

リポジトリ登録作業の半自動化

馬場 謙介, 星子 奈美, 工藤 絵理子, 吉松 直美, 伊東 栄典

九州大学附属図書館
qir@lib.kyushu-u.ac.jp

概要：本稿は機関リポジトリへの文献登録をサポートするシステムを紹介している。機関リポジトリは学術情報へのオープンアクセスを実現する重要なサービスであるが、登録されている文献数は十分とは言えない。著者らは、機関リポジトリへの文献登録作業のうち、文献の著者とリポジトリの管理者による作業を部分的に自動化するシステムを開発した。このシステムは、論文誌や学術会議等で発表された文献についての外部データベースを利用し、該当する著者へ本文の登録を促す。また、各文献の出版社等の著作権ポリシーについても、外部データベースを参照し自動的に確認を行う。本稿は、現行の機関リポジトリの問題点を挙げ、その解決方法として開発されたシステムの詳細を述べている。このシステムにより、文献著者とリポジトリ管理者の文献登録の手間を省くことができ、機関リポジトリの登録文献数の増加が期待できる。

1 はじめに

学術情報への「オープンアクセス[17]」とは、学術論文等の研究成果を無料で利用できることを実現するものである。公的機関からの資金を受けた研究については、その研究成果を社会に広く還元すべきである。実際に、米国国立衛生学研究所によるポリシー[9]等、多くの研究機関でこの考えが広まっている[4]。オープンアクセスを実現するための方法のひとつは「セルフアーカイビング」であり[16]、研究成果を蓄積・公開するシステムがリポジトリである。ある研究機関における成果のためのリポジトリは**機関リポジトリ**と呼ばれる。各研究機関において機関リポジトリを充実させることで、学術情報へのオープンアクセスを実現することができる。

研究成果の機関リポジトリへの登録率は、一般的に非常に少ないと推測される。例えば、九州大学の機関リポジトリ (QIR) [2]については、同大学で執筆された学術論文数に対する QIR の登録文献数の割合は高々30%程度である[13]。また、QIR の登録文献数についての他の機関リポジトリとの比較から、多くの機関リポジトリにおいて同じ状況であることが予想される。この状況を改善するには、文献の著者はもちろん、リポジトリの管理者の登録作業の労力をできる限り少なくする必要がある。

機関リポジトリへの研究成果登録を義務化することができれば、登録文献数の問題は解決できる。しかし、多くの機関ではこの解法を即座

には実行できないと予想される。文献著者の登録の手間を省くアプローチとしては、外部の学術論文データベースから当該機関の研究者による文献の書誌情報を取得し、本文の URI へリンクを張る方法が考えられる。しかし、この方法では本文を所持する出版社等との契約が前提になっており、根本的なオープンアクセスの状態を実現していない。著者が所属機関において直接本文を蓄積することが重要である。もうひとつのアプローチは、文献著者にとっての登録の動機をつくることである。我々は、QIR の登録文献へのアクセスログの解析を行い、文献著者へ有用な情報をフィードバックすることで、文献登録を促すシステムの開発を行っている[11,14]。また、九州大学研究者情報データベース[1]との連携により、文献著者の登録の手間を省くシステムの開発を行った[12]。しかし、現在のところ、QIR の登録数の増加傾向に顕著な変化は見られていない[15]。

我々は、文献著者とリポジトリ管理者の2つの観点から文献登録作業の整理を行った。そして、QIR の運営作業を行う中で、登録数増加の妨げとなる部分はそれぞれ以下の点であると考えた。まず、文献著者については、書誌情報等の入力作業の量よりも、登録の「きっかけ」が重要であると考えた。従来、文献登録は著者が自発的に行っているが、登録作業の簡略化だけでは登録への動機として影響が少ないと予想される。次に、リポジトリ管理者については、その処理能力がボトルネックになり得ると考えた。

管理者の登録作業のうち最も手間のかかる作業のひとつは著作権ポリシーの確認である。これに対し、我々は以下の機能を持つシステムを開発した[10].

