

Semi-automated Paper-registration System for Institutional Repository

Baba, Kensuke
九州大学

Hoshiko, Nami
九州大学

Kudo, Eriko
九州大学

Yoshimatsu, Naomi
九州大学

他

<https://hdl.handle.net/2324/20276>

出版情報 : Proc. the 3rd International Conference on Awareness Science and Technology (iCAST 2011), 2011-09. IEEE

バージョン :

権利関係 :



Semi-automated Paper-registration System for Institutional Repository*

Kensuke Baba[†] Nami Hoshiko Eriko Kudo Naomi Yoshimatsu Eisuke Ito

Abstract

The paper introduces a Web system which supports registration of scholarly papers to institutional repositories. Institutional repository is an important service for academic societies, however the number of archived papers is not enough in most of institutions. The authors of the paper developed a system that supports authors of papers (that is, researchers) and managers of repositories in registration of papers. By connecting with external databases of published research papers, the system urges researchers register their papers to repository. Additionally, the system automatically searches the copyright policy for each paper also using external databases. This paper makes clear the problem of current institutional repositories, and then describes the detailed idea of the developed system as a solution. The system is expected to decrease the efforts of researchers and managers for paper registration, and increase the number of archived papers in institutional repositories thereby.

keyword: Institutional repository; Web database; library.

1 Introduction

“Open access [16]” to scholarly information provides free availability of research outputs such as scholarly papers. Generally, it seems to be reasonable the opinion that the research outputs funded by public institutions should be returned to society. Actually, in 2008 the National Institutes of Health (NIH) showed their policy which requires the researchers funded by NIH to open their research outputs [9]. One of the methods to realize the idea of open access is “self archiving” [15]. Then, a *repository* is a system to archive and open research outputs, and a repository for outputs in an institution is called an *institutional repository (IR)*. By improving the IR in each institution, open access to scholarly information will be realized.

As a problem of IR, the number of archived papers is estimated to be extremely small compared to the total number. For example, the papers in the IR of Kyushu University (QIR) [3] is at most about 30 % of the papers produced in the university. By comparing with other IRs about the number of the archived papers, we can estimate that most institutions are in the same situation as the university [12]. To solve the problem, it is required to reduce the efforts for paper registration of “repository managers” in addition to that of the “authors of papers”.

If the registration of research outputs to IR is forced as a mandate, the problem of the number of archived papers may be solved. However, every institution cannot apply this solution immediately. An approach to the problem is to reduce the efforts of authors for paper registration. One of the straightforward solutions is to archive only the metadata and URI of the full-text in an external Web site. However, this solution requires an agreement of a subscription with the site of the full-text, which does not realize the situation of the basic idea of open access. Another approach is to make an incentive for authors. We are analyzing the access log of QIR and developing a system to feedback the result to authors to encourage to register their papers to the IR [10, 13]. Additionally, we developed a system to reduce the efforts of authors for paper registration to the IR by connecting with the researcher database of our university [11]. However, there seems to be no significant change in the increase trend of the number of paper registration to the IR [14].

We reconsidered the process of the paper registration to IR from the viewpoints of authors and IR managers. Then, we found that the obstacles are the following two points. First, for authors of papers, the effective element is the “trigger” to registration rather than the amount of the efforts of entering and uploading. The paper registration is basically made voluntary, and the reduction of the effort was not effective to the number of registration as we mentioned. Second, as for managers of IR, the hardest process is confirmation of the policy about copyright of each publisher. The process by IR managers can be a bottleneck if a lot of requests of registration are made from authors. To overcome the obstacles we developed a system which has the following two functions:

* An edited version of this report was published in: *Proc. the 3rd International Conference on Awareness Science and Technology (iCAST 2011)*, pp. 431–434, IEEE, Sep, 2011.

[†]Research and Development Division, Kyushu University Library, baba@lib.kyushu-u.ac.jp

- Offer of paper registration to IR is made to the authors by using results of search in external databases of published papers.
- Confirmation of copyright is automatically done for each paper by external databases of publishing policy.

In this paper, we generally reconsider the process of paper registration to IR and the problems against an improvement of the number of archived papers. Additionally, the developed system can be applied to general IR. Therefore, the proposed system is expected to be a method to increase the number of the archived papers in the IRs in the world.

2 Problem

This section makes clear the problems with increasing the number of archived papers in IR.

