

ブックマークエージェントによる WWW の URL 情報の共有

東京工業大学 大学院総合理工学研究科

森 幹彦 山田 誠二

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 [1][2]。本研究では、比較的小規模のグループで、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*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。

我々の提案するシステムにおいては、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 をも参照して、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、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。また、この分野の先行研究 [3][4] 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。

Information Sharing of WWW Using Bookmark Agent
Mikihiko Mori, Seiji Yamada
Tokyo Institute of Technology
4259 Nagatuta, Midori-ku, Yokohama, 226, Japan

*ここでは、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である。

2 システム構成

ブックマークエージェントのシステム構成を図 1 に示す。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：ユーザの明示的キーワードの入力、ユーザの使用開始、終了の合図に使われる。
- デモン：ターミナルに入力されたキーワードを受け取る。ブックマークのキーワード抽出を行う。エージェントの Web ページに類似ページのリストを与える。他のエージェントと通信する。
- Web ページ：類似ページのリストを表示する。

3 ブックマークエージェント

3.1 基本機能

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 をユーザに提示する。さらに、ユーザ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の URL 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

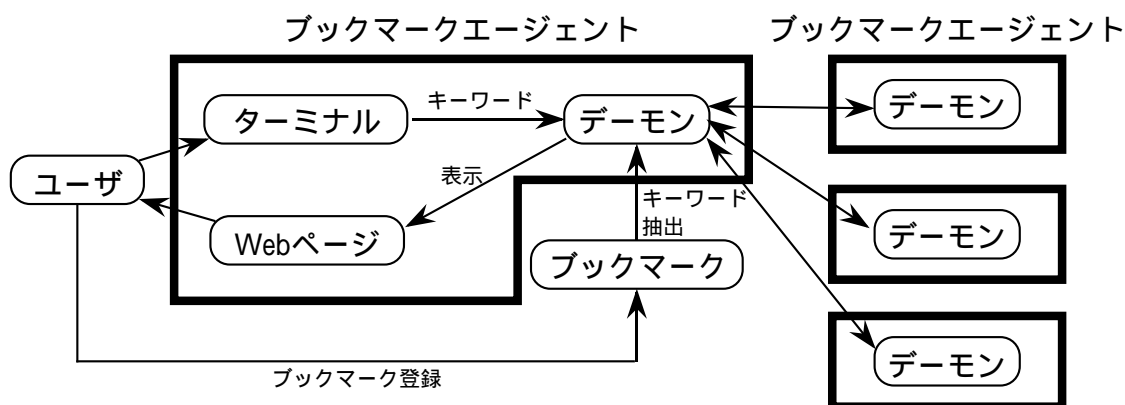


図 1: ブックマークエージェントのシステム構成

エージェントは、担当ユーザの 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。

- 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、担当ユーザの URL データベースで検索を行い、類似ページを返す。

3.2 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 [5] 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。

3.3 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。

ブックマークエージェントとユーザのインタフェースは、ユーザがブックマークエージェントの HTML ファイルをブラウザで見ることにより行われる。ページ上で、類似ページはハイパーリンクとなって提示されており、クリックするだけで、類似ページへ移る

ことができる。また、類似ページのしきい値は、ユーザが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みて変更することが可能である。

4 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] 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)
- [2] 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)
- [3] Lieberman, H.: Letizia: An Agent That Assists Web Browsing, *IJCAI-95*, pp. 924-929 (1995)
- [4] Joachims, T., Mitchell, T., Freitag, D., and Armstrong, R.: WebWatcher: Machine Learning and Hypertext *Fachgruppentreffen Maschinelles Lernen*(1995)
- [5] Salton, G. and McGill, M. J.: *Introduction to modern information retrieval*, McGraw-Hill (1983)