iSwitch: An Open-Access Paper Push and Forwarding Service System: Paper Distribution and Push

Authors: Qian Li, Shi Hongbo, Zhang Xiaolin, Liang Na, Qian Li

Date: 2016-01-27T00:00:00+00:00

Abstract

[Objective] To distribute and push successfully received and parsed open access papers to the repositories of author institutions and funding agencies. [Methods] Analyze the iSwitch system technical framework, design functional modules for paper distribution and push services, and implement paper distribution and push using task scheduling agents and the FTP protocol. [Results] The iSwitch system can implement paper distribution and push services, and has completed the distribution and push of 34,332 article records from Web of Science. [Limitations] Currently, paper distribution and push is only completed based on a single data source, and potential issues in distribution and push services for larger volumes of data from multiple data providers have not been sufficiently considered. [Conclusion] Experiments demonstrate that the workflow mechanism of the distribution and push service is correct, and the distribution efficiency meets future service requirements.

Full Text

Router Service Engine iSwitch for Open Access Articles: Pushing and Routing

Li Qian[1],[2], Shi Hongbo[1], Zhang Xiaolin[1], Liang Na[1]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

[Objective] To route successfully received and parsed open access articles to the institutional repositories of authors' institutions and funding agencies. **[Meth-**

ods] We analyzed the iSwitch system architecture, designed the service modules for article routing, and implemented article distribution and pushing using task scheduling agents and the FTP protocol. [**Results**] The iSwitch system successfully routed 34,332 articles from Web of Science. [**Limitations**] Currently, routing is based on only a single data source, which may not adequately address challenges that could arise when handling larger volumes of data from multiple providers. [**Conclusions**] Experiments demonstrate that the workflow mechanism for pushing and routing is correct and its efficiency meets future service demands.

Keywords: Open Access; iSwitch; Pushing and Routing; FTP; Institutional Repository
Classification Number: G250; TP393

1 Introduction

To support open access to research articles, promote the development of institutional repositories [1], avoid the cumbersome processes, data loss, and version confusion associated with author self-archiving, and establish efficient channels for research dissemination, the National Science Library of the Chinese Academy of Sciences—an active proponent of the open access movement—proposed the Router Service Engine iSwitch for open access articles in 2014 [3]. In this context, we previously established specific technical requirements and standards for iSwitch system construction in "Router Service Engine iSwitch for Open Access Articles: Technical Workflows and Standards" [4], and analyzed the distribution and routing service workflow based on the needs of pushers, routers, and receivers. In "Router Service Engine iSwitch for Open Access Articles: Receiving and Parsing," we presented a detailed design for iSwitch' s receiving and parsing functions, implementing systems to receive open access articles from publishers via FTP [5] and SWORD [6] protocols. According to iSwitch' s fundamental framework and standard workflow, a routing and forwarding module is required to distribute successfully received and parsed articles to the institutional repositories of corresponding institutions and funding agencies. This paper addresses this need by designing and implementing the distribution and routing functionality for iSwitch in accordance with unified technical specifications.

2 iSwitch Article Distribution and Routing Technical Framework

The iSwitch article distribution and routing module implements communication and transmission of distribution data using the FTP protocol. Its core functions include: creating distribution tasks, executing distribution tasks (pushing articles to corresponding institutional or funding agency FTP directories for

periodic harvesting by target receivers), distribution auditing, and distribution information configuration. The overall technical framework is illustrated in [Figure 1: see original paper].

2.1 Task Creation

iSwitch incorporates an automated "article distribution task"generator (taskCreator) that automatically creates routable task sequences based on the list of correctly parsed and institution/funding-mapped open access articles from the previous day (up to midnight of the current day). These tasks are saved to a pending distribution task list for periodic invocation by the distribution trigger.

2.2 Task Execution

The task execution function employs the iSwitch "article distribution task" distributor (taskRouter) to immediately push tasks in a routable state to the corresponding institutional directories on the FTP server. Receiving institutions and funding agencies harvest these according to their scheduled frequencies. Upon successful harvesting, iSwitch automatically updates the distribution status of the articles to "distributed" for subsequent auditing and to prevent duplicate distribution to the same target institutions. During execution, the "pending distribution article data ZIP package" —prepared during the receiving and parsing phase—contains the full-text files (PDF/XML), metadata description XML files, and any additional data files provided by the pusher. The metadata XML files conform to the JATS [7] standard format.

2.3 Distribution Auditing

Distribution auditing examines the results of pushed articles from three dimensions: data source, audit type (distribution task, author institution, and funding agency), and audit scope (time point or period). It tracks both successful and failed distribution records: (1) For successful distributions, the system displays the target institution, the article' s URL in the institutional repository, and the funding agency name; (2) For failed distributions, the system automatically identifies potential failure causes. After correction, iSwitch automatically invokes its exception handling mechanism to retry the distribution.