- 外部の学術論文データベースから情報を取得することで、機関リポジトリから文献著者へ登録の催促を行う。
- 各文献の出版社等の著作権ポリシーについて、外部データベースへ自動的に問い合わせを行う。

本稿では、文献登録数増加へむけての問題点を、特定の機関リポジトリについてだけでなく、一般的に整理している。また、これらの問題点の解決方法として開発したシステムも、他機関のリポジトリへ適用できるものである。このシステムは、全世界の機関リポジトリの登録文献数増加を実現するひとつの方法となることが期待できる。

2 問題点

この章では、機関リポジトリの登録文献数増加へむけての問題点を明らかにする。

2.1 登録文献数

QIR の登録文献数は、2011年10月現在、約17,000件である。九州大学の研究者情報データベース(DHJS)に登録されている学術論文の書誌情報の数は約70,000件であり、複数の著者による重複した登録を考慮しても少なくとも約50,000件の異なる論文が登録されている[13]。つまり、九州大学で執筆されている多くの論文がQIRには登録されていないということである。QIRの登録文献数を増やすには、まず、これらの埋もれた論文の登録を促すことが効果的だと考えられる。

執筆された論文全体に対する登録率が低いという状況は、QIRに限ったものではないと推測される。世界の機関リポジトリのランキング[3]によると、2011年7月現在、QIRの登録文献数は約2,000機関中51位である。九州大学で執筆される学術論文が特別多いのでなければ、QIRより下位にランクされる機関リポジトリにおいて同じ状況であることが予想される。

2.2 文献著者の作業

文献著者によるQIRへの登録方法としては、

- リポジトリ管理者へ電子メール等で依頼する方法と、
- QIRのアカウントを取得し、登録フォーム

より行う方法

がある。我々は、文献著者による作業を軽減するために、DHJSとQIRとを連携させるシステムを開発した[12]。DHJSに登録されている文献のリストからQIRの本文へのリンクを作成し、本文が無い場合は登録フォームへと誘導できるようにした。この際、入力情報として、DHJSの書誌情報が再利用される。

しかし、2008年6月から2010年10月までのQIRの登録数7,411件のうち、このシステムによるものは多くとも8%程度にとどまっている[15]。我々は、登録の作業量自体を減らすことに加え、著者に対し登録の「きっかけ」を与えることが重要であると考えた。従来、文献登録は著者により自発的に行われてきたが、オープンアクセスや機関リポジトリに対する積極的な関心が無い著者については、いくら登録作業を簡略化しても登録が行われないと予想される。オープンアクセスの思想に対する賛同を得られないまでも、登録可能な論文を著者に知らせることで登録を促すことができるのではないかと考えた。

2.3 リポジトリ管理者の作業

QIRの文献登録においては、前述の2つの登録方法のいずれについても、各文献に対して少なくとも一度はリポジトリ管理者による確認が行われる。著者による文献登録の依頼が大幅に増えたとしても、リポジトリ管理者の処理能力が対応できなければ登録文献数増加は望めない。この状況は、特に、機関リポジトリ管理へ人員を割くことが難しい、規模の小さな機関において問題となると思われる点である。

リポジトリ管理者の登録作業のうち最も手間のかかる作業のひとつは「著作権ポリシーの確認」である。学術論文等を出版する出版社や学術協会は、それぞれ、機関リポジトリでの論文の公開についての著作権ポリシーを持っている。機関リポジトリでの公開の可否に加え、論文のバージョンや開始時期についての条件を定めたものもある。この著作権ポリシーの確認は、一般に、図書館職員等のリポジトリ管理者が手作業で行っており、この作業の効率化は個人の経験や工夫に頼っているのが現状である。

