

横断検索エージェントシステムの検索先推薦機能とその評価

3P-8

柳本 豪一 松田 勝志 谷 幹也 市山 俊治
 NEC ヒューマンメディア研究所

1. はじめに

現在数多くのオンライン図書目録（OPAC：Online Public Access Catalog）サービスが、インターネット上で公開されている。これらは、独自の運営がなされており、検索インターフェイス、検索サービスの内容、および蓄積されているデータの量に差がある。ユーザがこのようなサービスを利用する際、それぞれの OPAC サービスの検索方法を知る必要がある。横断的に検索するシステムの開発により、この違いを意識する必要をなくすることができるが、全ての OPAC を検索対象とすると、ネットワークに負荷をかけたり、応答時間が遅くなってしまう。あらかじめ検索先を絞り込んで検索を行えば解決できるが、ユーザが図書館を絞り込むためには、図書館の蔵書傾向などを把握する必要がある。

これらの問題点を解決し、ユーザに統一したインターフェイスを提供する WWW ベース横断検索システム¹⁾を開発してきた。本システムでは、操作性の違いを吸収する統一した検索画面を提供し、図書館ごとに異なる検索結果の表記を統一することで、検索結果の比較を容易にした。また、ユーザが検索先の選択を行なう負担を軽減するため、ユーザの検索要求に合わせて自動的に検索先を絞り込むようにした。

本稿では、マルチエージェント構成で実装することで、検索先の追加・修正を容易にするとともに、入札による競合状況の調整を行なって、検索先を絞り込む横断検索エージェントシステム²⁾について、その構成と検索先推薦精度について述べる。

2. システムの構成

横断検索エージェントシステムの構成を図 1 に示す。本システムは、サーバエージェント、図書館エージェント、検索エージェントの3種類のエージェントから構成されている。本システムのサービスは WWW で提供されており、ユーザはサービスを利用するために専用のツールを使う必要はない。

サーバエージェントは、横断検索エージェントシステム内に一つだけ存在し、ユーザとの通信を一括して管理する。また、ユーザが行なった操作の履歴管理や、検索先を絞り込む入札処理の管理を行なう。

図書館エージェントは、本システムから検索対象とし

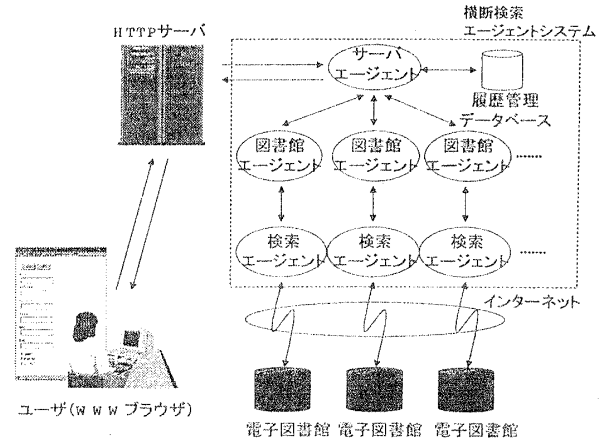


図 1 システム構成

て利用できる図書館の数だけ存在しており、担当図書館の蔵書傾向、利用可能時間などの固有情報を管理し、入札値の決定を行なう。

検索エージェントは、図書館エージェントごとに存在しており、ユーザが入力した検索条件を用いて、実際に OPAC を検索する。このため、OPAC を検索する手順を示した知識や、検索結果の表記を統一するための整形知識を持っている。

ユーザは、本システムの検索画面で検索条件を入力し、検索要求として送信する。検索条件を受け取ったサーバエージェントは検索条件を入札条件として図書館エージェントとの間で入札処理を行ない、検索先を絞り込み、その図書館名をユーザに推薦する。ユーザは表示された図書館名から検索先を選択し、検索を要求する。ユーザにより選択された図書館を担当する検索エージェントが、図書館と通信を行ない検索結果を取得する。

3. 推薦方式

本システムにおける検索先の推薦方式を、図 2 に示す。本方式では、漏れの少ない検索を行なうことを目標とし、ユーザが入力したキーワードに対する検索結果を多く返すと推測される図書館を選択する。図書館エージェントは、サーバエージェントから送られてきた入札条件と、担当図書館の蔵書傾向を比較し、蔵書傾向が一致する度合いが大きい入札条件ほど大きな入札値を返す。蔵書傾向は、15 の分野に対して、あらかじめそれぞれの分野に含まれる代表的な単語を用いて図書館の検索を行ない、得られた検索結果の件数に応じて蔵書傾向の値を設定している。図書館エージェントでは、入札条件にあるキーワードが含まれる分野を、分野推定知識を用いて特定し、管理している蔵書傾向でその分野に割り当てられている値を入札値としている。

Selection of target library for Cross Domain Search Agent System.
 Hidekazu YANAGIMOTO, Katsushi MATSUDA,
 Mikiya TANI, and Shunji ICHIYAMA
 Human Media Research Laboratories, NEC

