

# 利用者と図書のプロフィールを用いた図書検索手法の提案\*

竹 内 さおり

## 1. はじめに

図書館では、図書・雑誌・CD-ROM、Web 情報等の一次情報を検索するための手段として、目録、索引、リンク集等の二次情報を提供している。旧来、これらの二次情報は、冊子体やカード型の目録等の物理メディアに記録された形式で提供されてきたが、計算機技術の進展に伴い、電子化された書誌データとして OPAC (Online Public Access Catalog) が利用できるようになった。初期の OPAC サービスは、通信回線の遅さ、コマンド入力 of 煩雑さ、ユーザインタフェースの不親切さから、一般の利用者には扱いにくく利用頻度の低いものであった。

ところが、インターネットの爆発的な普及に追隨して Web ブラウザに対応した WebOPAC が出現し、マウス操作や簡単なキー入力による図書検索が可能となりはじめるにつれて、利用者が急激に増加している。WebOPAC の普及が利用者の利益となっているように思われる反面、却って利便性が損なわれているという指摘も否めない。つまり、図書検索の手段がインターネット上で利用できるようになったことは十分に評価できるが、利用者をサポートする部分での対応が充実していないため、利用者が満足する結果を保証することができていないのが現状である。

従来の図書館では、図書の探し方を理解していないために欲しい図書が入手できない、どこに必要な図書があるのかわからないというような問題が生じた場合、司書が手助けをするサービス (レファレンスサービス) が確立している。レファレンスサービスでは、司書が利用者と対面して、利用者の要求を直接聞き出し分析しながら、相応しいと思われる図書を提示し、利用者の反応をその場で確かめることができる。ネットワークを介して WebOPAC を図書館外から利用する機会の増加に伴い、検索の実行中に行き詰まった場合の対処法や司書が行うレファレンスサービス等の利用者支援の機能がより重要になってくる。既存の WebOPAC に、検索の進行をガイドする機能や利用者の要求を分析できる機能が追加され、利用者自身が WebOPAC とインタラクティブに検索を行うことができれば、検索における利用者の負担を軽減し、利用者ごとに異なる条件にも柔軟に対応した効率的な図書検索が可能となるであろうと予測する。

## 2. 研究の背景

本研究の基本的なアイデアは、従来から図書館で行われているレファレンスサービスにお

---

\*筆者は、本学国内研修制度の利用により、平成 12 年 4 月から平成 13 年 9 月まで奈良先端科学技術大学院大学情報科学研究科へ派遣され、恵まれた環境で学ぶ機会を得ることができた。本稿では研究成果の一部を報告する。

るレファレンスインタビューに基づく。司書が利用者と対話的に問題解決を行う過程に着目した。豊富な経験と蓄積したノウハウを活用しながら、司書が具体的なキーワードのヒントを提示したり、利用者の要求と図書の内容をマッチさせて利用者の求める情報が得られるように探し方をガイドするしくみを WebOPAC に組み込むことができないかと考えた。レファレンスサービスの役割を図 2.1 に示す。

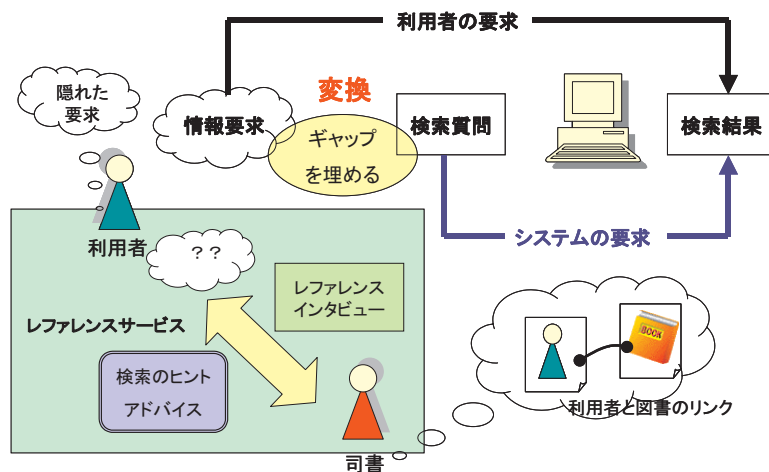


図 2.1 レファレンスサービスの役割

Yahoo<sup>†</sup>、Google<sup>‡</sup>等のサーチエンジンの成功に示されるように、求める情報をその内容から検索することに対する利用者の要求は高く、WebOPACにおいても同様の要求が高くなってきている。WebOPACでは、利用者が求める図書を引き出すために、自らの要求をキーワードで表現しなければならない。利用者がキーワードを適確に選択できないと意図に沿った検索が行えないため、いかに適切なキーワードを選択するかが重要な問題である。このときに、検索の対象が既知資料（存在することが自明な図書）の場合と未知資料（存在自体も明らかでない図書）の場合で、キーワードの選び方が異なる。既知資料の場合は、利用者の記憶が曖昧であっても書名や編著者名の一部をキーワードに選択して、意図する図書が見つけれられる場合もある。しかし、未知資料の場合は「〇〇についての本」のように、図書の内容から探さなければならないため、図書の内容を1語または数語のキーワードで表現しなければならない。内容から図書を探す手法として、図書館では件名目録と分類目録が使われてきた。いずれも、件名や分類の知識を必要とする専門的な検索方法である。

主題検索では、利用者が思いついたキーワードを手当たり次第に入力したとしても、有用な結果が得られるとは限らない。なぜなら、図書に付けられた件名という図書の内容や主題を表す索引語が、件名標目表に基づく統制語であるため、利用者が思いついたキーワードを入力しても、

<sup>†</sup> <http://www.yahoo.co.jp>

<sup>‡</sup> <http://www.google.co.jp/>

この件名と合致しない限りは結果が返されないからである。図書にどのような件名が付けられているのかを調べることができる WebOPAC は、現在のところ国立国会図書館が提供している WebOPAC だけで、他の WebOPAC では件名を調べることができない。また、件名付けの基になっている件名標目表がどのようなものかさえ知らない利用者も多く、件名標目表を活用することも難しい。しかしながら、書誌情報として件名のデータは蓄積されており、件名標目表を使わなくとも図書にどのような件名が付与されているかを知ることができる。例えば、ある利用者がこれまでに借りて読んだ図書の書誌情報を貸し出し履歴から得ることは可能であり、それらの件名だけを取り出して蓄積しておき、検索時に再利用することができれば、キーワードの選択を支援することも可能である。

