

A Copyright Management Service Model for National Research Reports in Korea

[KwangHo Song, JiHong Min, Yoo-Sung Kim*]

Abstract— A service model that manages the copyrights of national research reports in Korea is designed. When a new research document is submitted to the copyright management server, it is checked whether the document has similar expressions to the already published documents or not. If it has the plagiarized expressions, the service helps authors to prepare the reference list and also to attach the citation mark in proper locations of the new document. Only when the new document is free from the plagiarism suspicion, it can be registered as a copyright holding report. To download and view the registered reports, customers should register and sign on as member users. Only the registered users can download and see the encrypted documents through the dedicated document viewer. When the viewer opens a registered document, a log record is generated and transmitted to the server. After collecting all log records at the server, the copyright management service can analyze them also to mine usage patterns of research reports. With the proposed copyright management service, we hope to increase the utilization of research reports and make a foundation for resolving copyright-related disputes, and finally to contribute to revitalization of the national research and development efforts.

Keywords—copyright management, plagiarism detection, automatic citation-marking, log records, big-data analytics

I. Introduction

In 2009, a violation event of research ethics became a major social issue in Korea. In the case, a professor has acquired the research proposal document for new selection of National Research Laboratory (NRL) in 2008, and then in 2009 he submitted a new research proposal which has very similar idea to the previous one so that he was suspected of plagiarizing the previous proposal document. (1) From that event, the research ethics committees are installed and operated by research institutes, respectively to prevent recurrence of such violations, but that's not eradicated.

Actually according to a survey on research ethics evaluation taken in 2012, plagiarism and duplication of national research and development(R&D) reports produced by research institutes were serious in Korea.(2) As shown in the survey, only regulations of research ethics are not enough to prevent the copyright infringement problems.

According to (3), for the full-fledged copyright management services for national research reports, we need not only the copyright protection service but also copyright utilization service. For copyright protection, a new copyright management service is essentially required by which when a research report is written it should be testify to guarantee that it is free from plagiarism suspicion. Of course, if we have a researcher supporting system by which technical document writers can get support to check whether new document has some plagiarized contents in advance and to generate automatically the reference list and to put citation marks in the proper locations of the new document. For copyright utilization, when research reports are accessed by general users, the personal rights of the users should be checked first and log records which say when and who accesses some specific research reports to be collected to audit and to analyze the usage patterns of the national research reports later.

Up to now, some previous studies have published on the copyright protection issue (4)(5) and others have dealt with copyright utilization issue(6)(7). Up to our knowledge, there is no full-fledged service model which can cover both copyright protection issue and copyright utilization issue systemically and technically. So we need to design a service model to catch up both sides of copyright management of national technical reports in Korea.

Therefore, in this study, we design a new service model that combines multiple components to protect and utilize the copyright of national research reports. First, it contains the plagiarized area detection system and reference citation-mark attaching system for copyright protection. Before researcher submits a report as research result, the report should be checked by plagiarized area detection system. And when the plagiarized area is detected, the report is rejected for registration. But through the reference citation-mark attaching system, the reference list of the report is automatically generated from the document information which is considered as source documents of the detected plagiarized areas, respectively, of the new document. Second, it can control users' access rights to access the national research reports and can collect and analyze the log records to monitor and predict the usage patterns of national reports. From this service, we hope to increase the utilization of national research reports and make a foundation to resolve a copyright-related disputes, and finally we look forward to contribute to revitalization of the national R&D.

KwangHo Song

Dept. of Information and Communication Engineering, Inha University
Incheon 402-751, Korea

JiHong Min

Dept. of Information and Communication Engineering, Inha University
Incheon 402-751, Korea

Yoo-Sung Kim*

Dept. of Information and Communication Engineering, Inha University
Incheon 402-751, Korea

