する。
 3. キーワードデータベース構築のタイミングは、crontab の形式で指定することができる。

5 実行例

実行例を示す。ここでは、ブックマークファイルに、人工知能関連の国際会議のホームページが30個程度登録されている。

図3は、ユーザの現在見ているページである。図4は、ブックマークエージェントの Web ページであり、ユーザの現在見ているページの類似ページを提示している。この提示例では、ユーザは "Second International conference on Multiagent Systems ICMAS

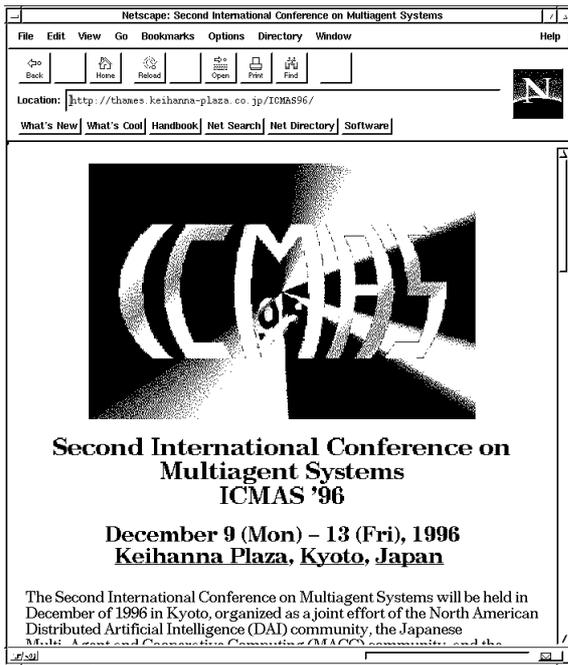


図 3 ユーザの現在見ているページ

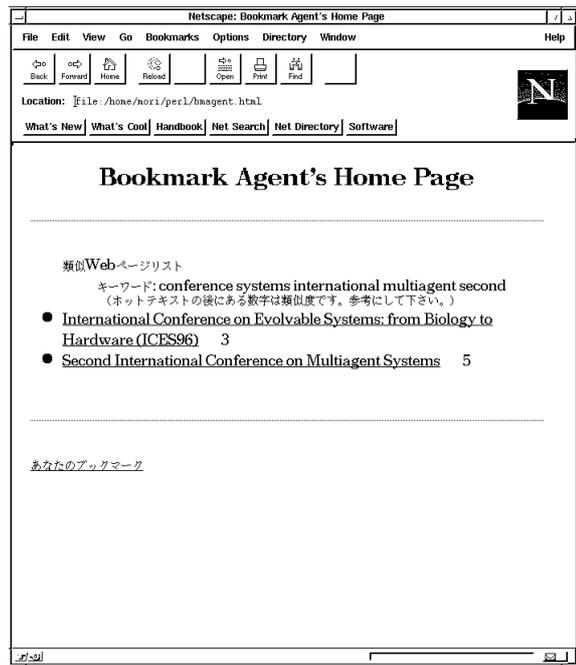


図 4 ブックマークエージェントによる提示

'96" のホームページを閲覧しており、その類似ページとしてエージェントは "International Conference on Evolvable System(ICES '96)" と "ICMAS '96" 自身を提示しており、直観的にも類似したページを提示しているのがわかる。

次に、ユーザが明示的にキーワードを与えた場合の例を図5に示す。ここでは、キーワードとして 'agent' を与えている。一方、ユーザ登録型検索エンジンの例として yahoo に、Web ロボットによる自動検索型検索エンジンの例として goo に同じキーワードを与えたのが、図6, 7である。

ここで注意すべきは、キーワード 'agent' に対して yahoo が 1683 個のサイトを、goo においてはなんと 594733 個ものサイトを示しているのに比べ、ブックマークエージェントは、3 個のサイトしか示していないということである。yahoo や goo では確かに、多くの必要とされるサイトが登録されているが、そのことはあまりに膨大なサイトの中から、ユーザの要求に見合ったサイトをユーザ自らが捜し出さなければならないことをも意味している。また、例えば研究のための調査を考えたとき、旅行代理店や会計業務代行等についての URL の提示は、かえって本来の情報を埋もれさせ、必要な情報にたどり着くことを阻んでいる。

一方、ブックマークエージェントは、エージェント関連の研究目的のグループに属していれば、それ以外の旅行代理店の URL といった余計な情報が入り込むことは少なくなり、ユーザは目的の情報に素早くたどり着くことができることは明らかである。

6 まとめ

本報告では、URL 情報であるブックマークを用い、情報検索を支援するブックマークエージェントを提案した。ブックマークは、ユーザによってあらかじめフィルタリングされた URL 情報であると考えられ、その情報を用いているブックマークエージェントは、単なる探索エンジンとは違う、適度にフィルタリングされた情報をユーザに提供できることが期待される。

現在、ブックマークエージェントは、開発中である。今後、実際に使用してみて、エージェントの提示する URL が、ユーザの要求する URL に見合ったものか否かを調査する必要がある。

