http://www.calsec.or.kr/jsebs
ISSN: 2288-3908

The Journal of Society for e-Business Studies
Vol.19, No.4, November 2014, pp.61-72
http://dx.doi.org/10.7838/jsebs.2014.19.4.061

전자학술정보 이용통계 유통을 위한 표준 플랫폼 구축

Implementation of Standard Platform for Distributing Usage Statistics of Digital Scholarly Information

정영임(Youngim Jung)*, 김재훈(Jayhoon Kim)**,
김광영(Kwangyoung Kim)***, 김환민(Hwanmin Kim)****

초 록

전자학술정보의 활용이 급속히 증가하고 대규모 이용로그 정보의 처리가 가능해지면서 전자학술정보의 이용에 대한 분석이 다양한 각도에서 이루어지고 있다. 그러나 각 출판사에서 각기 다른 형태와 방법으로 생성한 이용통계를 표준 포맷으로 통합 수집하여 국가 차원의 이용량 정보를 제공하는 통합 플랫폼은 없다. 따라서 본 논문에서는 국내 520여 기관의 전자학술정보 이용통계를 통합 수집하고, 이용통계 데이터 표준 프로토콜인 SUSHI(Standardized Usage Statistics Harvesting Initiative)에 기반하여 유통할 수 있는 플랫폼 모델을 제안한다.

ABSTRACT

Recently usage on digital scholarly information has been analyzed with various respects by various parties since the rapid expansion of digital scholarly information use and the increasing availability of large-scale log data. Nevertheless, no standard platform for distributing usage statistics of scholarly information at the national scale has been suggested so far. Therefore, this paper suggests a generalized SUSHI (Standardized Usage Statistics Harvesting Initiative) platform for distributing usage statistics of digital scholarly information at the national scale.

키워드 : 학술정보, 이용통계, 통합, 유통, COUNTER, SUSHI, 스크린 스크래핑
Scholarly information, Usage Statistics, Consolidation, Distribution, COUNTER, SUSHI, Screen Scraping

본 논문은 한국과학기술정보연구원 '핵심과학기술 전자정보 공동활용체제구축(K-14-L02-C01-S02)' 과제의 지원을 받아 수행되었음.

* First Author, Department Overseas Information, Information Service Center, KISTI(acorn@kisti.re.kr)
** Department Overseas Information, Information Service Center, KISTI(jay.kim@kisti.re.kr)
*** Department Overseas Information, Information Service Center, KISTI(glorykim@kisti.re.kr)
**** Corresponding Author, Department Overseas Information, Information Service Center(mrkim@kisti.re.kr)

2014년 04월 03일 접수, 2014년 08월 25일 심사완료 후 2014년 09월 25일 게재확정.

1. 서 론

전자학술정보의 유통과 활용이 급속히 증가하고 대규모 이용로그 데이터 처리와 분석이 가능해지면서 이용통계 활용의 중요성이 나날이 증가되고 있다. 도서관에서는 자관의 실정에 맞는 전자저널을 구성하고 관리해 나가기 위해 전자저널에 대한 자관 이용자들의 요구와 이용 현황을 파악하는 것이 무엇보다 중요해졌다[8]. 전자정보 컨소시엄 주관기관에서도 컨소시엄 내에서 유통되는 학술자원의 이용통계에 대한 분석을 통해 국가 차원의 전자학술저널의 유통 현황 파악 및 수요자 중심의 정보수집 정책 개발이 필수적이다[6]. 출판사에서도 자사에서 출판되는 저널 중 어느 저널이 많이 이용되는지를 분석하고, 이용량에 기반한 새로운 가격 모델을 제시하고 있다[13]. 또한, 최근 북미 연구자들과 세계표준기구를 중심으로 기존의 인용 기반 영향력 지수를 보완하기 위해 이용량에 기반한 저널의 영향력을 평가할 수 있는 지수의 개발 연구를 진행하고 있다. 인디애나 주립대의 Bollen 교수는 미국 Andrew W. Mellon 재단으로부터 3년(2006~2008), 미국 국립과학재단(National Science Foundation)으로부터 4년(2009~2012), 총 7년간 연구비를 지원받아 MESUR 프로젝트를 진행하였다. 이 프로젝트에서는 학술영향력 지수로서의 이용데이터의 가능성을 살펴보고, 이용량에 기반해 개발한 지수는 다양한 형태의 학술 커뮤니케이션 평가에 새롭고 광범위한 시각을 제공한다고 결론 내렸다[1]. 전자학술정보 이용통계 데이터 양식의 표준 지침인 Counting Online Usage of NeTworked Electronic Resources (COUNTER)를 개발한 COUNTER 그룹에서도 2008년부터 Journal Usage Factor(JUF) 개발 프로젝트를 야심차게 진행하고 있다[14]. 또한, SNS(Social Network Service) 기반 추천(페이스북, 트위터, 블로그, 멘델레이, 북마크) 및 논문 다운로드 수를 종합적으로 지수화한 altmetrics 개발 및 적용 연구가 진행되어[15] 2011년부터 ACM Web Science Conference의 한 분과인 Altmetrics12 Workshop에서 관련된 주체(연구자, 연구지원 에이전시, 출판사, 표준 담당자)들이 모여 altmetrics의 유효성, 잠재적 가치, 약점, 기존 방법들과의 관계를 논의하고 있다[4, 5].

