

## 효율적인 정보검색 및 관리를 위한 학술정보가공모델 연구

Design of Data Processing Model for Efficient Retrieval and Management of Science and Technology Information

이석형, 강남규, 김한기, 윤희준, 한성근, 윤화목  
한국과학기술정보연구원 정보유통기술개발실

Lee Seok-Hyoeng, Kang N.G., Kim H.G., Yoon H.J.,  
Han S.G., Yoon H.M.

Korea Institute of Science and Technology  
Information, Dept. of Information Technology

### 요약

일반적으로 데이터베이스를 구축하고 서비스하기 위한 대부분의 방식은 상용 DBMS를 이용하여 정보의 관리를 수행하고 사용자 검색은 정보검색 시스템을 사용하는, DBMS와 정보검색 시스템의 연동방식을 채택하고 있다. 그러나 이러한 방법은 DBMS와 정보검색 시스템을 이중으로 운영해야하는 불편함이 있고, 데이터의 관리 및 처리를 위한 과정이 중복으로 이루어지는 단점이 있다. 따라서, 본 논문에서는 과학기술 정보 데이터를 관리하고 원문과 함께 이용자에게 정보서비스를 위해 필요한 KRISTAL-2002 정보검색관리시스템 기반의 DB 구축과 정보검색 및 관리 모델을 제시한다.

### Abstract

It constructs a database generally and service it does, to use a commercial business DBMS that manages information and to use a retrieval system that supports a search, it adopts the gearing method of the DBMS and information retrieval system. But like this method is inconvenient must operate the DBMS and information retrieval system with duplication to be, there is a weak point where the data management and processing become accomplished with duplication. It presents the DB construction and information search and the management model of the KRISTAL-2002 information retrieval and management system base for the scientific and technical information data management and processing from the dissertation which it sees.

## 1. 서론

인터넷의 발달로 인해 사용자들은 웹을 이용하여 많은 정보를 습득할 수 있게 되었다. 또한 정보 수가 증가함에 따라 다양하고 양질의 정보를 제공하기 위한 많은 방법이 연구되고 활용되고 있다[1].

그 중에 과학기술정보는 한 국가의 과학기술발전의 중요한 밑거름이 되기 때문에 이에 대한 효율적인 정

보의 제공이 절실한 실정이다. 이에 한국과학기술정보연구원에서는 과학기술 종합정보시스템을 개발하여 국내 과학기술자에게 과학기술정보를 서비스 하였고, 2004년 11월 원내 주요 DB를 대상으로 하는 통합검색서비스를 제공하고 있다. 과학기술 통합검색 서비스는 국내외에서 발간된 학술잡지, 회의자료, 연구보고서, 학위논문, 특허기술, 분석 및 동향정보 등의 서지정보를 KRISTAL-2002 기반의 검색 데이터

베이스를 구축하고, 사용자가 이를 검색하고 필요한 정보에 대해 원문 신청까지 이루어질 수 있도록 하는 서비스이다.

한편, 이러한 고급정보를 종합적으로 관리하고, 이용자에게 빠르고 편리한 검색서비스를 제공하기 위해 정보 시스템에서의 정보의 저장 및 검색부분에 관한 많은 연구가 이루어지고 있으며 이러한 연구를 바탕으로 한 응용시스템들이 개발 및 이용되고 있다[2]. 일반적으로 데이터베이스를 구축하고 서비스하기 위한 대부분의 방식은 상용 DBMS를 이용하여 정보의 관리를 수행하고, 사용자 검색은 정보검색 시스템을 사용하는 DBMS와 정보검색 시스템의 연동 방식을 채택하고 있다. 이러한 방식은 DBMS의 장점인 대용량 데이터 트랜잭션의 용이성과 정보검색 시스템의 장점인 빠른 정보검색 시간 제공이 가능하다는 점에서 많이 이용되고 있으나, DBMS와 정보검색 시스템을 이중으로 운영해야하는 불편함이 있으며, 두 시스템간의 데이터의 일관성을 유지하는데 많은 어려움이 따른다[3].

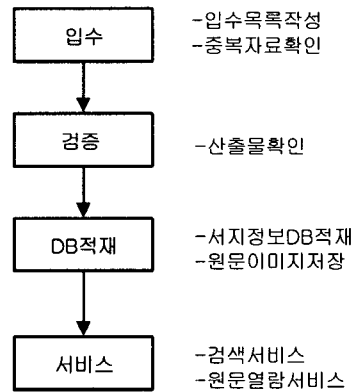
이에 본 논문에서는 서지 및 원문 정보를 포함하는 국내외 연구보고서 정보를 테스트 데이터로 선정하여, 서지정보의 관리 및 이용자 서비스를 위한, KRISTAL-2002 정보검색관리시스템 기반의 DB 구축, 정보검색 및 관리 모델을 제시하고, DBMS와 정보검색 시스템 연동방식과의 데이터 처리과정에 있어서의 성능을 비교 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 국내외 연구보고서의 데이터 처리과정과 DBMS와 검색 시스템을 이중으로 사용하는 현재의 관리시스템에 대해 살펴보고, 3장에서는 KRISTAL-2002의 특성과 이를 이용한 정보검색 및 관리 모델을 제시하며, DBMS와 정보검색 시스템 연동방식과의 데이터 처리과정에 있어서의 성능을 비교 분석하고 4장에서 결론을 맺는다.