2.1 Number of Archived Papers

QIR is an IR based on Dspace [2]. The number of the archived papers (full-text) in the IR is about 17,000 as of July 2011. The number of the metadata of papers registered in the researcher database in Kyushu University (DHJS) [1] is about 86,000, and the number of distinct papers is at least 68,000 if we consider duplication by co-authors [12]. That is, a large number of papers produced in Kyushu University are not archived in the IR. We considered that it is effective for the improvement of the number of archived papers to encourage researchers to register such buried papers to IR.

The previous situation, the number of archived papers is small compared with the number of the produced papers, is not special to QIR. According to Ranking Web of World Repositories [4], the number of the full-texts in the IR ranks 51st in about 2,000 institutions as of July 2011. If the number of the papers produced in the university is not extremely large, most institutions (especially, institutions rank under 51st) are expected to be in the same situation.

2.2 Efforts of Authors

There exist two ways to register a paper to QIR by an author.

- Create an account of QIR and register by the registration form.
- Submit the metadata and the full-text to the managers of QIR by Email.

In the registration form in QIR, the author has to fill the metadata (such as, the title and the authors) and upload the full-text. To reduce the efforts of registration, we developed a system which connects QIR with DHJS [11]. The system links the metadata of a paper in DHJS to the full-text in QIR, and leads the author to the registration form in the case where the corresponding full-text is not archived in QIR. In this time, the metadata in DHJS is reused for the paper registration to QIR. However, the number of paper registration by the link system from June 2008 to October 2010 was only 8 % of the total number 7,411 [14].

The previous result seems to show that reducing the efforts of paper registration is not so effective to increase the number of registration. Basically the paper registration to IR is made by researchers voluntary, however some researchers who are not interested in open access or IR may not register their paper even if the efforts of the registration be reduced drastically. It is necessary to inform widely about IR to researchers as a possible choice in addition to reducing the efforts, although it is difficult to obtain the consent to the idea of open access from every researchers.

2.3 Efforts of IR Managers

In the current process of paper registration to QIR, repository managers have to confirm the correctness of the metadata, the copyright policy of the publisher, and so on for both of the two ways in the previous subsection. This handwork process may become a bottleneck if the requests of paper registration from researchers increase. The situation is concerned especially for small institutions which cannot use enough workers to the service.

One of the hardest processes of repository managers for paper registration in IR is “confirmation of publishing policy”. The publishers or academic societies who publish scholarly papers respectively have their policy about copyright, that is, whether a paper can be open in IR. Usually, the policy depends on the versions of the paper, pre-print and post-print version, which are defined by the time of a review. Additionally, some publishers ask for a grace for opening the paper in IR (so called “embargo”). Currently, the confirmation of this policy is made by handwork of repository managers, and the efficiency of this process depends on improvements by individual experience.

3 Solution

As a solution of the problems in the previous section, we developed a system which has the following two functions.

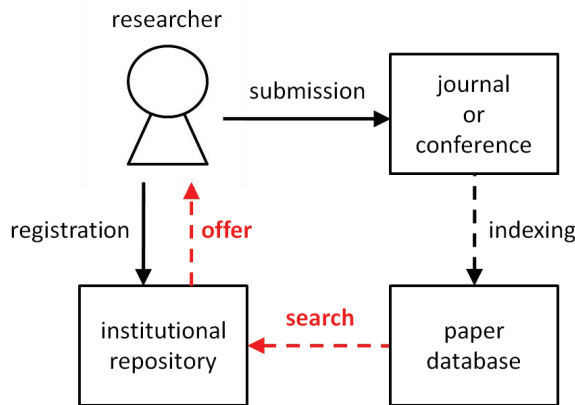


Figure 1: The idea of the function to offer paper registration to researchers by using an external database. A full line and a dotted line express transfers of the full-text and the metadata, respectively.

3.1 Offer of Registration to Authors

For the problem of the efforts of authors for paper registration to IR, we considered that it may be effective to urge authors to register their papers as a “trigger” in addition to reducing the efforts of registration.

The developed system can make an offer of paper registration to authors based on the result of search in external databases. The main idea is to move the trigger of paper registration from authors to IR. This idea is illustrated in Figure 1. In the figure, the full lines express transfers of the full-text of a paper. The problem of Subsection 2.2 is explained that the number of the full-text in “institutional repository” is extremely small compared with that in “journal or conference”. The dotted lines in the figure express transfers of the metadata of a paper. We are trying to increase the number of the full-text in “institutional repository” by using the metadata obtained from “journal or conference”.

There exist some databases of published scholarly papers which are widely used on the internet and producing API [5, 8]. While the paper registration to IR is currently made by researchers voluntary, by searching published papers written by researchers in our own institution, we can make an offer to the researchers to register the paper to IR. Actually, in the database of published paper Scopus [5], the number of the papers written by researchers in Kyushu University (the affiliation of at least one of the authors is Kyushu University) is about 4,000 in 2009. Since the number of paper registration to QIR is about 3,000 per year, a significant effect is expected if the system works well.