이와 같이 이용통계의 활용에 대한 중요성이 강조되고 있고, 다양한 측면에서 활용할 수 있는 방안에 대해 연구되고 있다. 그러나 복수의 정보공급사 이용통계를 통합적으로 자동수집하여 기관 단위, 컨소시엄 단위, 정보공급사 단위, 국가 단위의 대용량 이용통계를 유통할 수 있는 시스템은 제안되지 않았다. 특히 통합 표준 플랫폼의 적용은 데이터의 활용성을 증가시키며, 관련된 타 시스템과의 용이한 결합을 보장한다[2]

따라서 본 연구에서는 (1) 국내 학술연구기관의 정보공급사 전자학술정보 이용통계를 통합수집할 수 있고, (2) 통합 수집한 이용통계를 저널 단위, 컨소시엄 단위, 정보공급사 단위로 산출할 수 있으며, (3) 권한에 따라 이용통계를 유통할 수 있는 통합 표준 플랫폼 모델을 제안한다.

2. 선행 연구

2.1 이용통계 기계적 수집을 위한 표준 프로토콜

미국 정보표준화기구(National Information

Standard Organization, NISO)의 Standardized Usage Statistics Harvesting Initiative(SUSHI) Protocol은 웹 서비스 프레임워크에 기반하여 전자자원의 이용데이터를 자동으로 요청하고 응답하는 모델을 정의하였다. 도서관의 이용통계를 수집하기 위해 사서가 출판사 웹 사이트에 로그인하여 수집하거나 이메일을 통해 전달을 받는 기존의 시간 소모적인 방식을 대신해 SUSHI는 출판사 서버와 이용통계를 요청하는 클라이언트 간 기계적인 방식으로 대규모 이용통계의 신속하고 효율적인 전송을 목표로 한다. SUSHI를 통해 여러 종류의 이용통계 리포트를 검색할 수 있으며, 특히 COUNTER 표준 리포트 검색에 적합하다. SUSHI는 메시지를 주고 받기 위해 SOAP (Simple Object Access Protocol, SOAP)에 기반하여 설계되었다. 클라이언트는 'GetReport' 방식으로 'ReportRequest'를 전달하고, 서버는 출력 메시지로 'ReportResponse'를 반환한다. SUSHI 표준은 웹 서비스 네임스페이스와 작동을 기술하기 위한 웹 서비스 기술언어(Web Services Description Language, WSDL)를 포함하고 있으며 SUSHI 통사(syntax)규칙에 부합하는 일반적인 XML 스키마를 포함한다[12]. SUSHI 프로토콜은 일반적이면서 광범위하게 적용될 수 있도록 설계되었다고 하지만 SUSHI 서버를 구현하여 이용통계를 제공하는 자를 정보공급사로만 한정하였고 유통되는 이용통계 역시 단일 정보공급사의 이용통계 보고서로 한정하였다. 따라서 광범위한 목적으로 다양한 주체에 의해 전자학술정보 이용통계를 유통할 수 있는 더 일반화된 모델로 설계되어야 한다.

2.2 이용통계 수집 시스템

도서관 사서들이 복수 정보공급사의 웹 사이트를 일일이 방문하여 이용통계 보고서를 수집하고 통합하는 데 소요되는 시간과 노력을 줄이기 위해, 일부 도서관 전산화 솔루션 벤더에서 도서관 이용통계를 기관 단위로 수집할 수 있는 상용화 시스템 또는 서비스를 제공하고 있다.

Serials Solution에서는 360 COUNTER Data Retrieval Service를 통해 다양한 방법으로 도서관 이용통계를 수집하고 여러 종류의 학술자원에 대한 이용통계 리포트를 통합 제공한다. 이용통계 리포트를 자동으로 수집하기 위해서는 SUSHI 프로토콜을 이용하고 있다[16].

NISO SUSHI의 공동 의장이자 EBSCO 수석전략가 Oliver Pesch를 필두로 EBSCO에서는 이용통계 데이터를 수집, 통합하여 기존 구독 관리시스템과 연동하여 장서개발에 필수적인 이용-비용 분석을 지원하는 Usage Consolidation(UC)를 개발하였다[9]. 그러나 이와 같이 이용통계를 각 기관 단위로 수집하여 기관 단위로만 이용된다면, 저널의 국가적 이용량을 파악하기가 어렵고, 기관 단위로 수집된 이용통계를 다시 통합해야 하는 번거로움이 있다.

