

研究者業績情報システムと学術情報リポジトリの連携

*小野 真由美, *井上創造, *星子 奈美

*東京大学附属図書館, mono@lib.u-tokyo.ac.jp

*九州大学附属図書館, {nhosh, sozo}@lib.kyushu-u.ac.jp

概要

九州大学学術情報リポジトリ(略称 QIR)とは, 学内で生産された知的生産物を収集・蓄積・保存し, インターネットを通じて無償で配布するシステムである。

QIR と研究者情報の連携は, 各システムのユーザビリティの向上を目指すものであり, 連携により, 研究者情報の業績一覧からワンクリックで対応する文献の閲覧を可能にし, さらに研究者情報のデータを流用した QIR のデータ登録も実現している。

本論文では, システム連携に関する課題から設計・実装方法を説明し, 最後にデータの解析について述べる。

キーワード

学術情報リポジトリ, 研究者情報

Linking Institutional Repository and Researchers Database in Kyushu University

1. はじめに

近年, 学術雑誌の商業出版社による市場寡占化と価格高騰を契機に世界各国で学術機関リポジトリが構築されるようになってきている。

学術機関リポジトリとは, 学内で生産された論文などの知的生産物を収集・蓄積・保存しインターネットを通じて学内外に無償で配布するシステムであり, 九州大学でも九州大学学術情報リポジトリ(以下, 「QIR」と略称)[1]という名称でサービスを開始している。

一方, 九州大学研究者情報(以下, 研究者情報)[2]は, 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システムである。QIR のように内容までは持たないが, 著書, 原著論文, 学会発表論文の一覧という QIR と共通するデータを保持している。また研究者情報のデータ源である大学評価情報システムは, データ提出率 99% (平成 18 年 11 月 1 日現在)[3]と, データが非常に充実していることが特徴である。

本論文では, この QIR と九州大学研究者情報の連

携を述べる。

QIR と研究者情報の連携を行うことにより, 以下のようなメリットが考えられる。

- ・ 研究者情報から, リンクをたどって文献の内容を閲覧できるようになる
- ・ 学内研究者による QIR へのデータ入力の手間を省くことができる

ただし, この連携には, (1) 双方の文献情報を一意に紐づけることができない, (2) 利用者や管理者の手間を極力増やさない, (3) 研究者の意思をできるだけ反映する, (4) 既存の両システムにできるだけ変更を加えない, といった課題を解決する必要があった。我々は, システム間連携を静的なリンクではなく, 検索による動的なリンクにより紐付けのためのデータベースを極力小さく押さえながら, 研究者の意図する紐付けだけはデータベースで管理する, という方法をとることで, 各課題を解決した。

本研究で開発した, 研究者情報と QIR を連携するシステムは, 既存システムにほとんど手を加えなかったことで 3 ヶ月という短い開発期間でシステム間

連携が実現し、開発コストもかなり低く抑えることができています。

研究者情報と QIR のように検索ターゲットとなるデータベースが検索対象のデータベースのデータ量を大きく上回る場合、「検索の正確性より検索漏れの防止を図ることが重要であるが、運用開始後のデータ解析で、多数の検索漏れが発生していることが分かった。このことは、当初に設計した検索性能では、不十分であると考えられるため、利用者に失望感を与えないためにも、検索パラメータの調整や機能の拡充を行い検索性能の向上させることが今後の課題になっている。

以降では、2 章で九州大学 QIR と研究者情報の連携の必要性を述べ、3 章で要求の分析を行う。4 章ではシステムの設計と実装を述べ、5 章でデータの解析を行い、6 章でまとめる。

2. QIR と研究者情報

2.1 QIR の概要

九州大学では、平成 18 年 4 月より QIR の運用を開始した。

QIR の収集コンテンツは、ポストプリント、プレプリント、ワーキングペーパー、テクニカルレポート、学内発行の紀要、学位論文、会議資料、単行図書であり、あらゆる電子データが対象である。平成 19 年 9 月現在、紀要論文を中心に約 5400 件の電子データが格納されている。

機関リポジトリシステムは、EPrints[4] や DSpace[5] に代表されるオープンソースのものから各社ベンダーが開発する有料ソフトウェアまで数多く存在するが、九州大学では、DSpace を採用している。DSpace を採用した理由は、大規模総合大学に適したソフトウェアであるということ[6]と、DSpace が BSD オープンソースライセンスの下でソースが公開されているため、運用に合わせたシステムのカスタマイズが可能であるためである。

