第 26 回年次大会予稿

OPAC 利用ログに基づく文献検索システムの試作と評価
An Evaluation of a Bibliographic Information Retrieval System Using OPAC Usage Logs

高久雅生[1*] 小幡将司[2†] 江草由佳[3]
Masao TAKAKU[1*] Masashi OBATA[2†] Yuka EGUSA[3]

1 筑波大学 図書館情報メディア系
Faculty of Library, Information and Media Science; University of Tsukuba
E-mail: masao@slis.tsukuba.ac.jp
2 筑波大学大学院 図書館情報メディア研究科
Graduate School of Library, Information and Media Studies; University of Tsukuba
3 国立教育政策研究所 研究企画開発部教育研究情報推進室
Office for Educational Resources Research Promotion, Department of Research Planning and Development; National Institute for Educational Policy Research

*連絡先著者 Corresponding Author

図書館資料を対象とした文献検索における適合度順による検索結果ランキングを改善する手法として，OPAC 利用ログを用いた，A）書誌情報の閲覧回数の重み付けによるリランキング手法，B）同一セッション内で使われたクエリキーワードを加味する検索手法の 2 つの手法を提案した．筑波大学附属図書館 OPAC における約 46 ヶ月間にわたる OPAC 利用ログを用いて，提案手法を適用して評価した．評価にあたっては，春日ラーニングコモンズにおける質問事例を参考にした評価用クエリセット 16 件を作り，それぞれの手法による検索結果を人手により適合判定して評価した．結果，手法 A と B はいずれもベースライン（書誌項目を用いた BM25F ランキング手法）に比べて高い評価が得られ，さらに，手法 A よりは手法 B のほうが高い評価が得られることが分かった．

We proposed retrieval methodologies based on usage logs of library OPAC (Online Public Access Catalog) for bibliographic information held in a library. The following methodologies were proposed: A) re-ranking methodology based on the frequency of page visits and B) ranking methodology incorporating query keywords within the same session. We applied and evaluated the methodologies using the usage logs recorded at the library OPAC in University of Tsukuba over a period of about 46 months. For the evaluation, we built a set of 16 queries based on the past questions and answers at Kasuga Learning Commons, and assessed the topical relevance for search results of each method by manual. As a result, we showed that both the methods A and B were evaluated better than the baseline (BM25F ranking methodology using bibliographic fields) and the method B was better than the method A.

キーワード: OPAC, アクセスログ, 書誌情報, 情報検索システム
Keywords: OPAC, usage logs, bibliographic information, information retrieval system

1 はじめに

現在，ウェブ上での情報探索は，サーチエンジン等を用いて，適合度順の検索結果ランキングを閲覧しながら，求める情報を収集する形で行われている．図書館の所蔵文献等を対象とした情報探索においても，ディスカバリーサービスと呼ばれる商用サービス群の普及が進んでおり，これらのディスカバリーサービスではウェブサーチエンジンと同様に，適合度順による検索結果が提供されている [1].

一方，これまで図書館所蔵資料を探す際に用いられてきた OPAC（Online Public Access

†現在，アネックスインフォメーション株式会社

Catalog）では，適合度順の検索結果を提供する例はさほど多くない．

研究や学習等にともなって新しいことがらを学ぶ際に，関連する文献により素早くアクセスできるようになることは重要である．とりわけ，適合度順の検索結果は，情報収集型の探索タスクにおいて，効果的に適合文献にあたれるので有用である．

本研究では，適合度順による検索に加え，過去の利用ログを用いる手法を加味して検索結果を改善する手法を提案する．これにより，利用者が効果的な情報収集が行えるようにし，学習等に貢献することを目指す．具体的には，利用ログから抽出した過去の利用者の探索過程の情報から，書誌詳細情報の閲覧回数の情報や，同一のセッション内で使われた検索キーワードを検索項目に加えた検索ランキング手法を提案し，評価する．

2 関連研究

2.1 OPAC 利用ログを用いた検索改善

図書館における過去の利用者の行動を用いた研究には，1) 貸出データに基づく調査研究や情報推薦手法の研究，2) OPAC 利用行動の調査研究が見られる．

貸出データに基づく調査研究では，蔵書コレクションの状況を把握する方法論の提案や調査分析を行った調査研究 [2][3][4][5][6] に加え，利用者の動向把握を測る調査研究 [7][8] がある．また，貸出履歴を元にした文献推薦手法の提案も多くみられる [9][10][11]．さらに，小野らは貸出履歴を用いて利用者の読書経験を共有するサービス Shizuku2.0 を提案している [12]．

