

SUSHI 기반 학술정보 이용통계 수집 모델 연구

A Study on the Acquisition of Usage Statistics based on SUSHI Project

김선태, 임석중*

KISTI 선임연구원, KISTI 선임연구원*

Kim Sun-Tae, Lim seok-Jong*

Kim Sun-Tae KIST, Iim Seok-Jong KIST*

요약

최근 이용통계는 온라인 정보제공자로부터 폭넓게 제공 받고 있다. 그러나 그 통계는 여전히 일치된 데이터 수집 기로 이용할 수 없으며, 개별 정보제공자들의 관리 비용도 높은 상황이다. 표준화된 이용통계 수집을 주도하는 SUSHI는 정보제공사에서 제공한 이용통계를 도서관에 전자적으로 저장하기 위해 COUNTER 프로젝트를 실행할 자동화된 질의 응답 프로토콜을 개발하고 있다. SUSHI는 도서관들이 COUNTER 프로젝트의 통계를 활용함으로써 관리상의 업무부하 절감과 함께 의사결정을 지원할 것이다. 전자자원에 대한 이용통계의 기록과 교환에 있어서 출판사들이 COUNTER의 실무 규칙을 준수함으로써 최초로 저널과 데이터베이스들에 적용된다. 업체들은 COUNTER의 표준을 이용하여 엑셀이나 CSV 파일로 도서관 고객에게 제공할 수 있다. 그 결과 다양한 학술출판사와 정보서비스 기관에서 생성되고 있는 이용통계 데이터를 일관성, 신뢰성, 용이성을 확보할 수 있게 되는 것이다. 이에 본 연구에서는, 해외 전자저널 공동구매 컨소시엄인 KESLI의 전자저널 이용통계 수집 및 처리 절차를 분석하여, SUSHI에 기반한 전자저널 이용통계 수집 모델을 제안하고자 한다.

Abstract

Recently Usage statistics are widely available from online content providers. However, the statistics are not yet available in a consistent data container and the administrative cost of individual provider-by-provider downloads is high. The Standardized Usage Statistics Harvesting Initiative (SUSHI) is developing an automated request and response protocol for moving Project COUNTER (Counting Online Usage of Networked Electronic Resources) Code of Practice usage statistics from providers to library electronic repositories. SUSHI will help libraries make better decisions by reducing the administrative overhead of using Project COUNTER statistics. Publishers in the recording and exchange of usage statistics for electronic resources, initially journals and databases. By following COUNTER's Code of Practice, vendors can provide library customers with Excel or CSV (comma delimited) files of usage data using COUNTER's standardized formats and data elements. The result is a consistent, credible, and compatible set of usage data from multiple content providers. On this study, We propose the acquisition model of usage data based on SUSHI for KESLI that is overseas electronic journal consortium in Korea.

I. 서론

온라인 전자정보 자원의 활용이 급속히 늘어나고 있으며, 전자정보 자원 사용량에 대한 일관된 통계 구축 및 활용에 대해 정보 공급자 및 수요자의 인식이 널리 확산되고 있다. 기관에서는 사용권을 구매한 전자정보 자원이 다양한 정보 접근 점을 통해서 어떻게 활용되고 있는지 알고자 하며, 출판사 또한 자신들이 생성해 서비스 하고 있는 자원들이 어떻게 접근되고 있는지 알고자 한다.

이에 따라 출판사 및 전자 콘텐츠를 제공하는 제공자 측은 서로 상이한 유형의 통계데이터를 기관에 제공하는 수준을 벗어나고자 노력했다. 그 결과 출판사의 전자자원 이용통계를 제

공하기위한 표준포맷 COUNTER(Counting Online User Networked Electronic Resources, 이하 COUNTER)가 개발되었다. 이후 일정한 형식으로 생성된 통계데이터를 요청하고 전달하는 것에 대한 표준이 SUSHI(Standardized Usage Statistics Harvesting Initiative, 이하 SUSHI)이다. 즉, SUSHI는 이용통계 데이터의 전송 규칙으로 국내외 출판사들에 의해 소개되고 있는 ERM(Electronic Resource Management, 이하 ERM) 솔루션에서 사용하고 있는 기반 프로토콜이다.

