

利用者のプロフィールを考慮した連想検索 OPAC の構築

當 山 仁 健*

OPAC での検索の際、複数の利用者が同一の検索質問を用いたとしても、利用者のプロフィール（知識世界、背景等）が異なれば、検索質問で表現を試みている概念は異なり、情報要求も異なることが多い。そこで、検索結果を利用者のプロフィールに合わせて調整することにより、利用者満足度の高い OPAC の構築が期待できる。また、従来の文字列検索による OPAC では柔軟性に乏しく、情報検索の道具としては不十分である。これらを踏まえ、沖縄国際大学図書館において、適合性判断基準の一つとして貸出履歴を利用した OPAC を実装した。本システムでは柔軟な検索結果が得られるよう、従来の文字列検索ではなく連想検索技術を用いている。

キーワード：OPAC、情報検索、貸出履歴、貸出情報、連想検索、利用者志向、沖縄国際大学図書館

1. はじめに

インターネットの普及により、人々の扱う情報の量は爆発的に増加した。これに対処するため情報検索関連技術が盛んに研究されている。新しい技術が次々と提案され、さまざまな分野の情報検索システムで実用化されている。一方、一般的な OPAC は、出現当初から殆ど変化がなく、いくつかの問題を抱え続けている。本稿では、従来の OPAC の不十分な点を補うことを目標として構築した「利用者のプロフィールを考慮した連想検索 OPAC」の概要について述べる。

2. 従来の OPAC の抱える問題

一般的な OPAC では、書誌事項に検索質問がマッチする資料を検索結果として提示する。このようなシステムは探している資料が明確に特定できており、ISBN やタイトル等資料の同定に必要な書誌事項が分かっている場合には有効である。しかし、どの資料に必要な情報が載っているか特定できていない状態で情報を探する場合には不向きである。実際、利用者は必要な情報が載っている資料が明確でない状態で、問題解決に必要な未知の情報を求めて図書館を利用することも多い。OPAC を情報検索の道具として捉えたと、現在、一般的に利用されている OPAC には以下のような問題点がある。

（問題点 1）検索者に厳密な検索質問を要求する

利用者はまず、自分の探している情報に関連しそうな検索語を選択し、検索式を駆使して検索質問を作成し、OPAC に入力する。しかし、実際には未知の情報を求めている利用者が、対象についての十分な知識があり、適切なキーワードを検索語として選択できるとは限らない。また、検索語はシステム内の索引が持っている情報と一言一句違いのな

い表現である必要がある。さらに、同意語が存在する場合には、そのすべてを入力しないと検索漏れが発生してしまう。

（問題点 2）検索結果の表示順に意味がない

検索結果は検索語が検索対象に「含まれている」か「含まれていない」かで判断され、順不同なリストとして利用者に提示される。利用者は自分の必要とする情報かどうか検索結果を確認する必要がある。しかし、検索結果の量が多くなると確認作業自体が負担になる。これを絞込むためには AND や NOT による絞り込み検索が利用されるが、この過程で検索漏れが発生してしまうケースも多い。

（問題点 3）利用者のプロフィールを考慮しない

利用者のプロフィール（知識世界、背景等）により、個々の利用者の求めている情報は異なる。この情報要求の違いは検索語だけから読み取ることは難しい。例えば、「アルコール」という検索語で検索する場合を考える。検索する利用者が医学専攻の学生であれば、アルコールの依存性や薬効についての情報を求めていることが多いだろう。それに対して化学専攻の学生の場合、アルコール自体の性質や構造に関する情報を求めているケースが考えられる。このような場合、化学専攻の学生にとって「アルコールの依存性」についての文書が検索結果として提示されるとノイズになる。このように、利用者のプロフィールが異なれば、同一の検索質問であっても表現している概念は異なり、求めている検索結果も異なることが多い。しかし、現在の OPAC は利用者のプロフィールを考慮せず、一様に検索結果を提示する。

3. 本システムのアプローチ

沖縄国際大学図書館では、これら問題に対処するために利用者のプロフィールを考慮した連想検索 OPAC を実装し、2005 年 10 月 1 日より実運用している。システムは Web アプリケーションとして実装されており、学内からであればブラウザを利用してすべての端末から利用することが可能である。