このように、利用者の知的活動を支援することは、電子図書館においても同様に行われなくてはならない<sup>1</sup>と考えられ、資料や目録の電子化だけではなく、図書館におけるサービスの電子化についても早急な対応が必要である。また、電子資料の提供が進む昨今、電子情報の検索にはコンピュータやソフトウェアの操作技量が問われることも多い。マルチメディア資料などの教育用ツールの普及に伴い、情報設備を活用するための教育や支援が必要となってくる<sup>2</sup>。このような現状から、レファレンスサービスは電子情報を扱うための情報リテラシー教育とも深く関わり、利用者教育との連携が必要であると考えられる。

2002年9月現在、Web上で利用できる WebOPAC を検証したところ、司書が行うレファレンスサービスのように、利用者をインタラクティブに支援するしくみは実現されていない。一般的に、システムの設計は利用者のモデルを想定して行われ、システムの用途によっては、利用者の習熟度に応じたユーザインタフェースを工夫することは可能である。WebOPAC の場合は利用者の要求自体を扱うため、仮想の利用者をモデルとすることは困難であると言われてきた。利用者個々に適応したユーザインタフェースを提供することは確かに難しいことであるが、図書の情報や利用者の情報を別に蓄積して検索時に利用する方法は実現の見込みが高いと思われる。

### 3. 提案手法

本手法では、仮に、司書が利用者と図書検索システムの仲介役として介在できない場合でも、利用者が結果を得る前に検索を中止することなく最後まで検索を継続して行い、利用者の要求に合った満足度の高い結果を得られるように利用者を補助することを目標とした。以下に示す二つの手法を組み合わせて用い、利用者を支援するしくみを提案する。提案手法の概要を図 3.1 に示す。

1. シソーラスの階層構造を利用して、検索範囲の拡大や検索結果の絞込みのために、上位語あるいは下位語の追加によって問合せを拡張し、利用者と図書検索システムが対話的に検索を続行できるようなキーワードの提案を行う。
2. キーワードと書誌情報のマッチング以外に、利用者の要求や検索意図を反映できる方法として、利用者が借りた図書に対して難易度や感想を残し、利用者と図書の両方のプロフィールを作成して、検索結果のフィルタリングに利用する。

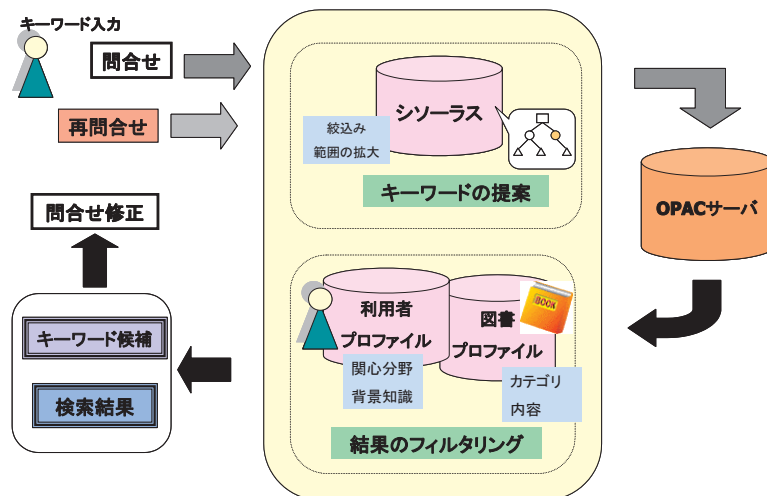


図 3.1 提案手法の概要

これらの機能により、利用者は図書検索システムと対話的に検索を進め、要求を満たす度合いが高いと思われる図書を優先的に得ることができると考えられる。

キーワードの提案を行う部分では、ユーザーが入力したキーワードから得られた検索結果の件数が多すぎる、あるいは少なすぎる場合に、シソーラスを基に条件に合ったキーワードを提案することで、利用者と図書検索システムが対話的に検索を進行できる。また、プロフィールによる検索結果のフィルタリング部分では、利用者の理解度と図書の評価から図書の難易度を特定して、キーワードだけでは反映できない利用者の要求を付加することにより、適合した満足度の高い結果が得られる。

### 3.1 シソーラスを用いた問合せ拡張に基づくキーワードの提案手法

シソーラスは言葉の意味的な辞書であり、言葉の概念的な意味と語の位置付けを階層的に表している。WebOPACにおいて、ユーザーが入力したキーワードによって得られた結果の件数が、多すぎるあるいは少なすぎる場合には、さらにキーワードを追加して再検索を行う必要が生じる。このような場合に、検索結果の件数を分析し、語の上位下位の概念的な関係を利用して、次にどのようなキーワードを追加するの適当であるかを示す。つまり、ユーザーが入力したキーワードに上位語を追加すると、検索範囲が拡大し結果の件数を増やすことができ、下位語を追加すると、結果の件数を絞り込むことができであろうという予想に基づいたキーワードを提案するしくみである。問合せ拡張とキーワードを提案するしくみを図 3.2 に示す。具体的には、次のように利用者とシステムが対話的に問合せ拡張を行う。

- (1) 利用者がキーワードを入力する。
- (2) システムがユーザーの入力したキーワードと検索結果の件数を分析して、キーワード候補を提示する。
- (3) 提示されたキーワード候補の中から、利用者が適当なキーワードを追加して再検索する。