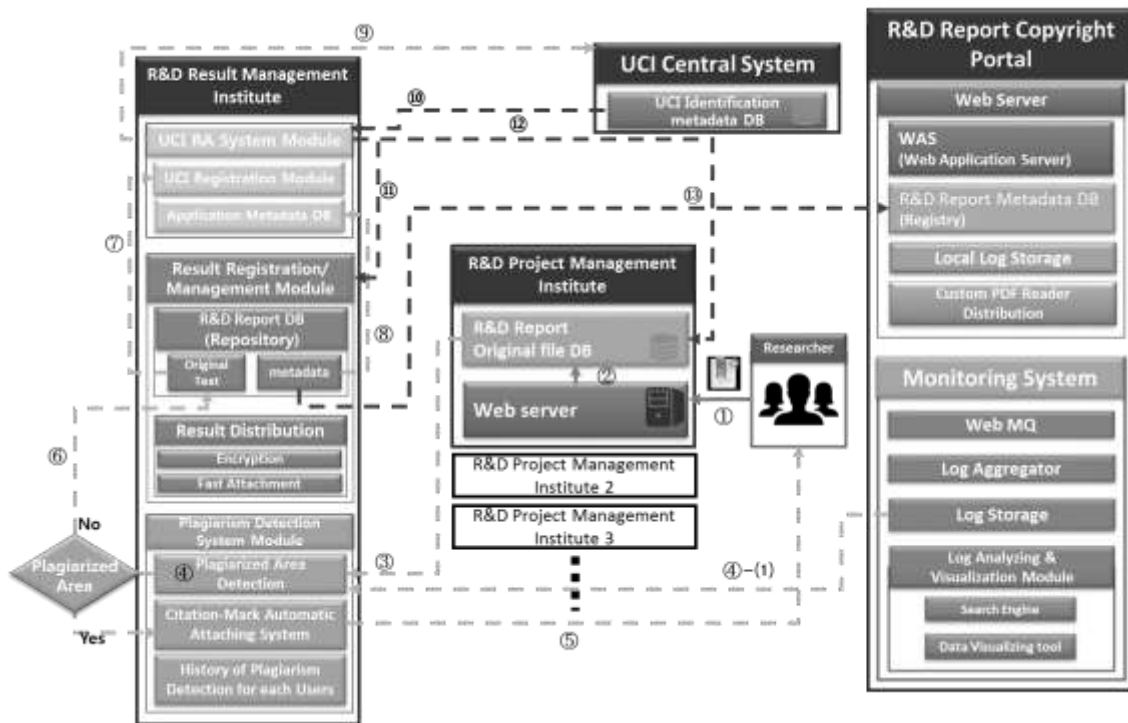


Figure. 1. Research report registration process

This paper is organized as follows. Chapter 2 introduces the previous service models. And Chapter 3 describes the overall service model for copyright management of national research reports. And, in Chapter 4 we conclude this study with the suggestion of further works.

have a goal that protect document copyright by plagiarism detection and vitalize document copyright utilization by supporting self-plagiarism prevention and providing report distribution log aggregation and analysis.

II. Related work

In previous studies (4) and (5), they suggest the bridging method between Universal Content Identifier (UCI) and Integrated Copyright Number (ICN) and the other identifiers that internationally used. And also they suggest the service scenario to contribute to the utilization and protection of contents copyright. They point out the limitations of UCI and ICN. Because UCI is just the identifier, it can't represent the copyright of the content while since ICN is just copyright registration number, so it can't be used in identifying each copy of the same content. So they propose a bridging method between ICN and UCI. But they include only a conceptual design in an abstract level and also they are too biased towards the copyright protection issue.

In previous studies (6) and (7), they suggest the improvement for utilization of copyright even if they are targeting different content types. So, they analyze the current legal or technical utilizing system represented by copyright consignment system and point out the flaw and propose the solution to make current status better. But the solutions are of just abstract system proposal.

As shown in the previous studies, most of the existing systems are concentrated in copyright protection issue. But the new copyright management system should catch up both protection and utilization issues. So, we propose a new service model to manage copyright of national research reports which

III. Copyright management service for research reports

As previously mentioned, we have designed the new copyright management service model to achieve both the copyright protection and utilization goals. This service model is made up five main components and proceeded by three main process. Main components are R&D project management institute, R&D result management institute, R&D report copyright portal, UCI central system, and Researcher and/or User. A research report is submitted to the authorized R&D project management institute and is delivered to R&D result management institution. R&D result management registers it and controls the distribution of the report as a repository. R&D report copyright portal is web portal service to provide research report metadata as the registry and to allow of access and downloading the reports and to monitor the utilization of the reports. UCI central system is provided the metadata about each research report and approves the UCI code issuance. Although UCI is just an identifier, it can be used as the key mean to identify each contents from another content and to trace content distribution for copyright management.

