

KRISTAL-2002를 이용한 학술정보의 관리를 위한 연구

김병규⁰, 강무영, 최선희, 박재원
한국과학기술정보연구원
(yourovin⁰,kmy,shchoi,ilonetos)⁰@kisti.re.kr

A Study on Management of Academic Content Using KRISTAL-2002

Byung-Kyu Kim⁰, Moo-Young Kang, Sun-Hee Choi, Jae-Won Park
Dept. of Domestic information, Korean Institute of Science & Technology Information

요 약

과학기술 분야 학회에서 생산하고 있는 학술자료(학회지, 논문지 및 학술회의자료 등)를 수집, 가공하여 학술정보 DB를 구축하고, 정보이용자들이 원하는 학술정보를 언제 어디서든지 인터넷을 통해 서지사항은 물론 디지털 원문까지 획득할 수 있도록 할 수 있는 학술정보 종합 관리 시스템이 필요하다.

본 논문에서는 데이터의 유효성 검증과 타 기관의 상호대차 및 원문 교환을 손쉽게 처리할 수 있는 메타데이터의 표준인 XML과 데이터의 저장관리와 빠른 검색을 제공하는 KISTI의 최신 IRMS인 KRISTAL-2002을 기반으로 하여 학술정보 관리를 연구하고 시스템을 설계하였다.

1. 서론

국내 과학기술 분야 학회는 662 개에 달하며 학회의 학술활동 및 교류를 통해서 생산하는 고급 학술정보는 그 양이 방대하다. 또한 일반 연구자들에게는 그 효용 가치가 매우 높다고 할 수 있다. 따라서 이들의 학술자료를 디지털화하여 부존자원화 하고 연구자가 쉽고 빠르게 접근 할 수 있는 정보 서비스체제를 구축 및 제공하는 것은 매우 중요한 일이다.

학술자료를 수집하여 가공하고 가공된 정보들을 효율적으로 관리 및 효과적으로 서비스하는 일은 쉽지 않은 일이다. 따라서 대량의 학술정보를 전문적으로 관리할 수 있는 전용 시스템이 절실히 필요하다..

학술정보 관리를 위한 시스템의 요구사항을 살펴보면 첫째는 복잡하지 않은 단순한 프로세스, 둘째는 다양한 검색 기능 제공, 셋째는 학술정보 즉 논문에 대한 메타데이터의 정의가 타 시스템과의 상호교환 및 운용성이 가능한 표준형식을 사용해야 하는 것이다.

본 논문에서는 효율적인 학술정보의 관리와 편리한 검색서비스를 제공할 수 있는 시스템을 설계한다. 이를 위해데이터 저장 관리 및 검색 기능을 제공하는 KISTI의 KRISTAL-2002[1]을 사용하였으며 데이터의 유효성 검증과 타 기관의 상호대차 및 손쉬운 원

문 교환을 위해 학술정보의 메타데이터의 문서 표준인 XML[2]를 사용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를, 3장에서는 학술 정보 관리를 위한 시스템의 설계 내용을, 마지막으로 4장에서는 결론 및 향후 연구 방향에 대해여 살펴본다.

2. 관련연구

2.1 XML

학술정보 관리 시스템에서는 학술 정보를 기술하기 위한 언어로 XML을 선택하였다. XML은 인터넷 상의 데이터 교환을 목적으로 하는 문서를 기술하기 위한 국제 표준 언어로서, W3C(World Wide Web Consortium)에서 표준안을 제안하였다. XML은 인터넷 상의 문서를 기술하기 위한 국제 표준 언어로서, 이를 이용하여 문헌 정보를 기술한다면 전 세계 어느 곳에서도 별다른 노력 없이 학술 정보를 공유할 수 있을 것이다. XML은 인터넷에서 기존에 사용되던 HTML(HyperText Markup Language)의 한계를 극복하고, SMGL(Stand-dard Generalized Markup Language)의 복잡함을 단순화함으로써, SGML과 HTML 양쪽 모두와의 상호 운용성과 용이한 구현환경을 제공한다.[3] XML은 DTD를 기반으로 작성된다. DTD는 XML문서를 기술하기 위해서 미리 정의

하는 문서로 XML로 기술된다. DTD에는 XML 문서를 기술하기 위한 정보들이 기술되며, XML 문서에서 어떤 항목이 어떤 의미를 갖는지의 정보가 기술된다. 그러므로 사용자는 DTD만 알고 있다면 DTD 기반의 어떠한 XML 문서라도 의미를 파악할 수 있다. 또한 문헌 정보 XML 파일을 기술하는 식자공의 입장에서 DTD만 알 수 있다면 학술 정보 XML 파일을 쉽게 구성할 수 있다.