さらに，OPAC 等の検索サービスの利用ログを利用者による探索過程を示すものとして，その利用動向や検索過程を調査した研究も多くみられる．利用されている検索アクセスポイントや主題検索の状況に関する研究 [13][14][15]，検索語の種類や利用者特性の関係の研究 [16][17][18][19][20][21] などがある．また，大規模な利用ログを共通データセットとして分析する試みに，CLEF（Conference and Labs of the Evaluation Forum）評価ワークショップにて 2009 年から 2011 年まで LogCLEF トラック [22] の実施例がある．

これら過去の利用者の行動を用いた研究の多くは利用傾向の把握や利用者行動の理解を目的としており，検索手法の枠組みや具体的なシステム改善手法の提案に踏み込んでいる研究事例はさほど多くない．本研究の貢献は，実際の利用ログデータを元に検索手法の提案を行い，その評価と考察を示す点にある．

2.2 検索クエリログを用いた検索改善

前節でも述べた通り，利用ログ分析はトランザクションログとも呼ばれ，従来からかなり広く行われてきた．2000 年代以降に顕著になっている研究領域の一つは，クエリログおよびクリックスルーログの活用である．ウェブ検索エンジンの普及とともに，検索エンジンの性能改善において，サーチエンジン自身で検索されたキーワードの頻度情報等をもとにしたクエリログに基づく検索支援手法 [23][24] や，検索結果一覧ページにおいてクリックされたページの頻度の情報をもとにしたクリックスルーログに基づく手法 [25][26] など，さまざまな研究がなされてきた [27]．

これらの研究はウェブ検索エンジンや単一ウェブサイトにおける検索を対象としたものが多く，OPAC における図書館資料のような構造化された項目からなるメタデータ検索を基本とする書誌情報検索とは異なる要素が多い．本研究の貢献は，OPAC の書誌情報を対象とした検索において，利用ログの活用を行う手法を提案し，評価する点にある．

3 ログ分析

3.1 対象データ

本研究の対象データは，筑波大学附属図書館所蔵文献の書誌情報，筑波大学附属図書館公式サイトへのアクセス記録の 2 点からなる．

筑波大学附属図書館所蔵文献の書誌情報は約 150 万件ある．また，OPAC 利用ログは，筑波大学附属図書館における現行システムの稼働期間ほぼ全体をカバーする 2014 年 3 月～2018 年 1 月の約 46 ヶ月分の期間のアクセス，約 3 億 6 千万件を対象とする．なお，対象となるアクセスログには，OPAC 以外にも附属図書館公式サイト，機関リポジトリ「つくばリポジトリ」（2015 年 3 月まで），展示コンテンツ等，OPAC 利用以外のコンテンツへのアクセスも含む点に注意す

図 1　筑波大学附属図書館公式サイトのトップページ（2018 年 3 月時点）

る必要がある．

図 1 に，対象としている筑波大学附属図書館公式サイトのトップページを示す．トップページ上部に設置された検索フォームは，ディスカバリーサービスおよび OPAC の検索サービスへの入り口である．トップページに掲載された検索フォームは，標準では，ディスカバリーサービスを対象とした検索となっており，OPAC 検索を選択するにはラジオボタンの選択肢を選ぶ必要があり，かつ，ディスカバリーサービスは，外部クラウドサービスにより提供されており，今回の分析の対象に含まれていないことに注意する必要がある．ただし，後述するように，ディスカバリーサービス上で検索した結果から OPAC に移ってきた際に，クエリの情報が参照元ページの URL に表現されている場合，これをクエリログの一種として用いることとした．

図 2 に，OPAC 利用ログの一部の抜粋例を示す．利用ログは，1) アクセス元の IP アドレス，2) タイムスタンプ，3) リクエスト内容（リクエスト種別（GET 等），リクエストリソース（対象ページ），プロトコルバージョン），4) HTTP ステータスコード，5) 送信バイト数，6) 参照元ページ（Referer），7) クライアントソフトウェア名から構成される．図 2 は 5 回のアクセス記録からなる利用ログの例となっている．このうち，1 行目では筑波大学附属図書館のトップページにアクセスし，2 行目では「企業経済学」というキーワードで検索を行っている．3 行目でレコード ID「1372568」（小田切宏之著『企業経済学』東洋経済新報社）の書誌詳細ページにアクセスしている．また，4 行目では別のキーワード「産業組織論」を用いて再び検索を行っており，5 行目でレコード ID「1176819」（新庄浩二編『産業組織論』有斐閣）の書誌詳細ページにアクセスしている．

3.2　利用ログ中のクローラアクセスの除去

OPAC 利用ログには，検索エンジン等のクローラ，すなわち機械的なプログラムによるアクセスが記録されており，過去の利用者の行動を活用する観点からは不要なため，これを特定して除去する．