2.4 Distribution Configuration

To facilitate standardized and convenient article pushing from publishers and reliable reception by institutions and funding agencies, iSwitch employs flexible personalized configuration strategies. Key configurations related to distribution services include: (1) Institution-repository mapping configuration, supporting scenarios where multiple institutions merge into one or multiple institutions share a single repository; (2) Distribution frequency configuration, allowing personalized settings based on each institutional repository' s article volume; (3)

Additional configurations including repository IP addresses and website details to ensure distribution security and provide verification information.

---

3 iSwitch Article Distribution and Routing Module Design

To ensure scientific and rational design of the distribution module while guaranteeing effectiveness and efficiency, the iSwitch article distribution and routing module adheres to four principles: (1) Loose coupling between functional modules and high cohesion within functional units to facilitate invocation and maintenance; (2) Compliance with workflow execution standards, with functional units designed and encapsulated as service interfaces for iSwitch system invocation; (3) Automated execution of distribution functions; and (4) Automatic exception handling mechanisms.

Based on the overall technical framework, the distribution module comprises four components: a scheduled task creation module, an immediate task execution module, an article distribution status update module, and an automatic exception handling module. The overall structure is shown in [Figure 2: see original paper].

3.1 Scheduled Task Creation Module

This module automatically creates distribution tasks for correctly parsed and institution/funding-mapped open articles through agent-based scheduling. It addresses three key issues: (1) **Agent-based automatic scheduling**: To enhance automation, the distribution module implements a "task agent scheduling" function that automatically invokes the task creation process daily according to predefined workflows. (2) **Retrieving distributable articles by institutional harvesting frequency**: Upon activation, the agent automatically identifies correctly parsed but undistributed articles from before the current day (prior to midnight). It retrieves the list of institutions or funding agencies requiring distribution, calculates whether their harvesting frequencies are satisfied, and includes eligible articles in the distribution task list while filtering out those that do not meet the frequency requirements. (3) **Creating routable tasks**: Task creation uses the institution ID from the frequency analysis as an identifier, generating a new "routable status" task number (TaskID) for each eligible institution and updating the TaskID for all corresponding undistributed articles.

3.2 Immediate Task Execution Module

This module executes routable tasks immediately (iSwitch is configured to scan for pending tasks every 5 seconds), distributing article packages to the FTP directories of author institutions and funding agencies. The execution flow involves: (1) locating the author institution; (2) retrieving the distribution batch

number; and (3) obtaining and distributing all pending article data ZIP packages under that batch number. Following this path, the compressed packages are distributed and stored on the server via FTP for periodic harvesting and parsing by author institutions.

3.3 Article Distribution Status Update Module

This module is published as a Web Services interface for institutional repository platforms to invoke, enabling updates to the distribution status of pushed articles. This prevents duplicate distribution while supporting iSwitch' s auditing and exception handling capabilities. When articles fail to push due to uncertain causes, the system automatically updates their status to ensure the task agent reschedules them for eventual delivery to the target institutions or funding agencies.

3.4 Distribution Task Exception Handling Module

iSwitch distribution tasks have four status codes: 1 (routable), 2 (distribution complete), 3 (partial failure), and 4 (complete failure). When a task encounters exceptions that prevent successful pushing, it is marked as 3 or 4 (determined after task completion). The system automatically reschedules such tasks twice; if failures persist, the task is removed from automatic scheduling and requires manual auditing. After resolving the issues, administrators can manually trigger the "distribution task exception handling module" to complete the push for undistributed articles, as shown in [Figure 2: see original paper].

---

4 iSwitch Distribution Auditing Module Design

The iSwitch distribution auditing function examines and statistics the results of distributed articles based on distribution task outcomes, ensuring reliable delivery and improving service quality while providing further validation of publisher data. The audit module design covers four dimensions—publisher, author institution, funding agency, and distribution task—and performs real-time auditing within time windows. It tracks the total number of articles requiring distribution (based on iSwitch receiving and parsing results), successfully distributed articles (status: distributed), and failed distributions (status: failed). The audit module structure is illustrated in [Figure 3: see original paper].

---

5 iSwitch Article Distribution and Routing Experiments

To validate the effectiveness of iSwitch's article pushing and routing services and ensure high-quality future operations, we conducted experiments. The following describes the experimental setup and results analysis.