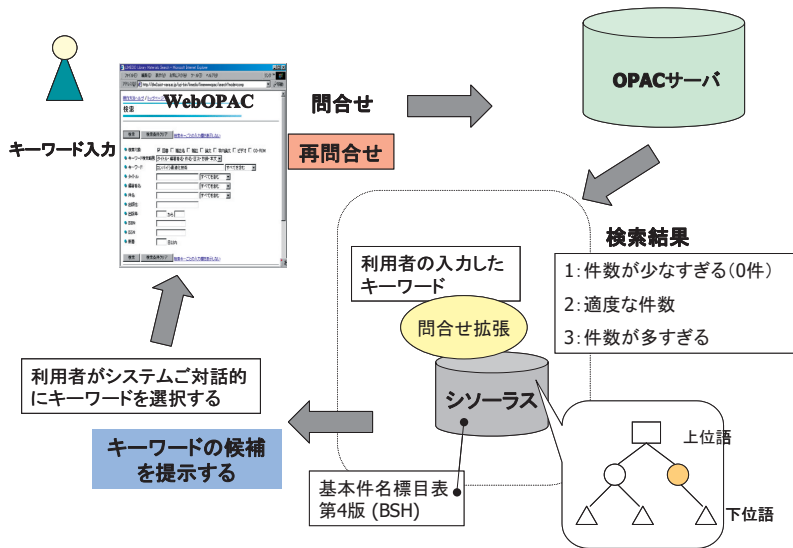


図 3.2 問合せ拡張とキーワードの提案

本研究では、『基本件名標目表』をシソーラスとして使用することとした。前述のとおり、個々の図書には、内容を表す件名（統制語）が付けられており、この件名付与のもとになっている『基本件名標目表』<sup>3</sup>を使用することで、検索精度が高くなると考えたからである。『基本件名標目表』は図 3.2 のように木構造で示ことができ、上位下位の関係が明確に示されていることを利用して、検索範囲の拡大や検索結果の絞込みを補助することが可能であると予想する。ただし、件名には最も特定な語を付与することになっているため、upword-posting（その上位の語も追加して付与する）を行わなければならない。例えば、図 3.3 に示す「データベース」を件名として付与するならば「情報検索」や「情報科学」も追加しておく。

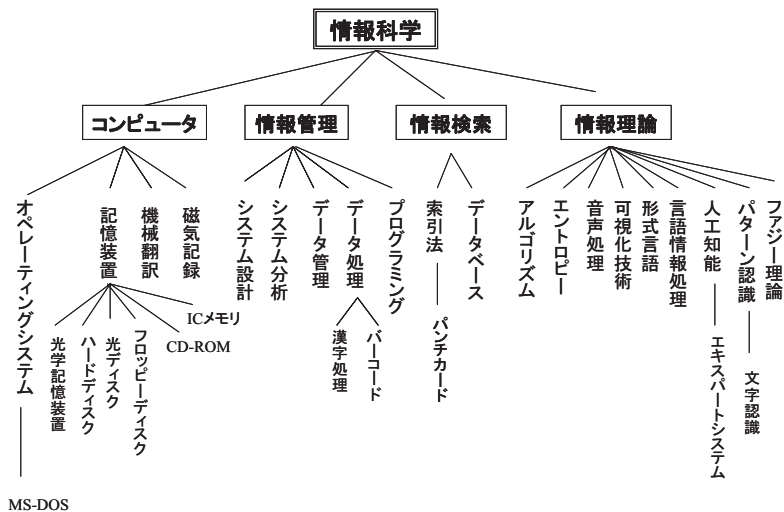


図 3.3 基本件名標目表の例

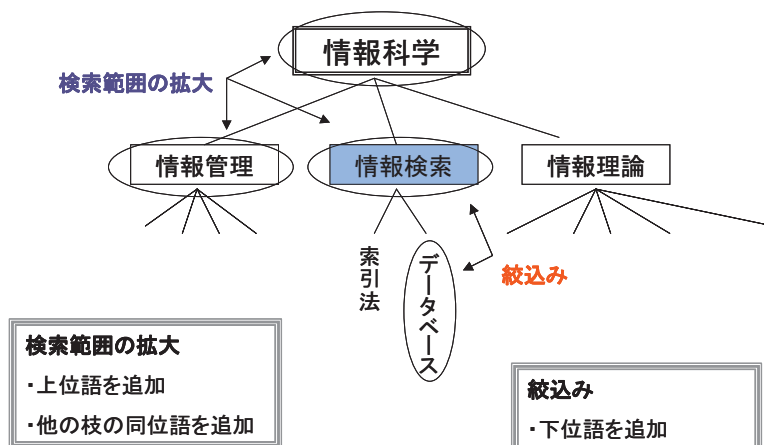


図 3.4 検索範囲の拡大と絞込み

### 3.2 検索範囲の拡大と絞込み

利用者はシステムと対話的に問合せ処理を繰り返す。システムは問合せ拡張の技術を利用して、徐々に利用者の意図にキーワードに近づけるようにする。利用者がシステムにキーワードを与えて検索を実行したときに、検索条件に合致する結果の件数によって以下のような場合が考えられる。

- A) 件数が少なすぎる (0 件)
- B) 利用者が一覧できる適度な件数 (50 件程度)
- C) 件数が多すぎる (100 件以上)

ここで、問題となるのは A) と C) の場合である。A) は問合せの条件が厳しすぎるため、より一般的なキーワードを使用して再検索を行う必要がある。C) の場合は、より詳細なキーワードで再検索するか、新たなキーワードを追加して件数を絞り込む必要がある。

検索条件に合致する件数が少なすぎる場合、検索条件を緩めることによって合致する件数を増やすことができる。『基本件名標目表』においては、階層構造を上にとどることによって、より一般的な語を得ることができる。図 3.3 に示すように、最初のキーワードに「情報検索」を選んだとすると、検索結果が少なすぎる場合には上位にある「情報科学」を追加する。また、他の枝にある同位語の「情報管理」を追加するのが良いと思われる。逆に、件数が多すぎる場合には、下位のより特化した語を追加し件数を絞り込む。具体的には、「情報検索」の下位語には「索引法」や「データベース」が存在するので、これらの語を追加するのが妥当であると考えられる。

### 3.3 実験

『基本件名標目表』を利用した問合せ拡張が、どの程度有効に働くかを検証するために、実験を行った。実験には、奈良先端科学技術大学院大学電子図書館の WebOPAC を使用した。実験環境では、検索対象に、図書・雑誌・Issue title・Papers・Thesis・ビデオ・CD-ROM があり、この中から図書と Papers を選択した。さらに、キーワード選択対象には以下の 3 とおりがあり、