まず，OPAC 利用ログ中から，検索エンジン等のクローラによる機械的なアクセスによる記録を除去する．図 2 に示す通り，今回用いるアクセスログには，アクセス元の IP アドレス，利用していたクライアント名称が記録されているため，これらを手掛かりとして除去を行う．あわせて，リクエスト記録のパターンをもとに，目視を通じたヒューリスティックを導入して，機械的なアクセス分を除去した．除去対象とするクローラ等のリストは一般的に利用ログ分析ツールで用いられているもの [28] を用い，クライアント名称に当該クローラ名称が含まれるかを基準として除去した．また，月間 10 万件以上のアクセス数があるものを中心に目視による判断で，追加の除去対象となるクローラ名称を加え，除去処理を行った．

元の利用ログ中に記録されたアクセス数の総数は 363,462,913 件あるが，クローラアクセスを除去した後のアクセス数は 47,133,975 件と，全体の約 12.9% となる．図 3 に，月ごとのアクセス数とクローラアクセスを除去したアクセス数を示す．このクローラアクセス分を除去した，月間約 100 万件前後のアクセス記録が，以降の処理対象となる利用ログの母集団である．

3.3　セッション，クエリ，文献アクセスの抽出

次に，OPAC 利用ログから，同一利用者によるアクセス群を特定して，1 回または複数のアクセス記録をまとめて「セッション」として定義し，これを抽出する．セッションの定義および抽出方法には，さまざまなものが提案されている [29] が，本研究ではひとまず以下の定義を用いて分析する．セッションとは，同一 IP アドレスかつ同一クライアントによるアクセスで，アクセス間隔が 30 分以内であるものをまとめ

```
1 130.1**.**.*** - - [01/Nov/2016:16:46:49 +0900] "GET /lib/ HTTP/1.1" 200
    ↪ 20347 "-" "Mozilla/5.0"
2 130.1**.**.*** - - [01/Nov/2016:16:47:16 +0900] "GET /mylimedio/search/
    ↪ search.do?keyword=企業経済学 HTTP/1.1" 200 14212 "http://www.tulips.
    ↪ tsukuba.ac.jp/lib/" "Mozilla/5.0"
3 130.1**.**.*** - - [01/Nov/2016:16:47:45 +0900] "GET /mylimedio/search/book.
    ↪ do?bibid=1372568 HTTP/1.1" 200 15635 "https://www.tulips.tsukuba.ac.jp
    ↪ /mylimedio/search/search.do?keyword=企業経済学" "Mozilla/5.0"
4 130.1**.**.*** - - [01/Nov/2016:16:48:45 +0900] "GET /mylimedio/search/
    ↪ search.do?keyword=産業組織論 HTTP/1.1" 200 19349 "https://www.tulips.
    ↪ tsukuba.ac.jp/mylimedio/search/book.do?bibid=1372568" "Mozilla/5.0"
5 130.1**.**.*** - - [01/Nov/2016:16:49:05 +0900] "GET /mylimedio/search/book.
    ↪ do?bibid=1176819 HTTP/1.1" 200 14706 "https://www.tulips.tsukuba.ac.jp
    ↪ /mylimedio/search/input-result-find.do" "Mozilla/5.0"
```

図 2　OPAC 利用ログに保存されたアクセス記録の例（抜粋，一部改変）

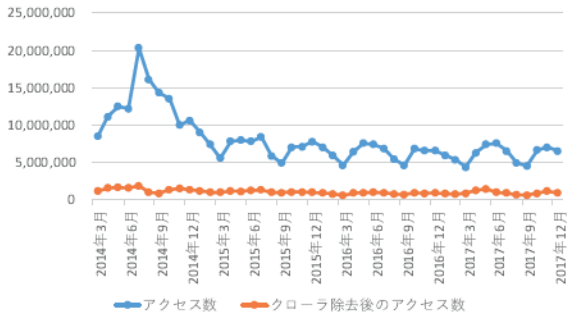

図 3　月ごとのアクセス数およびクローラ除去後のアクセス数

て，1 つのセッションとする．

まず，セッションの抽出にあたっては，上記の定義に加え，前節の単純な方式では除けなかったクローラセッションをはじくため，一定の速度以上で大量のリクエストが送られたセッションをクローラアクセスとして除去する．今回は 200 件以上のリクエストがあったセッション，かつ，そのリクエスト間隔平均が 5 秒未満のものをクローラによるセッションとして除いた．

さらに，本研究の目的は情報検索への応用にあるため，セッションの抽出の際には，文献アクセスが 1 回以上あるセッションに限定して抽出を行う．

抽出された利用者によるセッションは，約 88 万セッション．異なりクエリ発行回数は約 145 万件，文献アクセス回数は約 255 万回であった．抽出されたセッションに関するサマリーを表 1 に示す．

4　検索手法

所蔵資料に対する検索手法として，A）書誌事項の各項目のうちタイトルに重み付けを加えた BM25F ランキング手法，B）閲覧回数の重み付けを行った検索リランキング手法，C）クエリログのキーワードを検索対象に加えた BM25F ランキング手法の 3 手法を検討し，それぞれ評価を行う．なお，手法 A をベースラインとし，手法 B，手法 C，さらに手法 B と C を組み合わせた手法をあわせて比較検討する．