*とうやま よしたけ 沖縄国際大学

〒901-2701 沖縄県宜野湾市宜野湾 2-6-1

Tel. 098-892-1111

(原稿受領 2006.8.24)

3.1 連想検索による自然文での検索

本システムでは、一般的な OPAC で利用されている文字列検索ではなく、WebcatPlus¹⁾等で用いられている連想検索²⁾を利用している。連想検索では検索質問は自然文で入力することが可能であり、利用者に厳密な検索質問を要求しない。また、検索質問と検索対象文書の類似度を算出することが可能である。この類似度の順に検索された文書を表示することで、利用者の検索質問に内容的に近い順で検索結果を提示することが可能である。類似度は、検索質問と検索対象の文書と検索質問の単語の出現頻度を基に算出される。連想検索機能の実装には汎用連想計算エンジン GETA (Generic Engine for Transposable Association)³⁾を用いた。

3.2 書誌情報には現れない情報の算出

利用者のプロフィールが異なれば必要とする資料は異なる。また、同じ主題を取り扱っている資料でも、それぞれの資料の評価（人気）には違いがある。これらは利用者が決定することであり、書誌情報のみを客観的に分析するだけでは算出できない。しかし、これらの情報も検索結果の表示順を決定する際の重みとして反映できれば、利用者にとって有益と感じられる資料を検索結果の上位に提示することが可能となり、利用者の負担の軽減が期待できる。

本システムでは利用者をグループ化し、利用者の所属するグループの個々の資料に対する貸出件数を基に利用者のプロフィールに基づく重みを算出している。そして、この数値を検索結果の選択、表示順を決定づける材料として利用することで、書誌情報だけでは読み取れない情報を検索結果に反映させている。

3.2.1 利用者のグループ化方法の検討

利用者のグループ分けの方法を設計する際、まず考えられるのはそれぞれの利用者の利用傾向（貸出履歴）を分析し、似ている利用者をシステムの判断で動的にグループ化する手法である。これは暗黙的に利用者モデルを構築する手法といえる。この手法を用いれば、個々の利用者の実際の嗜好（読書履歴）を考慮してグループ分けが行えるので、柔軟なグループ分けが期待できる。実際に暗黙的に利用者モデルを構築する手法を採用し、運用することを考えると、システムは全利用者の貸出履歴を保有する必要がある。また、利用の際には個人を特定するための認証作業が必要になる。

これに対し、システムの設計者が前もっていくつかのグループを用意しておき、利用者自身に自分の属するグループを選択させる手法が考えられる。これは明示的に利用者モデルを構築する手法といえる。この手法では暗黙的な手

法ほど柔軟なグループは作れないが、認証を必要とせず、それぞれのグループのそれぞれの資料に対する評価のみを保有すればよいと、個人と貸出履歴情報が結びつくような情報を保持しない状態⁴⁾で実装できる。

貸出履歴には個人の思想、信条が読み取れるようなセンシティブな情報を含まれているため、そのままの状態ですべて保持し、利用することは問題となる恐れがある。さらに、利用の度に認証処理を要求することは利用者の負担になる。これらの理由や実現可能性を考慮し、本システムでは利用者自身が明示的に所属するグループを選択する手法を採用している。用意するグループの単位としては、学部、学科、学年、専攻などの利用が考えられるが、あまり大きなまとまりだとその効果が実感できなくなってしまう。本システムでは、保有しているデータ量から、効果が実感できる単位を検討した結果、学科単位を採用している。

3.2.2 利用者のプロフィールに基づく重み

単純に貸出件数を利用者のプロフィールに基づく重みとして利用することはできない。所蔵期間が長い資料は貸出件数が多いのは当然であり、そのような資料の重みは下げ必要がある。本システムでは「利用者の属するグループでの貸出回数」を「所蔵期間」で除算することで所蔵期間の長い資料の重みを下げている。