〔タイトル・編著者名・件名〕, 〔タイトル・編著者名・件名・目次・抄録〕, 〔タイトル・編著者名・件名・目次・抄録・本文〕の中から〔タイトル・編著者名・件名〕を選択した。また、問合せは検索キーワードの語数と組合せから 18 個用意した。2 語以上の組合せでは、図 2.3 の例に従って、「情報検索」を基準にした場合の上位語にあたる「情報科学」、下位語にあたる「データベース」、他の枝の同位語にあたる「情報管理」を and または or で追加した。このような検索条件によって実験を行う理由は、上位語や他の枝の同位語を追加することにより検索範囲が広がり、下位語の追加により絞込みが行われるであろうという仮説に基づいている。

- ◆ 検索対象：図書・Papers
- ◆ キーワードの検索対象：タイトル・編著者名・件名
- ◆ 検索条件 1：「情報科学」「情報検索」「情報管理」「データベース」の各々 1 語だけで検索する。
- ◆ 検索条件 2：「情報検索 and または or 情報科学」「情報検索 and または or 情報管理」「情報検索 and または or データベース」の 2 語による組合せで検索する。
- ◆ 検索条件 3：「(情報検索 and または or 情報科学) and または or データベース」「(情報検索 and または or 情報管理) and または or データベース」の 3 語による組合せで検索する。

検索結果を表 3.1 に示す。実験結果から、下位語を and で追加すると件数を絞り込むことができ、上位語を or で追加すると件数を増やすことができるといえる。つまり、追加するキーワード（上位語あるいは下位語）と論理演算子の組合せから、検索結果の絞込みと検索範囲の拡大を可能にすることを証明した。

問合せ	結果(図書)	結果(Papers)
情報科学	137	33
情報検索	44	50
情報管理	32	60
データベース	110	413
情報検索 and 情報科学	6	0
情報検索 or 情報科学	175	82
情報検索 and 情報管理	1	0
情報検索 or 情報管理	75	109
情報検索 and データベース	16	2
情報検索 or データベース	138	461
(情報検索 and 情報科学) and データベース	1	0
(情報検索 and 情報科学) or データベース	115	413
(情報検索 or 情報科学) and データベース	17	2
(情報検索 or 情報科学) or データベース	268	492
(情報検索 and 情報管理) and データベース	0	0
(情報検索 and 情報管理) or データベース	111	413
(情報検索 or 情報管理) and データベース	16	6
(情報検索 or 情報管理) or データベース	169	516

表 3.1 実験結果



### 3.4 複数のシソーラスの使用

『基本件名標目表』は、図書館で扱う広範な分野の図書をカバーしなければならないという理由から、非常に粗く分類されている。これは、『基本件名標目表』があらゆる図書館（公共図書館から専門図書館まで）で利用されることを前提に作成していることに起因する。しかし、実際問題としては、このような粗い分類は専門図書館では役立たない。なぜならば、偏った分野の図書を多く扱う専門図書館では、さらに細かな分類に基づく件名付けを行わなければ、内容から検索することができないからである。つまり、主題検索に使用する件名（統制語）自体が少なければ、利用者のニーズには対応しきれない。このような『基本件名標目表』の弱点を補うためには、分野ごとの詳細なシソーラスと統合的に使用するのが適当であると考えられる。例えば情報科学の分野では、用語の意味を網羅的におさえ、その位置付けを階層構造で明確に記した『岩波情報科学辞典』<sup>4</sup>があり、両者を併用することで効果の向上が見込まれる。

## 4. 利用者と図書のプロフィールを利用した検索結果のフィルタリング

利用者の要求に関わる詳細な情報すべてを1語または数語のキーワードに置き換えることは困難である。いかに高度な WebOPAC であっても、少ない手がかりから利用者の要求を完全に満たす結果を返すことは難しい。つまり、WebOPAC ができることは、キーワードとして入力された語または文字列が出現する箇所を検索して提示することであり、結果が利用者の要求に合っているかどうかまでは保証しない。しかし、利用者にとっては要求に適合していることが最も重要であり、要求と結果のギャップから不具合が生じてしまう。このような不具合を防ぐために、利用者のプロフィールと図書のプロフィールをそれぞれ作成し、検索結果のフィルタリングに使用する。これにより、利用者の理解度と図書の難易度が対応付けられ、利用者の要求を満たす適合性が高いと思われる図書（例えば、前提知識をもたない初心者には専門用語の説明も含まれている入門書）を優先的に結果として返すことができる。

フィルタリングのしくみを図 4.1 に示す。ここでは、同じ関心分野（情報科学）において理解度の異なる利用者（A）、（B）、（C）が同じキーワード「データベース」を入力して検索を行う場合を考える。通常、WebOPAC はキーワードの整合性を基準にするので、3人の検索意図や求める図書が異なっても、同様の結果を返してしまう。ここで、キーワードに加えて利用者自身の情報や図書の内容に関する情報が利用できれば、検索結果を利用者の背景知識や理解度に応じてフィルタリングすることが可能となる。

### 4.1 フィルタリングの概要

本節では、利用者の意図をより正確に反映した検索結果のランキングを実現するための手法について述べる。前述のとおり、既存の WebOPAC では、利用者がシステムに与えたキーワードと図書がもつ書誌情報のマッチングだけに基づいて検索結果を返すため、利用者の関心分野や理解度に相応しくない図書が含まれてしまうことがある。このような問題を解決するために、利用者の関心分野やその分野に関する理解度と図書の難易度を比較して、検索結果のフィルタリング