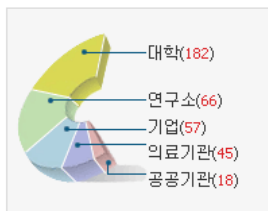
이에 본 연구에서는 SUSHI 프로토콜의 내용과 함께 현재 KESLI 시스템의 통계수집 및 처리 절차를 정리해 보고, 국제

표준의 이용통계 수집프로토콜인 SUSHI를 적용해 국내 각급 기관에서 활용할 수 있는 통계시스템을 구축할 수 있는 모델을 제시하고자 한다.

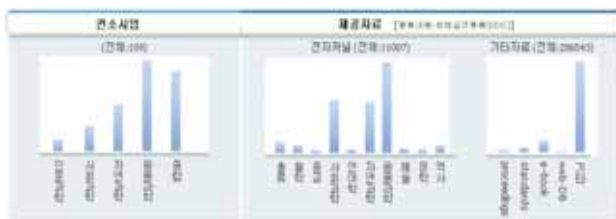
II. KESLI 이용통계 제공 현황

현재의 KESLI 시스템은 2005년 국가과학기술전자도서관(NDSL) 시스템과 통합되도록 재개발 되어있다. <그림 1>과 같이 KESLI의 368개 기관이 참여하고 있으며, 매년 증가추세에 있다.

● 참가기관 현황 [총 368개 기관]



▶▶ 그림 1. KESLI 참가기관 현황(2007. 3)
출처-<http://www.kesli.or.kr/index.html>



▶▶ 그림 2. KESLI 컨소시엄 현황(2007. 3)

또한, 100여 개 이상의 컨소시엄에 1만 중 이상의 전자저널이 유통되고 있다.

KESLI 시스템은 협력기관에게 다양한 통계데이터를 제공하기 위해 KESLI에서 만든 이용통계 포맷을 구축하고 있다. <표 1 참조>.

[표 1] KESLI 이용통계 유형

통계명	목적
컨소시엄별 이용통계	컨소시엄별-월별-1저널평균 이용통계 제공
기관별 이용통계	기관별-컨소시엄별-월별-1저널평균 이용통계 제공
저널별 이용통계	저널별-월별-년 평균 이용통계 제공
주제별 이용통계	컨소시엄별-주제별-월별 이용통계 제공
컨소시엄별 이용량 기반 효과분석	컨소시엄별로 참가기관의 컨소시엄 지출금액과 이용건수, PPV대비 비용효과데이터 제공
기관별 이용량 기반 효과분석	기관별로 컨소시엄 지출금액과 이용건수, PPV대비 비용효과데이터제공

<표 1>과 같이 이용통계와 관련된 서비스는 6 가지 이다. 개별기관의 자발적인 참여를 통한 통계데이터의 수집이 어려울 뿐만 아니라 한정된 인력으로 각 기관의 이용통계를 수작업으로 처리하는 것도 쉽지 않은 일이다. 이러한 이유로 KESLI 시스템을 통해 일관성있는 전체 이용통계의 확보가 어려운 상황이다.

현재 수집되고 있는 이용통계의 항목은 원문 다운로드 수에 제한되어 있으나, COUNTER의 다양한 항목을 추가될 수 있도록 개선되어야 한다.

III. SUSHI 프로젝트

3.1 SUSHI의 출현 배경

전자자원 이용통계의 표준인 COUNTER는 사서들과 출판사, 정보중개자에게 다음과 같은 혜택이 있다.

- 1) 사서들의 혜택: 다른 벤더들의 이용통계를 비교할 수 있음. 온라인 전자자원 선정 시 다양한 통계데이터를 활용할 수 있음
- 2) 출판사들과 정보중개자의 혜택: 고객이 원하는 형태의 통계데이터 제공. 서로 다른 서비스 채널의 관련된 이용현황 비교. 다양한 서비스 채널을 이용하는 고객을 위한 통계데이터 수집. 이용자들의 이용패턴에 대한 연구 가능

위와 같은 다양한 혜택이 가능하기 위해서 생성된 통계정보의 수집과 배포가 용이해야 한다. 이를 위해 SOAP를 기반으로 한 SUSHI가 나오게 되었다.