3.3 連想による類似度と利用者プロフィールによる重みの検索結果の表示順への反映

本システムではまず連想計算を行い、検索質問との類似度順に最大 200 件の書誌の ID とそれぞれの書誌の類似度を取得する。それぞれの書誌の類似度に対し、利用者プロフィールに基づく重み²⁾を算出し、掛け合わせた数値を適合度とする。適合度の高い順に最大 50 件の資料を利用者に検索結果として提示している。

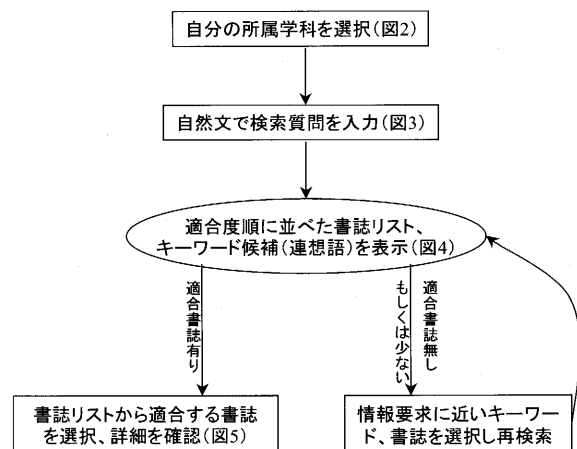


図1 利用者の検索作業手順

¹⁾本システムでは、それぞれのグループのそれぞれの書誌に対する貸出件数をグループの資料に対する評価として保有し、利用している。利用者 ID 等の個人の特定につながる情報は保持していない。

²⁾所属するプロフィールでの該当する書誌の貸出件数に基づく値。1以上の値をとる（所属するプロフィールでの貸出が無い場合、利用者のプロフィールによる重みは1になる）。