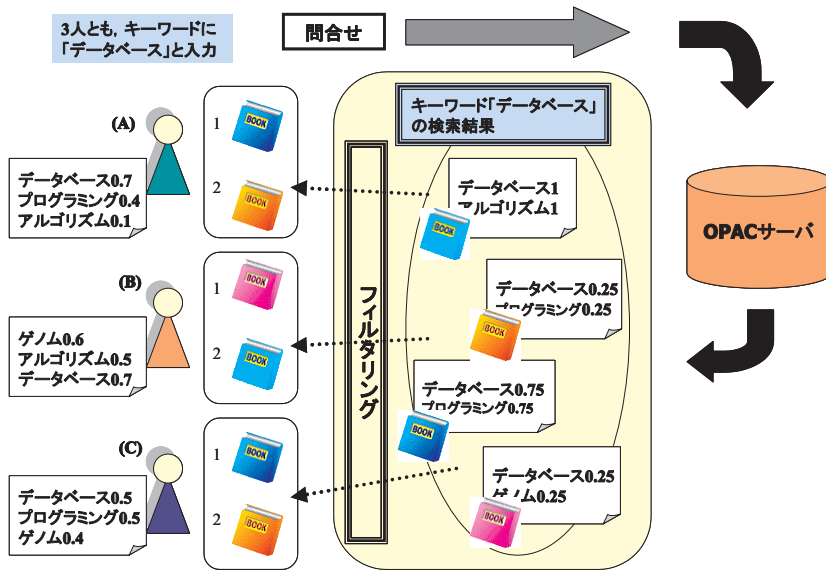


図 4.1 フィルタリングのしくみ

とランキングに利用する。以下では、利用者の関心分野およびその分野に関する理解度を利用者のプロフィール、図書館が対象としている分野およびその分野における難易度を図書のプロフィールと呼ぶことにする。

本手法では、利用者と図書のプロフィールを作成するために、両者の双対の関係に着目した。まず、図書館に所蔵されているすべての図書について、それが対象とする分野とその分野における難易度が予め決まっていると仮定する。このとき、ある利用者が借りた（読んだ）図書が判明すると、その図書の情報から、その利用者の関心分野と理解度を導出することができる。例えば、情報科学に強い関心をもつ利用者は、その分野に関連した図書を数多く借りると予想できる。このとき、情報科学分野の中でも難易度の高い図書を中心に借りているならば、その利用者は高い理解度をもつと考えられる。逆に、入門書を数多く借りているならば、情報科学分野では初学者であると予想できる。利用者の関心分野とその分野における理解度がわかっていると仮定すると、これらの情報から図書のプロフィールを作成することができる。例えば、前述の例に従うと、ある分野における専門家によく読まれている図書は、その分野では難易度の高い図書であると予想でき、初学者によく読まれている図書は入門書であると予想できる。

さらに、利用者が読んだ図書に対して残した評価（書評・感想）を利用して、情報を補う。利用者の残した評価を利用することから、図書のプロフィールをより実状に即したものにあらためることができるばかりか、利用者が学習を進めるにつれて理解度が変化する様子をプロフィールに反映させることが可能となる。

提案手法では、予め充分な量のプロフィールがデータベースに蓄積されていることを前提としている。運用初期では、これらの情報は十分に集まっていないので、代替手段よって得られるおおよその値を利用する。利用者のプロフィールを計算するためには、関心分野と理解度を得るた

めの簡単なテスト（アンケート）が利用できる。図書のプロフィールに関しては、一般の書誌情報に加えて、書籍案内や出版情報に示されている編著者や出版者が設定した難易度の情報が利用できる。

## 4.2 利用者と図書の対応

利用者と図書の関係を図 4.2 に示す。利用者と図書の間には双方向の関係が成り立つ。利用者の図書に対する評価は、利用者の理解度を基準とする図書の絶対評価であるとともに、図書を基準とする利用者の相対評価である。すなわち、図書のプロフィールには図書の内容を表す件名やカテゴリを表す分類の情報があり、利用者が選んだある図書から、利用者の関心分野を推定することができる。また、利用者の残した評価のデータから図書の内容と利用者の知識を比較し、利用者のその分野における理解度を推測する。これは、司書の行うレファレンスインタビューに相当すると考えられる。図書の内容が変化することはないので、図書のプロフィールも変化することのない固定のものである。一方、利用者は徐々に理解度が深まり知識が増加していくと考えられ、以前「難しい」と思った図書を「易しい」と思うようになることも起こり得る。つまり、利用者のプロフィールは固定のものではなく、時間の経過とともに変化する。このような利用者と図書のプロフィールの関係を利用すると、そのときの利用者の要求に合った図書を検索結果として提示できる。

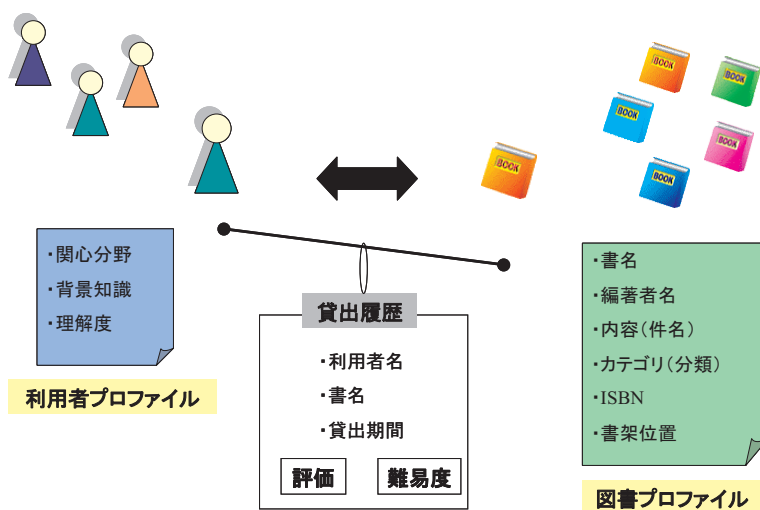


図 4.2 利用者と図書の関係

## 4.3 プロファイルの構築

利用者のプロフィールには、関心分野・背景知識・関心分野における理解度等の情報を、図書のプロフィールには、書名・編著者名・内容（件名）・カテゴリ（分類）・ISBN・配架位置等の情報を格納する。

図書のカテゴリは、図書分類に基づいて同じカテゴリに属するものについて処理を行う。以下では、議論を簡単にするために、「情報科学」分野の図書に限って話を進めたい。分類のカテゴリは前述の「情報科学」の一例を見てもわかるように、かなり広範であり、利用者の関心分や図書の記述内容を明確に定義するためには、さらに分類を行う必要がある。本手法では、件名および書名に含まれる自立語を利用する。換言すると、ある図書の内容は、その件名および書名に含まれる自立語の頻度情報によって表現されているとみなす。