이러한 XML의 특징을 활용하기 위하여 학술정보 데이터베이스에서는 문헌 정보를 기술하기 위한 언어로 XML을 선택하였다. XML 문서의 관리를 위한 방법은 현재 여러 가지가 존재하지만, 각 응용에서 XML 문서의 특성을 파악한 후 응용에서 최적으로 사용할 수 있는 데이터베이스 구조를 정의하여 사용하는 것이 가장 효율적이다. 이러한 것이 가능한 이유는 XML 문서를 위한 DTD정보가 XML 문서에 기술된 각 항목들의 의미를 포함하고 있으므로 가능하다. 또한 XML 문서의 검색을 위해서는 현재 XPath라는 XML 문서 국제 표준 검색어가 존재한다.[4] 이를 이용한다면, 전세계의 어떠한 XML 문서라도 질의가 가능하므로 향후, 여러 다양한 형태의 학술 정보를 통합하여 관리하기 용의할 것으로 사료된다.

2.2 KRISTAL-2002

KRISTAL-2002는 KISTI에서 개발한 정보검색관리 시스템(Information Retrieval & Management System, IRMS)으로, 윈도우 및 유닉스 기반에서 데이터에 대한 저장,관리,검색에 관한 처리를 효율적으로 수행한다.

KRISTAL-2002의 특징을 살펴보면 첫째, 데이터의 생성부터 관리, 정보 서비스까지 단일 플랫폼으로 제공하면서 Flat data, XML 정보를 모두 수용할 수 있고 둘째, 소규모 정보서비스부터 대규모 정보 서비스까지 이용가능 하며 셋째로 대용량 정보의 분산 수용 및 검색이 가능하다는 것이다.

KRISTAL-2000 데이터베이스는 검색엔진(IR)과 데이터베이스(DBMS)를 결합한 확장형 정보검색관리시스템(IRMS)으로서 상용 DBMS과 비교하였을때 대용량 데이터를 신속하게 적재할 수 있으며 더욱 빠르고 정확한 검색을 지원하며 멀티 프로세스를 이용한 다수 사용자 지원과 멀티쓰레드를 이용한 동시 검색을 제공하며 트랜잭션 처리를 통한 온라인 데이터관리를 통해 데이터 안정성 및 복구 지원을 제공한다.

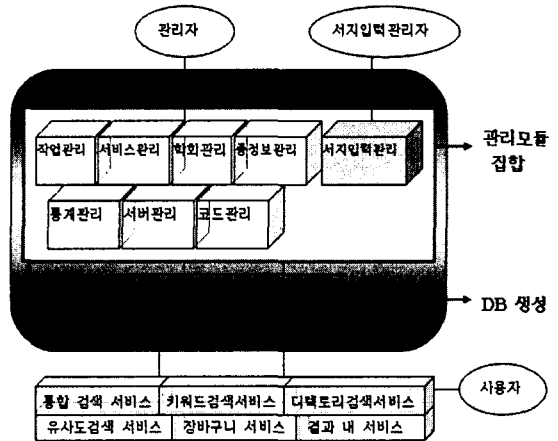
3. 학술정보 관리 시스템의 설계

학술정보 관리 시스템의 목표는 국내 연구자들에게 고품질의 학술정보를 제공하기 위해 수집한 학술자료의 서지정보와 초록 및 원문을 데이터베이스로 구축하고 관리하며 인터넷을 통해 학술정보를 일반 검색

사용자들에게 다양한 검색 서비스를 제공하는 것이다.

