

外部連携サービスによる機関リポジトリの潜在需要の解析

池田 大輔
九州大学大学院 システム情報科学研究院
〒819-0395 福岡市西区元岡 744
Tel: 092-802-3781, Fax: 092-802-3786,
E-Mail: daisuke@i.kyushu-u.ac.jp

星子 奈美
九州大学 附属図書館 コンテンツ整備課
〒812-8581 福岡市東区箱崎 6-10-1
Tel: 092-642-2342, Fax: 092-642-2330,
E-Mail: nhosh@lib.kyushu-u.ac.jp

井上 創造
九州大学 附属図書館 研究開発室
〒812-8581 福岡市東区箱崎 6-10-1
Tel: 092-642-4422, Fax: 092-642-2330,
E-Mail: sozo@lib.kyushu-u.ac.jp

概要

機関リポジトリの導入が進みつつある中で、機関リポジトリの評価としてアクセスログ解析の重要性が高まりつつある。しかし、機関から発信される教育研究関連のコンテンツに対し、機関リポジトリはまだその一部を掲載したにすぎず、アクセスの絶対数もそれほど多いとは言えない状況にある。そのため、現時点でのログ解析では利用状況は分かるものの、機関リポジトリそのものに対する十分な評価とはなりにくい。そこで、本研究では機関リポジトリに未掲載の成果に対する潜在的な需要解析を試みる。現時点でのアクセスが十分でなくとも、潜在需要により将来的な機関リポジトリの必要性や重要性を明らかにできると期待できる。また、潜在需要の存在は、機関リポジトリの維持管理やコンテンツ所有者のデポジットへの強力な動機づけとして期待できる。通常のアクセスログ解析では、すでに掲載されているコンテンツへのアクセスしか分からないため、本研究では機関内の独立サービスとして運用されている業績データベースとの連携機能から得られるログを用いて解析を行う。これにより、解析期間中においては9割程度の需要に応えられていないことが明らかになった。

キーワード 機関リポジトリ, ログ解析, 潜在需要, 連携サービス

Potential Demands Analysis of an Institutional Repository with a Linked External Service

Daisuke IKEDA
Department of Informatics,
Kyushu University
744 Motoooka, Nishiku, Fukuoka
819-0395, JAPAN
Phone: +81-92-802-3781, Fax:
+81-92-802-3786,
E-Mail: daisuke@i.kyushu-u.ac.jp

Nami Hoshiko
Kyushu University Library
6-10-1, Hakozaki, Higashiku, Fukuoka
812-8581, JAPAN
Phone: +81-92-642-2342, Fax:
+81-92n-642-2330,
E-Mail: nhosh@lib.kyushu-u.ac.jp

Sozo INOUE
Kyushu University Library
6-10-1, Hakozaki, Higashiku, Fukuoka
812-8581, JAPAN
Phone: +81-92-642-4422, Fax:
+81-92n-642-2330,
E-Mail: sozo@lib.kyushu-u.ac.jp

Abstract

As the institutional repository has become popular, the access analysis of repositories is becoming more important as their evaluations. However, institutional repositories preserves only a small part of academic works of the institutions and so the number of accesses to repositories is not so much large. Therefore the evaluation based on the access analysis does not reflect precise values of repositories although the access log shows the current usage. In this paper, we analyze potential demands to academic works which are not on an institutional repository. We can expect that existence of potential demands reveals necessity and importance of the institutional repository, even if the current usage shows its poor popularity, and provides strong incentives to managers of the repository and contents holder in the institution. Since accesses to a repository only show explicit demands to academic works which have already contained on the repository, we use access logs of a linkage system between an institutional repository and a database

of academic activities for self-inspection and self-evaluation. We found that our repository does not contain papers for about 90% of requests.