학술정보가 전자화되고, 저널이 패키지로 묶여 빅딜로 구독되면서 국가별, 지역별로 도서관 컨소시엄이 구성되었고, 전 세계적으로 200개가 넘는 컨소시엄이 구성되어 운영되고 있다. 따라서 컨소시엄 운영 주체가 컨소시엄을 통해 유통되는 전자학술정보의 이용현황을 파악하고, 컨소시엄 멤버 기관에 원스톱

이용통계 조회 서비스를 제공하기도 한다. 영국 정보시스템합동위원회(Joint Information Systems Committee, JISC)는 저널 이용통계 포털(Journal Usage Statistics Portal, JUSP)을 설립하여 대학 도서관이 쉽고 빠르게 여러 출판사의 저널 이용량을 비교할 수 있는 서비스를 제공하고 있다[10]. JUSP는 JISC Collection, 맨체스터 대학의 Mimas 등의 컨소시엄이 공동 운영하고 있다. JUSP를 통해 148여 개 도서관의 30여 개 출판사의 이용통계를 수집하여 원스톱으로 이용통계를 조회하고 다운로드 받을 수 있다. 수집 대상 출판사 중 SUSHI를 지원하는 American Chemical Society, Karger 출판사 등의 이용통계 데이터는 자동으로 수집하고, SUSHI를 지원하지 않는 출판사의 이용통계 데이터는 각 플랫폼에서 수동으로 수집하고 있다.

그러나 SUSHI 프로토콜을 통해 수집된 결과물이 XML(eXtensible Markup Language, XML) 파일이기 때문에 사람이 보기에 편한 엑셀 파일 등으로 변환해야 한다. 또, 표준이라고는 하지만 여전히 이용자 인증 방식, WSDL 지원 여부, 오류 처리 방식이 출판사별로 다르다는 문제를 안고 있다. 예를 들어, ProQuest는 웹 서비스 보안을 적용하여 ID/PW 및 SUSHI ReferenceID, CustomerID의 4가지 값을 입력해야 하고, Optical Society of America는 IP 인증 방식을 취하고 있다. 대부분의 SUSHI 채택 출판사에서는 WSDL을 지원하나 Oxford University Press 등은 WSDL을 지원하지 않아 프로그램 유지가 어렵다. 무엇보다 Elsevier, Wiley and Sons 등 대형 출판사를 포함한 대부분의 출판사에서 SUSHI를 적용한 이용통계 서비스를 제공하지 않고 있고, SUSHI를 지원한다는 출판사 중에서도 Jstor, Brill, Emerald 등과 같이 SUSHI를 통한 이용통계 서비스가 실제로는 제공되지 않는 경우도 많아 SUSHI 프로토콜에만 의지해 전자학술정보 이용통계를 수집하기에는 현실적인 어려움이 적지 않다[6].

2.3 이용통계 유통 시스템

NISO의 SUSHI 표준 프로토콜에 정의된 이용통계 제공자는 학술정보 이용통계를 생성하고 서비스는 정보공급사 및 출판사로만 한정되었다. 그래서 서버 사이드 규약은 단일 출판사가 학술정보 소비자인 개별 기관이나 컨소시엄 관리자가 요청하는 기관의 이용통계를 유통하는 모델이다. 그러나 서론에서 기술한 바와 같이 전자학술정보 이용통계를 유통하고 활용하는 주체가 출판사와 그 소비자를 넘어서서 다양한 주체들에 의해 학술정보 이용통계가 활용되고 있다. COUNTER 그룹에서는 중앙 레지스트리를 통해 저널의 글로벌 이용지수 통합 및 유통의 필요성을 언급했고[3] 국내에서도 교육부 출연 한국학술정보원에서 전국대학의 도서관 전자서비스 이용통계를 취합하여 매년 대학교육통계(대학도서관통계)를 내고 있다[11]. 그런데 각 대학마다 이용통계를 개별적으로 수작업으로 제출하고 있어 각 대학도서관 담당자와 교육부 담당자의 낭비되는 시간이 막대하다. 이에 국가적으로 공신력이 있는 제 3의 기관에서 국내 연구기관의 이용통계를 통합 수집하고 이를 필요한 기관에 표준 프로토콜을 이용해 제공할 수 있는 모델이 필요하다. 즉, 통합 수집된 복수 기관의 복수 정보공급사 이용통