<표 2>는 벤더들이 COUNTER 리포트를 제공하기 위하여 SUSH 서버를 구축한 현황이다. <표 2>를 살펴보면 현재 EBSCO Information Service, Swets Information Service, Thomson Scientific, Project Euclid, MPS ScholarlyStats에서 서버를 구축하고 있다. 또한 서비스 제공자는 10개 기관이 참여하고 있다.

3.2 웹 서비스 기반의 SUSHI

나온 이후 RSS(Really Simple Syndication)가 XML의 가장 성공적인 비즈니스 모델이라고 한다.

이것은 데이터 교환을 매우 간단하게 구현함으로써 이용자들에게 편리한 서비스를 제공할 수 있기 때문이다. XML을 이용한 도 하나의 비즈니스 모델은 웹 서비스라 할 수 있다. 이것은 서로 다른 기종(예: UNIX와 NT)간에 XML 메시지를 주고 받는 통신용 프로토콜이라 할 수 있다.

[표 2] SUSHI 서버 구축 현황

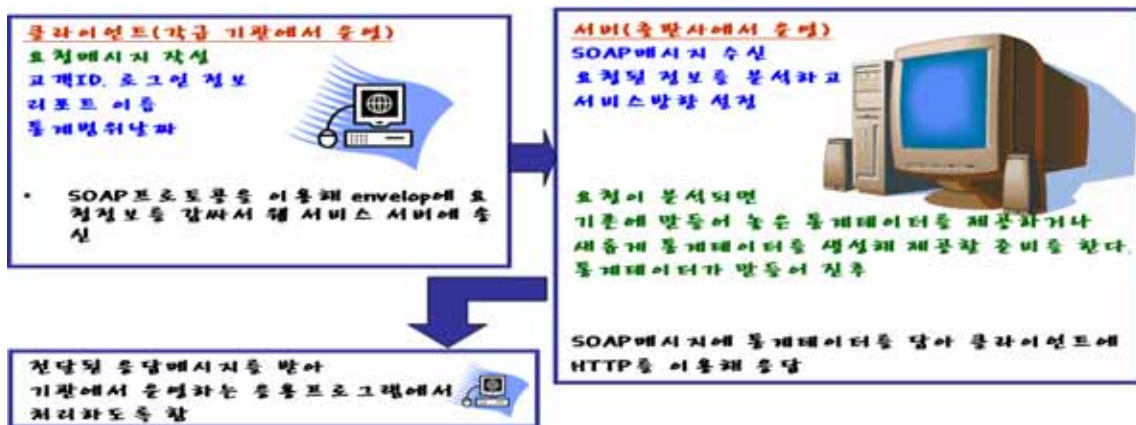
Organization	Service Provider	Data Provider	Deployment Status
Innovative Interfaces, Inc.	X		Member of SUSHI Working Group. SUSHI protocol originator. Actively involved with development.
EBSCO Information Services		X	Member of SUSHI Working Group. SUSHI protocol originator. Actively involved with development.
Swets Information Services		X	Member of SUSHI Working Group. SUSHI protocol originator. Actively involved with development.
Ex Libris	X		Member of SUSHI Working Group. SUSHI protocol originator. Actively involved with development.
Thomson Scientific	X	X	Member of SUSHI Working Group.
OTTO HARRASSOWITZ GmbH & Co. KG	X		Will begin investigating the protocol in early 2006
Endeavor Information Systems	X		Will begin investigating the protocol when developer documentation is available
Florida Center for Library Automation	X		Active: Building SUSHI support into their homegrown statistics collector
SirsiDynix	X		Q2 2006 development planned for Horizon 8.0 and Corinthian products.
OCLC	X		Active: Investigating protocol
Serials Solutions	X		Actively involved with development.
Project Euclid		X	Will implement protocol in Q1 2006.
College Center for Library Automation (CCLA) from the State of Florida Community Colleges	X		Will implement protocol in early 2006.
MPS ScholarlyStats		X	Actively supports SUSHI and are implementing it to deliver reports collected through ScholarlyStats to third parties for integration with other data.

출판사 및 정보중개자는 COUNTER 포맷의 이용통계 데이터를 제공하는 데 웹 서비스를 이용해 제공할 수 있다. 각 기관은 웹 서비스 클라이언트를 구현해 통계 데이터를 수집해 활용하면 된다. 이러한 일련의 과정을 도식화하면 <그림 3>과 같다.

