eGate 데이터 분석과 활용에 관한 연구

A Study of eGate Data Analysis and Usage

최원준, 김재훈, 김순영, 김정환
한국과학기술정보연구원

Won-Jun Choi, Jay-Hoon Kim, Sun-Young Kim, Jeong-Hwan Kim
Korea Institute of Science & Technology Information

요약

KISTI의 학술 자원 인프라(eGate)는 국내와 해외의 학술출판사들과 국내 도서관들이 협력하여 만들어 가고 있는 학술정보 공동 활용 인프라이다. KISTI는 2000년부터 현재까지 약 18년간의 eGate 구축 및 품질 관리 노하우를 바탕으로 전 세계 학술정보를 지속적으로 확대 · 수집하고 있으며 정보 요소들을 유기적 결합함으로써 활용 가치를 높여가고 있다. 본 논문에서는 eGate에서 지금까지 수집된 데이터를 분석하여 효용 가치를 판별하고 활용성과 진보성에 대해 연구해 보고자 한다.

Ⅰ. 서론

국내뿐만 아니라 해외에서도 학술정보 유통을 위한 과학기술정보 데이터 서비스 사업에 대한 지원이 확대되고 있다. 한국과학기술정보연구원(KISTI)에서는 전 세계의 과학기술 정보를 수집하여 국내 연구기관에 제공함으로써 첨단 R&D를 위한 길잡이를 해 왔고 과학기술의 정책 방향을 결정할 수 있도록 맞춤형 정보분석 서비스를 제공해 왔다. 과학기술정보가 날이 갈수록 증가함에 따라서 과학기술정보의 빅데이터에 대한 정보분석 요구가 증가하게 되었고 정보 분석에 필요한 기반이 되는 과학기술정보의 실시간 지원이 필요하게 되었다. 따라서 KISTI는 학술자원의 인프라를 효율적으로 관리 및 지원하기 위해서 국가과학기술 정보 통합 서비스 플랫폼에 기반이 되는 학술 자원 인프라(eGate)를 설계하여 운영하게 되었다. eGate의 목적은 국내 협력기관에서 인쇄 또는 전자형태로 소장/구독/이용 가능한 전 주제분야에 걸친 국가 가용 학술지 정보에 대해 메타정보 및 인쇄소장, 전자/라이선스정보, 주제정보, 다양한 고품질 서비스로의 연계정보를 구축 함으로써, R&D 연구자들이 보다 쉽고 빠르게 원문에 접근할 수 있게 하고 R&D 연구자들이 원문뿐만 아니라 관련 전문정보로도 접근할 수 있게 하는 것이다. 또한, 협력기관이 자관 이용자들의 이용실태 분석을 통해 장서개발 등에 활용할 수 있도록 하고, 협력기관이 자체적으로 고품질 서비스를 개발할 수 있도록 국가 학술콘텐트 인프라를 구축 및 공유하는 것이 주요 쟁점이다. 현재 대부분의 연구자들은 eGate가 무엇인지 어디에 사용되고 있는지 잘 모르는 부분도 있고 eGate의 중요성을 인지하지 못하고 있는 부분도 있어 본 논문에서는 eGate의 데이터가 어떤 내용이며 어떻게 활용될 수 있고 어떻게 발전 가능한가를 논의해 보고자 한다.