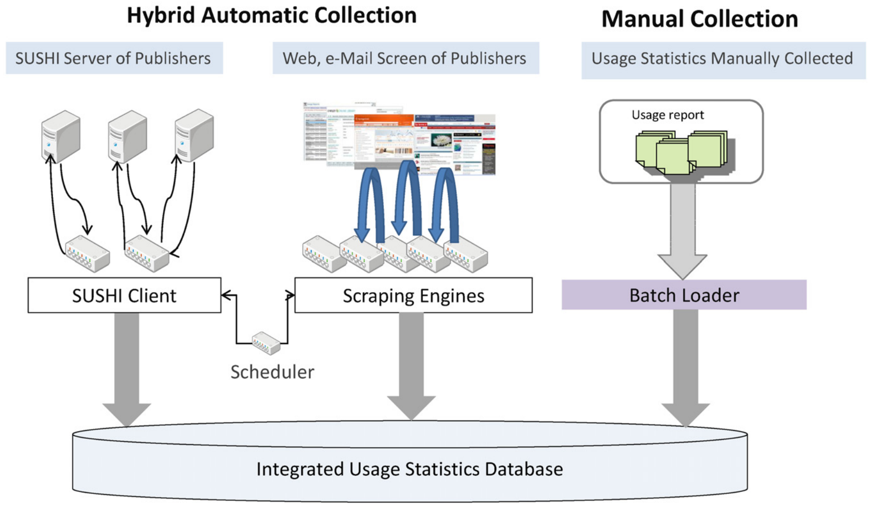
계에서 저널 단위, 컨소시엄 패키지 단위, 정보공급사 단위로 전체 이용량을 산출하고, 해당 데이터의 이용 권한을 확인하여 이용통계를 표준화된 방식으로 유통할 수 있는 플랫폼이 마련되어야 한다. 그러나 국내외 선행연구에서 이러한 통합 유통 플랫폼에 대해 제안된 바가 없다.

3. 국가 차원 이용통계 통합 플랫폼 모델

선행연구에서 살펴 본 바와 같이 국가 차원에서 이용하는 전자학술정보의 이용통계를 모두 수집하기 위해서는 국제 표준인 SUSHI 프로토콜을 지원하는 출판사, SUSHI를 지원하지 않는 출판사, 수동으로 이용통계를 제공하는 출판사의 이용통계를 통합적으로 수집할 수 있는 플랫폼을 구축해야 한다. 또한, 수집된 이용통계를 데이터 표준 포맷에 따라 설계된 통합 데이터베이스로 구축해야 한다. 제 3장에서는 <Figure 1>과 같이 이용통계를 제공하는 모든 출판사의 이용통계를 수집할 수 있는 복합적 방식의 이용통계 통합 플랫폼을 기술한다.

3.1 전자학술정보 이용통계 자동수집 모듈

SUSHI 프로토콜 표준을 채택한 출판사 서버에 클라이언트의 이용통계 요청 메시지(Request Report)가 접수되면 서버에서 이용권한을 확인한 후 결과값으로 요청한 이용통계 데이터를 담아 응답 메시지(ResponseReport)를 전송한다. American Chemical Society, Annual Reviews, Cambridge University Press, Royal Society of Chemistry 등 SUSHI 채택 출판사의 이용통계 수집을 위한 SUSHI 클라이언트를 개발하였다. SUSHI 서버를 구현하지 않은 출판사의 이용통계를 수집하기 위해서는 스크린 스크래핑 기술을 적용하였다. 본 플랫폼에



〈Figure 1〉 Consolidation Platform of National Usage Statistics

서 스크린 스크래핑 방식으로 이용통계를 수집하는 출판사는 Springer, Wiley and Sons, Elsevier 등 대형 출판사를 포함한 32개사이다. 스크래핑 엔진에서는 확보된 이용통계 조회 계정 정보를 이용하여 각 정보공급사 통계조회 사이트에 자동 로그인하고, 자동수집 일정에 따라 스크래핑 엔진을 실행하여 전자정보 이용통계 데이터를 추출한다. 이용통계 보고서의 파일 포맷은 HTML(HyperText Markup Language), CSV(Comma-Separated Values), PDF(Portable Document Format), XML, XLS (Excel Binary File Format) 등 다양하며, 출판사별로 이용통계 보고서 양식이 다르다. 따라서 각 출판사 이용통계 보고서를 분석하여 이용통계를 수집한다[6].

SUSHI 클라이언트와 스크래핑 엔진은 스케쥴러에 기록된 자동수집 일정에 따라 자동 실행되며 자동수집일에 수집이 실패하면 다음 정해진 2차 일정, 3차 일정대로 수집이 성공할 때까지 수집을 시도한다[7]. 그리고 실패 원인을 기록하여 시스템 관리자가 실패 원인을 파악하고 문제를 해결할 수 있도록 하였다. 네트워크 오류 및 출판사 서버 등의 일시적인 오류에 의한 수집 실패는 일정 시간이 지난 후 대부분 복구되기 때문에 2, 3차 시도에서 수집이 성공적으로 이루어지며, 출판사 통계 시스템 변경, 통계조회 계정 오류 등에 의한 수집 실패는 시스템 관리자가 해당 변경 사항을 적용하거나 오류를 복구한 후 수집을 바로 실행할 수 있도록 비정기 수집 기능도 지원한다[6].

