

분산 저장된 과학기술정보 서비스를 위한 검색 데이터베이스 관리 도구의 설계 및 개발

Development of the Management Tool for S&T information in distributed retrieval database

이석형, 윤희준, 여일연, 최성필, 윤화목
한국과학기술정보연구원

Lee Seok-Hyoung, Yoon Hee-Jun, Yeo Il-Yeon,
Choi Sung-Pil, Yoon Hwa-Mook
Korea Institute of Science and Technology
Information

요약

본 논문에서는 과학기술정보의 관리에 필요한 여러 기능을 포함하는 GUI 기반의 검색 데이터베이스 관리 도구(K-Manager)를 제안한다. 일반적으로, 웹 기반의 과학기술정보 서비스를 위해서는 검색 데이터베이스의 구성이 필수적이나, 검색 데이터베이스의 관리 기능이 관계형 데이터베이스와 같은 기능을 모두 지원하지 않을 뿐만 아니라 TOAD나 GOLDEN과 같은 데이터베이스 관리 도구를 지원하지 않기 때문에, 콘텐츠 관리자나 시스템 관리자가 손쉽게 정보를 처리하기가 어려운 실정이다. 이 도구에는 과학기술정보를 관리하는 콘텐츠 관리자가 필요한 기능과 검색 데이터베이스를 관리하는 시스템 관리자가 필요한 기능이 GUI 기반으로 통합 구성되어 있어 콘텐츠 관리자가 원하는 데이터베이스에 접근하여 각종 데이터 작업을 수행할 수 있고, 데이터베이스 관리자가 데이터베이스의 관리도 수행할 수 있다. 특히 K-Manager는 대용량 과학기술정보의 빠르고 정확한 검색 서비스와 안정적인 관리를 위해 구성된, 분산 검색 데이터베이스에 저장된 정보를 처리하는데 효과적이다. 본 관리 도구는 크게 콘텐츠 관리기와 데이터베이스 관리기로 그 기능을 분류할 수 있으며, 국내 최대 과학기술 포털사이트인 Yeskisti의 콘텐츠 및 데이터베이스 관리를 그 기본 모델로 하였다.

Abstract

In this paper, we suggest the GUI Management Tool, named K-Manager, for management and service of science and technology information that stored in distributed retrieval databases. Generally, it must be adapted retrieval database system for web based S&T contents service. But, It is inconvenient contents manager or the system administrator controls information easily, because it does not support the S&T information management process like TOAD, which can use for the relation database, in information retrieval database system. Using K-Manager, content manager can process the S&T content and system manager can manage the databases easily. The proposed tool ctive controls information effectively which is stored in the distributed retrieval database which guarantee the safety management of the contents stored in database and operate retrieval with efficient performances. Our tool consists of two sub systems, one is content manager, the other is database manager for YESKISTI based on KRISTAL-IRMS.

I. 서론

연구자나 과학기술자등이 제공하거나 작성하는 논문, 연구 보고서, 뉴스 정보 등의 과학기술정보는 또 다른 연구결과를 낳을 수 있는 중요한 정보이다. 정보 이용자 또는 연구자의 입장에서 연구 업적을 평가하고 재활용하기에 충분한 데이터를 얻을 수 있고, 진행 중인 연구들의 추세를 확인하여 연구방향 설정, 연구과제 도출업무 등에 필수적인 정보를 얻을 수 있다[1]. 현재 국내외 대표적인 과학기술정보 제공 사이트로 한국과학기술정보연구원의 과학기술정보포털(yeskisti.net)[2],

국가과학기술전자도서관(ndsl.org)[3], 한국교육학술정보원(keris.or.kr)[4], British Library[5], Science Direct[6] 등이 있다. 이러한 과학기술정보를 서비스하는 시스템들은 내부적으로 검색 데이터베이스의 구성이 필수적이다. 그러나, 검색 데이터베이스의 관리 기능이 관계형 데이터베이스와 같은 기능을 모두 지원하지 않을 뿐만 아니라[7], TOAD, GOLDEN과 같은 GUI 기반의 관리 도구를 지원하지 않는 것이 대부분이기 때문에 양질의 콘텐츠를 제공하는데 어려움이 있다. 이에 본 논문에서는 과학기술정보의 효과적인 관리와 이를 포함