4.1　BM25F ランキング手法

BM25F は Robertson らによる提案 [30][31] であり，構造化文書を対象とする情報検索ランキング手法の代表的なものの一つである．BM25F は，全文検索手法 BM25 の拡張であり，文書のフィールドごとに重み付けを行える特徴がある [31]．本研究では，この BM25F をベースライン手法として用いる．

本研究で対象とする情報は，図書館 OPAC における書誌情報であり，検索対象の文書はタイトルや著者名，出版者，主題件名など，多くのフィールドをもつ．今回は，タイトルの重みを強調するような重みを設定することにより，ベースライン手法とした．

下式 1 に示すように，ベースライン手法のスコア S_{base} は，各フィールドの重みとして $W_{field1}, W_{field2}, \ldots$ をそれぞれ設定したうえで，特に，タイトルに対する重み W_{title} を設定

表 1　セッションあたりの異なり件数

	リクエスト数	クエリ発行回数	文献アクセス数
平均	16.922	1.645	2.898
標準偏差	49.631	3.354	16.443
中央値	9.0	1.0	1.0
最小	1	0	1
最大	19,298	278	5,267

図 4　筑波大学附属図書館 OPAC における書誌情報の詳細画面

することとする．

$$S_{base} = BM25F(W_{title}, W_{field1}, W_{field2}, \ldots) \quad (1)$$

なお，予備実験の結果より，タイトルの重み付けには $W_{title} = 10.0$ を，その他のフィールドの重み付けは $W_{field1} = W_{field2} = \cdots = 1.0$ とした．

4.2　閲覧回数の重み付けによるリランキング手法

利用ログを用いることにより，ある書籍の情報がどれほど閲覧されたかを知ることができる．本研究では，この閲覧回数は，該当書籍の一種の人気度を示すものとして扱う．本研究で対象とする筑波大学附属図書館 OPAC における閲覧回数としては，図 4 に示すような書誌詳細画面を閲覧した回数を指すものとする．書誌詳細画面には，その書籍がどの所蔵されている分館や書架配置図などの詳細が記載されており，実際にその書籍を欲しいと感じた利用者によるアクセスがあると考える．

前節で述べたベースライン手法等によるランキングに対して，当該書籍に対応する書誌情報に対する閲覧回数を用いてリランキングする．つまり，ベースとなる手法によって得られたスコアと，閲覧回数の重み付けをブレンドしたスコアを最終的なスコア S_{usage} として算出してランキングを生成する．リランキング手法によるスコア S_{usage} は，閲覧回数 $Usage$ と元となるスコア $S_{original}$ とを下式 2 のように係数 α を用いてブレンドして算出する．

$$S_{usage} = \alpha \times Usage + (1 - \alpha) \times S_{original} \quad (2)$$

なお，閲覧回数 $Usage$ は，単に利用ログに含まれるアクセス回数そのものを使うと，同一セッション内で同一ユーザが当該ページを再読み込みしたり，何度か見直したりしたようなアクセスが含まれる場合には閲覧した回数としては過大に計数される場合があることから，同一セッション内の複数アクセスは重複してカウントせず，異なりアクセスセッション数としてカウントすることとした．

4.3　クエリログを用いた検索手法

利用ログ中で記録されたクエリのキーワードを，同一セッション内でアクセスされた書誌詳細画面の書籍群と関連したものと仮定し，その関連を BM25F を用いて重み付けする．この仮定は，ウェブ検索エンジンにおけるクリックスルーログの活用でも見られた関連付けであり，OPAC 利用ログの領域でも一定の妥当性をもつものと思われる．

たとえば，図 2 に示した利用ログ中のセッションでは，以下の系列で，2 つのクエリキーワードの発行と 2 つの書誌詳細画面へのアクセスが記録されている．

1. クエリ発行: 企業経済学
2. 書誌詳細アクセス: id=1372568
3. クエリ発行: 産業組織論
4. 書誌詳細画面アクセス: id=1176819

この時，最初のキーワード「企業経済学」と直後の書誌詳細画面 id=1372568 へのアクセスは直接的な関係，すなわち，キーワードでヒットした文献情報の確認を行っていると思われる．同様に，2 つ目のキーワード「産業組織論」と直後の書誌詳細画面 id=1176819 へのアクセスも同様である．よって，これらのキーワードと直後のアクセスは強い関連関係があり，かつ，利用者がその詳細情報を確認したいと思うような一定のニーズを満たすような関係があると思われる．さらに，同一セッション内においては，一定の関連ある検索行動が記録されていることが想定される．すなわち，上記のアクセス系列においても，最初のキーワード「企業経済学」と次の「産業組織論」の間には，一定の関連があると想定すると，最初のキーワード「企業経済学」は 2 回目にアクセスされた書籍 id=1176819 に対しても一定の関連付けを持つことが想定される．同様に，2 つ目のキーワード「産業組織論」も最初にアクセスされた書籍 id=1372568 と一定の関連があると想定できる．