研究者情報との連携の際にもオープンソースの利点を生かし、DSpace のデータベースや認証機能、データ登録機能を使用している。

QIR では、DSpace の権限管理機能や Web インタフェースを利用し、九州大学の研究者や大学院生が自分で文献のデータを登録できるようになっている。

現在、学内での周知と著作権処理方法を確立することでデータ登録を促進している。

2.2 九州大学研究者情報の概要

九州大学では、大学の教育・研究水準の向上を目指して、大学評価情報システムを設置している。大

学評価情報システムは、学内全ての研究者の基本情報および教育・研究・社会活動を収集するシステムであり、収集されたデータは大学の活動状況の評価と改善に役立っている。

九州大学研究者情報は、その大学評価情報システムに格納されたデータの内、学内研究者の教育研究活動を学内外に広く公開できるデータを抽出したデータベースである。

これらのデータは全て研究者自身の手によって入力されたもので、データ入力は「大学の員として大学運営に協力する義務のひとつ」[7]と位置づけられているため、入力率が非常に高いデータベースとなっている。

研究者情報は、現在、約 15 万件の原著論文、報告書等の研究業績の文献情報が掲載されている。月平均アクセス数は 10 万件を超え、九州大学の公式なサイトとして多くの人に活用されている。

平成 19 年 4 月には大学評価情報システム入力画面のリニューアルを行い、研究者情報の方も運用に併せて日々進化しているシステムである。

2.3 連携の意義（必要性）

QIR と研究者情報の連携は、それぞれのシステムのユーザビリティの向上を目指すものである。具体的には、研究者情報の業績一覧のページを訪れたユーザが、クリックひとつで対応する文献の内容を閲覧できるようにしたり、学内研究者が QIR へ内容を登録する際に、研究者情報のデータを流用できるようにし、データ入力の手間を省くことを実現している。

そのような連携が必要であった理由は、これまで、研究者情報と QIR のデータを横断的に利用しようとするユーザにとって、不便な状況がうまれていたためである。

研究者情報と QIR は、それぞれブラウザ上で利用できる独自の検索インタフェースとデータ入力インタフェースを持っているシステムで単独で使用するには不自由なく利用することができる。

しかし例えば、研究者情報で検索してヒットした文献の本文を QIR で閲覧しようとした場合、従来は研究者情報から検索キーとなる文献情報をコピー&ペーストして再び QIR において探さなければならなかった。ましてや QIR の存在を知らないユーザにとっては、その検索も行うことができない。

一方で、学内の研究者は、QIR に文献を登録する際、研究者情報に既に業績として入力したのと同じようなデータを QIR にも入力しなければならなかった。

学内には研究者情報や QIR の他にオープンコースウェア (OCW) [8] や数理学系のプレプリントサーバ [9] など学術情報を提供する同種のサービスが存在しており、同様の事態が今後も問題になると考えられる。

QIR と研究者情報との連携は、なるべく少ない手間で問題を解決し、ユーザにとって使いやすい環境を提供することで、今後の Web システム間連携の手本となることを目的としている。

3. 連携にあたっての考慮点や想定される課題

3.1 利用者の要求

本研究で開発した、研究者情報と QIR を連携するシステムを論文リンクシステムと呼ぶことにする。論文リンクシステムは、ターゲットとなるユーザを 1) 研究者情報を訪れる一般ユーザと 2) 研究成果の生産者である学内研究者と定義している。それぞれのユーザの要求は以下のとおりである。

- 1) 研究者情報を訪れる一般のユーザの要求：
研究者情報の研究業績一覧から、文献それぞれに張られたリンクをたどって文献の内容を閲覧できるようにすること。このリンクをたどることを、リンクアウトと呼ぶ。
- 2) 研究成果の生産者である学内研究者の要求：
データ登録の手間を省くこと。

3.2 技術的課題と性能要求

論文リンクシステムの具体的な連携方法については、中間データベースシステムを構築する方式と直接連携する方式を検討した。

中間データベースシステムを構築する方式とは、研究者情報と QIR のデータを結びつける情報を QIR と研究者情報の間に構築したデータベースのテーブルに書き込んでおき、研究者情報からのリンクアウト時に、そのテーブルを直接参照し、QIR の該当文献を表示するというやり方である。