3.2 수동수집 이용통계 데이터 업로더

네트워크 오류나 출판사 웹 사이트의 일시적인 오류는 금방 복구가 되어 스케쥴러의 수집 일정에 따라 이용통계가 자동으로 재수집된다. 그러나 출판사의 웹 사이트 플랫폼 혹은 SUSHI 서버가 변경되는 경우 통상 6개월 이상의 시간이 소요되어 해당 기간 동안 이용통계 데이터를 수집할 수 없다. 또, 이용통계를 자동으로 수집하는 못하는 출판사도 여전히 많다. 주로 군소 출판사나 비서구권 출판사의 경우 해당 출판사의 저널을 구독한 기관들이 이용통계를 조회할 수 있는 웹 사이트를 제공하지 않고, 기관이 요청을 하면 이용통계 데이터를 이메일로 전송하기 때문이다. 이러한 경우 출판사로부터 이용통계를 수동으로 수집하고 이를 통합 데이터베이스에 배치로 업로드 한다.

3.3 통합 이용통계 표준 데이터베이스 구축

<Figure 1>에 도식화된 바와 같이 복합적 자동수집 방식으로 수집된 이용통계 데이터와 수동으로 수집된 이용통계 데이터는 COUNTER의 “Code of Practice COUNTER” Release 4의 표준 양식에 맞춰 가공한다. 가공된 데이터는 컨소시엄 패키지 정보와 저널의 서지 정보를 연계하여 통합 이용통계 표준 데이터베이스로 구축한다.

4. SUSHI 표준 기반 통합 이용통계 유통 플랫폼

제 3장에서 기술한 바와 같이 다양한 경로를 통해 구축된 국가 차원의 이용통계를 다양한 목적으로 다양한 주체 간 유통하기 위해서 제

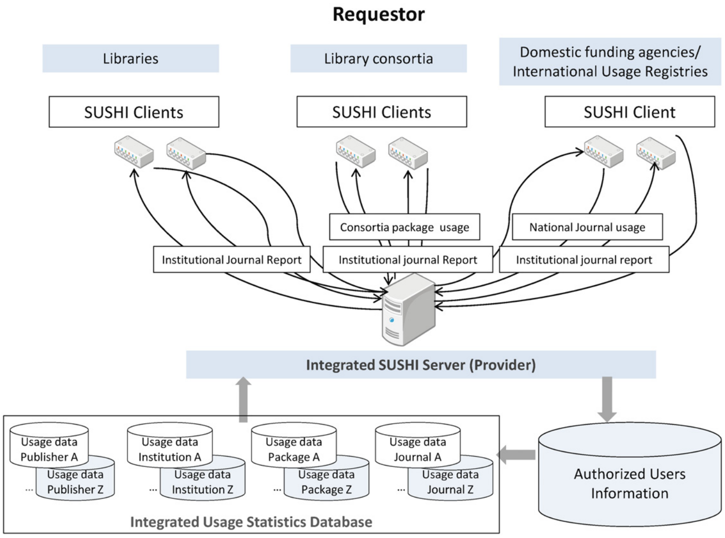
4장에서는 SUSHI 표준을 수정하여 기계 간 복수 출판사의 통합 이용통계를 다양한 수준으로 유통할 수 있는 모델을 제안한다.

4.1 SUSHI 서버 구성

SUSHI는 출판사에서 제공하는 COUNTER 기반의 이용통계 데이터를 기계 간 주고 받을 수 있는 프로토콜이며, 이용통계를 제공하기 위한 서버의 데이터 규약에 대해서도 잘 기술되어 있다. 본 플랫폼에서는 NISO의 1.5버전에 기반하되, 단일 서버에서 복수 출판사의 이용통계를 제공할 수 있도록 SUSHI 서버 표준 규약을 <Figure 2>와 같이 수정하였다.

NISO의 1.5버전에서는 데이터를 요청하는 주체(Requestor)를 개별기관 도서관과 도서관 컨소시엄 운영자로만 가정하였다. 그러나 본 연구에서 제안하는 통합 SUSHI 서버에서는 Requestor를 도서관, 도서관 컨소시엄 운영자, 국내 연구 예산 책정 및 심의 기관, 글로벌 중앙 이용통계 레지스트리 등으로 다양하게 정의하였고, 주체별로 각 이용통계 데이터에 접근할 수 있는 권한 및 계정 정보를 데이터베이스로 구축하였다. 또 NISO의 서버 규약은 단일 출판사가 자사의 이용통계만을 제공하는 모델을 제공하지만, 본 통합 SUSHI 서버는 복수 출판사의 다양한 이용통계를 저널 단위, 컨소시엄 단위, 출판사 단위로 산출하여 제공한다. 일반화된 통합 SUSHI 서버는 Requestor의 계정 정보와 권한을 확인하여, 해당 기관이 요청하는 이용통계 데이터 또는 오류 메시지를 회신한다. 본 통합 SUSHI 서버를 구성하는 모듈은 다음과 같다.

- Sushi.jsp : SUSHI Server 시작 페이지이다.
- Request.getReader : ReportRequest 메시지를 받아들인다.