上記で述べたようなクエリキーワードの集合を書誌情報項目の一つとみなして，BM25F による検索ランキングの生成に用いる（式 3）．

$$S_{query} = BM25F(W_{query}, W_{title}, W_{field1}, W_{field2}, \ldots) \quad (3)$$

5 評価実験

前節で述べたランキング検索手法が効果的なものかどうかを検証するため，大学図書館における学習支援の状況を想定した評価用クエリセットを設定したうえで，それらの適合判定を行い，検索手法の有効性を検証するオフライン評価を実施した結果について述べる．

5.1 評価用クエリセット

評価実験に使用する評価用課題として，筑波大学知識情報・図書館学類が運営するラーニングコモンズ「春日ラーニングコモンズ」[32][33] に記録された質問事例集を用いた．2016 年 7 月時点で春日ラーニングコモンズのブログ [32] に記録された質問事例集の中から，科目内容に関連しない質問を除き，質問件数が特に多かった科目上位 3 件と，質問数が少ない科目から無作為に 1 件ずつ取り出し，合計 20 件の科目に対応する質問を抽出した．抽出された質問例を図 5 に示す．

図 5 春日ラーニングコモンズに寄せられた質問例（課題 014）[34]

次に，抽出したそれぞれの質問に回答できると思われる書籍等を導き出すためのクエリを人手で作成した．

5.2 検索手法の実行と適合判定

前節で作成した評価用クエリセットを対象として，4 節で述べた各検索手法を用いた検索を実行した．

まず，ベースラインの BM25F 手法において検索対象としたフィールドは，タイトル（TR），その他のタイトル（VT），内容著作注記（CW），著者名（AL），出版者（PUB），注記（NOTE），親書誌タイトル（PTBL），件名（SH）である．

さらに，検索手法ごとのパラメータ設定としていくつかのバリエーションを試行した．

閲覧回数リランキング手法では $\alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ のそれぞれのパラメータによる検索結果を作成した．

クエリログに基づく検索手法では $W_{query} = 1, 2, 5, 10, 20, 50, 100$ のそれぞれのパラメータによる検索結果を作成した．

これらの複数の手法のそれぞれのパラメータ設定による検索結果から上位 50 件を取り出して，適合判定用プール文書集合とした．プール内の各文献がクエリ作成の元となる質問の情報ニーズを満たすかを基準として，人手により適

合判定を行った．適合レベルは，高適合，適合，部分適合，不適合の 4 段階とした．

表 2 に各評価用クエリごとの適合判定結果を示す．

表 2　適合判定結果

No.	クエリ	適合度ごとの文書数			
		0	1	2	3
001	学術論文情報	212	8	0	0
002	processing	213	11	7	0
003	XSLT	9	1	4	1
004	データベース	223	7	0	0
005	自己組織性	168	25	7	0
006	デンドログラム	0	1	1	0
007	十進分類法	152	27	13	0
008	CaboCha	0	0	0	0
009	2 進数	225	4	1	0
010	漢文	158	61	11	0
011	SQL	31	17	40	0
012	SQLite3	198	1	8	0
013	開発環境	207	3	0	0
014	対数関数	192	17	21	0
015	表記	182	0	0	0
016	5W1H	257	0	0	0
017	Python	31	44	31	0
018	画像	181	27	5	0
019	文字列	220	1	0	0
020	条件式	209	0	0	0

表 2 からもわかる通り，クエリ 008, 015, 016, 020 は，適合度によらず，適合文書がまったく見つからなかったため，以下の評価結果では，これらの 4 クエリ分を除く，16 クエリ分の評価結果を報告する．

5.3　評価指標

評価には評価指標 nDCG[35] を用いた．指標値の算出にあたっては以下の式

$$DCG_{pos} = \sum_{rank=1}^{pos} \frac{rel_{rank}}{log_2(rank+1)} \quad (4)$$

において，rel_{rank} をランキング中のランク $rank$ 番目の文献の適合度に応じた重み，DCG_{pos} をランク pos における DCG 値とする．さらに，下式

$$IDCG_{pos} = \sum_{rank=1}^{|R|} \frac{rel_{rank}}{log_2(rank+1)} \quad (5)$$

において，$|R|$ を適合文書全体を適合度の降順でならべたリストとするときのランク pos における最大値評価値 $IDCG_{pos}$ とする．このとき，評価値 $nDCG$ は以下の式

$$nDCG_{pos} = \frac{DCG_{pos}}{IDCG_{pos}} \quad (6)$$

により，算出できる．

本研究では，適合度に応じた重みを下式の通り

$$rel_{rank} = \begin{cases} 0 & (\text{不適合}) \\ 1 & (\text{部分適合}) \\ 10 & (\text{適合}) \\ 20 & (\text{高適合}) \end{cases} \quad (7)$$