## II. 자료관리시스템 - 연구보고서

연구보고서는 대학교 및 산업, 연구기관에서 수행되는 국가연구개발사업의 산출물로 수행 내용과 결과가 담겨 있다. KISTI는 한해 약 1만여 건의 연구보고서의 서지정보와 원문을 가공하여 이용자에게 서비스 하고 있으며, 현재 약 15만여 건의 국내외 연구보고서 정보를 웹을 통해 원문과 함께 제공하고 있다 [4].

### 1. 연구보고서 정보 처리절차

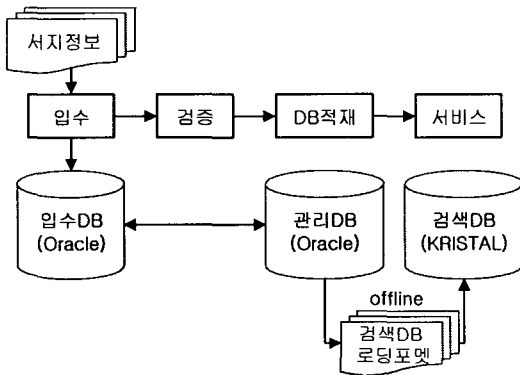


▶▶ 그림 1. 연구보고서 관리절차

연구보고서의 처리절차는 [그림 1]과 같이 이루어진다. 연구보고서의 DB관리자는 서지정보와 원문정보를 포함하는 연구보고서 데이터를 파일형태로 입수한다. 입수시에는 입수날짜, 입수건수, 처리형태 등이 기록되어야 관리가 용이하다. 입수된 데이터는 검증 단계를 거쳐 DB로 적재되는데 자료 검증 요소는 중복데이터 체크, 오류정보 추출 등이 있다. DB는 오류가 없는 데이터를 저장하는 정상DB와 오류 데이터를 저장하는 오류DB로 나뉘며, 서비스 대상이 되는 DB는 정상 DB이다. 오류 DB를 별도로 두는 이유는 오류 데이터의 관리를 용이하게 하기 위함이다. 서비스는 검색DB를 별도로 구성하여 이용자에게 제공하는데 이 때 서비스의 연속성을 보장하기 위한 DB 구성이 필요하다.

## 2. 연구보고서 자료관리시스템

[그림 1]과 같은 연구보고서의 관리를 위해 연구보고서 자료관리시스템이 개발, 운영되고 있다. 현재는 DBMS와 정보검색 시스템이 이중으로 운영되고 있는 형태로, 입수 및 데이터 검증에 대한 관리DB로 Oracle-DBMS를 이용하고 있다. 또, 작업이 끝난 정보는 KRISTAL 기반의 검색 DB로 적재하여 이용자 서비스를 제공하고 있다. 기존의 연구보고서 자료관리시스템의 구성도는 [그림 2]와 같다.



▶▶ 그림 2. 기존 연구보고서 자료관리시스템 구성도

## III. 연구보고서 자료관리모델

본 절에서는 KRISTAL-2002의 특징과 이를 이용한 정보검색 및 관리 모델을 제시하며, DBMS와 정보검색 시스템 연동방식과의 데이터 처리과정에 있어서의 성능을 비교 분석한다.

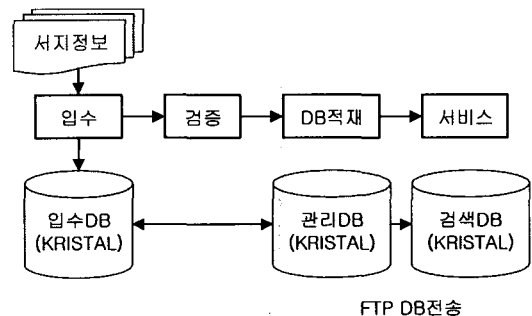
### 1. KRISTAL-2002

KRISTAL-2002는 KISTI에서 제공하는 정보검색 및 관리를 위한 시스템으로 데이터의 저장 관리, 검색에 관한 처리를 효율적으로 수행한다. 특히 구조 문서인 XML 정보를 수용할 수 있어, 정형데이터의 처리가 손쉽게 이루어 질 수 있다. KRISTAL-2002는 검색엔진과 데이터베이스관리시스템을 결합한 확장형 정보검색관리시스템으로 상용 DBMS의 장점과

검색엔진의 장점을 동시에 제공한다[3].

### 2. KRISTAL-2002 기반의 연구보고서 관리시스템

본 논문에서 제시하는 연구보고서 관리모델과 2장의 연구보고서 관리시스템의 가장 큰 차이는 KRISTAL-2002만을 이용한 단일시스템 기반의 관리모델이라는 점이다. 따라서, 2장에서 설명된 관리DB도 KRISTAL-2002 기반이며 입수DB 또한 KRISTAL 기반의 동일한 시스템으로 구성된다. [그림 3]은 KRISTAL-2002 기반의 국가연구보고서 관리시스템 구성도이다. [그림 4]와 같이 제안된 관리시스템은 서지정보를 관리DB로 저장하기 위한 방법이 온라인과 오프라인이 모두 지원되며, [그림 2]에서 보였던 검색DB로의 offline 적재 대신 관리DB와 검색DB 서버간의 FTP DB 전송으로만 검색DB구성이 끝나므로 검색DB 구성을 위한 로딩 작업등이 필요하지 않다.