各々のプロフィールは、その図書に対する「評価（難易度）」と「感想（推薦するかしないか）」を追加する。基準は次のように決める。

(1) 評価（難易度）

3：難しい（難しすぎて不相当）

2：適当（ちょうど良い）

1：易しい（易しすぎて不相当）

(2) 感想（推薦する／推薦しない）

1：推薦する

0：推薦しない

利用者のプロフィールについては、利用者の関心分野ごとの理解度を求める。利用者のレベルは図書の難易度との関連性があるため、次の方針に基づいて評価値を求める。すなわち、同じ「難しい」という評価であっても、その分野の専門家（上級者）の「難しい」と初学者（初級者）の「難しい」とでは基準が異なるため、そのままの評価を評価値とすることはできない。そこで、図4.3に示すように相対的な評価値を導入する。

5：非常に難しい（上級者が「3：難しい」と評価）

4：難しい（上級者が「2：適当」、中級者が「3：難しい」と評価）

3：適当（上級者が「1：易しい」、中級者が「2：適当」、初級者が「3：難しい」と評価）

2：易しい（中級者が「1：易しい」、初級者が「2：適当」と評価）

1：非常にやさしい（初級者が「1：易しい」と評価）

このように、利用者が残した評価と図書の難易度から5段階の評価が得られる。具体的なプロフィールの算出方法を以下に示す。

#### 4.3.1 利用者プロフィールの算出

利用者が借りて読んだ図書に関して、図書分類に従ってカテゴリ分けを行う。同じカテゴリに入った図書に関して、カテゴリごとに以下の処理を行う。まず、同じカテゴリの中での詳細な関心分野の特定を行う。これには、図書の書誌情報から抽出した「書名」と「件名」の情報を用いる。具体的には、書名および件名から形態素解析によって得られた自立語の頻度ベクトルを関心分野とみなす。

図書館には、 $n$ 冊の図書  $b_i (i = \{1, 2, \dots, n\})$  が存在するとし、利用者が  $m$ 人  $u_j (j = \{1, 2,$

- 利用者のレベルと評価から  
図書の難易度を推測する

		利用者のレベル		
		3	2	1
図書の難易度	5	3		
	4	2	3	
	3	1	2	3
	2		1	2
	1			1

- 図書のレベルと評価から  
利用者の理解度を推測する

		図書のレベル		
		3	2	1
利用者の理解度	5	1		
	4	2	1	
	3	3	2	1
	2		3	2
	1			3

図 4.3 相対的な評価値

...,  $m$ ) であるとする。また、すべての自立語の集合を  $W = \{w_k \mid k = \{1, 2, \dots, l\}\}$  とする。

利用者が借りた図書の履歴を表す  $m \times n$  行列  $R$  を考える。  $R$  の  $(j, i)$  成分  $r_{j,i} (\in \{0, 1\})$  は、利用者  $u_j$  が図書  $b_i$  を読んだか否か（読んでいれば  $r_{j,i} = 1$ ，読んでいなければ  $r_{j,i} = 0$ ）を示す正整数である。また、利用者が図書に残した評価（理解度）を示す  $m \times n$  行列  $E$  を考える。行列  $E$  の  $(j, i)$  成分  $e_{j,i} (\in \{1, 2, 3\})$  は、利用者  $u_j$  が図書  $b_i$  に残した評価である。次に、利用者の関心分野における理解度を  $C$  とする。これは  $n \times l$  行列であり、行成分が利用者を列成分が自立語を表している。すなわち、行列  $C$  の  $(j, k)$  成分  $c_{j,k}$  は利用者  $u_j$  の自立語（関心分野）  $w_k$  における理解度を  $[0, 1]$  の実数値（数値の大きいほうが理解度が高い）で表したものである。次に、図書のタイトル（件名）に含まれる自立語との対応関係を行列  $Wtitle$  ( $Wsubject$ ) で表すものとする。これらは  $l \times n$  行列となる。以下ではタイトルだけを対象として議論するが、実際には件名についても同様の処理を行う。これは、  $R \times Wtitle^T$  によって求めることができる。結果の各行が利用者の関心分野を表している。

次に、利用者が図書を読んで与える評価（3：難しい，2：適当，1：易しい）を利用して、関心分野における理解度を算出する。これは前述の  $C$  を求めることに相当する。それぞれの関心分野（自立語）に関して、理解度の平均を求める。計算の手順を以下に示す。

- (1)  $E \times Wtitle^T$  を計算する。結果として得られる行列の  $(j, k)$  成分は、利用者  $u_j$  の自立語（関心分野）に対する評価値の総和を表す。
- (2) 各成分を、  $E$  の各行の非ゼロ成分の数で割って平均値を求め、さらに  $[0, 1]$  に正規化する。以上から  $C$  が求められる。

例を図 4.4 に示す。利用者 (A) は 3 冊の図書を借りて読み各々を評価する。これによって、利用者 (A) の関心分野は、データベース，プログラミング，アルゴリズムであることがわかる。また、利用者(A)は、データベースの難易度が 0.7 でアルゴリズムの難易度が 0.2 の図書に対して、

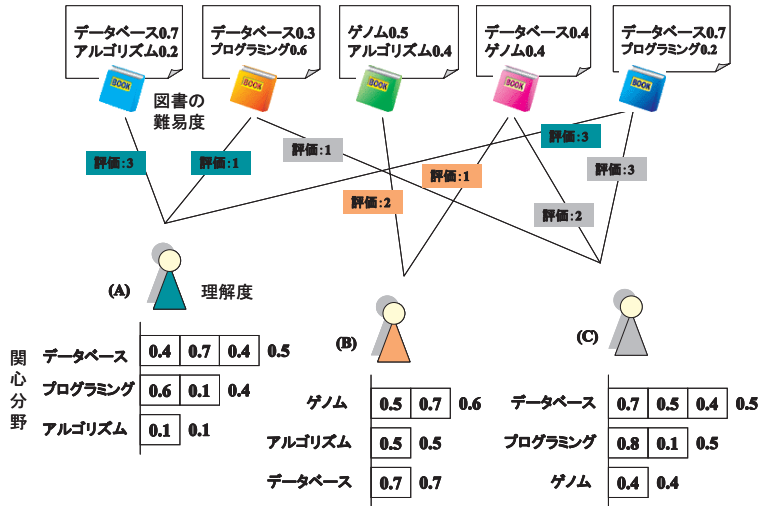


図 4.4 利用者プロフィールの算出

「3：難しい」と評価したことから、データベースの理解度が 0.4 でアルゴリズムの理解度が 0.1 であることがわかる。他の 2 冊においても同様の計算を行い、平均をとって利用者 (A) のプロフィールを算出する。