Figure 2 is an example of the interface of the developed system. The system are connected with an external database of published scholarly papers. The list in Figure 2 is the set of the metadata of papers ob-

tained from the database as the result of a search for papers produced by researchers in Kyushu University. A user of this system (that is, a repository manager) can correct the information in the columns of “Title” and “Authors” if necessary, and then can send mail of an offer of registration of the paper to the address in “Email”. The details of the offer mail can be edited in another interface. In this interface, in addition to the processes of sending the offer mail and receiving answers for it, we can manage the progress of registration processes about copyright policy as “Status” which is mentioned in the following subsection.

3.2 Automated Confirmation of Publication Policy

We reconsidered the process of the paper registration to IR especially from the viewpoint of repository managers. Practically, we made an investigation of the efforts of repository managers in several universities, and then formalized the process. The flow of the efforts is as follows.

- Obtain a request of paper registration (in one of the two ways in the previous subsection).
- Confirm the copyright policy of the publisher of the paper.
- Create (or correct) the metadata of the paper.
- Upload the full-text and the metadata of the paper to IR.
- Inform the registration of the paper to the authors.
- Operate revisions required from the authors.

Then, we found that the process of confirmation of copyright policy is the hardest effort.

In the developed system, the process of confirmation of publishing policy is semi-automated by using external databases about the policy [6, 7]. In the interface of Figure 2, the column named “Color” corresponds to the policy of the publisher of each paper. The policy can be categorized into four types based on the versions of a paper we mentioned in Subsection 2.3. The categories of the policy are generally expressed by colors as follows [7]:

- Green: IR can open pre-print and post-print version (or publisher version),
- Blue: IR can open post-print version (or publisher version),
- Yellow: IR can open pre-print version,
- White: IR cannot open any version,