3 解決手法

前節の問題点に対する解決方法として、次の2

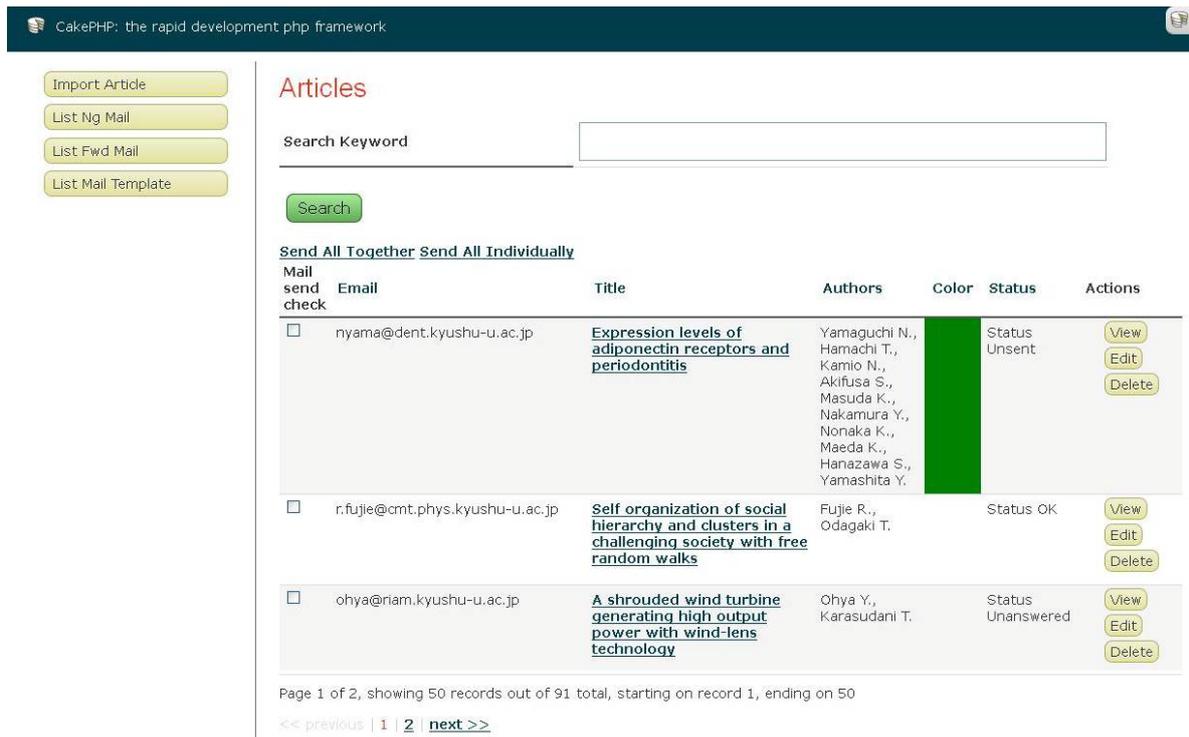


図 2：文献登録半自動化システムの管理画面

つの機能を持つシステムを開発した。

3.1 文献著者への登録の催促

文献登録のきっかけを研究者からリポジトリ側へ移すことで登録数が増えるのではないかと考えた。このアイデアの概要を図 1 に示す。従来、研究者が論文誌や国際会議に投稿し採択された学術論文は、研究者自身の判断で機関リポジトリへ登録されている。論文誌や会議録等は出版社や学術協会等から出版され、それらを横断的に検索できるデータベースが広く利用されている[5,8]。学術論文データベースから該当機関の研究者による文献の情報を取得し、文献著者へ登録の催促を行うことで、登録のきっかけをリポジトリ側へ移すことができる。

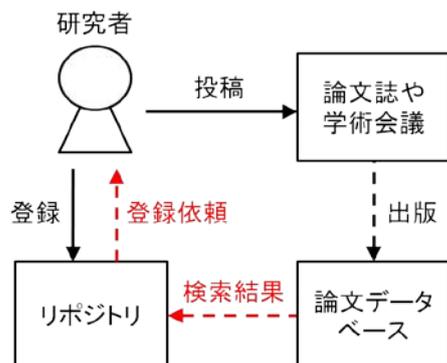


図 1：文献登録依頼機能の概要

図 2 は本システムの操作画面の例である。外部データベースから九州大学の研究者による論文の書誌情報を取得し、リストとして表示している。リポジトリ管理者は、必要ならば 3 番目と 4 番目の項目の「Title」や「Authors」等の修正を行い、2 番目の項目の「Email」へ登録を催促するメールを送ることができる。催促メールの内容は、別の画面で編集することができる。この画面では、メール送信までの作業の進捗管理と同時に、後述の著作権ポリシーに関する各論文についての作業の進捗を「Status」として管理することができる。

