

# 전자저널 이용통계 자동수집기를 이용한 컨소시엄 서비스 방안

## Consortium Service Using Automatic Collecting System of E-Journal Usage Statistics

정 영 임, 김 정 환, 류 범 종\*  
한국과학기술정보연구원 지식기반실

Youngim Jung, Jeong Hwan Kim, Beom-Jong You\*  
Dept. Knowledge Resources at Korea Institute of  
Science and Technology Information

### 요약

온라인 정보자원의 활용이 급속히 증가하면서, 정보자원 이용량에 대한 일관적이고 표준화된 통계 구축의 필요성에 대한 인식이 널리 확산되고 있다. 그러나 대부분의 도서관에서는 정보공급사로부터 이용통계를 수작업으로 수집하고, 정보공급사별로 상이한 통계 보고서를 재가공하여 분석하기에는 많은 어려움이 따른다. 본 논문에서는 전자정보 컨소시엄에 참가하는 개별 도서관을 대신하여 컨소시엄 주관기관에서 각 정보공급사별 이용통계를 자동으로 통합 수집하여 표준화된 포맷으로 구축하고, 이를 다양한 형태와 정보로 재가공하여 서비스 방안을 제안하였다.

## I. 서론

핵심 학술저널이 전자화되어 컨소시엄을 통해 패키지로 되어 유통되고 있고, 이에 대한 활용이 급속히 증가하면서 도서관에서는 자관에서 구입되는 전자저널이 얼마나, 어떻게 활용되고 있는지에 대한 관심이 증가하였다. 전자저널의 이용통계는 저널을 생산하고 공급하는 출판사에서 생성하고 있다. 각 출판사에서 생성된 이용통계는 웹 사이트에서 각 도서관 담당자가 직접 조회를 하거나 혹은 도서관의 요청에 따라 출판사에서 e-mail을 통해 전달하기도 한다. 그러나 각 도서관에서는 다수 출판사의 전자저널을 구독하고 있기 때문에, 담당자가 월별/분기별/연별 등 주기적으로 각 출판사 사이트를 일일이 방문하여 이용통계를 수집하기가 번거롭고, 수집한 이용통계 데이터를 표준화된 포맷으로 재가공하여 일괄분석하는 데 드는 시간과 비용의 낭비가 적지 않다.

따라서 본 논문에서는 각 도서관의 대용량 이용통계 분석 기반 마련을 위해 스크린 스크래핑 기술을 활용한 전자저널 이용통계 자동 수집 방안과 분석정보 서비스 방안을 소개하고자 한다. 본 논문의 구성은 다음과 같다. 먼저 제2절에서 관련 연구에 대해서 소개한다. 제3절에서는 본 이용통계 자동 수집 시스템을 설명하고, 본 시스템을 통해 개별 도서관에 어떤 이용통계 정보를 제공할 수 있는지에 대해 소개한다. 마지막으로 제 4절에서는 본 연구의 결론과 향후 과제를 논의하며 끝을 맺는다.

## II. 선행연구

[1]에서는 전자정보를 제공하고 지원하는 비용이 도서관 전체 예산의 대부분을 차지함에 따라 도서관 운영의 분석, 보고 및 의사결정을 지원하기 위해 전자정보 이용

통계가 중요함을 지적하였다. 그럼에도 불구하고, 이용통계의 활용에는 많은 한계가 있다. [2]에서는 국내 48개 대학 도서관을 대상으로 전자저널 이용통계가 제공되는 현황 및 활용 현황을 설문조사를 통해 파악하였다. 주로 도서관 측 요청에 의해 정보공급사로부터 이용통계 데이터가 제공되고 있었고, 제공되는 데이터의 내용 및 형식에 대한 만족도는 반수에 가까운 기관이 매우 불만 혹은 다소 불만(47.8%)을 표하였다. 불만족 이유는 ‘출판사마다 제공되는 데이터 형식이 달라 비교가 어려웠다’는 응답이 61.1%를 차지해 정보공급사별 이용통계 리포트 세부 항목의 차이와 형식의 비표준화가 이용통계 활용에 가장 큰 걸림돌이 됨을 파악할 수 있다.

## III. 본문

### 1. 정보공급사별 전자저널 이용통계 제공 현황

국내에서 최다 구독되는 11개 정보공급사에서 이용통계 데이터를 제공하는 현황은 <표 1>과 같다.

표 1. 정보공급사 이용통계 제공 현황

정보공급사	기관수	저널종수	제공방식	제공데이터		데이터포맷
				통계보고서	분석자료	
ACS	110	51	Email	○	×	표
AIP	81	81	웹	○	×	표
APS	81	8	웹	○	×	표
Elsevier	249	3,424	웹	○	×	표
IEEE	78	1,018	웹	○	×	표
IOP	72	217	웹	○	×	표
JSTOR	79	1,089	웹	○	×	표
Nature	120	171	웹	○	×	표

Science	93	3	웹	○	×	표
Springer	142	2,702	웹	○	×	표
WB	180	2,049	웹	○	×	표

11개 정보공급사는 in-house format의 이용통계 보고서를 표 형식으로 제공하고 있으나 이를 시각화하여 재가공한 자료나 타기관 대비 이용량 및 시기에 따른 전자정보 이용 추이 등의 분석자료는 제공하고 있지 않다.