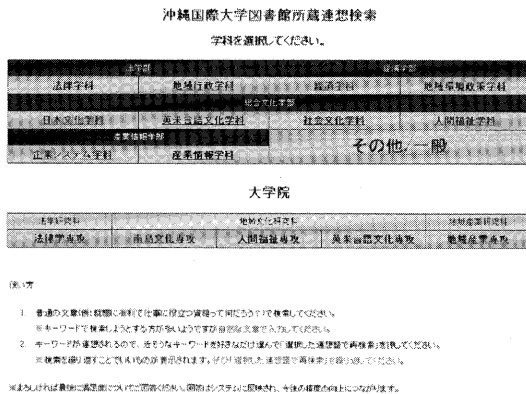


図 2 学科選択画面

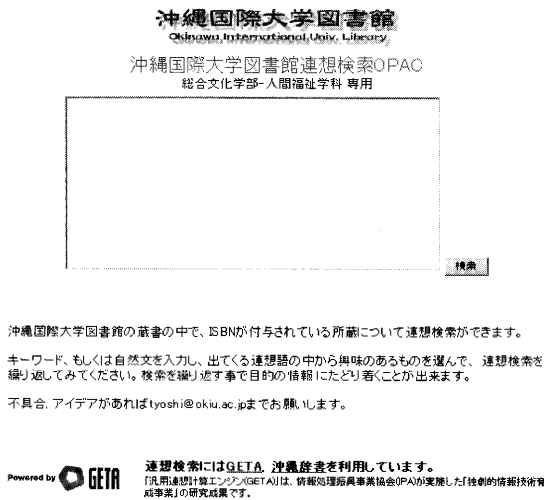


図 3 検索質問入力画面

人間福祉学科

福祉・介護業界で働くためのベスト視覚 20
福祉の進学・就職ガイド
社会福祉士受験ワークブック
資格試験のすべて
臨床心理士になろう

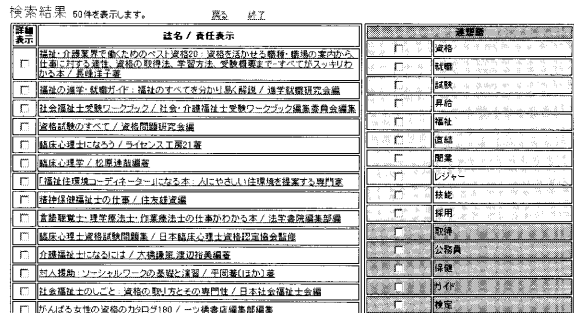


図 4 検索結果リスト表示画面例

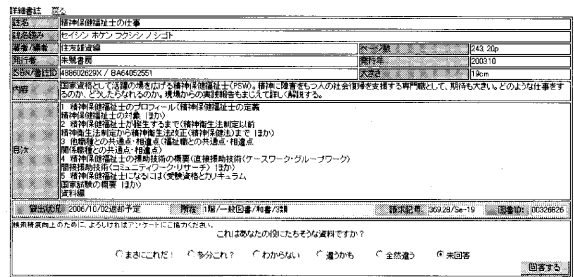


図 5 詳細書誌表示例

経済学科

就職・転職のための最強の資格
資格試験のすべて
資格が面白いほど取れる本
ここが知りたい！FP ファイナンシャルプランナー
ツアーコンダクターになるには

図 6 検索結果例（検索語：就職に役立つ資格）

4. システムの動作例

利用者は図 1 のような手順で操作を進める。インタフェースは WebcatPlus を参考にした。利用者はまず自分の所属する学科を選択する（図 2）。このとき、学科による偏りが不要な場合は「その他/一般」を選択する。次に図 3 のような該当する学科専用の検索質問入力画面で自然文で検索質問を入力する。検索結果はリストとして表示される（図 4）。検索結果の中に情報要求に適合しているものがあれば、「詳細表示」をチェックすることで、選択したすべて

の書誌の詳細を画面に表示し、比較検討することができる（図 5）。また、適合する資料に類似している資料を検索することもできる。思うような結果が得られなかった場合は関連しそうな関連語にチェックをつけて再度検索することができる。

4.1 検索結果例

検索例として、実際に所属学科を選択した上で「就職に役立つ資格」という検索質問で検索した場合の上位 5 件の検索結果を図 6 に示す。利用者の所属を人間福祉学科とし

表1 ソフトウェア構成

動作 OS	Linux (Fedora Core4)
Web サーバ	Apache 2.0.54
データベース	PostgreSQL 8. 0. 4
連想計算エンジン	GETA 第2版
形態素解析エンジン	ChaSen 2.3
辞書	IPA 辞書/沖縄辞書
書誌データ	日外アソシエーツ BOOK

表2 ハードウェア構成

CPU	Pentium4 2.2GHz
メモリ	640MB
HDD	40G

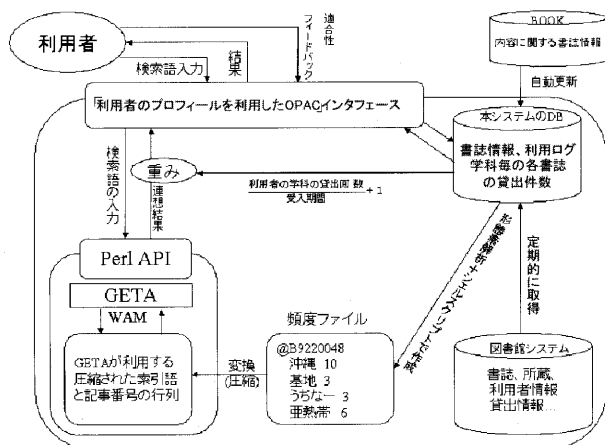


図7 システム構成図

た場合、実際に福祉に関係している資料が検索結果として表示されている。また、経済学科学士の立場で検索を行った結果には一般的な資格試験に関する書籍も表示されているが、下線の図書のように、実際に学科に関係すると思われる資料も検索結果の上位に表示されている。このように、利用者のプロフィールによって柔軟に検索結果を調整できている。

5. システム環境と利用した道具

システムは表1、表2のような環境で実装した。この章では利用したツールのいくつかについて説明する。実際のシステムの構成は図7に示す。

5.1 汎用連想計算エンジン GETA

文書の類似度による連想検索の必要性は古くから認識されていたが、実装には高度な理論とプログラミング能力が必要だった。しかし、GETAの登場に伴い、高速で柔軟な連想検索が実装できるようになり、さまざまなサービス⁴⁾が実用化されている。GETAの提供する機能はPerlを利