とし，上位 100 件時点の評価値 $nDCG_{100}$ を用いて評価する．

5.4　評価結果

図 6 に閲覧回数を重み付けとした検索手法の評価結果である．ベースライン手法の nDCG 値が 0.412 である一方，$\alpha = 0.2$ および $\alpha = 0.5$ のときの評価値がもっとも高く，0.459 となった．

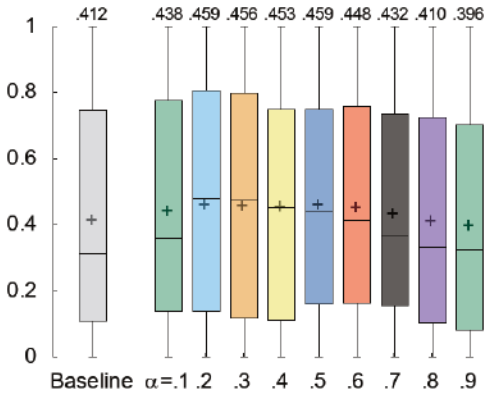

図 6　評価結果: 閲覧回数を重み付けとした検索手法 ($\alpha = 0.1, 0.2, \ldots, 0.9$)．図中の “+” は各手法の平均値を示し，その値を図の上部に示す．

図 7 にクエリログによる検索手法の評価結果を示す．ベースライン手法の nDCG 値が 0.412 である一方，$W_{query} = 10$ のときの評価値がもっとも高く，0.512 となった．

最後に，クエリログによる検索手法のベストスコアだった $W_{query} = 10$ と，閲覧回数に基づく検索手法のベストスコアだった $\alpha = 0.2$ とを組み合わせた手法による検索結果を評価したものを図 8 に示す．評価値は 0.542 であり，2 つの検索手法単体による結果よりもやや高く，元

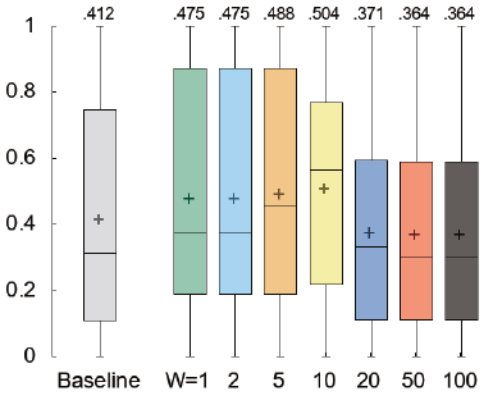

図 7　評価結果: クエリログによる検索手法 ($W_{query} = 1, 2, 5, 10, 20, 50, 100$)．図中の "+" は各手法の平均値を示し，その値を図の上部に示す．

となった閲覧回数に基づく検索手法よりも 8.3 ポイント高く，元となったクエリログによる検索手法よりも 4.5 ポイント高い評価となった．

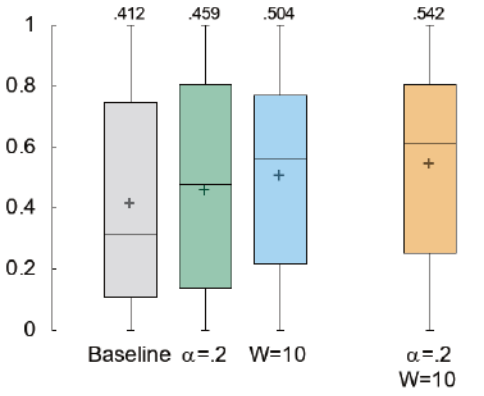

図 8　評価結果: 閲覧回数による手法 ($\alpha = 0.2$)，クエリログによる手法 ($W_{query} = 10$)，両者の混合手法 ($\alpha = 0.2, W_{query} = 10$)．図中の "+" は各手法の平均値を示し，その値を図の上部に示す．

これらの評価結果は，ベースライン手法，閲覧回数による検索手法，クエリログによる検索手法，両者の混合手法という順に検索有効性が高くなることが分かる．このような結果を示す理由としては，閲覧回数による検索手法はクエリキーワードそのものの情報を利用しない静的な手法であり，キーワードクエリの重み付けを考慮できるクエリログに基づく検索手法のほうがより優位になりうることと思われる．

