

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

森 幹彦[†] 山田 誠二[†]

Bookmark-Agent: Sharing of Bookmarks for Search Assists

Mikihiko MORI[†] and Seiji YAMADA[†]

あらまし ブックマークエージェントは、ユーザが興味をもった URL を登録してあるブックマークファイルを参照し、ユーザ間で情報の共有を行うシステムである。このエージェントは、既存の検索エンジンよりも効果的に情報の検索を行うことができる。ユーザがある Web ページをブラウジングによって探そうとしているとき、このエージェントはユーザの要求していると思われる Web ページをブックマークの中から検索して、ユーザのブラウザにハイパーリンクとしてのリストを提示することができる。更に、エージェントは他のユーザのエージェントと交信することにより、他のユーザのブックマークの検索も行うことができる。小規模で興味の似たユーザからなるグループにおいてこのエージェントを用いることで、ブックマークというユーザによってあらかじめフィルタリングされている情報を利用できるため、ブックマークエージェントは既存の検索エンジンよりも適合率の高い Web ページの提示が可能である。このことは、ブックマークエージェントの性能評価のため 6 人のユーザを使った実験により確認された。

キーワード 協調的情報フィルタリング、ソフトウェアエージェント、ブックマーク、the World Wide Web

1. ま え が き

近年、World Wide Web (WWW) 上の文書は急激な増加を続けている。したがって文書検索において、ブラウジングのみで要求している情報にたどりつくことは難しい。そこで、一般に検索エンジンと呼ばれる、キーワードを与えることで Web ページを検索するシステムを利用することになる。しかし、これらの検索エンジンは多数の不必要な検索結果を提示する傾向がある [11], [12]。このことは検索の更なる絞り込みを要求し、ユーザに負担となっている。ユーザが検索語としてある単語を選択したとき、暗黙のうちにその検索語はそのユーザの興味に沿って意味付けされている。しかし、検索エンジンはそれを察知できないため、ユーザの興味外の意味で使用されている場合のページも一律に表示することになる。

一方、有効な情報を得たい場合、その分野の情報をよく知っていそうな知人に尋ねることも多い。そのよ

うな尋ねられる人はあらかじめその分野に関する知識をもち、有効だと思われる情報を蓄えているエキスパートと考えられる。すなわち、その人は特定の分野の興味と有用性によりフィルタリングした情報を多くもっていることになる。しかし、肝心のエキスパートが知人にはいない、いても遠慮のため尋ねられない、不在であるなどの理由で常時尋ねられるとは限らない。

本研究では、ブックマークエージェントと呼ぶシステムを提案する。特定領域のエキスパートに直接尋ねることなく情報を得られる。このシステムは協調的情報フィルタリング [12] の考え方をもとにしている。共有される情報として、ブックマーク^(注1)を用いる。ブックマークはユーザの興味を反映しているので、興味の領域に応じて数に違いがある。また、ブックマークは WWW 上の文書中で有用だと思われる文書のみをフィルタリングにより抽出した結果であると考えられる。

ブックマークエージェントは各ユーザのブックマークを必要に応じて共有することで特定領域の興味をもつ情報を共有し、ユーザの情報検索を支援するシステムである。このシステムは興味がある程度共通して

[†] 東京工業大学 大学院 総合理工学研究所 知能システム科学専攻、横浜市

CISS, IGSSE, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama-shi, 226-8502 Japan

(注1)：ユーザが一度見た興味ある Web ページの URL の記録のこと。

いるグループに適用することを想定している。このグループは小規模であらかじめ集められているとする。したがって、ユーザ間での共通の領域における有効な情報を利用できるので、一般の検索エンジンよりも効果的な絞り込みが期待できる [7]。

ブックマークエージェントはそれぞれのユーザのもとで起動する。したがって、ユーザは共有の参加や脱退をその時々で自由に選択できる。ユーザが現在訪れている Web ページをもとに、そのページに類似したページをユーザに提示し、ユーザの検索を支援する。類似ページはユーザのブックマークから得られた Web ページから検索する。ただし、他のエージェントと通信することで、他のユーザのブックマークも検索することができる。また、ユーザはエージェントに対してキーワードを明示的に与えて検索することもできる。