参考文献

- [1] Rucker, J., Polanco, M. J.: Sitieseer: Per-

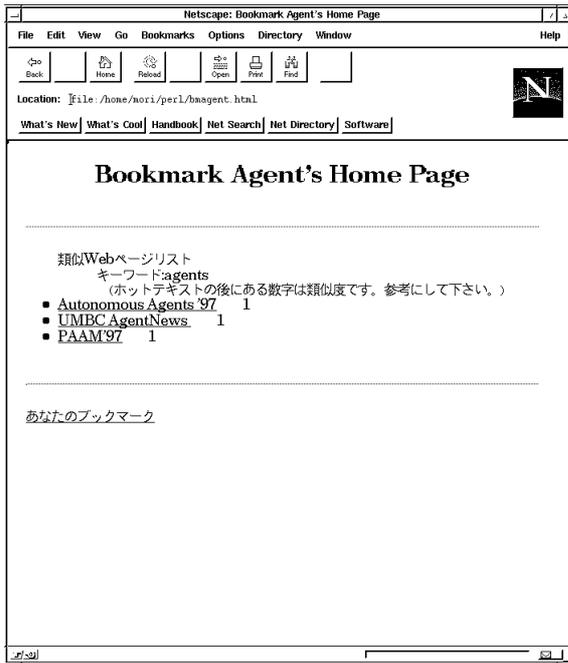


図 5 キーワードを与えた場合のブックマークエージェントによる提示

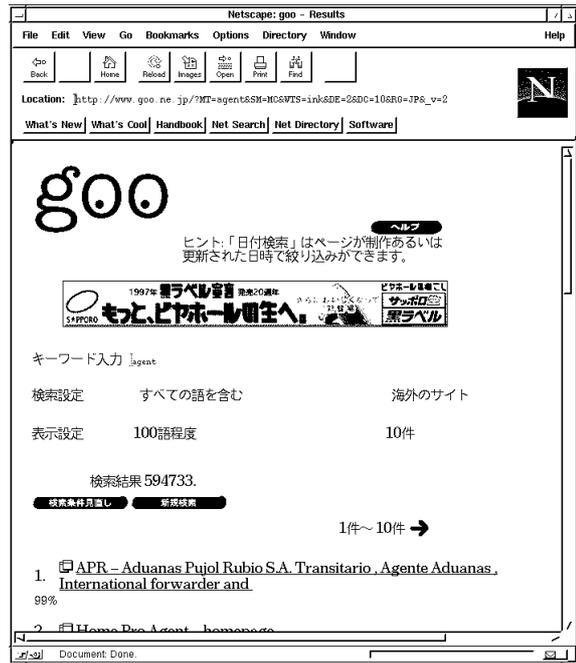


図 7 goo による検索結果

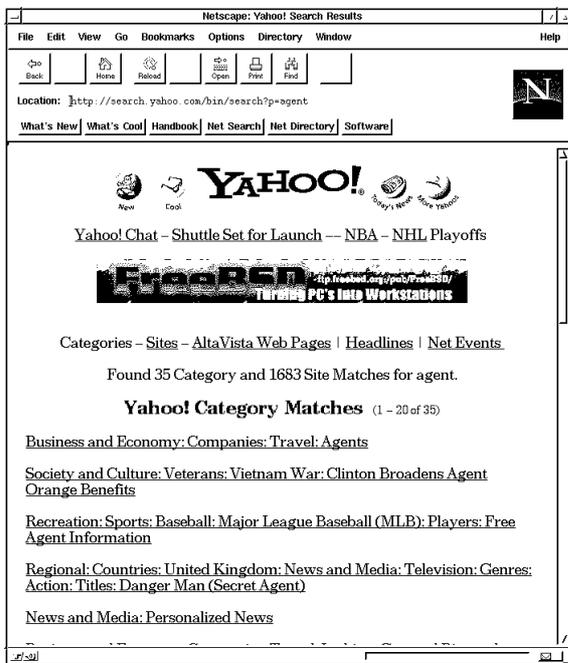


図 6 yahoo による検索結果

sonalized Navigation for the Web, *Communications of ACM*, vol. 40, No. 3, pp. 73-75 (1997)

- [2] 武田: ネットワークを利用した知的情報統合, *人工知能学会誌*, Vol. 11 No.5, pp. 680-688 (1996)
- [3] 森田, 速水: 情報フィルタリングシステム, *情報処理*, Vol. 37 No.8, pp. 751-757 (1996)
- [4] Lieberman, H.: Letizia: An Agent That Assists Web Browsing, *IJCAI-95*, pp. 924-929 (1995)
- [5] Joachims, T., Mitchell, T., Freitag, D., and Armstrong, R.: WebWatcher: Machine Learning and Hypertext, *Fachgruppentreffen Maschinelles Lernen* (1995)
- [6] Salton, G. and McGill, M. J.: Introduction to modern information retrieval, *McGraw-Hill* (1983) *Fachgruppentreffen Maschinelles Lernen*(1995)