〈Figure 2〉 Generalized SUSHI Server for Providing Multi-layered Usage Statistics

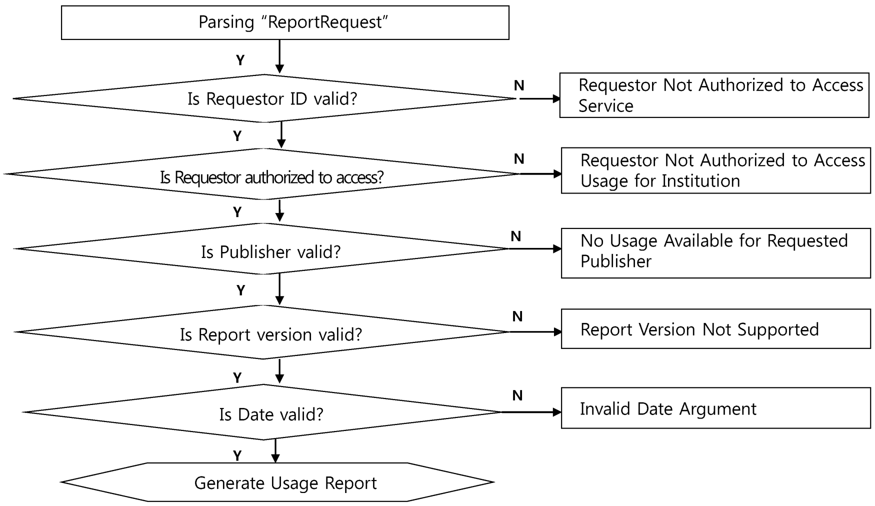
〈Table 1〉 Suggested Warnings and Error Messages for Integrated Error Handling

Exception	Severity level	Exception Number	Invocation Conditions
Partial Data Returned	Warning	3040	Request could not be fulfilled in its entirety. Data that was available was returned.
No Usage Available for Requested Publisher	Error	4000	Service did not find any data for the publisher requested

- ReadXml : 파라미터를 읽어 요청 메시지를 분석하는 Bean
- DecisionException : 요청 메시지의 오류 여부를 판단하는 Bean
- GenerateErrorReport : 오류 유형에 따른 Error Report 생성한다.
- GenerateUsageReport : 요청하는 이용통계 데이터를 생성한다.

4.2 예외 처리

서비스 예외 사항 또는 오류에 대한 처리 방침은 SUSHI 표준 규약을 준수한다. 다만, 제 4.1절에서 밝힌 바와 같이, 본 연구에서 개발하는 SUSHI 서버는 복수 출판사의 이용통계 데이터를 제공해야 하므로, SUSHI 클라이언트가 요청하는 이용통계 데이터의 출판사 정보가 유효한지 판단하는 부분이 추가된다. 추가된 부분을 포함하여 본 연구에서 개발한 SUSHI 서버의 예외 처리 루틴은 <Figure 3>과 같다. SUSHI 클라이언트로부터 전달된 'ReportRequest' 메시지를 파싱하여 <Figure 3>과 같은 순서로 접근 권한이 있는 자의 요청인지, 유효한 출판사의 이용통계 리포트 버전, 기간의 데이터를 요청한 것인지 파악한다.



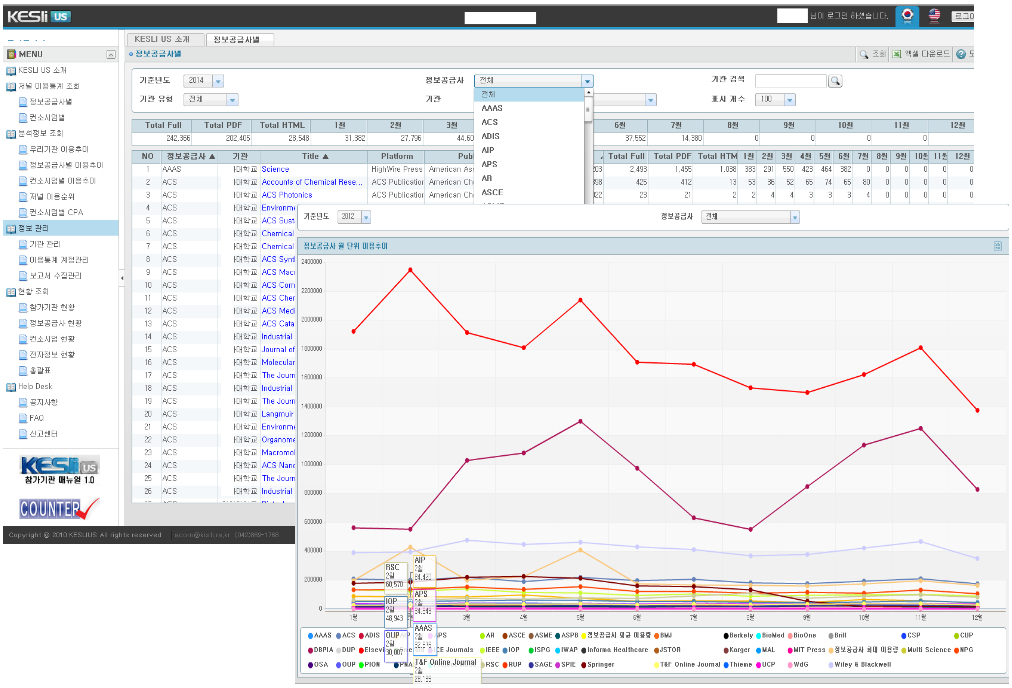
〈Figure 3〉 Error Handling in Suggested Integrated SUSHI Server