データベースに書き込んでおく情報は、例えば、研究者情報の論文 ID と QIR の論文 ID やそれぞれ論

文を指し示す URL である。

一方、直接連携する方式とは、中間システムが研究者情報からの問い合わせを仲介し、QIR へパラメータ付きの URL で問合せを行い、結果を返すことで連携を実現するというものである。

中間データベースシステムを構築する方式は、従来から行われているシステム連携の方法であるが、今後 QIR、または研究者情報に変更が生じた際に、中間データベースシステムのプログラムばかりでなくデータにも変更が必要になる可能性がある。機械的なデータ変換のみで対応できれば良いが、そうでない場合には非常に手間がかかる。特に、研究者情報の方では研究者が業績情報を編集した際に、その ID がずれたりして一貫しないことが考えられるため、データの保守が問題となる。

このため今回は、中間データベースシステムのデータが極力小さくなるようシステムを設計し、必要な情報はできるだけ動的に生成するようにした。具体的には、QIR の外部システムとして論文リンクシステムを構築し、論文リンクシステムでは、URL を解析し結果を返す処理を行うようにし、中間データベースシステムのデータベースには、著者が九州大学情報リポジトリと研究者情報の間の文献を紐付けした場合のみ、そのデータを保持するようにした。

このような、リクエストごとに QIR のデータを検索する方式をとっている論文リンクシステムでは、検索の性能が重要になる。

今回の連携の場合、「研究者情報格納データ 15 万件に対して、QIR のデータが、約 4000 件である」という状況から検討したところ、検索の正確性よりも検索漏れをなくすことを重視して構築すべきという結論に達した。図 1 は検索性能を検証した図である。QIR のデータ集合 (Q) に対してターゲットエントリ (T) の元になる研究者情報のデータ量が非常に多いため、検索結果は、図 1 の③ケースが大半になることが分かる。

このため、初期の実装においては検索語を「文献タイトル」のように荒く設定することにより、できるだけ漏れのない結果を返すことを意図した。実データに対する分析は 5 章において行う。

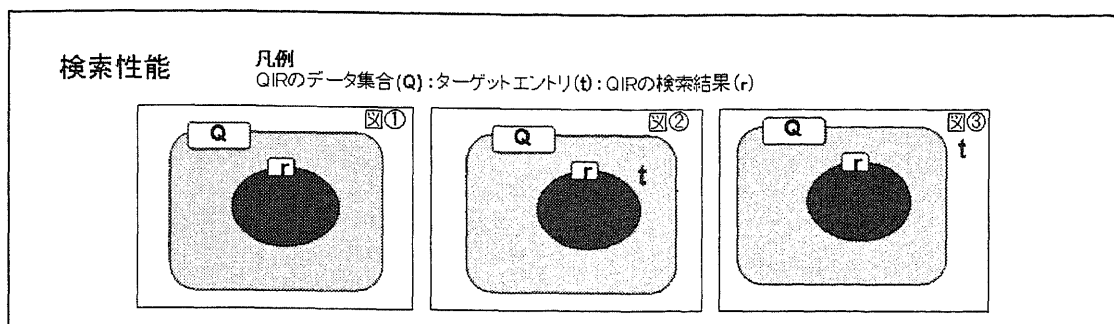


図 1 検索結果の分類と性能への要求

また、「研究者情報に格納されている文献情報は、掲載誌タイトルや掲載ページなどが同一フィールドに格納されており、研究者の好みの区切り記号で記述されているためにデータが正規化されていない」データの性質上、検索の正確性にはある程度限界があることが分かっている。

3.3 運用に関する要求

研究者情報は、2.2 節でも紹介した通り、格納されるデータ量が多く、操作性やユーザインターフェースもデータ入力者である研究者に浸透されたシステムであるため、今回の連携では、研究者情報への改修は極力避けることが条件であった。

QIR においても、サービス開始後、運用面で安定的な稼働が求められる時期であり、大幅な改修はなるべく回避する必要があった。

また、研究者情報へ格納されるデータは、研究者自身で作成されたものであるため、研究者の QIR へのリンクアウトを希望するかないかの意思を確認できるものでなければならないという条件も追加されている。

一方図書館において、システム運用を担当する人員は1名が限度であるため、研究者の意思は研究者自身で設定できるものでなければならない。

以上の条件をふまえて、研究者情報の改修は、以下の2点のみにとどめている。

- ・ 研究者情報の研究業績一覧ページに記載された文献タイトルに QIR へのリンクが貼られるようにした。
- ・ 九州大学大学評価情報システムへの研究業績情報入力時に、研究者が、QIR へのリンクを希望するかないかを選択できるフォームをデータ入力インタフェースに追加した。