Main process of the copyright management service consists of three components; report registration, report distribution, log collection & content utilization tracking. Therefore in this section, explanation will be proceeded by each component because it is too complicate to explain the overall service model at once.

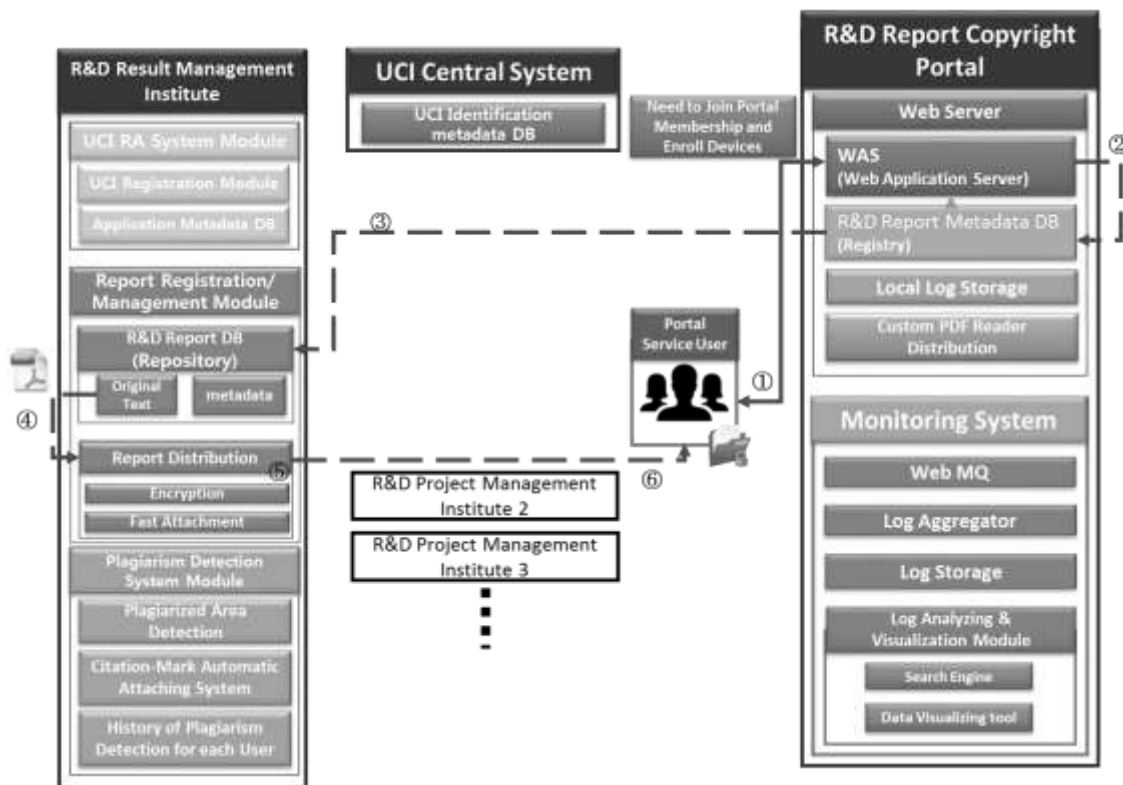


Figure. 2. Report distribution process flowchart

A. Research report registration

First process is for the registration of research reports. As shown in Fig. 1, if researcher submits a report to a R&D project management institute via its web site(step ①), the R&D project management institute saves the report into its DB(step ②). After then they deliver the report to R&D result management institute(step ③). And then R&D result management institute tries to detect plagiarism areas in the report to prevent preemptively the occurrence of plagiarism before a registration(step ④). For the testing, the log analysis data which is created by the user's action(downloading, viewing, etc.) can be used for weighting(step ④-1). So, if the plagiarized area is detected, the system rejects the registration request and provides a citation-mark automatic attaching service which retrieves the reference information about the detected plagiarized areas and generates the reference list including the retrieved reference information automatically and puts reference citation-mark in the proper locations of the document to support researchers feel free from unintended plagiarism suspicion(step ⑤). On the other hand, if the plagiarized area is not detected, then system registers the report into the R&D report DB(repository) and make metadata (step ⑥). Concurrently the system automatically applies a UCI registration of the new registered report to UCI RA(Registration Agent) (step ⑦) And the saved metadata in R&D report DB also hands over to UCI RA(step ⑧). In succession, UCI RA delivers the information for registration to UCI central system(step ⑨). If the registration approved, UCI central system notifies to UCI RA(step ⑩). After then

UCI RA sends the UCI code to R&D result management institute (step ⑪) and R&D project management institute(step ⑫). And last, R&D management institute updates the metadata DB of R&D report copyright portal(step ⑬).

B. Report distribution

Second process is for research report distribution. As shown in Fig. 2, if portal service user connects to the R&D report copyright portal website to search or to download the research report(step ①), then the web server accesses and retrieves the R&D report metadata DB using UCI code of the report to get the URL which is for the location saved the report file(step ②). After getting URL, the portal sends a download request to R&D report DB of R&D management institute with UCI code and URL(step ③). After getting request, R&D report management module submits the requested report file to the report distribution module(step ④). In the distribution module, the distributing information is attached at the report file and the file is going to be encrypted that can be unscrambled through custom PDF reader only(step ⑤). After then the file is delivered to user(step ⑥).

C. Log collection & content utilization tracking

Last process is for log collection & content utilization tracking. As shown in Fig.3, first of all to use encrypted PDF file, researcher should connect to the R&D report copyright portal website and join the membership and enroll device and download the custom PDF reader(step ①). When they connect