Siteseer [7] は、ブックマークがユーザの興味を表しているとしてユーザ間の情報を共有するシステムである。各フォルダ^(注2)の中にある URL がいくつか一致しているかによって、ユーザ間のフォルダを関連づけ、他のユーザの類似するフォルダの URL を推薦する。しかし、Siteseer ではフォルダがユーザによって完全に把握管理されていて分類されていることが前提である。また、フォルダ間の URL の一致した数のみを見て類似性を判断しているが、フォルダ中の URL 数によって一致数が変化することは明らかである。ブックマークエージェントはフォルダを利用せず、各 URL にある HTML ファイルから抽出されるキーワードをもとに類似度を測るため、一貫した基準で類似度を測ることができる。

WebWatcher [4] は、ユーザが最初に与えたキーワードとユーザの現在見ている Web ページを参考にしながらユーザの興味を学習し、ユーザに探索のヒントを与えるシステムである。ユーザの興味を示すものとして Web ページのハイパリンクに重点を置いており、既存のハイパリンクをアイコンで強調するというページの提案法を用いている。

Letizia [5] も、WebWatcher に似たシステムである。異なるのは、単一ユーザを対象としており、完全クライアントとして動作するところである。ユーザの探索時には同行してページを読みながら学習し、類似したページをユーザに提案する。

Syskill & Webert [6] もユーザの挙動を見ながらユーザの興味のあるような Web ページを提案するシステムである。各ユーザの各話題についてそれぞれ

プロファイルを学習する。手動で作ったインデックスページから探索を開始し、ユーザのプロファイルを学習していく。

WebWatcher, Letizia, Syskill & Webert は、明示的なキーワード検索が行えないため、ユーザは黙々とリンクをたどっていく必要がある。また、これら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、グループ内のユーザ間の情報共有に主眼を置いている。

WebMate [1] は TF-IDF を用いてユーザの興味を徐々に学習し、更新し続け、自動的にユーザの興味のあると思われるニュースを提供する。また、学習したプロファイルをもとに、ユーザが検索をするときにユーザの与える検索語を補完する検索語を自動的に追加することができる。WebMate はユーザの興味を知るためにユーザのブラウジングを監視するが、ブックマークエージェントはユーザの興味を凝縮したブックマークの変化を監視することでユーザの興味を知ることができる。

ShopBot [2] は、複数の商品取扱いサイトから商品情報を自動抽出するシステムである。自然言語処理をすることなく、商品の知識とホームページの規則性から帰納学習によって商品情報を得ることができる。しかし、ShopBot は興味の対象として商品しか扱えず、他の分野への応用が難しい。

Beehive [3] はメールシステムをもとに協調的情報フィルタリングを行うシステムである。あるユーザが興味をもった情報は、Beehive のシステムを通じて登録されたグループのメンバにメールとして送られる。メールのやり取りの少なかったメンバはある頻度を下回るとリストからはずされる仕組みになっている。また、Beehive システムはユーザ間の交信の内容には関知しない。このシステムはユーザが情報提供することを主眼としているのに対して、ブックマークエージェントはユーザの興味によりユーザに対して情報を提供することを主眼としている。

2. ブックマークエージェント

2.1 機能

図 1 にブックマークエージェントのシステム構成を示す。ブックマークエージェントは 1 人に対して一つ起動され、情報検索を行う。

(注 2): ブックマーク中で、URL を分類するためにある入れ物のこと。

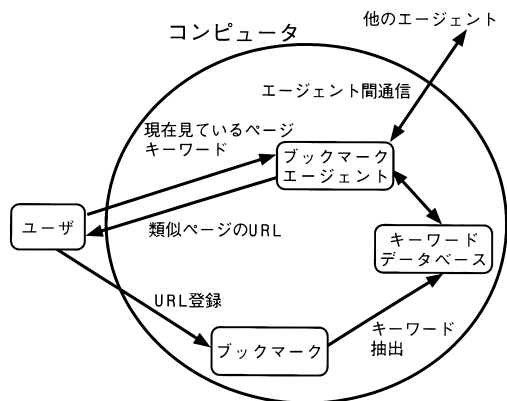


図1 ブックマークエージェントのシステム構成
Fig.1 The Bookmark-agent system.