QIR 側でも、研究者情報との連携を実現する機能は、基幹システムの外側に構築しているため、大幅な改修を行うことなく実現している。

4. 設計したシステムの概要と構成、動作

4.1 構成

論文リンクシステムは、論文リンク機能、論文登録機能、検索結果登録機能、タイムスタンプ更新履歴機能、アクセス履歴管理機能で構成される。

以下では、各機能について、簡単に説明する。

論文リンク機能は、取得したい論文情報のパラメータを含む URL で論文リンクシステムにアクセスすることで、学術情報リポジトリの検索結果を返す機能である。

例えば以下のような URL をブラウザからアクセスすると学術情報リポジトリの検索結果を得ることができる。

http://leda.lib.kyushu-u.ac.jp/link/link.php?pub_flag=p&au_id=00161&pub_id=3036×tamp=2007/05/01/%2015:42:01&au_f_name=%E6%A3%AE&au_l_name=%E9%9B%85%E7%94%9F&title=Functional%20Composition%20of%20Web%20Databases&au_all=Masao%20Mori,%20Tetsuya%20Nakatoh%20and%20Sachio%20Hirokawa&conf_name=The%209th%20International%20Conference%20on%20Asian%20Digital%20Libraries&conf_year=2006&conf_month=11&other=&conf_category=1&conf_journal=2&language=2

この例では、半角スペースが「%20」でエンコードされているため、少々読みづらいが「title= 論文タイトル」といった複数のパラメータを「&」で区切って論文リンクシステムに問い合わせを行っている。

研究者情報の論文一覧ページでは、このような URL を論文タイトルごとにリンクアンカータグ<a>の href 属性値に設定している。

論文登録機能は、上述の URL のパラメータ値を使って学術情報リポジトリに論文登録ができる機能である。

論文リンクシステムは、送られてきた URL のパラメータ値を学術情報リポジトリの登録用フォームに

予めセットしておくことで、論文の著者である研究者の入力の手間を省いている。

検索結果登録機能は、論文リンク機能で複数の検索結果が表示された後で、検索結果の中から紐付けデータを登録しておく機能である。

具体的には、論文の著者がQIRの検索結果である複数の候補の中から正しいリンク対象を選択しておく機能で、紐付けデータ登録後は、複数の候補を表示することなく、目的の論文に直接リンクアウトできるようにする。

論文リンクシステムでは、3.1で述べた通り検索漏れをなくすように設計されているため、論文リンクシステムを通して学術情報リポジトリを検索した結果はほとんどの場合、複数の候補が表示されることが考えられるが、検索結果登録機能により、利用者の利便性をより向上させることができるようになる。

研究者情報とQIRの論文を紐付けするためのデータとしては、研究者情報とQIRのそれぞれの論文ID及び研究者情報のレコードのタイムスタンプをデータベースに格納している。

タイムスタンプ更新履歴機能は、検索結果登録機能で一旦紐付けられた論文が、研究者情報側で変更された場合でも、QIRと同一論文であるという保証を保つための機能である。

例えば、研究者が研究者情報に格納された論文のタイトルを変更し、紐付けされたQIRの論文が全く別のものとなってしまった場合などを想定している。

タイムスタンプ更新履歴機能では、検索結果登録機能でデータベースに格納した研究者情報のレコードのタイムスタンプと、URLパラメータ値として送られてくる最新のタイムスタンプを論文リンクシステムが呼び出される度に比較し、値が異なる場合は、業務担当者へ確認を促す動作を行う。

業務担当者は、論文リンクシステムの管理画面を通して研究者情報に格納された論文タイトルとQIRの論文タイトルを確認し、間違った論文に紐付いている場合には、紐付けを解除できるようになっており、また、正しい論文に紐付けし直すこともできるようになっている。

アクセス履歴管理機能は、論文リンクシステムのアクセス数を収集・管理する機能である。これにより、論文リンクシステムの使用度や、学術情報リポジトリの各論文へのアクセス動向を調査することができる。