하는 데이터베이스의 관리 기능을 포함하는 GUI 기반의 검색 데이터베이스 관리 도구를 제안한다. 이 도구에는 과학기술정보를 관리하는 콘텐츠 관리자가 필요한 기능과 검색 데이터베이스를 관리하는 시스템 관리자가 필요한 기능이 GUI 기반으로 통합 구성되어 있어 콘텐츠 관리자가 원하는 데이터베이스에 접근하여 각종 데이터 작업을 수행할 수 있고, 데이터베이스 관리자가 데이터베이스의 관리도 수행할 수 있다. 특히 제안된 관리 도구는 대용량 과학기술정보의 빠르고 정확한 검색 서비스와 안정적인 관리를 위해 구성된, 분산 검색 데이터베이스에 저장된 정보를 처리하는데 효과적이다. 본 관리 도구는 크게 콘텐츠 관리기와 데이터베이스 관리기로 그 기능을 분류할 수 있으며, 기반 데이터베이스는 KRISTAL이 적용되었다.

본 논문의 구성은 다음과 같다. 2장에서는 과학기술정보의 관리를 위한 요소를 한국과학기술정보연구원의 과학기술정보포털(yeskisti.net)을 예로 설명하고, 3장에서 관리도구의 구조를 제시한다. 4장에서 관리도구의 주요 기능에 대해 설명한 후 결론을 맺는다.

II. 과학기술정보 서비스 관리요소

한국과학기술정보연구원에서는 국내최대의 과학기술정보포털 사이트인 yeskisti.net[2]을 운영, 이용자에게 양질의 콘텐츠를 제공하고 있다. yeskisti.net에는 국내외 학술논문, 연구보고서, 특허, 분석 및 동향, 인력 정보와 블로그, 커뮤니티 등의 이용자 콘텐츠를 서비스 하고 있으며, 그 건수는 3,000만건 정도에 이른다. 이를 관리하기 위한 관리요소를 시스템과 콘텐츠로 구분하여 분석한다.

1. 시스템 관리요소

과학기술정보 서비스를 위한 시스템 관리요소로는 기반 데이터베이스의 관리와 웹 및 컨테이너 서버의 관리등을 들 수 있다. 본 논문에서는 yeskisti.net의 기반 데이터베이스인 KRISTAL[8]을 중심으로 한 관리 요소를 살펴본다.

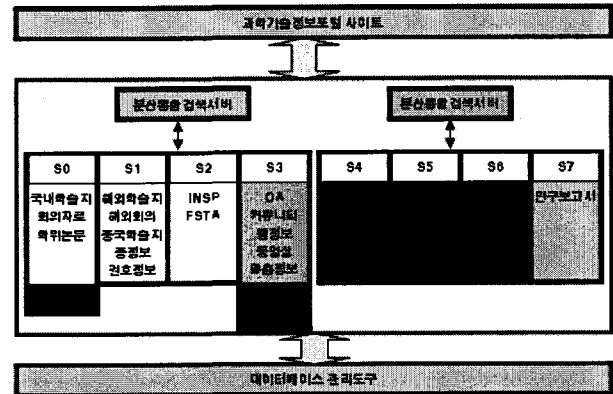
1.1 KRISTAL

KRISTAL은 한국과학기술정보연구원에서 개발한 정보검색관리시스템으로, UNIX 및 LINUX 기반에서 데이터에 대한 저장, 관리 및 검색에 대한 처리를 효율적으로 수행할 수 있는 시스템이다[8]. 최근 대표적인 포털사이트 등에서 자체 개발하여 서비스 하고 있는 검색엔진과 달리 KRISTAL은 정보검색엔진에 데이터베이스 관리기능(DBMS)을 추가한 정보검색관리시스템(IRMS)이다. 현재 KRISTAL은 국내의 대표적인 과

학기술포털사이트인 yeskisti 를 비롯한 국가평판위원회[9], 한국역사정보통합시스템[10]등 과학기술정보와 공공기술정보 서비스에 적용 중에 있다.

1.2 시스템 구조

KRISTAL 기반 과학기술정보포털의 데이터베이스 구조는 다음과 같다.