<그림 3>은 일반적인 웹 서비스 방식의 전형적인 흐름을 나타내고 있다. 통계데이터 요청에 대한 제한은 없으며, 단지 서버에서의 접근권한 체크 및 기타 정책적인 영향을 받을 수 있다. <그림 3>에서 주목할 것은 클라이언트에서 서버로 보내는 요청 메시지의 내용과 결과값으로 주어지는 통계데이터의 활용이다.

SUSHI가 유용한 이유는 다음과 같다.

첫째, 출판사의 웹사이트에 개별적으로 접속해 통계데이터를 검색하고 취합하는 반복적인 수작업이 불필요하다는 것이다. 둘째, 항상 통일된 포맷인 COUNTER 통계를 수집할 수 있다. SUSHI 이전에는 출판사별로 제공하는 통계데이터가 다양하였으며, 심지어 웹 화면을 통해서만 통계데이터를 제공하고, 기관에게 다운로드 서비스를 제공하지 않는 경우도 있었다. 그러나 SUSHI를 지원하는 출판사는 모두 COUNTER 포맷으로 파일을 제공할 수 있어야 한다. 다시 말하면, 표준 포맷의 이용통계 데이터의 자동수집이 가능해 졌음을 의미한다.



▶▶ 그림 3. SUSHI 웹 서비스

2006년 현재 NISO와 COUNTER의 협약에 의해 NISO SUSHI Working Group이 COUNTER 갱신에 책임을 지고 있다.

SUSHI버전 1.0(2006)을 기준으로 웹 서비스 클라이언트의 요청정보에 포함되어야 하는 것은 요청자 정보와 리포트 정보이다.

- 1) 요청자 정보: 기관(도서관) ID, 요청자명, 이메일
- 2) 리포트 정보: 요청하는 리포트이름, 요청범위(날짜)

웹 서비스 서버의 응답정보에는 요청자명, 리포트명, 리포트가 포함되어 있다.

SUSHI 1.0은 COUNTER의 모든 리포트 검색을 지원한다.

IV. SUSHI 적용모델

SUSHI를 통해 개별기관의 전자자원 이용통계 데이터를 수집할 수 있을 뿐만 아니라 컨소시엄 레벨의 리포트도 제공할 수 있다.

KESLI에서는 현재의 비 표준화된 이용통계 수집방법 이외에 SUSHI를 이용한 통계데이터 수집 및 배포기를 개발하여야 할 것이다.

국내에서 유통되는 ERM 솔루션 및 도서관자동화시스템에 쉽게 내장되어 활용될 수 있어야 하며, 타 시스템에 영향을 받지 않고 Stand-alone의 형태로 작동도 될 수 있는 솔루션이어야 한다.

KESLI에서 컨소시엄 레벨의 리포트를 수집하여 국내에서 유통되는 해외 전자자원의 이용통계를 다각도로 분석하면, 해외 전자자원 구독과 관련하여 국가차원의 정책결정을 지원할 수 있을 것이다. 또한 개별기관에서 사용할 수 있는 SUSHI 솔루션을 개발해 보급함으로써 개별기관별 이용통계 수집의 어려움을 극복할 수 있게 된다.

컨소시엄 레벨의 리포트 수집모델을 도식화하면 <그림 4>와 같다.

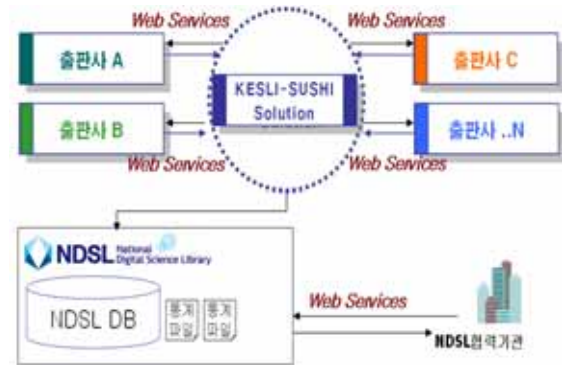
각 단계별로 살펴보면 첫째, 각 개별출판사의 웹 서비스 서버와 SOAP 통신을 통해 컨소시엄 레벨의 리포트를 수집하여 NDSL(KESLI) 시스템을 구축한다. 둘째, 수집된 리포트를 분석하여 데이터베이스에 구축하며, 원활한 활용을 위해 파일 형태로 통계데이터를 구축한다. 셋째, 개별기관에서는 표준 웹 서비스 방식을 이용하여 컨소시엄 레벨의 리포트 정보를 활용한다.

