Exploring Alternativity of Institutional Repositories to Subscription Resources Based on Users’ Accesses Log

Kuang-hua Chen
Department of Library and Information Science
National Taiwan University
1, Sec. 4, Roosevelt Rd. Taipei 10617, TAIWAN
khchen@ntu.edu.tw

Chien-chih Huang
National Taiwan University Library
National Taiwan University
1, Sec. 4, Roosevelt Rd. Taipei 10617, TAIWAN
chienchihhuang@ntu.edu.tw

ABSTRACT
Institutional repository (IR) has been widely deployed in a lot of universities and research institutions. Researchers all over the world share their research outputs in an open-access way. Not only does it speed up the dissemination of scholarly information, but it can also decrease the cost in acquiring commercial databases. This article investigates whether IRs are the viable alternative to subscription resources. The primary result shows that the satisfaction rate is 8.255% based on 124,357 real users’ experiences. Furthermore, the different institutional repositories demonstrated their capability to complement each other in fulfilling the users’ access requests. The promising end result suggests that future studies should include more types of open-access repositories.

Keywords
Access Log, Alternativity, Institutional Repositories, OpenURL, Subscription Resources

INTRODUCTION
Academic library has been the knowledge hub of universities and institutions that are responsible for collecting or creating academic resources and distributing them in an efficient and effective way. A lot of resources are regarded as “pull-based resources”, e.g., priced online databases, which means resources are acquired from outside of the university and institution and are deemed to be useful for researchers and faculty. While other resources are regarded as “push-based resources”, e.g., institutional repository, which are created by faculty, staff, and librarians of the university and pushed outside for demonstration of research outcomes and promotion of university. As we all know, the journals and online databases are becoming more expensive, while library budget are lessening each year. According to the research of Association of Research Libraries, serial expenditures increased by 227%, but the number of serials purchased increased only 9% (Kyrillidou & Young, 2003, p. 9). Even worse was that the serial expenditures increased by 402% in 2011 (ARL, 2012). It is necessary to resolve the critical problem and open access can be considered as one of the possible solutions.

Although, the concept of “open access” has been a practice to share research outputs for a long time, the term of “open access” was first formulated in Budapest Open Access Initiative in 2002. (Suber, 2012, pp. 7-8) As a matter of fact that the open access has been widely accepted and utilized by academia and libraries. Many well-known open-access subject repositories and institutional repositories were built by numerous research communities and academic libraries. Now is the time to explore whether the contents in the institutional repositories can be a viable alternative to the commercial database subscription services. If open-access repositories can satisfy many users’ access requests to subscription resources, we can then assert that it is a potential alternative to subscription resources. Furthermore, the low cost could relief some the stress from tight budget.

The exploratory work presented in this article would like to investigate the possibility that if institutional repositories, a type of open-access repositories, can serve as an alternative to subscription resources from the perspective of real users’ accesses in the first place. The study for other types of open-access repositories will be include in the future.

This article is structured as follows. “Related Work” reviews researches reported in literature, “Research Design” describes research structure, research data, and preprocessing procedure, “Methods and Results” discusses data gathering, data comparison, and result analysis, and “Conclusion” presents brief conclusions.

RELATED WORK
To the best of our knowledge, no studies like this have been documented in literature. Only a few related work were presented. Crum (2011) discussed the availability of research article service in a medical library based on 414 sample journal articles and the results showed that the availability is around 80%. This means that the requests for
research articles were not fully satisfied. Price (2007) and Nisonger (2009) also discussed availability of academic resources.

Some other works discussed the impacts of research outcomes and the productivity of researchers by using academic resources. Vakkari (2008) investigated the influences that electronic resources have on productivity of researchers. Bjork and Solomon (2012) discussed and compared the impacts of subscription resources and open-access resources.

As mentioned above, no similar work dealt with the issues of availability and alternativity. Crum (2011) and other researchers have explored availability of resources. However, their discussions were based on the perspective of resource supply. The work presented here explores alternativity of academic resources from the resource usage based on real users’ accesses to subscription resources.

RESEARCH DESIGN

Research structure
This study will explore the alternativity of institutional repositories to subscription resources. The log file of an OpenURL service will be regarded as the real users’ access requests to subscription resources in National Taiwan University Library. The contents of OAIster and TAIR (Taiwan Academic Institutional Repository) will be compared to examine the satisfaction rate of OAIster and TAIR against the users’ accesses log. Figure 1 shows the research structure.

Log file
We have collected user logs from OpenURL service (OneCate) in National Taiwan University Library for 3.5 months from 2013-05-17 to 2013-09-04. The OneCate is an articles-linking-aggregator to a variety of online databases and electronic journals. The OneCate runs on Apache Http Server, which records the OneCate users’ access on each article. There are more than 160,000 accesses to library acquired materials. Figure 2 shows the sample user’s log from OpenURL. Some instances in the log file show access to specific articles and some show access to journal titles only.