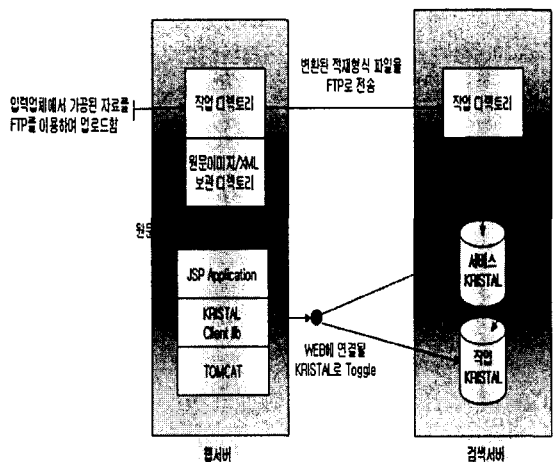
3.1 학술정보 관리 시스템의 기능 및 구조

학술정보 관리 시스템은 [그림 1]과 같이 이용자의 레벨을 3가지로 구성한다. 관리자와 서지입력관리자는 학술정보의 입수,가공,DB 저장을 담당하며 일반 사용자는 일반 검색서비스를 이용한다.



[그림 1] 사용자 별 관리 및 검색 기능

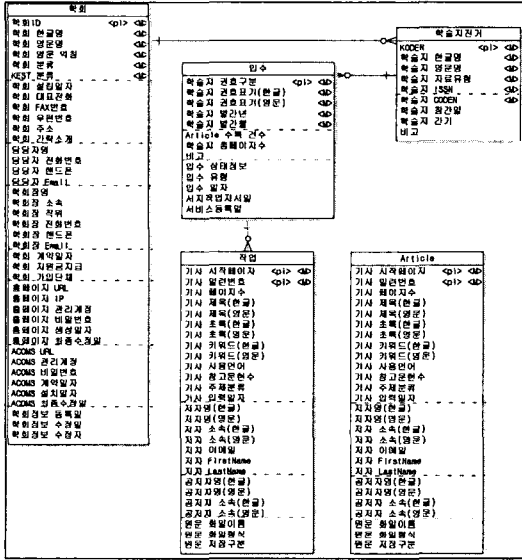
학술정보 관리 시스템은 시스템 장에 또는 관리실수에 적절히 대응할 수 있도록 [그림 2]와 같이 웹서버와 검색서버로 분리구성하고 KRISTAL DB를 듀얼로 관리하는 방식을 선택하였다. 또한 권 단위로 제작되는 대량의 XML문서를 빠르게 KRISTAL DB에 로딩하고 검색서비스를 통해 서비스 할 수 있는 배치적제 기능을 제공하도록 설계하였다.



[그림 2] 시스템 구조

3.2 학술정보 관리 시스템의 DB 설계

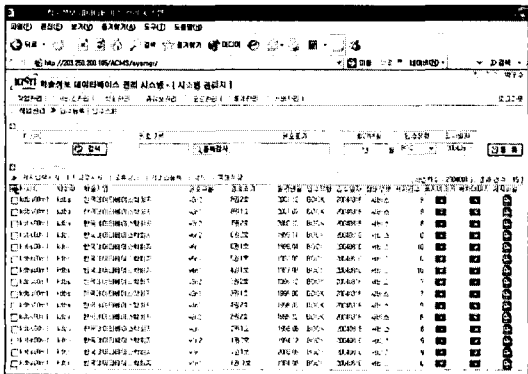
학술정보 관리 시스템의 DB는 학회정보, 중정보, 입수정보, 기사정보 테이블로 구성된다. [그림 3]는 DB 스키마를 ER-다이어그램으로 도식한 것이다.



[그림 3] ER-다이어그램

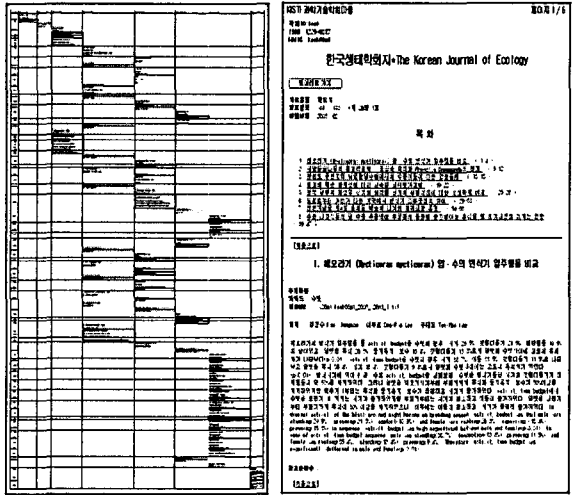
3.3 학술정보 관리 시스템의 동작원리

관리자는 입수된 학술정보를 시스템에 등록하고 서지입력자에게 권한위 XML문서로 작성 및 원문 이미지 스캔작업 명령을 내린다. 서지입력관리자는 XML데이터와 원문이미지를 제작하여 작업디렉토리에 저장한다. 서지입력관리에 관한 모든 작업 현황은 [그림 4]를 통해 확인할 수 있으며 생성된 데이터의 품질을 XML의 문법 체크와 원문이미지 뷰를 통하여 검증할 수 있다.