▶▶ 그림 3. KRISTAL-2002기반 연구보고서 관리모델

### 3. 성능 평가

제안시스템을 성능평가하기 위한 요소로는 트랜잭션 발생시 안정성, 빠른 데이터 처리시간, 시스템 확장성, 검색 및 서비스 시간 등이 있을 수 있다. 제안시스템은 트랜잭션 발생시 안정성은 로그 처리 등으로 기존 시스템과 별 차이가 없음을 확인하였다. 그 외에 제안 시스템의 성능평가를 위해 연구보고서 서지정보를 처리하는데 가장 중요한 요소인 처리 시간

과 서비스 시간을 비교하였다. 처리시간은 입수, 발주 및 검수절차를 거친 서지정보를 관리DB에 입력하는 시간으로, 초록정보가 포함된 10,000건의 서지정보 (약 60Mb)를 온라인으로 입력하는데 걸리는 시간을 측정하였다. 비교 대상이 되는 시스템은 Oracle DBMS를 이용하는 관리시스템으로, 일반적으로 대부분의 기관 및 이용자가 관리 및 운영에 있어 그 안정성을 신뢰하고 있는 상용 DBMS이다.

```
public static boolean insert(String tn, HashMap hm) {
    Iterator iter = null;
    String secName, secValue = null;
    ParameterType pOut = null;
    ParameterType pIn = new ParameterType();
    DocumentType docType = new DocumentType();
    ClientLIB cl = new ClientLIB(ip, port);
    iter = hm.keySet().iterator();
    while (iter.hasNext()) {
        secName = (String) iter.next();
        secValue = (String) hm.get(secName);
        docType.sections.add(new SectionType(secName,
        secValue));
    }
    pIn.documents.add(docType);

    pOut=cl.Request(KristalHeader.APPEND_DOCUMENT, pIn);
    //System.out.println(pOut.service_time);
    return true;
}
```

▶▶ 그림 5. 관리DB 적재를 위한 KRISTAL DB API

DB 스키마는 양 시스템 모두 동일한 조건으로 설계하였으며, 관리를 위한 색인(Indexing)항목은 제목, 저자, 관리번호, 발행년도 등으로 하였다. [그림 5]는 KRISTAL-2002 API를 이용하여 KRISTAL DB에 데이터를 온라인 적재하는 Java API의 일부이다.

[표 1] 관리시스템 문서적재 시간

비교 시간	문서적재(비색인)		문서적재&색인	
	Oracle	KRISTAL	Oracle	KRISTAL
	28m58s	15m35s	29m38s	26m14s

[표 1]과 같이 KRISTAL 기반의 관리시스템에서의 적재시간이 관리형 DBMS인 Oracle을 기반으로 한 관리시스템에서의 적재시간보다 약 55%의 시간 감소 효과를 가져옴을 알 수 있었다.

#### IV. 결론

본 논문에서는 연구보고서 서지정보를 대상으로 KRISTAL-2002 기반의 학술정보가공모델을 제시하고, 일반적인 데이터 관리 모델과의 성능 평가를 수행하였다. 수많은 정보가 발생되고 처리되어야 하는 현실 속에서 데이터 처리 속도는 관리시스템의 설계에 있어 중요한 요소가 되기 때문에, 본 성능 평가는 중요한 의미가 있다고 할 수 있다. 또한, 데이터 정제 과정에서 필요한 중복 데이터 체크 부분에서도 일반 DBMS가 가지는 검색의 한계를 KRISTAL-2002 기반의 관리시스템에서 수용하기 때문에 양질의 데이터를 제공할 수 있다. 따라서 제시된 구축 모델은 예로 든 연구보고서 자료관리시스템뿐만 아니라 기타 대용량의 정보검색 및 관리 모델로 활용될 것으로 기대된다.

#### ■ 참고 문헌 ■

- [1] 이석형, 강남규, 윤화목, 예용희, 김현 "Implementation of the XML based Science and Technology Information System using KRISTAL", The 7th IASTED International Conference, IMSA 2003, 2003.08
- [2] 이민호, 진두석, 김광영, 주원균, 서정현, 류범중 "정보 검색관리시스템 KRISTAL-2000 설계 및 구현", 정보 과학회 봄 학술발표논문집 Vol.30, No.1, pp.653-655, 2003.
- [3] 주원균, 정창후, 이민호, "KRISTAL-2002를위한 JAVA 사용자 API의 설계 및 구현", 정보과학회 가을 학술발표논문집 Vol.31, No.2, pp.433-435, 2003.
- [4] 한국과학기술정보연구원, "과학기술종합정보 서비스 시스템 개발 및 운영", 2004.