▶▶ 그림 1. Yeskisti 데이터베이스 구성 및 연계도

그림 1과 같이 데이터베이스는 콘텐츠의 양과 갱신정도, 관리부서에 따라 8개의 서버로 분산되어 저장된다. 서버 0~2까지는 논문 데이터베이스를 분산 저장하였고, 서버 4~6은 특허 데이터베이스를 분산 저장하였다. 논문 및 특허 데이터베이스의 콘텐츠 양은 각각 2,000여만건과 1,000여만건으로 그 양이 매우 크기 때문에 하나의 서버에서 이러한 용량을 처리하는 것이 매우 비효율적이다. 따라서, 각 콘텐츠를 특성에 맞게 분류한 후 분산된 데이터베이스에 저장한 후, 검색결과정보는 분산통합검색서버를 이용, 카테고리별로 통합하여 이용자에게 서비스한다. 그러므로, 그림 1의 하단부에 기술되어 있는 데이터베이스 관리 도구는 분산 저장된 콘텐츠를 종합적으로 관리할 뿐만 아니라, 이를 포함하는 데이터베이스의 모니터링 및 관리를 지원해야한다.

1.3. 관리요소

과학기술정보 서비스를 위한 시스템 관리요소 중 콘텐츠를 저장하는 데이터베이스의 관리요소는 다음과 같다.

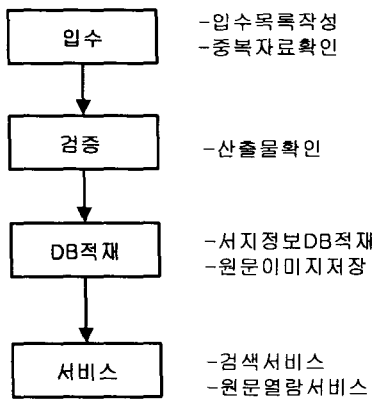
- 데이터베이스 동기화
- 데이터베이스 개방, 폐쇄
- 신규테이블 추가/삭제
- 데이터베이스 모니터링
- 자료 백업 및 복구

데이터베이스 동기화는 검색 데이터베이스의 특성상 일괄

적재시에 데이터베이스를 폐쇄해야하기 때문에 서비스의 연속성을 위해 이중으로 관리되는 데이터베이스의 일치를 의미한다. 일괄적재를 위해서는 데이터베이스의 개방 및 폐쇄가 필수적이며, 검색 중에 발생하는 서버 디스크의 로드 밸런싱을 위해서는 콘텐츠의 저장 정책에 따른 테이블의 추가/삭제 기능 또한 필요하다. 데이터베이스 모니터링은 분산된 데이터베이스중 하나의 데이터베이스가 시스템 오류로 인해 작동이 멈추었을 때 신속한 복구를 위해 필요한 요소이다. 자료 백업 및 복구는 서버 디스크의 오류 발생 시에 신속한 복구를 위해 필요하다. 이러한 요소들은 시스템 관리자가 서버에 직접 접근하여 실행해야하기 때문에, 매우 복잡한 절차에 따라 이루어진다. 따라서 손쉽게 시스템을 관리하기 위한 도구가 필요하다.

2. 콘텐츠 관리요소

2.1 콘텐츠 처리 절차



▶▶ 그림 2. 콘텐츠 처리 절차

콘텐츠 관리자는 서지정보와 원문정보를 포함하는 자료를 파일형태로 입수한다. 입수 시에는 입수날짜, 입수건수, 처리 형태 등이 등록 되어야 관리가 용이하다. 입수된 자료는 검증 단계를 거쳐 데이터베이스로 적재되는데 자료 검증 요소는 중복데이터 체크, 오류정보 추출 등이 있다. 데이터베이스는 오류가 없는 자료를 저장하는 정상 DB와 오류 자료를 저장하는 오류DB로 나뉘는데, 오류 DB를 별도로 두는 이유는 오류 자료의 관리를 용이하게 하기 위함이다[11]. 콘텐츠 처리 절차 중 입수 자료의 등록 및 검증은 각 자료별로 그 방법이 매우 상이하기 때문에 범용으로 도구화하기 어렵다. 따라서, 본 논문에서는 콘텐츠 처리 절차 중 입수와 검증과 관련된 부분을 제외한 자료의 적재와 적재된 자료의 수정 등에 관련된 부분만을 대상으로 한다.