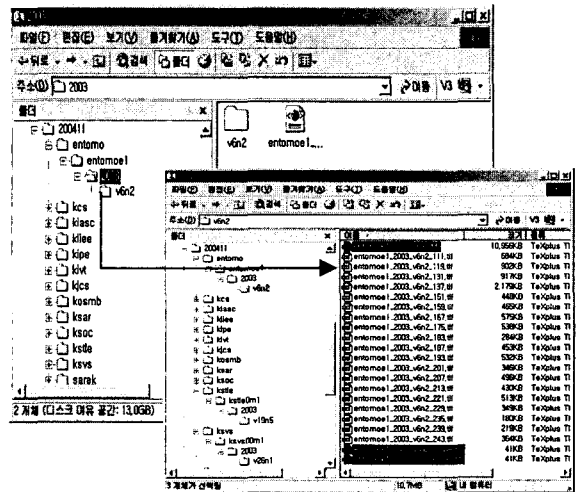
[그림 4] 서지입력 작업 관리 화면

또한 관리자는 [그림 5]과 같이 학술정보 XML DTD 구조에 맞게 XML문서를 XSL을 통하여 논문의 내용 확인 및 원문이미지의 가공상태를 확인할 수 있다.[5]



[그림 5] XML DTD 구조 및 XSL 적용 화면

데이터 가공상태를 확인한 관리자는 XML문서를 작업 크리스탈과 서비스 크리스탈에 배치적재하며 원문 이미지를 학회->종->발간년도->권호와 같이 구조적으로 파일시스템에 저장한다. 학회의 식별은 학회의 ID를 사용하며 종은 KISTI의 과학기술잡지 식별기호인 KOJIC을 사용한다[6]. [그림 6]은 학술정보 관리 시스템에서의 원문이미지와 XML문서의 저장 예이다.



[그림 6] XML문서 및 원문이미지 저장 구조

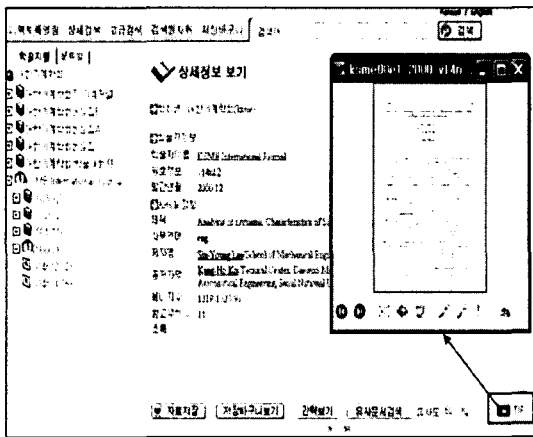
DB에 저장되는 서지 데이터의 키는 시스템 키 대신 위에서 정의한 저장구조+ 시작페이지를 사용함으로써 데이터 중복을 미연에 방지할 수 있도록 하였다.

키 구성의 예:

- 학회: 한국정보처리학회 (학회ID: kips)
- 종: 한국정보처리학회지(KOJIC: jbc11Z)
- 권호사항: 제 3권 3호
- 논문 시작페이지: 32page

=> kips_jbc11Z_v3n3_32

학술자료의 수집부터 가공, DB 구축이 모두 끝나면 일반 이용자들은 학술정보 관리 시스템을 통하여 다양한 검색 서비스를 제공 받을 수 있다. [그림 7]은 디렉토리 검색 서비스를 통해 학술정보가 제공되고 있는 화면이다. 사용자는 인터넷 상에서 고급 학술정보를 손쉽게 검색하여 일반 서지 사항 뿐 아니라 원문까지 열람할 수 있다.