以上が、論文リンクシステムの機能である。なお、論文リンクシステムでは、各機能で使用する

データをRDBMS（リレーショナルデータベース管理システム）で管理している。

以下は、論文リンクシステムのBaseURLと原著論文のパラメータ仕様である。参考までに紹介しておく。

表1 論文リンクシステムのBaseURLとパラメーター一覧

BaseURL <http://leda.lib.kyushu-u.ac.jp/link/link.php>

項目	パラメータ名
原著論文フラグ	pub_flag
教員ID	au_id
原著論文ID	pub_id
タイムスタンプ	timestamp
教員氏名(姓)	au_f_name
教員氏名(名)	au_l_name
論文題目	title
発表者氏名(全員)	au_all
学会又は雑誌名	conf_name
発行年	conf_year
発行月	conf_month
発表区分	conf_section
その他	other
論文形態	conf_category
論文形態(Journal)	conf_journal
使用言語	language

4.2 動作及び操作

本節では、論文リンクシステムに関わる画面の操作と動作について、前節で紹介した各種機能と対応付けて説明する。

まず、研究者情報の研究業績一覧のページを訪れたユーザは、図2の「FullTextQIR」というアイコンをクリックする。

すると、論文リンクシステムが呼び出され、論文リンク機能によりQIRの検索結果を表示する。図3がその例である。

図3の画面上段では、研究者情報からURLパラメータとして送られてきた検索条件を表示しており、画面下段では、論文リンク機能の検索結果を表示している。

図3の例では2件の論文がヒットしている。

図3の画面中央の「検索条件を変更」ボタンをクリックすると検索条件を変更する入力フォームが表示され(図4)、研究者情報からのリンクアウトだけでなく、手動でも検索結果を導き出せるようになっている。