2.2 관리요소

앞 절에서 살펴본 콘텐츠 처리 절차에 따른 관리요소는 다음과 같다.

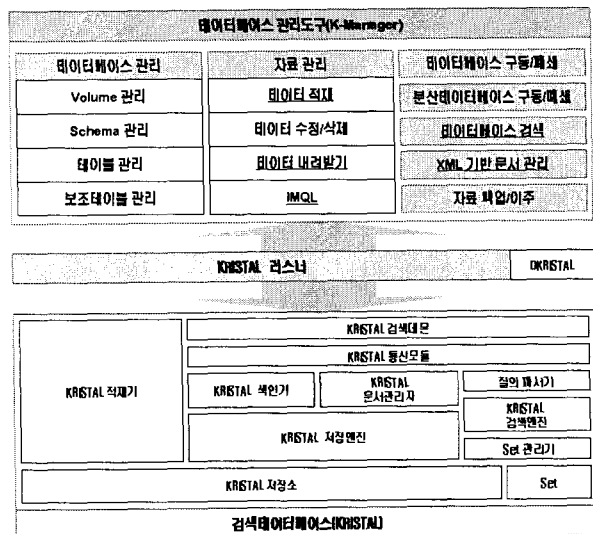
- 자료 적재
- 자료 갱신
- 일괄 내려받기
- 자료 검색
- XML 문서 관리

자료 적재는 가공이 끝난 자료를 일괄 적재 및 건 당 입력을 의미하며, 갱신은 적재된 자료의 속성을 변경하거나 자료를 삭제하는 과정을 의미한다. 주로 과학기술정보의 갱신은 원문이 생성되었거나, 권 호 정보 등 특정한 값의 일괄 변경, 비공개 속성 변경, 연구자 등의 요청에 따른 자료의 삭제 등이 빈번하게 발생된다. 또한, 타 기관으로의 콘텐츠 제공 및 재가공을 위해 특정 파일 형태로 내려받는 일괄 내려받기의 기능도 필요하다. 위와 같은 요소를 지원하기 위한 자료의 검색 기능과 정형문서의 관리 기능도 관리자 입장에서 필요한 요소가 된다.

III. 시스템 설계 및 개발

2장에서 살펴본 바와 같이 시스템 관리자와 콘텐츠 관리자를 위한 관리요소는 양질의 콘텐츠 서비스를 위한 중요하다. 이를 위한 데이터베이스 관리 도구(K-Manager)의 주요 내용은 다음과 같다.

1. 시스템 구성 및 개발 환경



▶▶ 그림 3. K-Manager와 데이터베이스 관계도

일반적으로 검색 데이터베이스에서는 검색이 중요한 기능이 되기 때문에 관리기능을 위한 프로세스가 별도로 필요하다. 이에 그림 3과 같이 데이터베이스 관리도구와 검색데이터베이스 사이에 리스너를 두어 관리 기능을 지원하도록 하였다. 리스너는 RDBMS에서 적용하는 리스너의 일반적인 기능뿐만 아니라 앞 절에서 살펴본 시스템 관리요소 및 콘텐츠 관리요소인 데이터베이스 개방/폐쇄, 테이블 관리, 백업관리, 모니터링 등 온라인 상으로 데이터베이스를 직접적으로 제어할 수 있는 관리 기능을 제공하는 서버 프로세스이다. 이 프로세스는 KRISTAL이 기본적으로 제공하는 검색 데몬 프로세스와 독립적으로 작동되며, 데이터베이스 관리 실행 요청을 받아 관리 API를 실행하고, 그 결과를 도구로 전송하는 역할을 수행한다.

시스템 개발환경은 Java 기반의 개발도구인 JBuilder 2005를 기반으로, J2SDK.1.4 버전을 이용하였으며, 본 도구는 Windows 계열의 운영체제에서 작동한다.