Keyword Institutional Repository, Access Log Analysis, Potential Demand, Linked External Service

1 はじめに

機関リポジトリとは大学等の学術機関内で生成された学術的な資料を保存発信するためのシステムである。機関リポジトリの導入は世界的な動きとなっており、OpenDOAR^{*1}によると、2009/2/15日現在、日本の72機関を含めて全世界で1300以上の機関リポジトリが稼働している。また、国立情報学研究所の「機関リポジトリ一覧」^{*2}によると、試験公開や共同リポジトリなどを含めて国内では91の機関リポジトリが稼働している。

これらの数字が示すように、機関リポジトリ普及は量による評価では一定の段階に到達したと言えるだろう。そのため、次のステップの課題として質の充実が考えられる。例えば、収録コンテンツの充実やまだほとんどの機関リポジトリでは定着しているとは言い難い研究者によるセルフアーカイピングの普及などである。このような質の充実には、現在の機関リポジトリを正確に評価することが重要であり、その重要な手段としてアクセスログの解析がある。

このようなログを用いた解析の動きとして、例えば、海外では英国情報システム合同委員会(JISC)による、機関リポジトリのログ等の情報を相互に交換するためのスキーマ開発の取り組みがある[1]。国内でも機関リポジトリのログ解析[3]や、これに基づく機関リポジトリの評価の取り組みがなされている[2]。

しかし、機関リポジトリ上のコンテンツ数はまだ十分とは言えず、機関リポジトリへのアクセス解析だけではその評価として十分とは言えない。というのも、アクセスログに残る情報は、既に機関リポジトリに情報に対する評価であり、機関が発信しているものの機関リポジトリには未掲載のログに対する需要についての情報は持たないからである。例えば、ある機関リポジトリへのアクセス総数は右肩あがりだとしたら、機関リポジトリとしては十分な成果と考えることもできるが、その機関から発信される学術コンテンツの総数と比較すると、より大きな機会損失をしている可能性もある。特に、今後の機関リポジトリにおいては、従来の学術コミュニケーション環境において流通しづらかったコンテンツ、例えば、紀要や教育資料などの拡充が期待される中で、ある機関がどの分野のコンテンツに注力すべきかを知るためには、このような潜在的な需要解析がもたらす成果は大きいと期待できる。

本研究では、機関リポジトリの潜在的な需要解析を試みる。このために、著者らの所属する九州大学における機関リポジトリ(九州大学学術情報リポジトリ^{*3})、研究者情報データベース^{*4}、及びこれら二つのシステムを連携させる中間システム[4]を用いる。機関リポジトリと研究者情報データベースの間は検索機能へのリンクで結ばれるが、仮に機関リポジトリに論文がない場合は中間システムへ誘導され、機関リポジトリへの登録の手続きへと利用者を導く。つまり、機関リポジトリに存在しないコンテンツへの需要が明らかになると期待できる。

機関リポジトリのログの解析により、まず研究者情報データベースからのアクセスが安定して存在すること

*1 http://www.open_doar.org/

*2 <http://www.nii.ac.jp/irp/list>

*3 <http://qir.kyushu-u.ac.jp/>

*4 <http://hyoka.ofc.kyushu-u.ac.jp/search/index.html>

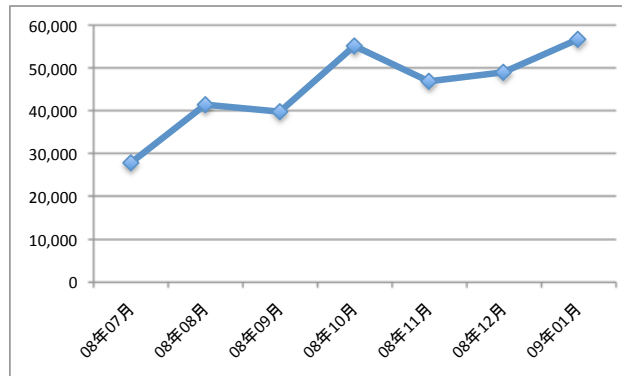


図1 Google Analytics を用いた QIR の月間ページビューのグラフ。

を示す。次に、中間システムのログもあわせて解析し、研究者情報データベースからのアクセスに対し、どの程度の論文が存在した、あるいは、存在しなかったのかを示す。アクセスの安定的な存在については、半年程度のログを解析できたが、論文が存在しなかったことを示すログについてはシステムの設定により1ヶ月程度のログでの解析だった。このような制限があるものの、この期間内において約700件程度の要求に対し、約1割程度しか論文が存在しておらず、約9割の論文が存在しないことが明らかになった。

2 解析対象データ

本研究で用いるデータは機関リポジトリ、業績データベース、及びこれらの中間システムから生成される。本節では、これらのシステムについて説明した後に、解析に用いるデータについて述べる。