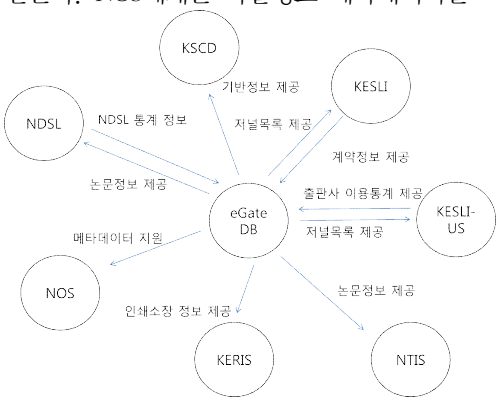
Ⅱ. 선행 연구

eGate의 인프라를 활용하여 과학기술정보를 활용하여 KISTI에서 운영중인 관련 시스템은 다음과 같다. 국가과학기술정보센터(NDSL)[1]은 NDSL은 산학연 연구자들에게 고품질 정보를 제공하여 국가 과학기술 혁신에 기여하고자 개발된 과학기술정보 서비스 플랫폼으로, 논문 · 특허 · 보고서 · 동향 · 저널/프로시딩 · 연구자 · 연구기관 등 약 1억건 이상의 콘텐츠에 대한 검색 및 콘텐츠 유형별 전문검색 서비스를 제공하고 있다. 학술정보통합관리시스템(KSCD)[2],는 KISTI 과학기술정보센터에서 구축 및 운영하는 국내 과학기술분야 학술콘텐트 데이터베이스이다. 전자정보국가컨소시엄(KESLI)[3]은 국내외 전자정보 공급사들로부터 다양한 학술연구분야 전자정보를 공동구매하고 국내 학연 산 기관들이 보유한 인쇄 정보를 공동 활용할 수 있는 체계를 구축 및 운영함으로 국내 연구자들이 필요로 하는 다양한 정보를 신속하게 제공할 수 있는 지식정보서비스를 제공하고 있다. 과학기술학회마을[4]은 학술정보 공동활용 사업의 대표적인 서비스 페이지로써 디지털화된 학술자료의 검색 및 활용이 가능하도록 국내학술지의 One-Stop 서비스를 지원한다. Korea Science[5]는 KISTI DOI(Digital Object Idetifier) 학술지의 DOI landing page이다. KISTI는 학술정보 공동활용 사업과 연계하여 학회의 학술지에 대해 DOI 획득을 위한 CrossRef 기탁 업무를 무상으로 지원한다. KSCI(Korea Science Citation Idex)[4] 는 국내 과학기술분야 핵심학술지에 수록된 연구논문에 대한 서지 및 인용정보인 KSCD를 기반으로 연구정보의 추적검색, 학술지 영향력 측정, 연구동향 분석 등 여러 가지 고부가가치 정보를 서비스하는 플랫폼이다. KPubs(The Korean Journal Publishig Service)[4]는 KISTI에서 운영하는 오픈 엑세스로 발행된 학술저널의 플랫폼이다. NOS(NDSL)[5]

는 과학기술분야 개방형 유통 플랫폼으로 과학기술정보의 유통에 도움을 주고 있다.

Ⅲ. KISTI eGate 인프라 시스템

eGate 가 보유하고 있는 데이터는 과학기술분야를 중심으로 이와 관련된 주제분야에 걸쳐 국가가용학술정보에 대한 메타콘텐트 및 연관정보를 망라적으로 포괄한다. 예를 들면, 학술저널/프로시딩 메타 및 초록, 참고문헌정보, 학술저널/프로시딩 원문(Fulltext), 국내협력기관 인쇄학술지 소장정보/전자학술지 라이선스정보, 전자원문링크정보 및 OA 저널정보, 고부가 콘텐트 연계정보, 주제분류, 저널명 전거, (저자명 전거, 기관명 전거), 주요학술지 등재정보, 국내 600여 협력기관의 전자저널 라이선스 및 인쇄저널 소장정보 공유, 전자정보 원문서비스 : KISTI 구매 전자저널, OA/FA 전자저널, 인쇄정보 원문제공서비스(NDSL NIDS, KERIS DDS(RISS)) 등이 있다. 이러한 정보는 대국민 서비스로 현재 사용 중이다. (서울대, 고려대, KAIST, 국립중앙도서관, 국회도서관, 특허정보원, 네이버 전문정보, JST 한국 학술지 서비스 등), NDSL은 eGate가 학술논문 정보를 제공하고 통계 정보를 제공받는다. KESLI에게는 학술저널 목록을 제공하고 문헌계약정보를 제공받는다. KESLI 통계정보시스템(KESLI-US)에게는 학술저널 목록을 제공하고 이용통계 정보를 제공받는다. NOS에게는 학술정보 메타데이터를 지원한다.