5.1 Experimental Setup

- **Open access article pusher**: Thomson Reuters
- **Data source**: Web of Science
- **Data volume**: 34,332 articles
- **Covered Chinese Academy of Sciences institutions**: 99
- **Covered funding agencies**: 2
- **Coverage period**: 2013
- **Open access article router**: iSwitch
- **Open access article receivers**: Institutional Repositories (IRs)
- **Routing communication protocol**: FTP

5.2 Results and Analysis

(1) Overall Results: By scheduling the iSwitch distributor periodically, we pushed all 34,332 open access articles. Detailed statistics are presented in , with an overview shown in [Figure 4: see original paper]. The results demonstrate successful distribution of all experimental data, including cases where a single article belonged to multiple institutions, which were correctly routed to the corresponding author institutions or funding agencies. Thus, iSwitch' s distribution functionality meets the project objectives.

(2) Overall Efficiency: iSwitch supports both FTP and SWORD protocols for distribution, with FTP offering higher efficiency for batch processing. This experiment employed FTP-based communication. Detailed efficiency analysis is shown in , revealing that iSwitch' s FTP-based routing service achieves high efficiency, enabling timely batch forwarding to author institutions and funding agencies' IRs, which can then rapidly collect their research outputs.

(3) Auditing Results: We audited the push results across three dimensions—data source, audit type, and author institution. As shown in [Figure 5: see original paper], all 99 institutions received the routed articles. The audit confirms that all 99 institutions involved in the experimental data correctly received iSwitch' s forwarded articles, further validating the correctness of the entire workflow.

We implemented iSwitch' s distribution and routing functionality according to its overall design framework, technical specifications, and workflow. Using FTP communication, the system securely and efficiently distributed article data to corresponding institutional FTP directories. The mechanisms for "automatic

task creation, automatic task execution, and exception handling" ensure that research institutions can collect their scholarly assets in a timely and effective manner, promoting IR development and the dissemination and utilization of open access articles. Future work will focus on: (1) expanding to multiple data sources and larger volumes of open research articles to further refine the technical implementation based on practical distribution outcomes; and (2) developing technical service mechanisms to validate and improve article information in institutional repositories based on iSwitch distribution services.

References

[1] 张冬荣, 祝忠明, 李麟, 等. 中国科学院机构知识库建设推广与服务 [J]. 图书情报工作, 2013, 57(1):20-25. (Zhang Dongrong, Zhu Zhongming, Li Lin, et al. Construction, Promotion and Service of CAS IRs[J]. Library and Information Service, 2013,57 (1): 20-25.)

[2] 张晓林, 刘细文, 李麟, 等. 研究图书馆推进开放获取的战略与实践——以国家科学图书馆为例 [J]. 图书情报工作,2013, 57 (1): 15-19, 48. (Zhang Xiaolin, Liu Xiwen, Li Lin,et al. The Strategies and Practices of Research Library to Support Open Access —Taking National Science Library as an Example [J]. Library and Information Service, 2013,57 (1): 15-19, 48.)

[3] 张晓林, 梁娜, 钱力, 等. 开放获取论文推送转发服务系统 iSwitch: 概念、功能与基本框架 [J]. 现代图书情报技术, 2014(10): 4-8. (Zhang Xiaolin, Liang Na, Qian Li, et al. Router Service Engine iSwitch for Open Access Articles: The Concept, Strategy, and Framework[J]. New Technology of Library and Information Service, 2014(10): 4-8.)

[4] 梁娜, 张晓林, 钱力, 等. 开放获取论文推送转发服务系统 iSwitch: 技术流程与标准 [J]. 现代图书情报技术, 2014(10): 9-13. (Liang Na, Zhang Xiaolin, Qian Li, et al. Router Service Engine iSwitch for Open Access Articles: Technical Workflows and Standards[J]. New Technology of Library and Information Service, 2014 (10): 9-13.)

[5] Serv-U File Server Administrator Guide [EB/OL]. [2014-12-25]. http://www.serv-u.com/Serv-U-Administrator-Guide.pdf.

[6] Simple Web-service Offering Repository Deposit [EB/OL]. [2014-12-25]. http://swordapp.org/about/.

[7] Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV [EB/OL]. [2014-12-25]. http://docs.rioxx.net/v2-0-beta-1/.

Author Contributions

Li Qian: Conducted requirements research for the iSwitch pilot system, developed the article distribution component, and wrote the paper.
Shi Hongbo: Conducted requirements research for the iSwitch pilot system and developed the article receiving and parsing component.
Zhang Xiaolin: Proposed and refined technical workflows and requirements, and reviewed the paper.
Liang Na: Refined technical workflows and requirements, and proposed reference technical standards.

(Corresponding author: **Li Qian**, ORCID: 0000-0002-0931-2882, E-mail: qianl@mail.las.ac.cn)

Note: Figure translations are in progress. See original paper for figures.