用して呼び出すこともできるので、Perlに関する知識があれば、連想検索の詳細な理論自体は知らずとも、とりあえず連想検索を実装することができる。公開サイトにチュートリアル⁵⁾が用意されているので、じっくりと読み進めていけば、一通りの連想検索の機能を実装することができる。

5.2 書誌データ、辞書データ

連想検索は単語の出現頻度を基に行われる。大学図書館でよく利用されているNACSIS-CATの書誌情報のみでは索引として情報が少なく、連想検索の効果を実感しづらい。そこで本システムでは、NACSIS-CATの書誌情報に加え、内容に関する書誌情報（目次と書籍の帯に記載されている要旨）を持っている日外アソシエーツの「BOOK」データベースを購入し、索引として用いている。

また、沖縄には通常の辞書には登録されていないような固有名詞や読みの漢字が多数ある。これらに対しては形態素解析の精度が悪くなり、検索に悪影響が出る恐れがある。これに対応するために形態素解析システム ChaSen⁶⁾の辞書として沖縄辞書⁷⁾を組み込み、利用している。

6. 利用者の満足度を基にしたシステムの評価

6.1 アンケートによる評価

図5の詳細書誌表示画面の中に「これはあなたの役に立ちそうな資料ですか?」という設問を表示し、任意で回答を求めた。回答の選択肢は「全然違う」「違うかも」「わからない」「多分これ」「まさにこれだ」の5つを用意した。「まさにこれだ」と「多分これ」と回答した利用者を検索結果に満足した利用者とし、図8に示したように約72%の利用者が満足したと回答している。

利用件数 (A)	5,747 件
アンケート回答総数 (B)	425 件
アンケート回答率 (A/B)	7.4%

質問:「これはあなたの役に立ちそうな資料ですか?」

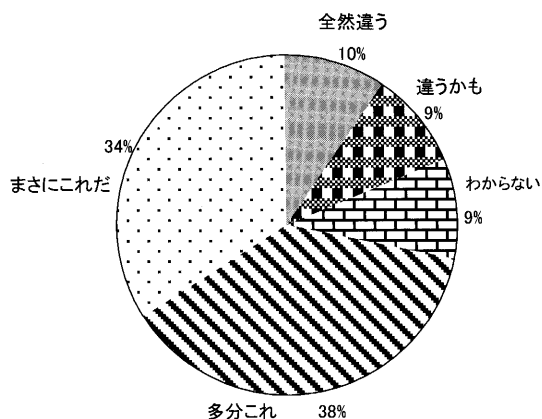


図8 アンケート回答状況 (2005/10/30～2006/2/4)

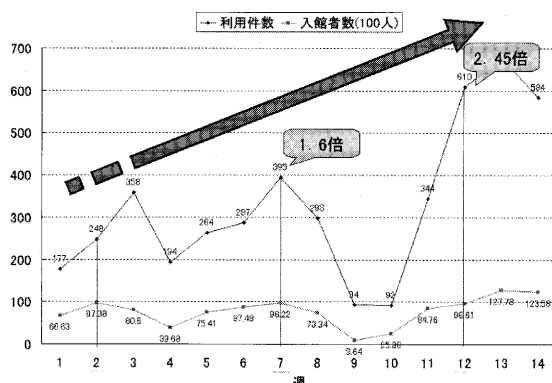


図9 利用者数の推移 (2005/10/30~2006/2/4)

6.2 利用者数の推移による評価

次に、図9の利用者数の推移からシステムの有用性に対する評価を考察する。一般的に、新たなサービスを導入すると、最初はものめずらしさで利用されるが、利用者が有益さを実感できるようなものでない限り、利用者は繰り返して利用しようとは思わない。つまり、導入後、ある程度の時間が経過した後まで利用者数が増加していれば、利用者はシステムの有用性を実感しているとみなすことができる。