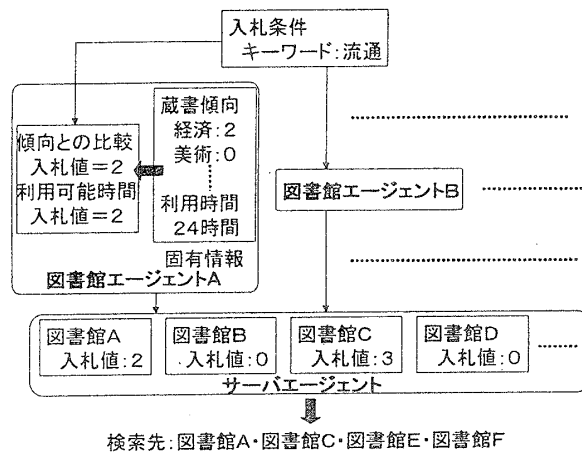


図2 検索先の推薦方法

入札値を決める際に、担当図書館が利用できる時間であるかどうかチェックし、利用できる時間帯である場合には、蔵書傾向より決定した入札値を送信し、利用できない時間帯である場合には、入札値を0としてサーバエージェントに送信する。

サーバエージェントでは、入札値の大きい順にあらかじめ設定した数の図書館を選択し、検索先としてユーザに推薦する。

4. 実験

適切な細かさで分野を設定すれば、それぞれの分野に含まれる代表的な単語に対する蔵書傾向は、その分野に含まれる他のキーワードに対する蔵書傾向と一致するという仮定のもとで検索先推薦機能を設計した。その仮定の妥当性の評価実験を行なった。まず、検索対象の図書館の決定、蔵書傾向の設定、実験用の単語の選択を以下のようにした。検索対象は、OPACサービスで良く利用されている7種類15ヶ所の図書館とした。蔵書傾向の設定では、15の分野から代表的な2単語を選び、合計30個の単語を用いて15ヶ所の図書館を実際に検索し、件数の多い図書館から順番に、0から3に正規化した値を設定した。実験用の単語は、キーワードより分野を推定する約7万2千語を持つ分野推定知識から、蔵書傾向の設定で利用していない111個のキーワードを選択した。分野推定知識に含まれない単語が入力された場合、本システムでは蔵書傾向を利用しないで検索先を推薦するので、これを防ぐため本実験で利用する単語は分野推定知識から選んだ。また、実験で利用する111個の単語の集合は15の分野をカバーしているため、十分な評価データが得られると考えている。

評価実験は以下のように行なった。実験用に設定したキーワードを用いて検索対象となる15ヶ所の電子図書館をすべて検索し、検索結果の件数を調べ、件数が多い上位5ヶ所図書館を正解の図書館とした。検索結果が得られた図書館が5ヶ所未満の場合には、検索結果が得られた図書館だけを正解の図書館としている。一方、本システムで、同じ単語を用いて検索先の絞り込みを行ない、5ヶ所の図書館を選択した。正解の図書館の中に、

絞り込まれた図書館が含まれる割合を正解率とした。図3はこの実験より得られた正解率のグラフである。横軸は正解率の範囲であり、縦軸はその正解率の範囲を満たすキーワードが、全キーワードのうちに占める割合である。

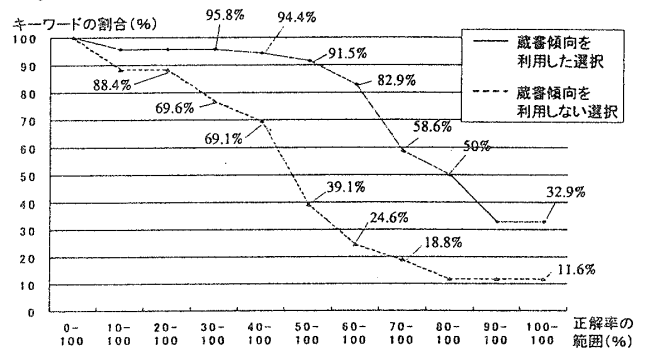


図3 自動選択の正解率

5. 考察

蔵書傾向を利用した検索先の選択では、80%以上のキーワードにおいて、絞り込まれた図書館の3ヶ所以上が、検索結果の件数が多い順位の5位以内に入っている。また、蔵書傾向を利用しない場合より、本方式を用いた検索先の選択の方が正解率が良い。これより、キーワード検索において、分野を代表する単語で設定した蔵書傾向を用いて、多くの検索結果を得られる図書館を選択する本方式が有効であることが確認できた。

選択を誤った主な原因は「分野の分割が粗い」ことである。本システムでは、蔵書傾向を指定するために、15の分野を利用している。しかし、現在の分類では同じ分野に含まれる単語で検索した場合にも多くの検索結果が得られる図書館の順位に違いがある。これは、本来分割すべき分野が統合されているためであると考えられる。これは、蔵書傾向を表す15の分野分類が、現実の図書分類に比べて粗いため生じた問題であると思われる。よって、推薦方式のアルゴリズムを変更することなく、分野の設定を検討することで対応ができると考えられる。

6. おわりに

本稿では、入札により検索先を選択して推薦する横断検索エージェントシステムについて述べた。また、本推薦方式は、分野を代表する単語を用いて設定した蔵書傾向を用いて、検索件数が多く得られる図書館を選択するという観点から考えると、有効な方式であることが確認できた。

本研究の一部はJIPDEC「次世代電子図書館システム研究開発事業」の一環として開発を行なっている。

参考文献

- [1]柳本 他,「検索先の自動選択を行うWWWベース横断検索システム」,情処第56回全国大会,1998
- [2]柳本 他,「競合型エージェントを用いた横断検索システムの開発」,第13回デジタル図書館ワークショップ,1998