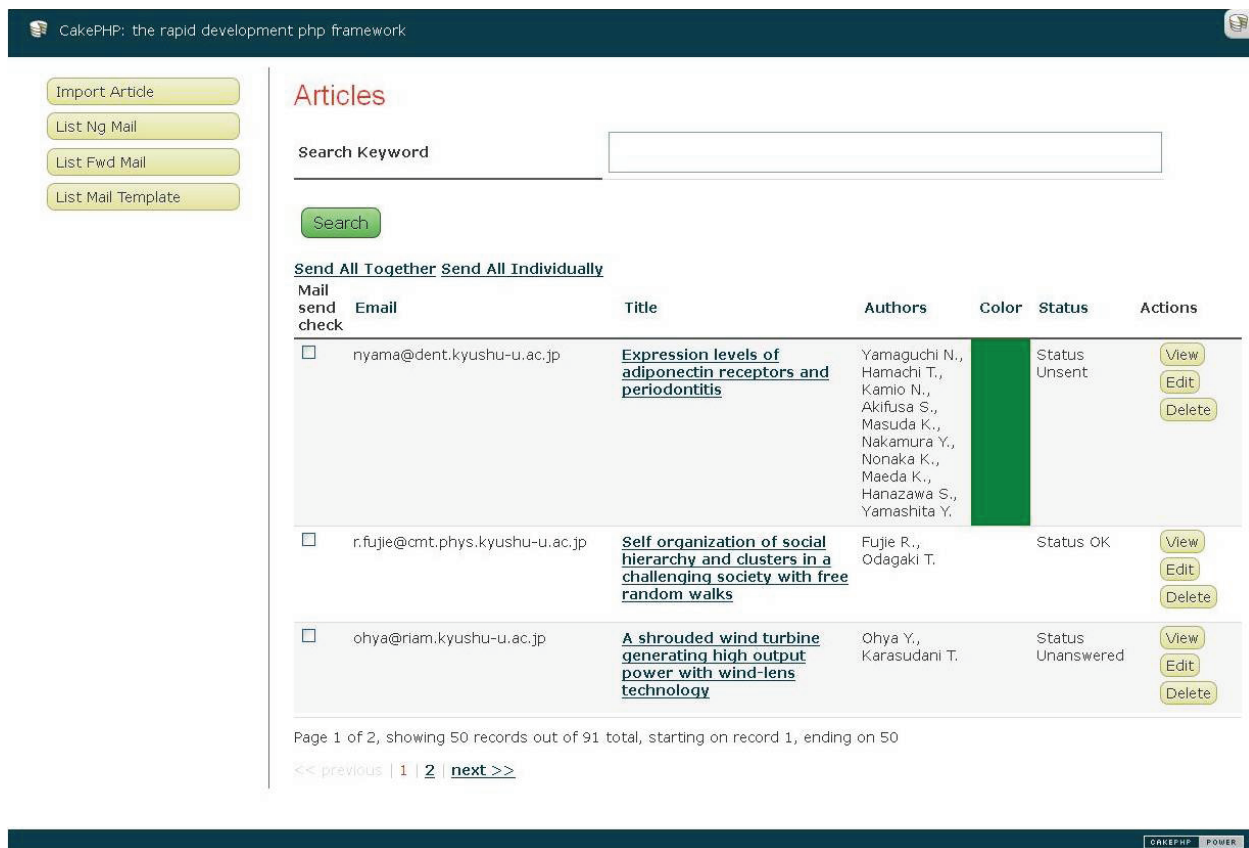


Figure 2: An example of the interface of the developed system.

and we are using also another category [6]

- Gray: the policy is not clear.

In the system, the result of a search of the publishing policy in databases is shown as one of the colors in the “Color” column.

Repository managers have to change their efforts for paper registration based on the kind of color. For some publishers, they have to manage the “embargo”. Additionally, the managers have to confirm also the policy of the authors of the paper. Currently, only the process of the policy confirmation for publishers is automated. We expect that other processes can be automated by generalizing the detailed efforts of paper registration by repository managers.

4 Conclusion and Future Work

In this paper, we introduced a system which supports registration of papers to IR. We examined the processes of the paper registration carefully from the viewpoints of researchers and repository managers, and then developed the system which has the functions to offer researchers to register their paper to IR and to confirm the copyright policy of each publisher

automatically. By the generalization of the problem and the versatility of the system, the developed system can be applied to other IRs.

One of our future work is evaluation of the effectiveness of the system. We are going to analyze the number of the paper registration to QIR in the period from the implementation of the system.

References

- [1] DHJS: Kyushu University Academic Staff Educational and Research Activities Database. http://hyoka.ofc.kyushu-u.ac.jp/search/index_e.html, [accessed 8 Aug, 2011].
- [2] DSpace. <http://www.dspace.org/>, [accessed 8 Aug, 2011].
- [3] QIR: Kyushu University Institutional Repository. <https://qir.kyushu-u.ac.jp/dspace/>, [accessed 8 Aug, 2011].
- [4] Ranking Web of World Repositories. <http://repositories.webometrics.info/>, [accessed 8 Aug, 2011].

- [5] Scopus. Elsevier. <http://www.scopus.com/>, [accessed 8 Aug, 2011].
- [6] SCPJ. <http://scpj.tulips.tsukuba.ac.jp/>, [accessed 8 Aug, 2011].
- [7] SHERPA/RoMEO. <http://www.sherpa.ac.uk/romeo/>, [accessed 8 Aug, 2011].
- [8] Web of Science. Thomson Reuters. <http://science.thomsonreuters.jp/products/wos/>, [accessed 8 Aug, 2011].
- [9] Analysis of comments and implementation of the NIH public access policy. The National Institutes of Health, 2008. http://publicaccess.nih.gov/analysis_of_comments_nih_public_access_policy.pdf, [accessed 8 Aug, 2011].
- [10] K. Baba, E. Ito, and S. Hirokawa. Co-occurrence analysis of access log of institutional repository. In *Japan-Cambodia Joint Symposium on Information Systems and Communication Technology*, pages 25–29, 2011.
- [11] K. Baba, M. Mori, and E. Ito. A synergistic system of institutional repository and researcher database. In *The Second International Conferences on Advanced Service Computing*, pages 184–188. IARIA, 2010.
- [12] K. Baba, M. Mori, and E. Ito. Identification of scholarly papers and authors. In *Networked Digital Technologies*, volume 136 of *Communications in Computer and Information Science*, pages 195–202. Springer-Verlag, 2011.
- [13] K. Baba, M. Mori, E. Ito, and S. Hirokawa. A feedback system on institutional repository. In *The Third International Conference on Resource Intensive Applications and Services*, pages 37–42. IARIA, 2011.
- [14] K. Baba, T. Tanaka, E. Ishita, M. Mori, E. Ito, and S. Hirokawa. Evaluation of link system between repository and researcher database. In *International Conference on Asia-Pacific Digital Libraries*. Springer-Verlag, 2011.
- [15] 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*, 30(4):310–314, 2004.
- [16] P. Suber. Open access overview. Open Access News, 2007. <http://www.earlham.edu/~peters/fos/overview.htm>, [accessed 8 Aug, 2011].