각 조건에서 처리 거절된 요청에 대해서는 예외 메시지를 생성한다. NISO의 SUSHI 표준 문서에서 정의한 예외 처리 메시지를 적용하되, 본 통합 SUSHI 서버에서는 통합 데이터베이스에 저장이 되지 않은 출판사의 이용통계 보고서 요청에 대한 예외 처리에 대해 추가로 정의하였고, <Table 1>과 같이 추가로 정의한 오류 메시지를 생성하였다,

예외가 발생한 일시, 예외 코드, 엄격도, 예외 내용 등이 포함된 예외 메시지를 xml 형식으로 생성하여 SUSHI 클라이언트에게 반환한다.

4.3 이용통계 리포트 생성

SUSHI 클라이언트에서 전송된 Request-Report 메시지가 예외 처리 루틴을 거치면서 요청된 기관, 출판사의 해당 이용 기간 동안의 이용통계 데이터를 이용통계 통합 데이터베이스에서 검색하고 해당 데이터에 대한 접근 권한이 있으면 검색된 데이터를 제공한다. 요청된 데이터가 COUNTER 보고서이면 검색 결과로 COUNTER 보고서를 반환하고, 요청된 데이터가 COUNTER 표준 보고서가 아닐 경우, 이를 COUNTER 보고서에 준하게 가공하여 xml 형태의 리포트로 제공한다. 요청을 받은 보고서 및 반환 메시지를 생성한 일시, 클라이언트의 Requestor, CustomerReference 정보를 포함하여 ReportResponse 메시지를 생성하고 이를 클라이언트에게 반환한다. 이상에서 기술한 이용통계 통합 유통 표준 플랫폼을 구현한 결과는 <Figure 4>와 같다. 본



〈Figure 4〉 Implemented Standard Platform for Usage Statistics Distribution

플랫폼에서는 다수 출판사의 이용통계 데이터를 복합적인 방법으로 통합하고, 통합 이용통계 데이터베이스를 논문 서지정보 데이터베이스 및 각 기관의 저널 패키지 구독정보 데이터베이스와 연계하였다. 또, 다년간에 걸쳐 수집된 데이터를 차트를 통해 제공함으로써, 다양한 분석정보 산출이 가능하도록 하였다. 단, 이용통계 데이터를 유통할 수 있는 SUSHI 서버는 기계 간 데이터 전송을 목표로 하여 이용자 인터페이스를 구현하지 않았다.

5. 결론 및 향후 연구

최근 다양한 주체에 의해 이용통계의 활용에 대한 중요성이 강조되고 있고, 다양한 측면에서 활용할 수 있는 방안에 대한 연구가 활발히 진행되고 있다. 그러나 본 연구에서 살펴 본 바와 같이 복수의 정보공급사 이용통계를 통합적으로 수집하여 기관 단위, 컨소시엄 단위, 정보공급사 단위, 국가 단위의 이용통계를 산출하고, 이를 표준 프로토콜에 기반하여 유통할 수 있는 시스템은 제안되지 않았다.

이에 본 연구에서는 국내 500여 학술연구기관의 국내외 출판사 전자학술정보 이용통계를 통합 수집하여 대규모 이용통계 데이터베이스로 구축하기 위해 국제 표준인 SUSHI 기반 이용통계 자동수집, 스크래핑 기술에 의한 자동수집 및 수동수집 데이터의 배치 로더를 결합한 이용통계 통합 플랫폼 모델을 제안하였다. 또, 구축된 기관별 저널 단위 이용통계를 가공하여 저널 단위, 컨소시엄 단위, 출판사 단위, 기관 단위 등 다양한 통계데이터를 산출할 수 있으며, 권한에 따라 복수 출판사의 복수 기관 이용통계를 유통할 수 있는 일반화된 통합 유통 표준 플랫폼을 구축하였다.

현재 대규모 이용통계가 구축된 데이터베이스를 안정적이고 효율적으로 운영하기 위해서는 기존 데이터베이스 관리 체제에서는 한계가 있다. 현재는 다수의 대규모 데이터베이스를 연동하여 다양한 분석 자료를 실시간으로 산출할 수 없기 때문에 과거 데이터의 경우 분석용으로 산출한 데이터를 다시 저장하여 서비스하고 있어 전체 데이터 규모가 더 증가한다. 또, 현재는 출판사로부터 전자저널의 이용통계 데이터만 수집하고 있으나, 전자책 이용통계 보고서, Web DB 이용통계 보고서 및 멀티미디어 이용통계 보고서 등 수집 범위를 확대하여 서비스하려면 빅데이터 운영 체제로 전환해야 한다. 이에 대한 연구도 향후 지속할 예정이다.