以下に、ブックマークエージェントの機能を挙げる。

- 新規 URL におけるキーワードの獲得：ブックマークファイルを監視し、ユーザにより新たな Web ページの URL が追加された場合、その URL の Web ページの HTML ファイルからキーワードを抽出し、キーワードデータベースに URL やタイトルとともに登録する。ユーザのプライバシーを守るため、他のユーザに公開するかどうかはブックマークごと、フォルダごとに指定できる。

- 類似ページの提示：ユーザが現在見ている Web ページから自動的にキーワードを抽出し、それをもとに類似ページの URL をユーザに提示する。また、ユーザがエージェントに明示的にキーワードを与えて、類似ページを提示させることも可能である。

類似ページの検索方法は、以下の 2 通りがある。

- エージェントが担当しているユーザのキーワードデータベースを検索する。
- エージェントは、他のエージェントと通信して他のユーザのキーワードデータベースの検索を依頼できる。グループごとにブックマークエージェントが通信できる相手はあらかじめ決めておく。各ユーザは自分のキーワードデータベースの各 URL に対して許可情報を与えることができ、他のユーザに一部若しくは全部の URL 情報を非公開にすることが可能である。

- 検索依頼の受領：他のブックマークエージェントから検索依頼があれば、担当のキーワードデータベースを検索し、見つかった類似ページを返信する。

2.2 キーワード抽出

テキストからのキーワード抽出は、出現頻度を用い

るものをはじめとして様々な方法が提案されている [8]。本研究では HTML という構造化された文章を扱うことから、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る。このような構造を用いる方法は、一般的に用いられる TFIDF [8] などに比較して、文書集合の必要がなく効率が良いという利点がある。

Web ページのキーワード集合を次のように定義する。HTML ファイルにおいて以下に述べる特定のタグが見つかった場合、そのタグ中に含まれている単語の語尾を正規化したあとその単語に対し、括弧内の重みを加算していく。そして、その重みの上位五つの単語をその Web ページのキーワードとする。

ブックマークエージェントは以下のタグを利用する。

- <META>中のCONTENT 属性 (10)
- <TITLE> (10)
- <Hn> (7 - n)
- (1)
- <U> (1)

2.3 類似度と類似ページ

Web ページの類似度は、二つのページのキーワード集合の積集合をとり、その要素数とする。したがって、類似度は 1 から 5 までの値をとる。

類似ページは、あるしきい値以上のページを意味する。ユーザは類似ページのしきい値を、ブックマークエージェントの提示する類似ページの内容を見て変更することが可能である。しきい値の初期値は 3 としている。

3. 実装方法

3.1 実装における問題

ユーザのブックマークを共有する手法として、以下の二つが考えられる。

(1) 各ユーザのブックマークを一つの中央データベースに登録して共有する。

(2) 各ユーザごとにデータベースを用意し、ユーザの参加表明によって共有する。

手法 (1) は共通のデータベースに登録するため、ネットワークの状態にかかわらず他のユーザのブックマークを参照することができる。しかしこの方法では以下の問題点が考えられる。まず、中央データベースの管理という仕事が増えることに加え、その管理者がすべてのデータにアクセスが可能になるプライバシーの問題がある。もう一つに、ユーザの共有の停止・再開のときのコンピュータの負荷の問題である。ユーザ

の停止時にはユーザから得た情報をすべて見つけ出して削除する必要があり、再開時にはユーザの情報をすべて登録する必要がある。このため、停止・再開が頻繁にあると、コンピュータの負荷は大きい。