6　おわりに

図書館資料を対象とした文献検索における適合度順による検索結果ランキングを改善する手法として，OPAC 利用ログをもとに，A）書誌情報の各項目のテキスト類似度による BM25F ランキング手法，B）書誌情報の閲覧回数の重み付けによるリランキング手法，C）クエリログを用いた検索手法の 3 つの重み付け手法を提案した．筑波大学附属図書館 OPAC における約 4 年弱の期間にわたる OPAC アクセスログを利用して，提案手法を適用して評価した．評価にあたっては，春日ラーニングコモンズにおける質問事例を参考にした評価用クエリセット 16 件を用い，それぞれの手法による検索結果を人手により適合判定して評価した．結果，手法 B と C はいずれもベースライン（A 手法）に比べて高い評価が，手法 B よりは手法 C のほうが高い評価が得られることが分かった．

今後の課題として，評価用のクエリセットは 16 件と少数であり，さらに対象となる探索ニーズも筑波大学のなかでも知識情報・図書館学類における学習支援を企図したものに限られているため，これらの結果が他の主題領域にも安定的に適用できるかはさらなる検討を要する．加えて，大学における授業等の学習情報を用いた検索手法も考えられ，これも今後の課題である．さらに，今回用いた情報源は OPAC であるが，ディスカバリーサービス等の別種の検索サービスが中心的に用いられていくなかで，これらの検索サービスやアクセスログをどのように活用かも検討する必要があると思われる．

なお，本研究において開発した利用ログの処理および検索システムのプログラムは Github 上で公開している [36]．

謝辞

本研究成果の一部は，JSPS 科研費 JP16H02913, JP17K00449 の助成を受けたものです．また，本研究は，2017 年度筑波大学附属図書館研究開発室第 13 プロジェクトの研究成果です．ここに記して感謝いたします．

参照文献

[1] 飯野勝則: 「図書館を変える! : ウェブスケールディスカバリー入門」. ジャパンナレッジライブラリアンシリーズ. ネットアドバンス, 270p., 2016.

[2] 星野雅英; 渡邉真由美; 風巻利夫; 原香寿子: 「東京大学総合図書館における入館・貸出統計データ分析の試み : 中央図書館としての役割を考えるために」, 大学図書館研究, Vol. 82, pp. 1–11, 2008.

[3] 松井朗; 磯野肇: 「「蔵書回転率」と「蔵書貸出率」を指標とする貸出データの分析調査–奈良大学における図書館資料利用の傾向について」, 奈良大学紀要, No. 34, pp. 177–190, 2006.

[4] 岸田和明: 「利用統計を用いた蔵書評価の手法」, 情報の科学と技術, Vol. 44, No. 6, pp. 300–305, 1994.

[5] 岸田和明; 逸村裕; 高山正也: 「大学図書館における館外貸出データの分析手法: オブソレッセンスと貸出頻度分布の分析を中心として」, 図書館研究シリーズ, No. 31, pp. 79–127, 1994.

[6] 原田隆史: 「大学図書館貸出データの計量的分析 : 上智大学図書館貸出データの分析を中心に」, 彦根論叢, Vol. 260, pp. 83–99, 1989.

[7] 塩沢千文; 玉置すみ子; 翠川舞; 永井貴子; 田中仁: 「図書館の貸出統計から見る学生像」, 飯田女子短期大学紀要, Vol. 25, pp. 191–200, 2008.

[8] 南俊朗: 「図書館利用者理解への試み–貸出データを通して探る利用者プロフィール」, 九州大学附属図書館研究開発室年報, Vol. 2010, pp. 9–18, 2011.

[9] 原田隆史: 「図書館の貸出履歴を用いた図書の推薦システム」, ディジタル図書館, No. 36, pp. 22–31, 2009.

[10] 原田隆史; 増田浩佑: 「貸出記録を用いた図書推薦システムにおける重みづけの変更」, ディジタル図書館, No. 38, pp. 54–66, 2010.

[11] 辻慶太; 黒尾恵梨香; 佐藤翔; 池内有為; 池内淳; 芳鐘冬樹; 逸村裕: 「図書館の貸出履歴を用いた図書推薦システムの有効性検証」, 図書館界, Vol. 64, No. 3, pp. 176–189, 2012.

[12] 小野永貴; 常川真央: 「Web 時代にあるべき未来の図書館サービスの胎動: 貸出履歴の議論を超えた Shizuku2.0 の実現へ」, 情報管理, Vol. 53, No. 4, pp. 185–197, 2010.

[13] 伊藤真理:「楽譜資料の主題検索:アクセス・ポイントの選定に関する調査」, *Journal of library and information science*, Vol. 14, pp. 39–42, 2000.

[14] 酒見佳世: 「統制語による検索の未来」, *Medianet*, No. 12, pp. 40–43, 2005.

[15] 佐藤翔: 「国立国会図書館サーチのアクセスログに基づくアクセスポイント利用状況の検討」, TP&D フォーラムシリーズ : 整理技術・情報管理等研究論集, No. 26, pp. 3–11, 2017.