九州大学学術機関リポジトリ（以下、QIR と呼ぶ）は九州大学の機関リポジトリであり、平成18年4月より運用を開始した。収録コンテンツは2009/1/23現在で10万件以上あり、ページビューは月間6万に迫ろうとしている。図1に、2008年7月以降のQIRの月間ページビューの推移を示す。

九州大学での業績データベースは、研究者情報データベースと呼ばれ、学内研究者の基本情報から教育・研究・社会活動をインターネット上で公開している。図2は研究者情報データベースのトップページである。2007年6月時点で、約16万件的原著論文、報告書等の研究業績の文献情報が掲載されている。これらはメタデータのみで本文を持たないが、2006年11時点でのデータ提出率99%が示すようにデータが非常に充実していること、また月平均アクセス数が10万件を超えていることなどが特徴である。

機関リポジトリと業績データベースは、本文の有無などの違いはあるものの、研究業績に関しては同じデータを収集していることもあり、連携している、あるいは連携を検討している機関も多い。連携により研究者の入力負担の軽減と、互いのシステム間で足りない情報を補完できるため外部利用者のアクセスも容易になると期待できる。九州大学でもこれらのシステムの連携を行っているが、研究者情報データベースが先行して運用開始しており、システムに大きく手を入れることが不可能だったため、以下に述べるように検索とリンクを主体とした動的な連携を行った[4]。

QIR と研究者情報データベース間は、互いの検索機能を利用したリンクによる連携を実現している。QIR から研究者情報データベースへのリンクは、論文の著者名を用いた検索になっている。逆に、研究者情報データベースからQIRへは、入力時に選択すれば論文へリンクを示すアイコンがつき、このアイコンをクリックすると、論文の情報を検索キーワードにしてQIRで検索を行う。このアイコンは、コンテンツが機関リポジ



図 2 九州大学研究者情報データベースのトップページ。右下に QIR のロゴと適当に選ばれた論文へのリンクがある。

トりに掲載されているか否かに関係なく、貼りつけることが可能である^{*5}

ただし、研究者情報データベースのデータが QIR よりも多いため^{*6}、QIR へのリンクによる検索は、そのままでは該当なしということになる場合が多い。そこで、該当する論文が QIR にない場合は、論文の登録へと導くような機能を実現し [4]、さらに各研究者ごとに、QIR に載っていない論文へのアクセスを SNS などで見られる『足跡』として見せることで論文の登録を動機づけるシステムを実現した [5]。

互いのシステムにほとんど変更を及ぼさずと可能であるという利点のために、このような連携の構成になったが、本研究の目的である潜在的な需要の測定に関しても利点がある。つまり、研究者情報ベースには多くの業績一覧があり、またこれを閲覧する人は多いという状況において、興味のある論文に連携によるリンクがある場合^{*7}は、これを押すことで欲しい論文に対する需要が測定できる。

解析に用いたデータは QIR のアクセスログと、研究者情報データベースと QIR の間に設置した中間システムのアクセスログである。どちらのアクセスログも Web サーバ apache のログである。解析に利用したログの採取期間は、中間システムが 2008/12/28 ~ 2009/02/05 であり^{*8}、QIR は 2008/7/1 ~ である。以後、特に明示しない限り、期間といえば中間システムの期間を表すことにする。

文献 [3] にも述べられているように、検索エンジンのクローラーや連続するクリックなどの重複を取除くことが重要である。本研究でもこのような前処理を施している。具体的にはクローラー分のログの排除、ファイルタイプによる選別を行い、また過度にアクセスの多い IP アドレスのログを重点的にチェックし、サーバの

*5 技術的には全てのコンテンツにリンクをつけることが可能だが、現在は研究者情報データベースへの入力時にリンクのアイコンを付与するかどうか利用者が選択できるようになっている。後述するように、この選択肢のためにコンテンツの数に対しリンクの数が十分でないと思われる。

*6 以下で述べる中間システムの導入前はこの差はさらに大きかったし、数の差が小さくなった現在でも、紀要が中心のリポジトリとコンテンツの重複はそれほど多くないと予想される。