これに対し、手法(2)ではユーザ側にデータベースを移すので上記の問題を解決できる。まず、データがユーザのマシンに存在するため、個々のユーザの管理下であり、全体のデータにアクセスできる者はいない。したがって、ブックマークエージェントがユーザの参加・脱退が自由であるオープンなシステムであること考えて(2)の分散型のシステムを採用した。

3.2 実装

ブックマークエージェントは、三つの部分から実装される。プログラム言語として perl5 を使用してシステムを構築した。ただし、WWW の閲覧や類似ページのリストの閲覧のためのブラウザには Netscape Navigator を用いる仕様になっている。

● ターミナル：ユーザの明示的キーワード等のコマンドを伝える役割や、コマンドの実行結果やエラーを表示するという、ユーザとエージェントのインタフェースの役割を担う。また、ユーザのエージェントの使用開始、終了の指示をする。

● サーバ：エージェントの実行本体。類似ページの検索、ブックマークからのキーワード抽出、そしてエージェント間通信を行う。各ユーザに1個ずつ起動される。サーバには以下の2種類がある。

－ ルートサーバ：キーワードデータベースサーバ同士との交信の中継をする。検索の依頼の中継時には、送り主の名前を伏せることを基本とするが、公開することもできる。名前を伏すのはユーザのプライバシーを守るためである。

－ キーワードデータベースサーバ：実際のブックマークエージェントの検索やユーザのブラウザ監視はここで行う。ユーザごとに起動される。限定的公開のための実装はここですが、現在その実装はされていない。

複数のキーワードデータベースサーバは、一つのルートサーバに星状に接続される。

● URL 提示ページ：ブラウザ上に類似ページのリストを表示する。

4. 実行例

ブックマークエージェントの実行例を二つ示す。この実行例では、ブックマークエージェントは6人の

表1 ユーザの有するブックマーク中の URL 数
Table 1 The number of users' URLs in each bookmarks.

ユーザ名	A	B	C	D	E	F	合計
URL 数	137	356	10	85	167	71	782

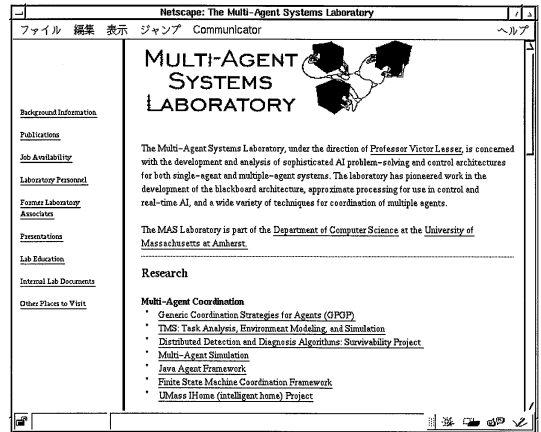


図2 ユーザの現在見ているページ
Fig. 2 A current Web page.

ユーザ(AからF)のブックマークを用いている。それぞれのユーザのブックマークファイル中のURL数を表1に示す。6人のユーザのブックマーク中で同一のURLは除いたため、合計欄は互いに異なるURLの総数となっている。

まず、エージェントが自動的に類似ページを提示する例を示す。図2は、あるユーザの現在見ているWebページを示している。図3は、図2のページに対する、ブックマークエージェントのURL提示ページを示している。この提示例では、ユーザは *The Multi-Agent Systems Laboratory* というWeb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 *Learning in Multi-Agent Systems* と *The Multi-Agent Systems Laboratory* 自身を提示している。このように、直感的にも類似したページを提示しているのがわかる。

次に、ユーザがブックマークエージェントに明示的にキーワードを与えた場合の例を図4に示す。ここでは、キーワードとして 'agent' を与えている。比較のために、同じキーワードを二つの検索エンジンに与えた。検索エンジンとしては、ユーザ登録型検索エンジンとして Yahoo^(注3) を、Webロボットによる自動登録型検索エンジンとして god^(注4) を使用した。