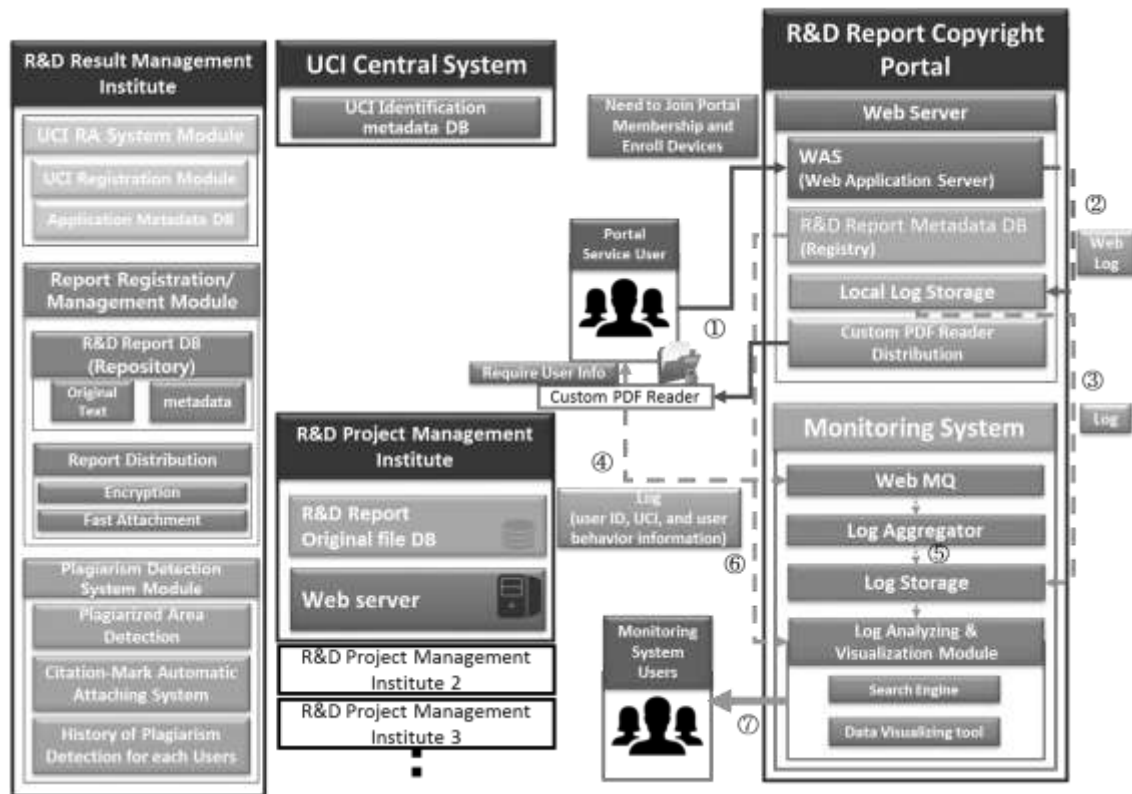


Figure. 3. Log aggregation & content utilization tracking process flowchart

to the website, WAS(Web Application Server) makes a web log and saves it into the local log storage(step ②). After then the system makes the log ,which includes user ID, UCI, and user behavior information(step ③). At this time user ID represents login user identifier and UCI represents the content identifier used by the user and the user behavior information represents what action is made by user for example downloading, searching, exposure, or viewing, and so on. And then, the log is saved in log storage of the monitoring system. meanwhile (step ④)when researcher view the encrypted PDF file though custom PDF reader, the reader makes a log and sends it to the monitoring system. Here, the log format is similar to the previous one. And then, the monitoring system receives the log and saves it into log storage(step ⑤). the logs in log storage are analyzed and visualized using the information in the R&D report metadata database(step ⑥). At last, the monitoring service user can use the visualized data and statistical information about research report utilization(step ⑦).

iv. Conclusion

In this paper, we proposed a new copyright management service model for Korean national research reports. The proposed model consists of three sub-processes to make up a goal which protects document copyright by plagiarism detection and vitalizes document copyright utilization by supporting self-plagiarism prevention and providing report distribution log aggregation and analysis. With this service, we

hope to vitalize the utilization of national research reports and make a foundation to resolve copyright-related disputes.

As the future study, we'll make detailed design about this model and make a prototype based on the detailed design.

Acknowledgment

This research project was supported by Ministry of Culture, Sports and Tourism(MCST) and from Korea Copyright Commission in 2015.

References

- [1] [Online] http://web.hallym.ac.kr/~de1351/magazine/KS_090928.pdf