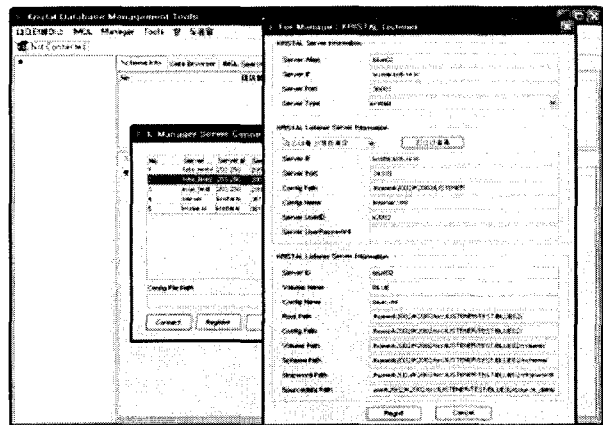
2. 시스템 구현

K-Manager는 아래와 같은 기능을 제공한다.

2.1 시스템 관리

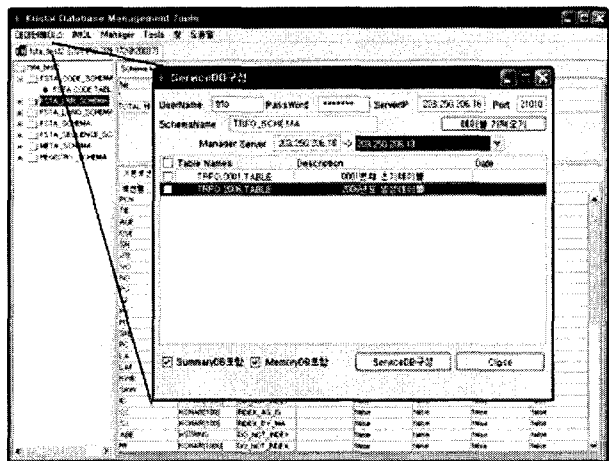
K-Manager를 통한 시스템 관리는 관리자가 관리하고자 하는 데이터베이스의 등록부터 데이터베이스 접속, 데이터베이스의 구동/폐쇄 뿐만 아니라 테이블의 생성과 분산검색서버의 모니터링, 데이터베이스의 동기화 등을 지원한다. K-Manager에서 시스템 관리를 위해 지원하는 주요 기능은 아래와 같다.

- 데이터베이스 생성 관리
 - 등록된 데이터베이스 정보 열람
 - 데이터베이스(Volume) 생성/삭제
 - 테이블 생성/삭제 기능
 - 스키마 편집 기능(테이블의 속성 변경)
- 데이터베이스 구동 관리
 - 데이터베이스 구동/멈춤 기능
 - 데이터베이스 모니터링 기능
- 데이터베이스 동기화
- 데이터베이스 백업/Migration



▶▶ 그림 4. K-Manager 데이터베이스 연결 및 리스너 등록

그림 4는 K-Manager에서 데이터베이스 접속과 데이터베이스 관리를 위한 리스너를 등록하고, 리스너 서버를 생성하는 화면이며.그림 5는 데이터베이스 볼륨 전송을 통한 데이터베이스 동기화를 위한 화면이다.



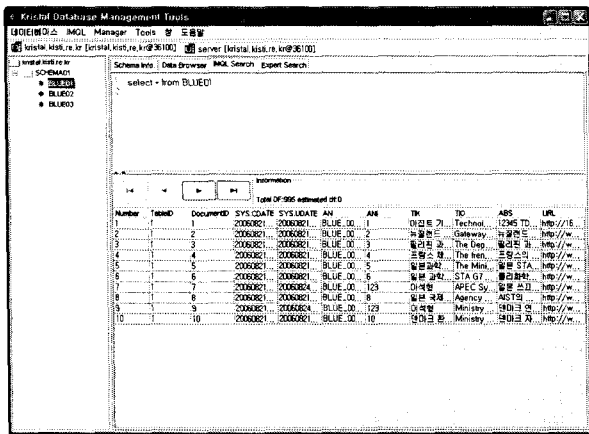
▶▶ 그림 5. K-Manager 데이터베이스 동기화

2.2 콘텐츠 관리

K-Manager를 통한 콘텐츠 관리는 자료의 적재 및 적재된 자료에 대한 수정 및 XML 문서 관리 등을 들 수 있다.