(注3) : <http://www.yahoo.com/>

(注4) : <http://www.goo.ne.jp/>

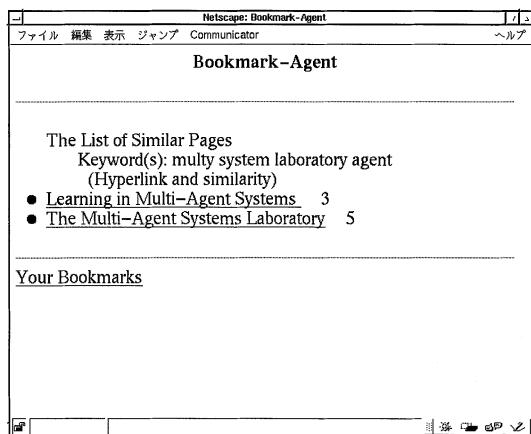


図3 キーワードを与えない場合の提示ページ
Fig. 3 Indicated page by implicit keywords.

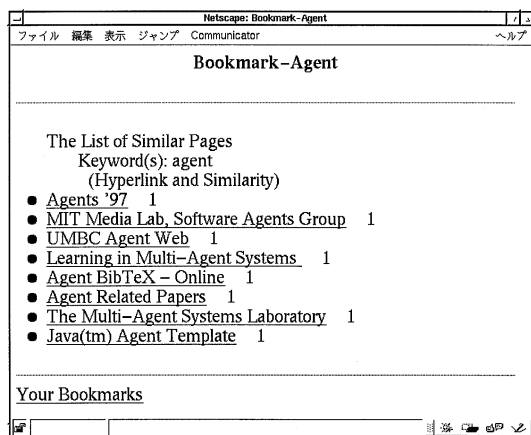


図4 キーワードを与えた場合のブックマークエージェントによる提示
Fig. 4 Bookmark-agent indicating page given keyword "agent".

その結果、キーワード 'agent' に対して、Yahoo が 2,485 個のサイトを、goo においては 1,811,013 個ものサイトを示したのに比べ、ブックマークエージェントは、有用な 8 個のサイトしか提示しなかった。

5. 実験

5.1 適合率の定義

情報検索の効率の定量的な評価方法に、再現率と適合率がある [10]。これらの定義は、以下のとおりである。

- 再現率：質問と適合する文書総数に対する検索

された適合文書の割合。

- 適合率：検索された文書総数に対するその中で
の適合文書数の割合。

ここで、検索の母集合となる WWW 上の適合するすべての文書数を測ることは現実的に不可能なため、再現率は求められないため算出をあきらめる。一方、適合率は検索された文書数を母集合とし、その中で正しく検索された文書数の率なので計算可能である。したがって、適合率を検索効率の指標とする。

検索エンジンの検索結果では時として数千から数万件の URL を提示することがある。この場合、適合率を調査することは困難である。そこで、検索エンジンでは URL を並び替え、与えられたキーワードの内容に適していると思われる URL を先に表示する機能を利用する。上位 20 個までに入った URL の適合率を求めることで近似的にそのキーワードに対する適合率とする。

通常、ユーザは上位数十個の検索結果しか見ないことが多い。見る数は必ずしも一定ではないが、検索結果が数千件を超える場合は、ユーザが実際に見る URL 数が多少増減してもごく一部のみを見ていることになり変わらない。したがって、このような条件を考慮し、本研究では上位 20 個を利用することにした。

ここで、上位 20 個の URL の適合率が検索結果全体の URL の適合率よりも一般的に高いことに注意してほしい。なぜなら、一般に検索エンジンはキーワードの頻度などに基づき、適合する確率の高い順に URL を提示するからである。したがって、完全ではないものの、検索エンジンを不利にすることのない条件で適合率を求めていることになる。

5.2 実験方法

ブックマークエージェントの性能評価のための実験を行った。ブックマークエージェントと比較するために検索エンジン goo を使った。ブックマークエージェントは協調情報フィルタリングを行っているため、検索エンジンとは異なる性質を示すことが期待される。この性質が結果にどう影響するかを調べた。また、評価基準として適合率を用いた。