入館者数は時期により変動するため、利用者数の推移を一律に比較することはできない。そのため、入館者数がほぼ等しい2, 7, 12週目の利用者に着目して比較する。入館者数9,738名、利用件数248件だった2週目に対し、同様の入館者数(9,822名)の第7週では利用件数が395件になっている。これは2週目の約1.6倍である。また、入館者数9,861名の第12週目には利用件数が610件と、第2週目の2.45倍の利用者がいる。このように、利用者数は右肩上がりが増えていくことから、本システムはものめずらしさで利用されたのではなく、有効性を実感し利用されているということが言える。

7. おわりに

OPACが単なる同定のためのツールという位置づけを越え、利用者の問題解決を手助けできるツールになるためには、利用者が真に必要な情報を提供する機能が必要である。そのためには、現在利用されているOPACのように、誰にでも一様な検索結果を返すだけでなく、利用者に

よって異なる情報要求を柔軟に解釈し、検索結果に反映したり、必要とする資料を推薦する仕組みが効果的である。

また、単純なパターンマッチングによる手法ではなく、連想検索のような利用者にやさしい検索手段も有効だと考える。

すでにマーケティングの分野では、購買履歴等の個人の利用履歴が顧客の要求分析に利用され、実際に顧客満足度の向上や利益の最大化に効果を発揮している。一方、図書館での貸出履歴をはじめとする個人の利用履歴の利用については、利用者の思想調査、身元調査に利用される恐れがあるなど、さまざまな問題の可能性が指摘され、議論されてきた経緯がある。そのため、現在の図書館界では貸出履歴を利用者分析に利用するどころか、保持する行為自体に批判的である。

しかし、図書館でも、事前に利用者に対し収集する情報の範囲や用途、さらには収集されることで考えられる危険性や得られるメリットについての説明責任を果たし、その上で理解と免責を得られた利用者に限定すれば、貸出情報を収集、蓄積し利用者の情報要求の分析に利用することは可能だと考える。また、その際には、貸出履歴等の個人情報を残したくない利用者のために、該当するサービスを利用しないという選択肢を用意する必要があるだろう。

近年、図書館(特に大学図書館)においては、学内統一認証システムの整備やMyLibraryのような個人向けサービスの基盤が整いつつある。これらの認証機能と組み合わせ、個別の利用動向のデータに対し、協調フィルタリングやその他のデータマイニング技術を図書館の利用履歴情報に適用することで、さらに柔軟な個人の情報要求を満たすことのできるOPACが構築できるのではないかと期待している。

参考文献

- 1) 高野明彦, 西岡真吾, 丹羽芳樹. 連想に基づく情報アクセス技術. 情報の科学と技術. Vol.54, No.12, p.634-639 (2004)
- 2) Webcat Plus: <http://webcatplus.nii.ac.jp> (accessed 2006-8-21)
- 3) 汎用連想計算エンジン GETA: <http://geta.ex.nii.ac.jp> (accessed 2006-8-21)
- 4) 連想出版: <http://rensou.info> (accessed 2006-8-21)
- 5) Perl インタフェース活用術
<http://geta.ex.nii.ac.jp/getaN2001/gdoc/geta/ext/wam/doc/tutorial.html> (accessed 2006-8-21)
- 6) 茶釜: <http://chasen.naist.jp/hiki/ChaSen/> (accessed 2006-08-21)
- 7) 沖縄辞書: <http://www.zukeran.org/o-dic/> (accessed 2006-08-21)

Special feature : Web2.0 and library. Associative OPAC based on user's profile. Yoshitake TOYAMA (Okinawa International University, 2-6-1 Ginowan, Ginowan city, Okinawa 901-2701 JAPAN)

Abstract : OPAC users may use the same search query, but depending on the users' profiles (knowledge, background), they might be trying to express different concepts and their information need might differ. We can expect to develop an OPAC system, which can obtain higher user satisfaction, by modifying search results view, sorted by users' profiles. Existing OPAC search system, which provides results by string search, lacks flexibility. Therefore, OPAC is insufficient as a search tool. Aiming to lessen the above mentioned problem, I

incorporated users' book borrowing histories into the system as a factor for making a proper judgment and developed a new OPAC system at Okinawa International University. The new OPAC system uses an associative search technology, not the conventional string search method, to achieve flexible search results.

Keywords : OPAC / information retrieval / lending history / lending information / association / associative retrieval / user oriented / Okinawa International University