*7 リンクがあっても論文が QIR に存在するとは限らない。

*8 中間システムではログの上書きをする設定にしていたため、これより古いログが残っていなかった。

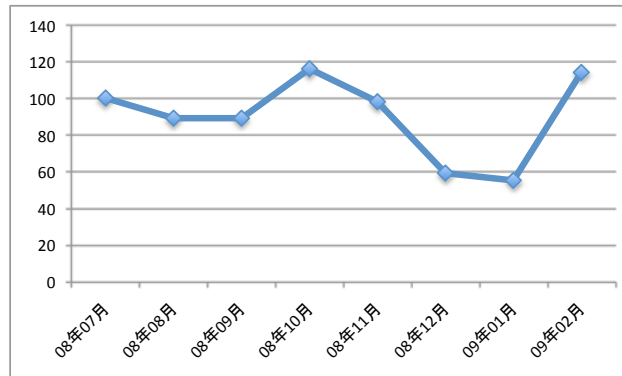


図3 九州大学研究者情報データベースの論文に付与されたリンクを経由して QIR に掲載されている論文にアクセスした数の月毎の推移。ただし、2009年2月は5日までの数値を6倍して使用している。

セキュリティホールを狙った不正アクセスなども削除している。ただし、連続クリックによる重複は排除しなかった。

3 解析結果

本節では、前述したデータを用いて研究者情報データベースからのアクセスがどの程度あるのかを明らかにする。

まず、この期間における QIR におけるアクセス全体を見積もる。ファイルのダウンロードに限定すると、69,529 件のアクセスがあった。ハンドル ID により誘導される論文のメタデータを記述したページへのアクセスもあわせると、その数は倍程度になり、120,779 件のアクセスがあった。

このうち、中間システムをリファラに持つものは 75 件であった。つまり、研究者情報データベースのリンクを経由して論文に到達できた数が 75 件である。この数は、中間システムではなく QIR のログから見積れるので、2008 年 7 月以降の各月における数の推移を示す（図 3 参照）。この図から安定的なアクセスがあるとは言えるが、QIR の論文へとアクセスしてきた数は QIR 全体の論文ダウンロードから見ると 0.1% 程度と非常に少ない。しかし、これは、研究者情報データベースの全ての論文にリンクが自動的に貼られるわけではなく、リンクの総数自体が少ないことも影響していると考えられる。

これを示す、つまり、リンクが十分にあれば研究者情報データベースからのアクセスが多いことを示すため、研究者情報データベースの個々の論文ではなく、トップページ（図 2）や検索結果のページ（図 4）に表示される QIR へのリンクからのアクセス数を計測した（図 5）。つまり、このアクセスは個々の論文のリンクではなく、研究者情報データベースが自動的に生成するリンクからのアクセスであり、個々の論文からよりも 1 ケタ多いアクセスを示している。

最後に、本研究の特徴である潜在的な需要に関する数値を示す。つまり、研究者情報データベースの個々の論文に付与されたリンクをクリックしたものの、論文が QIR にないため中間システムに誘導された数である。この期間、730 件が中間システムに誘導されている。上述したように、同じ期間に同じようにリンクを辿って論文に辿りついた数は 75 件であったため、この期間での結果からは 9 割の需要に対して答えられていないことになる。また、他のアクセス数のグラフが比較的安定していることから、定常的に 10 倍近い需要があり、

基本情報 | 研究活動 | 教育活動 | 社会活動 | ページを印刷

データ更新日：2008.12.25

池田 大輔 (いけだ だいすけ)

准教授
システム情報科学研究所 情報理学部門 発見科学講座

- ◎ 大学院(学府)担当
システム情報科学府 情報理学専攻
- ◎ 学部担当
理学部 物理学科情報理学 コアセミナー
理学部 物理学科情報理学 データ科学
理学部 物理学科情報理学 情報検索・データベース
- ◎ その他の教育研究施設名
附属図書館
- ◎ 電子メール
daisuke@i.kyushu-u.ac.jp
- ◎ ホームページ
http://i.kyushu-u.ac.jp/~daisuke/.
- ◎ 電話番号
092-802-3781
- ◎ FAX番号
092-802-3786

高純度高合金鋼の冷間再結晶挙動に及ぼすC, Nの影響に関する研究
フェライト系ステンレス鋼の高温変形挙動
微生物を利用した資源の効率的生産
光応答性分子エレクトロニクス素子の開発
光応答性超分子組織体の開発