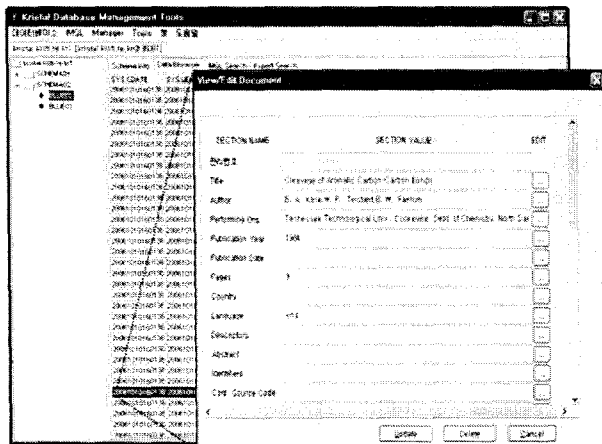
K-Manager에서 콘텐츠 관리를 위해 지원하는 주요 기능은 아래와 같다.

- 자료 관리
 - 온라인 삽입/수정/삭제,
 - 일괄 적재/색인/적출 기능
 - IMQL
 - 스키마 편집 기능
- 자료 검색
- XML 기반 문서 관리
- XML 기반 문서 열람 및 편집 기능



▶▶ 그림 6. K-Manager IMLQL

그림 6은 자료의 검색을 위해 지원하는 IMQL을 실행하는 화면이며, 그림 7은 검색된 자료를 수정하는 기능을 보이는 화면이다.



▶▶ 그림 7. K-Manager 자료 수정

IV. 결론

과학기술정보 서비스를 위한 검색 데이터베이스 관리 도구(K-Manager)는 KRISTAL 기반 데이터베이스를 구성하고, 저장된 콘텐츠를 관리하기 위한 도구로서, 관리자가 원격으로 데이터베이스 서버에 직접 접근하지 않고, 개인 PC에서 One-Click으로 데이터베이스 및 콘텐츠의 손쉬운 관리 및 운영을 지원하는 응용 프로그램이다. 기존의 데이터베이스를 운영하기 위해서는 복잡한 스키마 작성 과정 및 명령어 실행 과정이 필요했으나, 본 도구를 이용하면 데이터베이스 운영 및 콘텐츠 관리의 편의성을 제공할 수 있다. 또한, 본 도구는 분산검색서버 뿐만 아니라 단일검색서버도 지원하기 때문에, 모든 KRISTAL 기반의 데이터베이스를 관리하는데 효과적

일 것으로 기대한다. 향후 연구로는 초대용량 콘텐츠를 안정적으로 처리하는데 필요한 요소기술의 적용 및 사용자 편의성을 강화한 시스템 개선이 필요할 것이다.

■ 참고 문헌 ■

- [1] 주원균, 정창후, 이민호, “KRISTAL-2002를 위한 JAVA 사용자 API의 설계 및 구현”, 정보과학회 가을 학술발표논문집 Vol.31, No.2, pp.433-435, 2003.
- [2] 과학기술정보포털, <http://www.yeskisti.net>
- [3] 국가과학기술전자도서관, <http://www.ndsl.org>
- [4] 한국교육학술정보원, <http://www.keris.or.kr>
- [5] British Library, <http://bl.uk>
- [6] Science Direct, <http://www.sciencedirect.com>
- [7] 이민호, 진두석, 김광영, 주원균, 서정현, 류범중, “정보검색관리 시스템 KRISTAL-2000 설계 및 구현”, 정보과학회 봄 학술발표논문집 Vol.30, No.1, pp.653-655, 2003.
- [8] KRISTAL-IRMS, <http://www.kristalinfo.com>
- [9] 국사편찬위원회, <http://www.history.go.kr>
- [10] 한국역사정보통합시스템, <http://kh2.koreanhistory.or.kr>
- [11] 이석영 외, “효율적인 정보검색 및 관리를 위한 학술정보공모 델 연구”, 한국콘텐츠학회 2005 추계종합학술대회 논문집 Vol.3, No.2, pp.442-445, 2005.