- [2] National research council for economics, humanities and social sciences, 2012 research report ethic appraisal?, National research council for economics, humanities and social sciences, 2013
- [3] M.S.Park, "A study about copyright management reformation for improving protection and utilizing of copyright.", Master thesis, The graduate school of law, Yonsei university, Seoul, Korea, 2011
- [4] S.H.Oh, and J.J. Lee, "A study on the development of the digital content service scenario, An integrated structure of ICN and UCI identifier system, Journal of the Korean society for information management, vol. 26, no. 3, pp 317-333, September 2009
- [5] Y.H.Kim, "A study on development of test model and linkage method among international standard identification systems and UCI/ICN", The journal of society for e-business studies, vol. 19, no. 4, pp.15-168, November 2014
- [6] J.H.Hong, "The study on copyright limitations for activation of use of public works", Journal of Korean library and information science society, vol. 44, no. 2, June 2013

- [7] Y.J.Shin, “A Study on the policy for promoting the copyright management and the distribution of broadcasting contents”, Master thesis, The Graduate School of Law, Yonsei University, Seoul, Korea, 2009

About Author (s):



Kwangho Song
(2015) Bachelor of Information and Communication Engineering, Inha University.
(2015 ~) Master student at Department of Information and Communication Engineering, Inha University
Research Areas: database, copyright management, and bigdata.



Jihong Min
(2015) Bachelor of Information and Communication Engineering, Inha University.
(2015 ~) Master student at Department of Information and Communication Engineering, Inha University
Research Areas: bigdata and database.



Yoo-Sung Kim
(1986) Department of Computer Science, Inha University (B.S.)
(1988) Department of Computer Science, KAIST (M.S.)
(1992) Department of Computer Science, KAIST (Ph. D.)
(1992 ~) Professor at Department of Information and Communication Engineering, Inha University
Research Areas: multimedia information retrieval system, intelligent information system, bigdata analysis