Target repositories
In order to examine the alternative resources for users in our university, institutional repositories (IRs) around the world should be considered. However, it is difficult to cover all IRs. Instead, OAIster is considered in this study. OAIster is not an IR only but a union catalog of millions of records representing open access resources that was built by harvesting from open access collections worldwide using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Today, OAIster includes more than 30 million records representing digital resources from more than 1,500 contributors. (OAIster, n.d.)

The universities in Taiwan began building IRs from 2006, and since then more than 100 IRs have been created. TAIR (Taiwan Academic Institutional Repository), an umbrella
system of IRs in Taiwan, is a portal website for academic researches that showcases the collaborative academic achievements of all academic institutions in Taiwan. (TAIR, n.d.) The research contents stored in OAIster represent the international research achievements; while those in TAIR represent national research achievements.

Examining the contents of both repositories for coverage of users’ need in terms of downloading library acquired materials will make it clear that the commitment of IRs’ movement. Since TAIR is implemented by our team, we could compare their contents to users’ log directly. For the coverage of OAIster, we mimicked the user queries to OAIster by submitting article titles or journal titles of users’ log to OAIster and check whether the targets could be found. The mimic agent was written in Python and submitted query to OAIster every ten seconds to reduce potential overhead.

METHODS AND RESULTS

Data Gatherings

Preprocessing stage processes 161,458 users’ access requests to subscription resources in National Taiwan University Library (NTUL). 124,357 out of 161,458 accesses with title information were extracted from users’ accesses log. The procedure and result of preprocessing stage are shown in Figure 4.

Comparisons for TAIR Bibliographic Records

In order not to interfere with the routine services of TAIR for general public, we export bibliographic records from TAIR. In addition, since there are the possibility of extra annotations or typo between TAIR, OAIster, and OneCate titles, it is entirely possible that false negative cases would be registered if the exact match is applied. For example, the “exchange rate dynamics in a model of pricing to market” query of OneCate and the “exchange rate dynamics in a model of pricing-to-market” record in OAIster are only different in spelling and punctuation marks. Therefore, we decide not to compare OneCate titles to TAIR titles directly. Instead, SOLR, an Apache Lucene-based search software, is used. We import all the titles from TAIR into SOLR. SOLR then provides a web service interface for searching the indexed TAIR titles, which are imported in the previous stage. A python program is created to query each title from OneCate Access Log in this SOLR web service. SOLR then returns with possible titles that could match. We then apply Ratcliff/Obershelp pattern recognition algorithm (Ratcliff & Metzener, 1988) to compare the SOLR-returned titles with the queried title. The queried title with a similarity score above 0.95 (a strict threshold) is defined as “SAME.”

The results show that 2,845 OneCate titles are found in TAIR. This means about 2.288% of users’ accesses could be satisfied by TAIR. Please see Figure 5 for those data.

Comparisons for OAIster Bibliographic Records

In order to compare the users’ accesses with the records stored in OAIster, we created a python program to query each OneCate title in OAIster system. In addition, the query is submitted every 10 seconds to avoid overhead. OAIster then returns potentially matched titles. We then apply Ratcliff/Obershelp pattern recognition algorithm again to compare the OAIster-returned titles with the queried title. Using the same criterion used in comparison for TAIR records, the queried title with a similarity score above 0.95 would be considered as “SAME.”

The results show that 7,645 out of 124,357 titles (6.148%) submitted to OAIster system found their corresponding records in OAIster. Please take a look at Figure 6.
Brief Analysis
At first glance, it seems like only a few users’ accesses could be satisfied by TAIR (2,845 records, 2.288%) and OAIster (7,645 records, 6.148%). However, this work only examines two open-access institutional repositories. The satisfaction rate will increase when more open-access repositories are included. Another interesting question is the overlap ratio of TAIR and OAIster in terms of satisfying users’ accesses. We compare 2,845 records of TAIR and 7,645 records of OAIster. There are only 224 users’ accesses (about 224/124,357=0.180%) could be found in both TAIR and OAIster. This indicates that TAIR and OAIster are complementary at least from the perspective of real accesses of NTU’ users to subscription resources. As a result, we can conclude that the satisfaction rate of TAIR and OAIster together for real users’ accesses is (2,845+7,645-224)/124,357 = 8.255%

CONCLUSION
This work explores the satisfaction or alternative of open-access institutional repositories for NTU users’ accesses to subscription resources. Two federated IRs, TAIR and OAIster, are used for analyzing the alternativity. The results show that TAIR and OAIster together will satisfy 8.255% of users’ access requests. Although the satisfaction rate is not still good enough, this is the first step towards understanding how open-access repositories can aid in the distribution of academic information and help libraries in making decision for academic resource acquisition.

The work presented here is an exploratory study, since only two federated open-access institutional repositories were used. However, the results are encouraging and suggest that future study should include more types of open-access repositories, e.g., arXive, JAIRO, and Citeseer, to carry out much more detailed analysis.

ACKNOWLEDGMENTS
Parts of this work are supported by National Science Council under the grant number NSC102-2410-H-002-203. The authors thank assistants for their contributions to this work. The authors thank Professor Muh-Chyun Tang and Mr. Mickey Wu for their helpful suggestions.

REFERENCES