[그림 7] 디렉토리 검색 서비스 화면

학술정보관리 시스템은 다양한 검색 서비스를 제공한다. 서비스 내용으로는 통합검색, 유사도 검색, 결과 내 검색, 상세검색, 고급검색, 디렉토리 검색이 있으며 부가적으로 원하는 논문들의 기본 정보를 저장하여 활용할 수 있는 저장바구니와 같은 편리한 기능들을 제공한다. 뿐만 아니라 국외 검색 이용자들을 고려하여 한글 검색과 동일한 기능을 제공하는 영문 검색 서비스도 제공할 수 있도록 화면을 설계하였다.

작업 KRISTAL과 서비스 KRISTAL의 듀얼 체제 하에서 학술정보의 입수, 가공, 배치적재까지의 일련의 DB 구축에 관한 전 과정을 단계별로 정리하면 [표 1]과 같다.

구분	설명
1. 입수정보등록	입수관리자가 서비스에 해당 저수의 입수정보를 등록.
2. 작업 지시	해당 저수의 입수정보 입력완료후 작업지시를 내리 입력업체에서 작업을 시작.
3. 서지입력	입력업체에서 해당 저수의 XML생성 및 원문이미지를 스캔.
4. 자료 업로드	작업디렉토리에 해당 저수의 XML과 원문이미지를 FTP를 이용하여 업로드.
5. 입수정보 복제	해당 저수의 서비스K에 등록된 입수정보를 작업K로 복제.
6. 오류검증	해당 저수의 XML의 오류경사와 원문이미지의 유무를 조사하여 적재가능한지 확인.
7. 배치적재화일 생성	해당 저수에 대해 오류경중이 끝난 XML을 KRISTAL 적재화일 변환기를 통해 적재화일을 생성.
7.1.FTP전송	작업디렉토리가 해당 저수에 존재하지 않을 경우, 검색서버에서 FTP로 백서버로 전송하여 적재화일을 가지고 검색서버에서 적재.
8. 작업K 배치적재	해당 저수의 배치적재화일을 작업K로 배치적재.
9. 원문자료 복사	적재완료 성공시 XML과 원문이미지를 원문저장디렉토리로 복사.
10. 백서서비스 작업K로 전환	백서서비스 대상 KRISTAL을 서비스K에서 작업K로 전환.
11. 작업K 입수완료 설정	작업K에 해당 저수의 입수완료를 설정.
12. 서비스K 배치적재	해당 저수의 배치적재화일을 서비스K로 배치적재.
13. 백서서비스 서비스K로 전환	백서서비스 대상 KRISTAL을 작업K에서 서비스K로 전환.
14. 서비스K 입수완료 설정	서비스K에 해당 저수의 입수완료를 설정. 평상시의 백서서비스 상태로 돌아옴.

[표 1] 학술정보 데이터 DB구축 작업 순서

4. 결론

본 논문에서는 데이터의 저장관리와 빠른 검색을 제공하는 KISTI의 최신 IRMS인 KRISTAL-2002을 기반으로 하여 학술정보관리 방안을 연구하고 시스템을 설계하였다. 학술정보관리 시스템은 학술정보의 처리를 메타데이터의 표준인 XML을 사용함으로써 데이터의 유효성 검증과 타 기관의 상호대차 및 원문 교환을 손쉽게 하였다. 또한 학술정보의 입수, 가공, 처리의 전 과정을 간단하게 온라인으로 관리할 수 있도록 시스템을 구성하고 사용자 별로 기능을 세분화하였다. 일반 검색이용자들은 KRISTAL-2002에서 제공하는 다양하고 강력한 검색기능과 듀얼 시스템운영으로 시스템 장애 또는 예외상황에서도 일관된 서비스를 제공할 수 있게 되었다.

5. 참고문헌

- [1] 한국과학기술정보연구원, "사용자 메뉴얼", 2004.
- [2] W3C, "Extensible Markup Language 1.0," <http://www.w3.org/TR/1998/REC-XML-19980210>, 1998.
- [3] ISO 8879(1986). Information Processing - Text and Office Systems - Standard Generalized Markup Language(SGML), International Organization of Standardization.
- [4] W3C, XML Path Language (XPath) 2.0, Working Draft , November 2002.
- [5] W3C, XSL Transformations (XSLT), Recommendation, Oct. 1999. (<http://www.w3.org/TR/WD-xsl>)
- [6] 한국과학기술정보연구원, " The Standardization of the Identification Code on Sci-Tech Journals" ,2004.