#### 4.3.2 図書プロフィールの算出

利用者のプロフィールと貸し出し履歴から、図書のプロフィールを求める。具体的には、図書の書名や件名含まれる自立語に関して、利用者の理解度と評価値から導出し、平均を求める。ある自立語  $w_k$  に関する評価値の平均は以下のように求められる。

- (1)  $E' = E \times W_{title}^T$  を計算する。 $E'$  の  $(j, k)$  要素を  $e'_{j,k}$  とする。
- (2) 各図書の自立語  $w_k$  に関して  $\sum_j \frac{e'_{j,k} + 2 \cdot c_{j,k}}{4}$  を求める。

この結果、図書の属する分野における難易度が得られる。

例を図 4.5 に示す。利用者 (A) の関心分野と理解度が、データベース 0.7、プログラミング 0.4、アルゴリズム 0.1 の場合に、利用者 (A) がデータベースとアルゴリズムについて書いてある図書に対して「3：難しい」と評価したならば、この評価により、データベースとアルゴリズムの難易度がそれぞれ 1 であることがわかる。このように図書のプロフィールが算出でき、図 4.6 に示すように利用者の理解度から図書の難易度を推定することができる。

## 5. 関連研究

本節では、本研究の遂行にあたり参考とした関連研究にふれておく。

- ・目次情報の利用による図書検索システム

書誌情報だけによる図書検索システムの不具合については前述したが、キーワードを用いた検索で最も重要なことは、いかにして図書の内容を的確に表すキーワードを選択するかという点であ

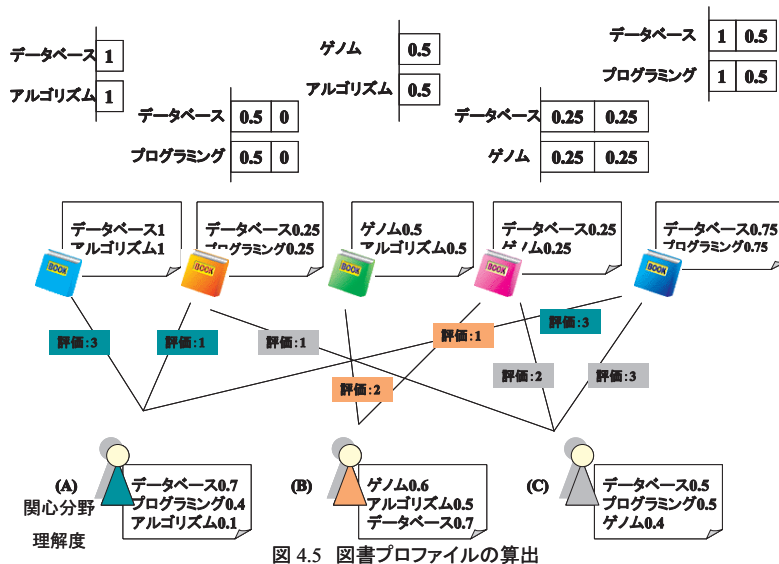
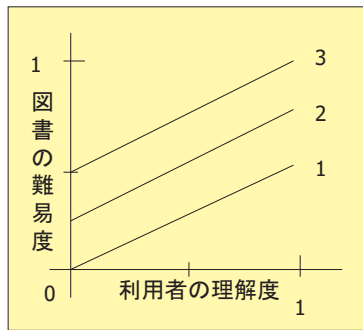


図 4.5 図書プロフィールの算出

• 利用者→図書

- 利用者の理解度から 図書の難易度を推定



• 図書→利用者

- 図書の難易度から利用者の理解度を推定

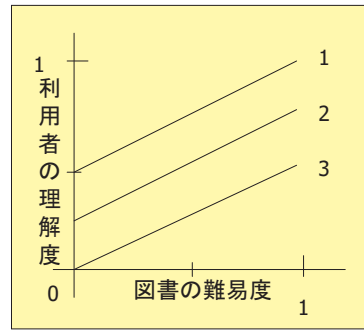


図 4.6 利用者の理解度図書の難易度の関係

り、内容を表すキーワードを何から選ぶかという問題があった。このような問題を解決するために、長尾等<sup>5</sup>は目次情報を利用した図書検索手法を提案し、黒橋等<sup>6</sup>により目次情報を利用した図書検索システムを構築した。

• レファレンスサービスの自動化

黒橋等<sup>7</sup>は、大学のコンピュータ実習室でのトラブルに対応する対話的ヘルプデスクシステムの構築を基に、自然言語処理技術を用いて、図書館のレファレンスサービスを自動化する研究<sup>8</sup>に取り組んでいる。このシステムは、知識ベースと対話管理部により構成される。知識ベースは、「～に関する図書を探している」等の典型的な質問に答えるために、① 国立国会図書館分類デー

データベースを使用して質問内容を図書分類のカテゴリへ一般化する部分と②広範な質問に対応するためのレファレンスサービスに関連した知識を辞書のような形式で記述した自然言語知識ベースから成る。

- ・ユーザ履歴を活用した検索システム

川前等<sup>9</sup>は、キーワード入力による検索システムを利用する際に、利用者が適切なキーワードを入力しなければ目的に合った文書を得られない事実から、キーワードの入力に対する利用者のストレスに着目し、利用者の検索を支援するキーワードディレクトリの作成と効果を示した。利用者は自らが入力したキーワードを基点に、キーワードディレクトリをたどることで、キーワードの変更や追加が容易に行える。