References

[1] Bollen, J. and Van de Sompel, H., "Usage Impact Factor : The effects of sample characteristics on usage-based impact metrics," Journal of the American Society for Information Science and technology, Vol. 59, No. 1, pp. 1-14, 2008.

[2] Chang, T., Yoon, S., and Lim, H., "Classification and Standardization of Master-Data of Supply Chain for Adopting Common Standard Platform," Journal of Soci-

ety for e-Business Studies, Vol. 17, No. 1, pp. 151-171, 2012.

[3] COUNTER, "The COUNTER Code of Practice for Usage Factors : Release 1" available at http://www.projectcounter.org/documents/UF_CoP_R1_Mar2014.pdf.

[4] Gunn, W. and Reichelt, J., "Social Metrics for Research : Quantity and Quality," presented at ACM Web Science Conference 2012 Altmetrics Workshop, Evanston, US, 2012.

[5] Holbrook, J. B. and Frodeman, R., "Peer review and the ex ante assessment of societal impacts," Research Evaluation, Vol. 20, No. 3, pp. 239-246, 2011.

[6] Jung, Y. and Kim, J., "Improving Efficiency of Usage Statistics Collection and Analysis in E-Journal Consortia," Journal of the Korean Society for Information Management, Vol. 29, No. 2, pp. 7-25, 2012.

[7] Jung, Y. and Kim, J., "Hybrid Standard Platform for E-Journal Usage Statistics Management," Lecture Notes in Electrical Engineering, Vol. 215, pp. 1105-1115, 2013.

[8] Kim, S., "Toward the Effective Utilization of Usage Statistics for the Management of Electronic Journals," Journal of Information Management, Vol. 41, No. 4, pp. 69-91, 2010.

[9] Kim, S., "Integrated Application for E-resource Usage, EBSCONET® Usage Consolidation," Digital Library, Vol. 65, pp. 79-93, 2012.

[10] Lambert, J. and Conyers, A., "Adding Value to Usage Statistics : the Journal Usage Statistics Portal(JUSP) Enhancement," The Serials Librarian, Vol. 63, No. 3-4, pp. 305-314, 2012.

[11] Ministry of Education, "Statistics of University Library" available at http://www.rinfo.kr/.

[12] NISO, "NISO Standardized Usage Statistics Harvesting Initiative(SUSHI) : Z39.93" available at http://www.niso.org/apps/group_public/download.php.

[13] Pastore, S., "KESLI-Supporting the Expansion of Korean Science," presented at 2013 KESLI Forum, Pyeongchang, ROK, 2013. 04.

[14] Peter, S., "Journal Usage Factor : a promising new metric," Serials, Vol. 24, No. 1, pp. 64-68, 2011.

[15] Priem, J. and Hemminger, B. M., "Scientometrics 2.0 : New metrics of scholarly impact on the social Web," First Monday, Vol. 15, pp. 7-5, 2010.

[16] Serials Solutions 360 Counter Data Retrieval Service available at http://www.serialssolutions.com/assets/resources/360CounterDRS.pdf.

저 자 소 개

정영임	(E-mail : acorn@kisti.re.kr)
2001년	부산대학교 영어영문학/문헌정보학 (학사)
2004년	부산대학교 인지과학협동과정 (석사)
2009년	부산대학교 컴퓨터공학과 (박사)
2009년~현재	한국과학기술정보연구원/선임연구원
관심분야	자연언어처리, 데이터마이닝, 의사결정시스템

김재훈	(E-mail : jay.kim@kisti.re.kr)
1999년	성균관대학교 경영학과 (학사)
2008년	연세대학교 문헌정보학과 (석사)
2014년~현재	충남대학교 문헌정보학과 (박사과정)
1999년~2005년	KAIST 국가전자과학도서관/연구원
2006년~현재	한국과학기술정보연구원/선임연구원
관심분야	도서관컨소시엄, 전자정보 라이선싱, 이용자 행태분석

김광영	(E-mail : glorykim@kisti.re.kr)
1996년	부산대학교 컴퓨터공학과 (학사)
2001년	부산대학교 컴퓨터공학과 (석사)
2011년	충남대학교 문헌정보학과 (박사)
2001년~현재	한국과학기술정보연구원/선임연구원
관심분야	정보검색시스템, 개인화 검색시스템, 디지털도서관

김환민	(E-mail : mrkim@kisti.re.kr)
1991년	성균관대학교 문헌정보학과 (학사)
2006년	충남대학교 문헌정보학과 (석사)
2012년	성균관대학교 문헌정보학과 (박사수료)
1992년~1995년	현대중공업
1996년~1999년	대우 고등기술연구원
1999년~2005년	KAIST 과학도서관
2006년~현재	한국과학기술정보연구원/책임연구원
관심분야	학술 커뮤니케이션