図2 研究者情報論文一覧画面

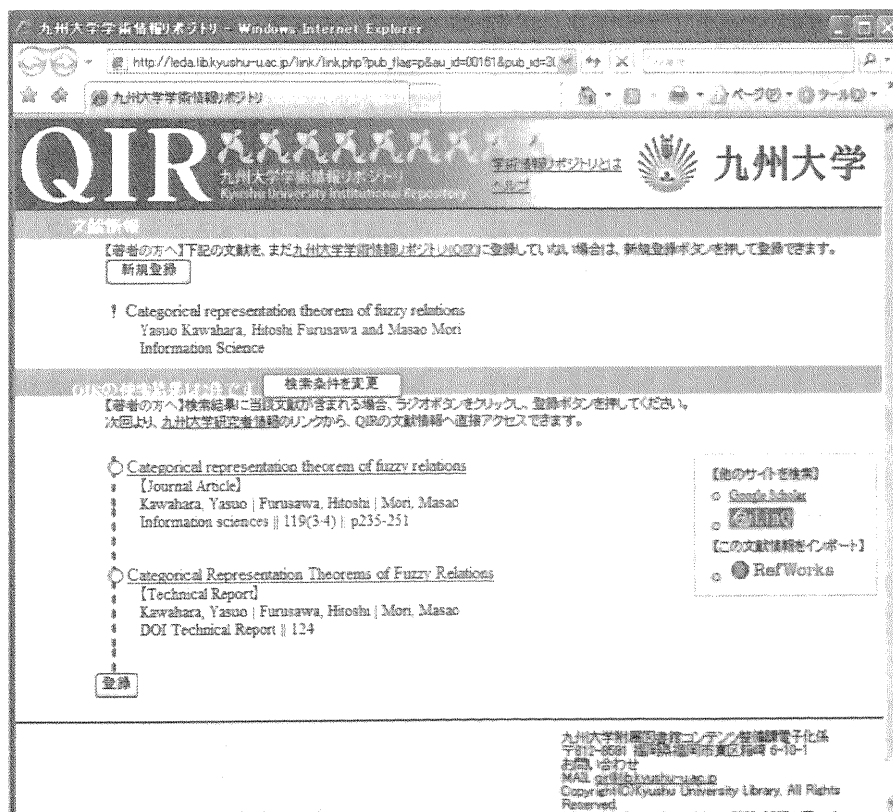


図3 論文リンクシステム検索結果画面

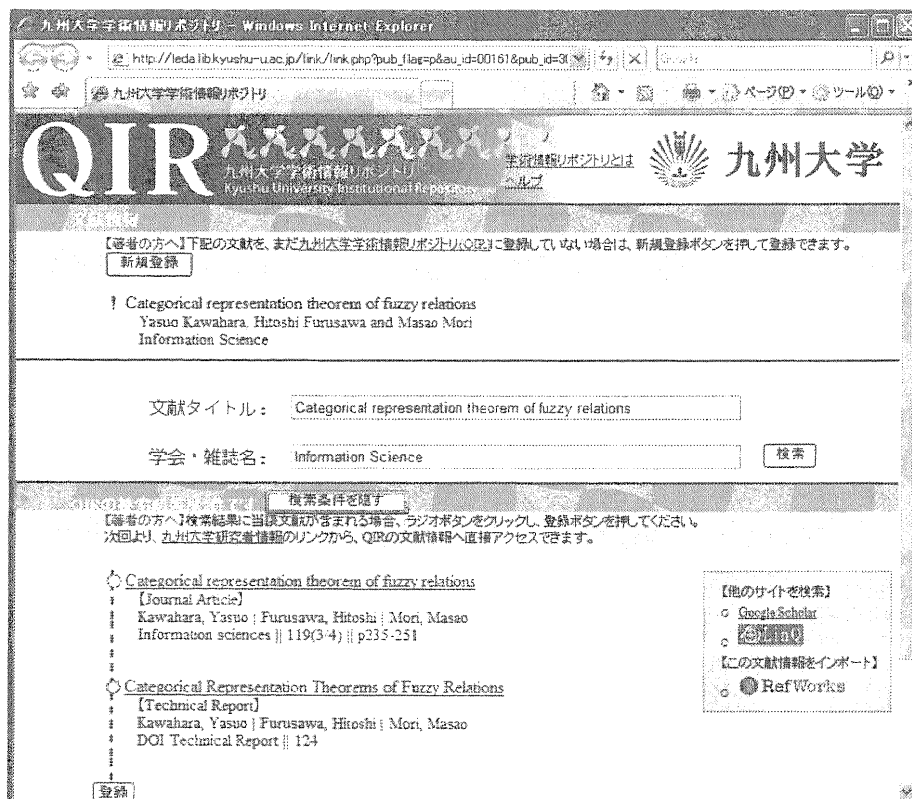


図4 論文リンクシステム検索条件変更画面

文献の著者が、図3の画面でヒットした複数の候補の中から正しい文献を選択し、タイトル横のチェックボックスにチェックを入れると、論文リンクシステムの検索結果登録機能が呼び出され、研究者情報とQIRの紐付け用のデータが登録される。

紐付けデータが登録されると、次回からは検索結果画面(図3)を経由することなく直接QIRの該当文献にリンクするようになる。

図3の画面で検索結果の候補に該当の文献が存在しなかった場合は、検索条件を引き継いで

GoogleScholar[10]やリンクリゾルバを利用した学術情報リンクサービスであるきゅうとLinQ[11]で検索できるようになっている。

また、文献の著者であれば、図3の「新規登録」ボタンをクリックすることで、QIRへ文献を登録する画面にログインすることができる。

このとき、論文リンクシステムでは、論文登録機能が呼び出されるため、図5の通り研究者情報から送られてきた文献情報をもとに、QIRの入力フォームの各項目があらかじめ入力された画面が表示される。



図5 QIR登録画面

なお、検索結果登録および文献登録の際のシステムへのログインに使用される認証機能や、論文登録機能は、QIR (DSpace) の機能を使用しているため、論文リンクシステムへのログイン後の画面操作は、すべてQIRに直接ログインした場合と同様の操作となっている。

5. データの解析

ここでは、2007年5月の時点での、QIRの3,904件のデータと研究者情報の151,903件のデータを元に、論文リンクシステムの効果を見積もる。

図6は、研究者情報中のタイトルでQIRを検索したときのヒット件数の分布である。この図から、1件ヒットするのは455個ある。また2件ヒットするのは3個、3件以上は0個と少ない。このことから、タイトルのみを入力した検索でさえも、QIRの件数が研究者情報の件数より現時点で少ないことを考慮しても、検索結果が0件の場合が多いことがわかる。情報が存在するにもかかわらず漏れているのか、情報が存在しないことによるのかはわからないが、少なくともユーザにとっては検索結果が0件なのは余計な失望感を与える可能性があるため、あいまい検索などの工夫により件数を上げる工夫が必要であろう。