## 2. 이용통계 자동 수집 및 분석 시스템

본 연구에서 개발한 전자정보 이용통계 자동 수집을 통한 통계 분석 정보 서비스 모델은 <그림 1>과 같다.



▶▶ 그림 1. 전자정보 이용통계 자동 수집을 통한 서비스 모델

본 서비스를 위해 (1) 정보공급사 웹사이트에 접속하여 각 사이트에 구축된 이용통계 조회화면을 스크래핑하여 이용통계 보고서를 추출하는 자동 수집 모듈을 개발하고 (2) 추출한 이용통계 데이터를 COUNTER JR1 표준 포맷[3]을 적용한 이용통계 데이터베이스를 구축하였으며, 통계 데이터와 결합하여 다양한 분석자료를 생성하기 위해 비용정보, 라이선스 구독정보 등의 데이터베이스도 구축하였다. 또한 (3) 구축된 데이터를 표, 그래프, 차트로 시각화하여 보여줄 수 있는 통계 분석 정보 산출 모듈을 개발하였다.

## 3. 이용통계 분석 정보 서비스 제공

본 시스템에서 제공하는 통계 정보 및 분석 정보는 <표 2>와 같다.

표 2. 이용통계 분석정보 제공

제공 정보	형식	속성	비고
저널별 이용통계 보고서	표	원자료	월별, 연도별
컨소시엄 품목별 이용통계	표, 차트	분석자료	월별, 연도별
주제별 저널 이용순위	표	분석자료	월별, 연도별
기관별 저널 이용순위	표	분석자료	월별, 연도별
전체기관 저널 이용순위	표	분석자료	월별, 연도별
컨소시엄 품목별 CPA <sup>1)</sup>	표, 차트	분석자료	연도별
기관별 컨소시엄 품목 이용추이 (전체평균 대비)	그래프	분석자료	연간

1) CPA = cost / number of articles accessed

본 시스템을 통해 자동수집되어 구축된 이용통계 보고서는 2011년 3월 기준 총 5,082,962건이다[4]. 출판사로부터 수집한 원자료인 이용통계 보고서 외에도, <표 2>와 같이 시각화된 자료로 산출되는 다양한 통계 분석 정보를 통해 각 도서관 담당자들은 자관의 이용자들이 어떤 전자정보를 얼마나 활용하고 있는지 직관적으로 파악할 수 있다. 이러한 분석자료를 통해 담당자들은 향후 수서 정책의 근거 자료로 활용할 수 있고, 행정부서에 제출해야하는 다양한 보고서 작성에 이용할 수 있다. 또한 컨소시엄에 참여하는 전체 기관의 전자정보 평균 이용량 대비 자관의 이용량을 확인할 수 있어, 전자정보 활용을 촉진하기 위한 정책을 수립할 수 있다.

## IV. 결론

본 논문에서는 전자저널 이용통계 자동 수집 및 분석 시스템을 통해 컨소시엄 참여 도서관을 위한 이용통계 분석정보 서비스 제공 사례를 보였다. 본 시스템을 통해 컨소시엄에 참가하는 각 도서관 담당자가 주요 출판사의 대용량 이용통계를 한 개 사이트에서 편리하게 표준 포맷으로 검색할 수 있으며, 다양한 통계 분석 정보를 그래프를 통해 직관적으로 확인할으로써 향후 수서 정책과 전자정보 활용 촉진 정책을 수립하는 데 큰 도움이 될 것으로 기대한다.

다음 연구로는 (1) 통계데이터 전송 표준 프로토콜인 SUSHI[5]를 적용하여 본 시스템의 이용통계 자동수집 채널을 다양화하는 연구와 (2) FLEX를 적용한 시각화틀을 보완하여 수집한 이용통계 정보를 더욱 다양한 형태로 제공하는 방안에 대한 연구를 진행하고자 한다. 이와 같은 일련의 연구를 통해 컨소시엄에서 유통되는 전자정보의 유통 현황을 보다 면밀하게 파악할 수 있고, 참여 도서관에는 이용통계에 대한 고품질의 분석 데이터를 더욱 편리하게 제공할 수 있을 것이다.

## ■ 참고 문헌 ■

- [1] 심원식 “전자정보 이용통계 활용 전략”, 정보관리학회지 제22권, 제2호, pp.5-21, 2005.
- [2] 황옥경 “대학도서관에서의 전자저널 이용통계 제공 및 활용현황”, 정보관리연구, 제38권, 제4호, pp.68- 87, 2007.
- [3] COUNTER “The COUNTER Code of Practice” [cited 2011. 3. 11] [available at [http://www.projectcounter.org/code\\_practice.html](http://www.projectcounter.org/code_practice.html)]
- [4] 정영민, 김정환, 류범중 (2011) “스크린 스크래핑 기반 전자저널 이용 통계 자동 수집 시스템 개발” 2011 한국정보과학회 컴퓨터종합학술대회 발표논문집 게재 예정
- [5] NISO “NISO Standardized Usage Statistics Harvesting Initiative(SUSHI): Z39.93”, [cited 2011. 3. 14] [available at [http://www.niso.org/apps/group\\_public/download.php/1473/SUSHIupdate\\_alamid09\\_cli.ppt](http://www.niso.org/apps/group_public/download.php/1473/SUSHIupdate_alamid09_cli.ppt)]