双方に与えるキーワードとして、各ユーザのキーワードデータベースを総合し、その中に存在する任意のキーワードを 1 から 3 個ずつ選んだ。ユーザ A が被験者になり、提示された URL の Web ページが適合するかどうかの判定を行い、適合率をそれぞれ求めた。

ブックマークエージェントの方の評価では、表 1 に

示す 6 名のキーワードデータベースを使用する。

このユーザグループではみな OS として Linux を使用しており、人工知能、エージェント、ロボットを研究しているということに注目する。

5.3 実験結果

表 2, 表 3 はそれぞれ 1, 2 個のキーワードを与えたときの適合率をまとめたものである。ここで、キ

ワード 2 個の場合は、キーワードの AND 検索を行っている。表の BA と SE はそれぞれブックマークエージェントと検索エンジンによる結果を表す。キーワードを 3 個与えた実験も行ったが、結果の表は省いた。なぜなら、2 個のときで既に提示 URL 数が 1 個に近いことから推測できるように、キーワード 2 個と 3 個の結果で大きな違いがなかったためである。

6. 考 察

6.1 単語の性質による適合率の違い

適合率は、ブックマークエージェントが検索エンジンにかかわらず、検索するキーワードの単語としての性質によって特徴が現れる。キーワードが形容詞として使われやすい単語のとき適合率は低い。例えば、‘web’, ‘english’ は名詞にもなるが、後ろに名詞をとることで ‘Web Agent’, ‘English Dictionary’ のように形容詞としても働く。このような単語のとき、単独で使用する場合の適合率よりも他の単語をとって熟語を作った場合の方が高い。

ブックマークエージェントにおいてこの傾向は、興味の領域に入っていない熟語のときに起こりやすい。‘English Dictionary’ や ‘Mailing List’ はユーザにとってそのもの自体に興味があるわけではない。したがって、表 2 と表 3 より、‘english, dictionary’ の適合率はブックマークエージェントで 100%, 検索エンジンで 70% なのに対して、‘english’ ではそれぞれ 37.5% と 5.0% となってしまう。‘mailing, list’ の適合率でも同様の結果が見られる。

6.2 ブックマークエージェントと検索エンジンとの比較

ユーザ間で興味の領域が共有されているかどうかで、ブックマークエージェントの URL 提示はどのように変化するかを考察する。ブックマークエージェントは、共通する興味で情報を集約するという文脈で、協調的情報フィルタリングを用いている。このフィルタリングの効果は、興味領域の共有の有無によって適合率が異なっているから観察することができる。

まず、ユーザ同士が興味の領域を共有している場合を考える。本実験におけるユーザは皆、一般的な科学には興味がないため共有していないが、人工知能、ロボット、エージェントシステムには興味があるため共有している。したがって表 2 より、ブックマークエージェントにおいて、‘ai’, ‘robot’, ‘agent’ といった単語の適合率は ‘science’ という単語の適合率よりも高い。

表 2 キーワード 1 個を与えた場合の適合率
Table 2 Precision when 1 keyword was given.

キーワード	呼出 URL 数		適合率 (%)	
	BA	SE	BA	SE
agent	8	1811187	100.0	5.0
robot	10	268591	100.0	30.0
intelligence	10	23622	100.0	10.0
ai	4	1134699	100.0	45.0
research	4	8465215	75.0	20.0
internet	15	13131781	26.7	15.0
web	14	128811	35.7	10.0
linux	22	1415551	100.0	70.0
unix	6	2708291	100.0	40.0
science	3	5477707	33.3	50.0
software	24	9113879	91.7	45.0
faq	7	3144108	100.0	50.0
perl	4	893256	75.0	75.0
java	16	2317347	100.0	45.0
sendmail	4	168982	100.0	65.0
search	13	13542226	84.6	55.0
mac	22	1974327	100.0	60.0
macintosh	16	1575040	100.0	85.0
weather	1	2705383	100.0	75.0
dictionary	6	11096	100.0	35.0
japanese	19	1400994	21.1	40.0
english	8	4262210	37.5	5.0
ml	4	540252	75.0	35.0
archie	5	282468	100.0	55.0
平均			81.5	44.2

表 3 キーワード 2 個を与えた場合の適合率
Table 3 Precision when 2 keywords were given.