컨소시엄 레벨의 리포트 수집 기능은 다음과 같다.

- 1) 국제 표준 웹 서비스 프로토콜을 활용해 구현된 클라이언트 기능
- 2) NDSL 시스템의 OpenAPI 방식을 이용한 Pushing 서

스 기능

- 3) 스케줄러에 의한 자동수집 기능
- 4) 수집결과 모니터링 및 후처리 기능



▶▶ 그림 4. 컨소시엄 레벨 리포트 수집 모델

<그림 5>는 개별기관에서 ERM 솔루션이나 도서관자동화시스템에 내장하거나 독립된 형태로 솔루션을 운영하는 모델이다. 개별기관에서 기관의 아이디를 이용하여 리포트를 수집하여 통계데이터를 NDSL 시스템에 자동으로 전송한다.



▶▶ 그림 5. 기관 레벨 리포트 수집 모델

기관 레벨 리포트 수집 솔루션 기능은 컨소시엄 레벨과 동일하며, 추가로 Stand-alone 서비스 운영과 독립된 모듈로 개별기관의 ERM이나 도서관자동화시스템에 내장될 수 있으면 될 것이다.

기존 NDSL 시스템에서도 NDSL 학술지와 매핑 작업을 통한 통계데이터 구축 기능을 구현하여 KESLI-SUSHI의 호출에 응답할 수 있도록 하여야 한다. 또한, 기 구축된 KESLI 시스템의 통계정보도 기능개선이 요구된다.

V. 결 론

개별기관에서는 제대로 된 학술정보의 수서업무를 위해 소속기관의 연구자들의 해외 학술정보 이용빈도를 조사해야 한다. 또한 출판사 역시 자신들이 제공한 전자자원이 어떠한 경

로 활용되는지 파악하고 있어야 하므로 COUNTER와 SUSHI 프로토콜에 대한 필요가 많은 이해관계자들의 참여가 이루어지고 있다.

2006년 중반 이후 국내에 ERM 솔루션이 소개되면서, 핵심 기능은 전자자원 수서 및 관리, 통계로 볼 수 있다. 이 중 통계와 관련해서 SUSHI를 이용한 이용통계 자동 수집을 위한 핵심 프로세스가 포함된다.

KESLI를 통하여 방대한 해외 전자자원이 유통되고 있다. 또한 링크서버 호스팅 사업을 통해 개별기관의 연구자에게 원문제공 서비스가 가능하게 되었다. 이렇게 전자자료의 유통 및 이용환경이 향상되고 있으며, 전자정보 이용에 대한 모니터링이 가능하도록 되었다. 이를 통해, 개별 기관 및 KESLI 차원의 리포트를 수집·분석함으로써 국내에서 유통되는 해외 전자자원 구독 결정을 위한 기초자료가 될 것이다.

■ 참고 문헌 ■

- [1] 심원식 “전자정보 이용통계 활용 전략”, 정보관리학회지, 제22권, 제2호, pp.5-21, 2005.
- [2] A Starter kit for sushi java client, 2006. [cited 2006. 3. 27]
<http://sushi.swetswise.com/material/SushiClnt.zip>
- [4] EBSCO Information Service, Automating Usage Statistics Harvesting Requirement, pp1-6, 2006 [cited 2006. 6. 26]
http://www.niso.org/committees/SUSHI/Automate_SUSHI.pdf
- [4] NISO Press, The Standardized Usage Statistics Harvesting Initiative(SUSHI) Protocol, 2006. [cited 2006. 3. 20]
<http://www.niso.org/standards/resources/z39-93-DSFTU.pdf>
- [5] SUSHI FAQ, 2006. [cited 2006. 3. 27]
http://docs.google.com/View.aspx?docid=d2dhjwd_63tkkwf
- [6] SchorarlyStats, ScholarlyStats: Description of Services, 2006 [cited 2006. 6. 20]
<http://www.solinet.net/emplibfile/Description%20of%20Services-12-11-06.pdf>