企業ニーズ解決システム
企業の悩みを、ズバリと解決。これが、企業ニーズ解決システム。お探しの先生、見つけます

ライフサイクル
カテゴリ

検索

QIR
九州大学学術情報リポジトリ
情報理学部門

- 新個人認証システム
Personal IDが変える
図書館の個人情報管理
：...
- Coloring for Pattern Detection

図4 九州大学研究者情報データベースの検索結果のページ。右下にQIRのロゴと検索結果にあうように選ばれたQIRの論文へのリンクがある。

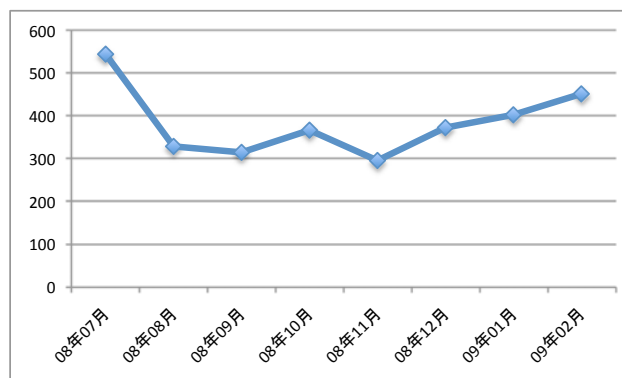


図5 九州大学研究者情報データベースのページに付与されたリンクを経由してQIRに掲載されている論文にアクセスした数の月毎の推移。ただし、2009年2月は5日までの数値を6倍して使用している。

そのうち1割程度の需要にしか応えられていないと推測できる。

4 まとめ

本研究では、九州大学の業績データベースである研究者情報データベースと機関リポジトリを用いて、九州大学で生産された学術コンテンツ（主に論文）に対する需要と、この需要に対しどの程度の割合で本文のアクセスまで提供できたかを解析した。研究者データベースの個々の論文につけられたリンクをクリックして論文に到達した人は今年の約1ヶ月程度で約70程度と少ないが、同様にクリックして論文がないために論文に到達できなかった数はこの十倍程度存在した。つまり、9割程度の需要に応えられていないことが分かった。

このような解析を可能にしたのは研究者情報データベースと機関リポジトリの間にある中間システムのログによるが、このログが一ヶ月程度しかなかったため、十分な解析ができたとは言い難い。そのため、より長期に渡る解析は今後の重要な課題である。一方で、研究者情報データベースからの貼られた他のリンク経由のアクセス数は安定して多く、また、本稿では述べなかったが逆(QIR から研究者情報データベース)のアクセスも一定の数が存在するため、今後の研究者情報データベースと機関リポジトリ間でのアクセスがしやすいように整備することも重要である。

クリックしたにもかかわらず論文が掲載されていなかった論文730件については、論文のタイトルや掲載誌を目視したところ、特に分野の偏りは見られなかったが、定量的な解析はできていない。分野ごとのコードなどは付与されていないため、定量的な解析が難しいが、どのような分野へのアクセスが多いかを明らかにすることも今後の重要な課題である。

謝辞

この研究の一部は『電子情報環境下において大学の教育研究を革新する大学図書館機能の研究』(科研費基盤(B):19300082),『研究者コミュニティが機関リポジトリに深く関わるための入出力活性化』(学術機関リポジトリ構築連携支援事業)によって得られた成果である。

参考文献

- [1] Christine Merk and Nils K. Windisch. Jisc usage statistics review: Final report. Project Report, 9 2008.
- [2] 佐藤義則. 機関リポジトリのアウトプット分析. 平成 19 年度 CSI 委託事業報告交流会, 6 2008.
- [3] 佐藤義則. 機関リポジトリの利用統計のゆくえ. カレントアウェアネス, No. 295, pp. 12–16, 2008.
- [4] 小野真由美, 井上創造, 星子奈美. 研究者業績情報システムと学術情報リポジトリの連携. デジタル図書館, 11 2007.
- [5] 井上創造, 藤井達朗, 小材健, 池田大輔. 学術情報リポジトリ活性化のための足跡機能. 九州大学附属図書館研究開発室年報, Vol. 2007/2008, pp. 17–22, 10 2008.