キーワード	呼出 URL 数		適合率 (%)	
	BA	SE	BA	SE
software agent	1	400238	100.0	30.0
web agent	1	541780	100.0	15.0
linux software	3	494729	100.0	35.0
free software	1	2502799	100.0	10.0
english dictionary	3	197152	100.0	70.0
java software	1	687154	100.0	30.0
java internet	1	766101	100.0	15.0
web robot	1	91377	100.0	55.0
artificial intelligence	6	264060	100.0	45.0
linux application	1	129544	100.0	20.0
tokyo university	1	207367	100.0	50.0
mailing list	4	2453832	100.0	75.0
fj news	1	40508	100.0	60.0
sports soccer	1	374736	100.0	35.0
平均			100.0	38.9

また、検索エンジンと比較した場合、例えばキーワード ‘agent’ の適合率は、検索エンジンが平均適合率の半分以下になっているのに対して、ブックマークエージェントは高い適合率を維持している。この現象は、‘agent’ が普通、‘代理人’ や ‘代行’ の意味で使われるが、人工知能についてのユーザグループであれば、その意味が “コンピュータを使ったエージェントシステム” という意味になるために起こっている。

次に、興味の領域を共有しないキーワードの場合、ブックマークエージェントの URL の提示数は共有するキーワードに比べて 1 けた少なく、絶対数としても非常に少ない。例えば、‘linux’ は共有するため 22 個呼び出されて提示されているが、‘perl’ はユーザ A によってのみ提供されていることがわかっており共有されていないため、4 個しか提示されていない。興味の領域を共有しない場合は URL 提示数が共有する場合に比べて低いことがわかる。また、実験中のデータにおいて検索エンジンの提示数に比べて常に $\frac{1}{100}$ 以下である。

興味深い結果もある。UNIX という単語は Linux という単語よりも一般的なもので、提示 URL 数は UNIX の方が多いはずである。しかし、ここで思い出さなければならないことは、本実験のユーザは皆 Linux のユーザであることである。したがって、キーワード ‘linux’ は ‘unix’ よりも多数の URL を提示している。

上記のことから、ブックマークエージェントは興味の領域を共有していない場合、及び一般的な検索語を与えた検索には不向きである。適合する Web ページをほとんど提示しない場合があるためである。対照的に、検索エンジンによる検索では、適合率こそ低いが大量の URL を提示するため、時間をかけてよく探せば要求に適合する Web ページを見つけられる場合がある。

一方、ユーザ間で興味の領域を共有するようなキーワードで検索する場合、ブックマークエージェントは適合率が高く、かつ少数の URL を提示をすることができ、有効である。すなわち、検索エンジンのような膨大な数でしかも適合率が高くない提示の中から、適合する Web ページを掘り起こす必要がない。このことは、興味の領域の共有による協調的情報フィルタリングが働いていることと考えられる。

6.3 ブックマークエージェントと検索エンジンの役割分担

各キーワードに対する適合率における、ブックマークエージェントと検索エンジン goo との有意差を調べるために、順位和検定 [9] を行った。この検定は、それぞれのキーワードの適合率を、ブックマークエージェントと検索エンジンの結果にかかわらず併せ、ソートし、順位に偏りがあるかどうかを調べるものである。よって、片方の検索結果の順位の分布が他方と異なる場合に有意差が生じる。キーワード数が 1 個、2 個、3 個のときに、エージェントと検索エンジンとの適合率を標本として検定を行った結果、それぞれ有意水準 5% で有意差が確かめられた。

この検定により、ブックマークエージェントは検索エンジンにはない URL 提示法を用いていることが明らかとなり、ブックマークエージェントの協調的情報フィルタリングが有効であることが示唆された。