▶▶ 그림 1. eGate 데이터 관계도

KERIS로는 인쇄소장 정보를 제공한다. NTIS에게는 논문정보를 제공한다.(그림. 1) 이러한 관계에서 eGate의 역할은 각 기관이나 서비스에게 기반이 되는 논문정보와 저널정보를 제공하여 과학기술 정보 활용과 공유를 확장하도록 지원하는 역할을 한다. eGate의 역할이 상실된다면 각 기관의 서비스가 제대로 되지 않고 각 서비스가 제공하는 통계정보도 정확하지 않게 된다. eGate의 장점은 첫번째로 대용량 데이터를 구축하고 있다는 것이다. 8,000여만 건의 전 세계 학술논문 정보를 입수하고 자체 구축, 저널 명 전거정보 국내에서 유일하게 보유하고 있으며, 국가 종합목록 서비스 지원이 가능하다. 또한, 신뢰적인 데이터를 제공한다. 전 세계 주요 30여 개 출판사로의 고품질 데이터를 정보원(소스 데이터)으로 활용하기 때문이다. 그리고 신속한 데이터 갱신 체제를 유지하여 최신성을 확보하고 있다. eGate DB의 관리 측면에서는 체계적 데이터 품질관리가 이루어 진다. 데이터 상시 모니터링:, 취약점 발견 및 성능 개선, URL 모니터링이 있으며, 데이터 표준화를 준수하여 국제 표준에 맞도록 데이터를 수집한다. 그림. 2는 eGate 데이터 보유 건수를 나타내 주고 있다. 대부분의 자료는 고품질의 데이터로써 각 기관에서 승인을 거쳐 서비스로 이어진 컨텐츠이다. eGate 데이터는 신뢰할 수 있는 기관에서 데이터를 수집 및 서비스하고 있다. 또한, 대국민 전자저널 서비스를 위해 해외 전자저널을 1295종 구매하여 구축하였다.

▶▶ 그림 2. eGate 데이터 보유 건수

Ⅳ. 결론

KISTI에서eGate 인프라는 18년 동안 축적된 대량의 고품질 데이터를 확보하여 국가 과학기술정보 서비스의 기반이 되는 고품질의 학술정보를 제공하고 있다. 36여만 종의 저널정보, 7천여만 건의 논문정보, 1.2억여 건의 참고문헌정보 등 8,000여만 건에 달하는 국내 유일의 전자저널 중심 학술정보 공동 활용 인프라이다. 이러한 국가 과학기술정보는 앞으로의 연구 개발에 필요한 자원을 제공하는 차원에서 그 효용 가치는 높으며, 이러한 정보자원을 활용하여 저자간의 네트워크 분석, 성과물 검증 등 다양한 서비스 활용이 가능할 것으로 보인다. eGate가 국내 과학기술 빅데이터 분석에 기반이 되는 대국민 정보시스템으로 거듭나길 기대한다.

■ 참 고 문 헌 ■

[1] 강남규, 신용주, 박근철, 주원균 외(2010), NTIS-NDSL 활용한 국가R&D논문 수집 방안에 관한 연구, 한국인터넷정보학회 학술발표대회 논문집, 2010.10, 231-232.
[2] 김병규, 강무영, 최선희 (2006), 협회 기술정보 관리 및 유통 시스템 구축에 관한 연구. 정보관리연구, 37(3), 117-137.
[3] 전자정보국가컨소시엄, Kesli.or.kr
[4] KSCD 부가가치 정보서비스, Ksci.ndsl.kr
[5] DOI서비스, Koreascience.or.kr