3.2 著作権ポリシーの自動問い合わせ

出版社や学術協会の著作権ポリシーに関するデータベース[6,7]を利用し、従来リポジトリ管理者が手作業で行っていた作業の一部を自動化する。図 2 の 5 番目の項「Color」は、各出版社等の著作権ポリシーを表している。一般に、この著作権ポリシーは機関リポジトリでの公開について、「可能」、「不可能」、「条件付きで可能」、「不明」等を、色を当てはめて区別している。リポジトリ管理者は、この色に応じてエンバゴ管理等の登録作業を変化させなければならない。

現在、本システムではこの部分の機能として

出版者に対応する著作権ポリシーの問い合わせのみを想定しているが、著作権ポリシー管理をさらに一般化することによって、より多くの部分を自動化できるのではないかと考えている。

4 おわりに

本稿では機関リポジトリへの文献登録をサポートするシステムを紹介した。文献著者に加え、リポジトリ管理者の観点から文献登録作業を見直し、該当する著者へ本文の登録を促す機能と、各文献の出版社等の著作権ポリシーの確認を行う機能を開発した。対象とする問題点の一般化とシステム汎用性から、本システムは他機関のリポジトリについても適用可能である。

今後の課題として、著作権ポリシー管理作業と、それに関わるエンバーゴ管理についてさらに詳しく一般化を行い、より多くの作業の自動化を目指す予定である。また、本システムの導入前後での文献登録の履歴を詳しく解析し、本システムの効果の検証を行う予定である。

参考文献

- [1] DHJS: 九州大学研究者情報データベース. http://hyoka.ofc.kyushuu-u.ac.jp/search/indin_e.html
- [2] QIR: 九州大学学術情報リポジトリ. <https://qir.kyushuu-u.ac.jp/dspace/>
- [3] Ranking Web of World Repositories. <http://repositories.webometrics.info/>
- [4] ROARMAP: Registry of Open Access Repository Material Archiving Policies. <http://www.eprints.org/openaccess/policysignup/>
- [5] Scopus, <http://www.scopus.com/>
- [6] SCPJ, <http://scpj.tulips.tsukuba.ac.jp/>
- [7] SHERPA/RoMEO, <http://www.sherpa.ac.uk/romeo/>
- [8] Web of Science, <http://science.thomsonreuters.jp/products/wos/>
- [9] Analysis of comments and implementation of the NIH public access policy. The National Institutes of Health, 2008. http://publicaccess.nih.gov/analysis_of_comment_s_nih_public_access_policy.pdf
- [10] K. Baba, N. Hoshiko, E. Kudo, N. Yoshimatsu, and E. Ito, "Semi-automated Paper-registration System for Institutional Repository", In Proc. the Third International Conference on Awareness Science and Technology, pp. 431-434, IEEE, 2011.
- [11] K. Baba, E. Ito, and S. Hirokawa, "Co-occurrence analysis of access log of institutional repository", In Proc. Japan-Cambodia Joint Symposium on Information Systems and Communication Technology, pp. 25-29, 2011.
- [12] K. Baba, M. Mori, and E. Ito, "A synergistic system of institutional repository and researcher database", In Proc. the Second International Conferences on Advanced Service Computing, pp. 184-188, IAIRA, 2010.
- [13] K. Baba, M. Mori, and E. Ito, "Identification of Scholarly Papers and Authors by Connecting Databases", Journal of E-Technology, vol. 2, no. 3, pp. 120-125, 2011.
- [14] K. Baba, M. Mori, E. Ito, and S. Hirokawa, "A Feedback System on Institutional Repository", In Proc. the Third International Conference on Resource Intensive Applications and Services, pp. 37-42, IAIRA, 2011.
- [15] K. Baba, T. Tanaka, E. Ishita, M. Mori, E. Ito, and S. Hirokawa, "Evaluation of link system between repository and researcher database", In Proc. the International Conference on Asia-Pacific Digital Libraries, Springer-Verlag, 2011.
- [16] S. Harnad, T. Brody, F. Vallieres, L. Carr, S. Hitchcock, Y. Gingras, C. Oppenheim, H. Stamerjohanns, and E. Hilf, "The access/impact problem and the green and gold roads to open access", Serials Review, vol. 30, no. 4, pp. 310-314, 2004.
- [17] P. Suber, "Open access overview", Open Access News, 2007. <http://www.earlham.edu/~peters/fos/overview.htm>