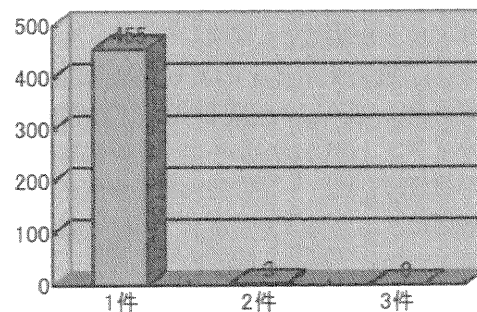


図6 研究者情報中のタイトルでQIRを検索した時のヒット件数分布

図7は、逆にQIR中のタイトルで研究者情報を検索したときのヒット件数の分布である。図6と比べて、2件以上ヒットすることが比較的多いことがわかる。これは、研究者情報からリンクアウトされるQIRの文献が、実はほかの研究者の情報からもリンクアウトされるべきであることを意味していると考えられる。たとえばある研究者が紐付けをしたら、その文献を共著するほかの研究者からも紐付けをするといった、複数の研究者にまたがる機能の有用性を示唆

している。

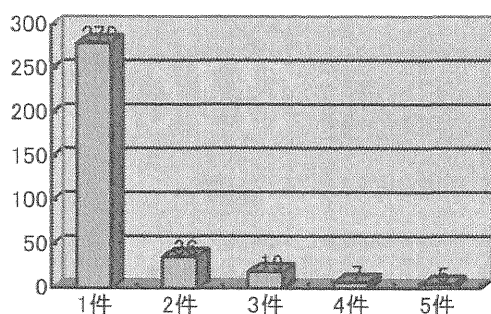


図7 QIR 中のタイトルで研究者情報を検索したときのヒット件数分布

ここで示したデータは運用初期のものであるため、今後データが増えていくにつれて、データ解析を行い、検索のパラメータの調整や機能の拡充を模索していく必要があると考えられる。

6. 運用開始後の課題

論文リンクシステムは、9月30日現在、アクセス数6971件、紐付けデータ登録数161件である。

論文リンクシステムの運用開始後、QIRの新規データ登録は、研究者情報の論文データを流用することで簡単に行うことができるようになった。

新規データについては、研究者情報からQIRへのリンクアウトも問題なく実現している。

しかし、論文リンクシステム構築以前より存在したデータについては、研究者情報の文献情報に記号やハイフン、括弧等がある場合検索漏れを起こす事例が見受けられた。このような検索漏れの場合は、図書館職員が検索条件を変更し、当該文献がヒットした時点で紐付け登録処理をおこなうという運用で対応している。

今後、QIRのユーザ登録数が増加すると、このような運用が難しくなることが予想されるため、データ解析と運用に即したシステム改善を重ねることで検索機能の向上と管理機能の充実を図ることが今後の課題になっている。

7. おわりに

本論文では、九州大学におけるQIRと研究者情報の連携を述べた。今後は、利用履歴や利用者からのフィードバックを元に新たな要求を見つけ解決することが課題である。

謝辞

本システム的设计と実装にご協力いただいた、大学評価情報室の小湊卓夫氏、金丸玲子氏、株式会社Fusicの浜崎陽一郎氏、宋孝氏に感謝いたします。

参考文献

- [1] 九州大学学術情報リポジトリ,
<https://qir.kyushu-u.ac.jp/>
- [2] 九州大学研究者情報,
<http://hyoka.ofc.kyushu-u.ac.jp/search/>
- [3] 大学評価情報システムの部局別入力状況
<http://hyoka.ofc.kyushu-u.ac.jp/hyoka-home/system/jyoho/statsframe.html>
- [4] Eprints:GNU Eprints Archive Software
<http://software.eprints.org/>
- [5] MIT Libraries & Hewlett-Packard Company.:
"Technology:DSpace Federation",
<http://dspace.org/technology/>
- [6] Raym Crow : Open Society Institute ,
http://www.nii.ac.jp/metadata/irp/osi_guide_3/
- [7] 九州大学 大学評価 『システムの概要』,
<http://hyoka.ofc.kyushu-u.ac.jp/hyoka-home/system/jyoho/jyohokoukai.html>
- [8] 九州大学オープンコースウェア(OCW),
<http://ocw.kyushu-u.ac.jp/>
- [9] 九州大学数理学研究院 :Kyushu University Preprint Series,
<http://www.math.kyushu-u.ac.jp/gakufu/>
- [10] Google Scholar,
<http://scholar.google.com/intl/ja/>
- [11] 片岡真 : リンクリゾルバが変える学術ポータル : 九州大学附属図書館「きゅうとLinQ」の取り組み
<http://hdl.handle.net/2324/2905>