[16] 野末道子: 「OPAC ログ分析による検索過程の類型化」, 2004 年度三田図書館情報学会研究大会発表論文集, pp. 41–44, 2004.

[17] 種市淳子; 逸村裕: 「短期大学図書館における情報探索行動 : 目次を付与した OPAC のログ分析と検索実験をもとにして」, 名古屋大学附属図書館研究年報, Vol. 5, pp. 57–68, 2007.

[18] 金田千寿; 村上晴美: 「大阪市立大学携帯 OPAC の 2005 年のログ分析」, 大阪市立大学学術情報総合センター紀要, Vol. 8, pp. 35–40, 2007.

[19] Walker, Kizer; Entlich, Richard; Green, Gregory; Hirtle, Peter; Rockey, Steve; Schnedeker, Donald; Stevens, Patrick; Tancheva, Kornelia: “Report of the Collection Development Executive Committee Task Force on Print Collection Usage Cornell University Library”, 2010. `http://hdl.handle.net/1813/45424` (参照 2018 年 3 月 11 日).

[20] Lau, Eng Pwey; Goh, Dion Hoe-Lian: “In Search of Query Patterns: A Case Study of a University OPAC”, *Information Processing and Management*, Vol. 42, No. 5, pp. 1316–1329, 2006.

[21] Blecic, Deborach D.; Bangalore, Nirmala S.; Dorsh, Josephine L.; Henderson, Synthia L.;

Koenig, Melissa H.; Weller, Ann C.: “Using Transaction Log Analysis to Improve OPAC Retrieval Results”, *College and Research Libraries*, Vol. 72, No. 2, pp. 39–50, 1998.

[22] Nunzio, Giorgio Maria Di; Leveling, Johannes; Mandl, Thomas: “LogCLEF 2011 Multilingual Log File Analysis: Language identification, query classification, and success of a query”, *CLEF 2011 Working Notes*, p. 8, 2011.

[23] Cui, Hang; Wen, Ji-Rong; Nie, Jian-Yun; Ma, Wei-Ying: “Query expansion by mining user logs”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, No. 4, pp. 829–839, July 2003.

[24] Shi, Xiaodong; Yang, Christopher C.: “Mining related queries from Web search engine query logs using an improved association rule mining model”, *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 12, pp. 1871–1883, 2007.

[25] White, Ryen W.; Bilenko, Mikhail; Cucerzan, Silviu: “Studying the Use of Popular Destinations to Enhance Web Search Interaction”, *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’07, pp. 159–166. ACM, 2007.

[26] White, Ryen W.; Bilenko, Mikhail; Cucerzan, Silviu: “Leveraging Popular Destinations to Enhance Web Search Interaction”, *ACM Trans. Web*, Vol. 2, No. 3, pp. 16:1–16:30, July 2008.

[27] Agosti, Maristella; Crivellari, Franco; Di Nunzio, Giorgio Maria: “Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction”, *Data Mining and Knowledge Discovery*, Vol. 24, No. 3, pp. 663–696, May 2012.

[28] “podigee/device_detector: DeviceDetector is a precise and fast user agent parser and device detector written in Ruby”. `https://github.com/podigee/device_detector`（参照 2018 年 3 月 8 日）.

[29] Spink, Amanda; Jansen, Bernhard J., editors: “Search Sessions”. “*Web Search: Public Searching of the Web*”. pp. 101–124. Springer Netherlands, Dordrecht, 2005.

[30] Robertson, Stephen; Zaragoza, Hugo; Taylor, Michael: “Simple BM25 Extension to Multiple Weighted Fields”, *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*, CIKM ’04, pp. 42–49. ACM, 2004.

[31] Robertson, Stephen; Zaragoza, Hugo: “The Probabilistic Relevance Framework: BM25 and Beyond”, *Foundations and Trends in Information Retrieval*, Vol. 3, No. 4, pp. 333–389, 2009.

[32]「KLC: 春日ラーニングコモンズ」. `http://klis.tsukuba.ac.jp/lc/`（参照 2018 年 4 月 8 日）.

[33] 松本紳; 逸村裕; 歳森敦: 「筑波大学情報学群知識情報・図書館学類について: 人材養成を中心に」, 大学図書館研究, Vol. 91, pp. 9–14, 2011.

[34]「8 月 6 日の業務報告 — KLC –春日ラーニングコモンズ–」, 2013. `http://klis.tsukuba.ac.jp/lc/20130806`（参照 2018 年 4 月 8 日）.

[35] Järvelin, Kalervo; Kekäläinen, Jaana: “Cumulated Gain-based Evaluation of IR Techniques”, *ACM Transaction on Information Systems*, Vol. 20, No. 4, pp. 422–446, 2002.

[36] “masao/tulips-lees: Tulips usage Log Enhanced Exploratory Search system”. `https://github.com/masao/tulips-lees`（参照 2018 年 3 月 10 日）.