- ・読書感検索システムの構築

宮田等<sup>10</sup>は、ナレッジマネジメントの手法を用いて、利用者の暗黙知を取り込んだ知識ベースを形成し、図書館利用サービスに活かそうとしている。つまり、個々の利用者がもっている図書館を利用するための知識＝暗黙知（図書館の利用方法、検索システムの利用技術、レファレンスサービスの活用法、読書感）を利用しようというものである。例えば、「元気の出るような本」や「感動するような本」というような、要求があまりに主観的すぎて主題分析が困難であり、既存の図書検索システムで対応しきれない場合を想定して、利用者の読書後の感情（面白かった、元気が出た、ドキドキした、悲しくなった等）を、利用者の分身（Avatar）としてつくりあげ、匿名性を備えた読書感検索システムを構築した。

- ・情報フィルタリングシステム

酒井等<sup>11</sup>の開発した情報フィルタリングシステム NEAT では、利用者が関心をもつであろうと予測した電子記事をメールで配信する。質の高い情報フィルタリングサービスを実現するためには、不要な記事を配信することや必要な記事を見落とすことを最小限に食い止めなければならない。関心があることとその記事が有用か不用かが、個々の利用者で異なることや時間の経過とともに利用者の関心が増減することに対応するために、適合性フィードバックによってプロフィールの修正を行う。

- ・情報推薦システム

浅川等<sup>12</sup>は、利用者の嗜好が多岐にわたっているにも関わらず、常に固定の嗜好を中心として、傾向の近い情報だけが推薦されてしまう不具合を改善するために、利用者の嗜好の部分的な類似性を利用して、利用者が新しく興味をもつことができる情報を推薦する手法を提案した。

- ・WebOPAC のカスタマイズ

渡部等<sup>13</sup>は WebOPAC のユーザインタフェースを携帯電話用に開発した。WebOPAC を携帯できる便利さは、巨大な書庫を有する図書館や複数の分館をもつ図書館などで有効であり、フィルタリング機能を付加することでカスタマイズが可能となれば、個人指向 WebOPAC が実現できる。

## 6. まとめと今後の課題

本稿では、計算機技術の進歩やインターネットの普及により、変化してきている図書検索手法



やこれまで図書館で司書が果たしてきた役割、今後ますます増加することが予想できる電子図書館での図書検索における利用者支援の在り方について考察したうえで、シソーラスによる問合せ拡張とキーワードの提案及び利用者と図書のプロフィールを用いた図書検索手法を提案した。

今後は、実験システムを構築し、提案手法の有効性を検証する。複数のシソーラスを統合的に利用すること、プロフィールの作成を容易に行うように工夫したいと考えている。さらには、同じ関心分野に興味をもつ利用者同士で検索履歴の共有や活用を行えるような知識共有システムへ発展できるように研究を進めていく予定である。

また、実際に図書館で運用することを念頭においた場合には、利用者の個人的な情報を利用することへの配慮が必要であると考えられ、プライバシーの保護についての検討も充分に行わなければならない。

### 《謝辞》

本研究の遂行にあたり、多くの方々に大変お世話になりました。御礼申し上げます。

奈良先端科学技術大学院大学 情報科学研究科 マルチメディア統合システム講座（現データベース学講座）の植村俊亮教授、吉川正俊助教教授（現名古屋大学情報連携基盤センター教授）、天笠俊之助手、波多野賢治助手には大変お世話になりました。厚く御礼申し上げます。

また、1年6ヶ月という長期間の研修を快諾いただいた本学関係者のみなさまに心より御礼申し上げます。

### 参考文献

- 1 原田 勝, 田屋裕之: 電子図書館, 勁草書房, 1999.
- 2 石坂憲司: 情報技術を使用した利用者教育・利用者支援: カリフォルニア大学バークレー校, スタンフォード大学訪問記, 大学図書館研究, 59号, pp. 63-70, 2000.
- 3 基本件名標目表第4版, 日本図書館協会, 1990.
- 4 長尾 真: 岩波情報科学辞典, 岩波書店, 1990.
- 5 長尾 真: 目次情報などを利用した図書・文献検索方式, 情報の科学と技術, Vol. 42, No. 8, 1992.
- 6 黒橋禎夫, 萩原典尚, 長尾真: 目次情報を利用した図書検索システム, 情報処理学会研究報告, 97-FI-45, pp. 27-32, 1997.
- 7 黒橋禎夫, 日笠 巨, 藤井綱貴: 入力質問と知識ベースとの柔軟なマッチングに基づく対話的ヘルプシステム, 情報処理学会研究報告, 99-NL-134, 1999.
- 8 白木伸征, 黒橋禎夫: 自然言語入力と柔軟な照合による図書検索システム, 情報処理学会論文誌, Vol. 41, No. 4, pp. 1162-1170, 2000.
- 9 川前徳章, 青木輝勝, 安田 浩: ユーザ履歴を活用した検索システム, 電子情報通信学会技術研究報告, データ工学, DE2000-37, pp. 113-120, 2000.
- 10 宮田聡子, 小林久恵, 桑田てるみ: 図書館空間におけるナレッジ・マネジメントシステム - 読書感検索システムの構築 -, 情報メディア学会第2回研究会発表要項, 2001.  
<http://wwwsoc.nii.ac.jp/jsims/2001-3-21genkou06.pdf>

- 11 酒井哲也, Gareth J. F. Jones, 梶浦正浩, 住田一男: ユーザの要求に応じた情報フィルタリングシステム NEAT のプロファイル生成, インタラクシオン '98 論文集, pp. 149-152, 1998.
- 12 浅川智文, 川口知昭, 山口陽一, 鎌原淳三, 下条真司, 宮原秀夫: ユーザの嗜好の部分的な類似性を利用した情報推薦手法の提案, 第 12 回データ工学ワークショップ (DEWS01), 2001.
- 13 渡部聡彦, 武井純孝, 杉本雅則, 中川裕志: I-mode, J-Phone 向け OPAC 検索システム  
- 東京大学附属図書館における実装 -, 情報処理学会第 62 回全国大会予稿集, 2001.