したがって、ブックマークエージェントと検索エンジンは検索したい内容により役割分担が可能であると考えられる。ブックマークエージェントは興味の領域を共有しているという条件において有効な手段である。一方、検索エンジンは安定した適合率は望めないものの、どのようなときでもユーザが時間を掛けさえすれば適合する Web ページを見つけ出すことができる。

また、ブックマークエージェントはブックマークに登録されない限り新しい URL を提示できないが、検索エンジンは自動的に獲得するため可能である。したがって、最初に特定の情報を探そうとするユーザは検索エンジンで調べなければならず、その恩恵をブックマークエージェントによって他のユーザも受けられるようになるので、二つのシステムの間には相補的な関係があることがわかる。

7. む す び

本研究では、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。また、6 人のユーザによる実験を行った。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る。したがって、検索エンジンがユーザが望まない Web ページを多数提示してしまうのとは異なり、ブックマークエージェントは適度にフィルタリングされ、少数であるがユーザの要求に見合った URL 情報を、協調的情報フィルタリングによって提供することができた。

ブックマークエージェントは前提として、ユーザの興味が類似しているとし、そのユーザグループ内でフィルタリングを行っている。今後、ユーザの興味の類似性を測定し、それをもとにグループ分けを行うことを考えている。また、類似性に着目したマッチメイキングのシステムを考案したい。

謝辞 本研究は、日本学術振興会 未来開拓学術研究推進事業研究プロジェクト「生物的適応システム」の支援のもとに行われた。記して、感謝致します。

文 献

- [1] L. Chen and K. Sycara, WebMate: A Personal Agent for Browsing and Searching, 1997.
- [2] R.B. Doorenbos, O. Etzioni, and D.S. Weld, "A scalable comparison-shopping agent for the world-wide web," Proc. 1st Int. Conf. on Autonomous Agents, pp.39-48, 1997.
- [3] B.A. Huberman and M. Kaminsky, "Beehive: A system for cooperative filtering and sharing of information," <http://www.parc.xerox.com/spl/groups/dynamics/www/beehive.html>, 1996.
- [4] T. Joachims, D. Freitag, and T. Mitchell, "Web-watcher: A tour guide for the world wide web," Proc. 15th IJCAI, vol.1, pp.770-775, 1997.
- [5] H. Lieberman, "Letizia: An agent that assists web browsing," Proc. 14th IJCAI, vol.1, pp.924-929, 1995.
- [6] M. Pazzani, J. Muramatsu, and D. Billsus, "Syskill & webert: Identifying interesting web sites," Proc. 13th AAAI and 8th IAAI, vol.1, pp.54-61, 1996.
- [7] J. Rucker and M.J. Polanco, "Sitereer: Personalized navigation for the web," Commun. ACM, vol.40, no.3, pp.73-75, 1997.
- [8] G. Salton and M.J. McGill, Introduction to modern information retrieval, McGraw-Hill, 1983.
- [9] 浅野長一郎, データとデータ解析, 日本放送出版協会, 1992.
- [10] 伊藤哲郎, 情報検索, 昭晃堂, 1986.
- [11] 武田英明, "ネットワークを利用した知的情報統合?" 人工知能誌, vol.11, no.5, pp.680-688, 1996.
- [12] 森田昌宏, 速水治夫, "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.

(平成11年7月5日受付, 11月9日再受付)

森 幹彦 (学生員)



1996 千葉大・工卒。1998 東京工業大学大学院総合理工学研究科知能システム科学専攻修士課程了。現在、博士課程在学中。WWW 上におけるソフトウェアエージェントやユーザプロファイルの研究に従事。人工知能学会会員。

山田 誠二 (正員)



1984 阪大・基礎工卒。1989 同大大学院博士課程了。同年大阪大学基礎工学部制御工学科助手。1991 同大学産業科学研究所講師。1996年4月より、東京工業大学大学院総合理工学研究科助教授、現在に至る。工博。人工知能、特に、WWW 上でのプランニング、ロボット学習に興味をもつ。人工知能学会、情報処理学会、AAAI、IEEE 各会員。