

메타데이터 정보를 이용한 분산 KRISTAL-2002 시스템의 통합 검색 시스템 개발

최 귀 자[†] · 김 재 곤^{**} · 서 정 현^{***} · 조 한 형^{**} ·
이 민 호^{***} · 정 창 후^{***} · 박 동 인^{***} · 남 영 광^{****}

요 약

본 논문에서는 메타데이터를 이용하여 분산된 KRISTAL-2002 기반의 시스템을 통합 검색하는 정보검색시스템의 구현 방법을 제안한다. 이 시스템은 서로 다른 분야에서 사용되고 있는 기계화된 정보검색시스템을 통합 검색하거나, 같은 분야의 시스템이라도 스키마가 다를 경우에 이를 사용자가 마치 하나의 시스템에서 검색하는 것과 같은 통합검색을 수행할 수 있도록 해준다. 본 시스템은 통합메타데이터베이스와 원천서버간의 스키마 매핑을 지원하는 원천서버관리기, 메타데이터를 등록하고 원천서버의 정보와 통합 메타데이터간의 매핑을 관리하고 지원하는 통합메타데이터 관리기, 통합검색을 위한 사용자 질의를 각 원천서버에 적합한 질의로 생성하여 처리하는 분산질의 처리기, 검색된 결과를 통합하여 사용자 화면으로 출력하기 위해 HTML 문서로 변환하는 분산 데이터 셋 관리기, 그리고 통합검색을 위한 통합검색 엔진으로 구성되어 있다. 통합메타데이터는 본 시스템의 일부분인 메타데이터 등록기를 이용하여 ISO/IEC 11179에서 정의된 표준화된 절차로 등록되었다고 가정하여 사용하였다. 사용자는 하나의 통합시스템에서 검색하는 것과 같이 통합검색 화면에서 검색대상 시스템을 선정하고 기본검색과 상세검색을 수행하며, 분야별 혹은 기관별로 검색하고 결과를 볼 수 있도록 하였다. 본 시스템은 KRISTAL-2002 시스템을 기반으로 Visual C++와 C++ CGI를 이용하여 리눅스 상에서 개발되었으며 6개의 서로 다른 데이터베이스를 이용하여 실험, 검증하였다.

Development of an Integrated Retrieval System on Distributed KRISTAL-2002 Systems with Metadata Information

Gui-ja Choe[†] · Jae-Gon Kim^{**} · Jung-Hyun Seo^{***} · Han-Hyung Cho^{**}
Min-Ho Lee^{***} · Chang-Hu Jung^{***} · Dong-In Park^{***} · Young-Kwang Nam^{****}

ABSTRACT

In this paper, we propose an integrated information retrieval system for distributed multiple KRISTAL-2002 systems by using the metadata information. This system integrates current systems for different areas or systems for the same area with the different schemas so that the users can get the answers by once from the whole systems. The proposed system comprises of the Source Server Manager(SSM) supporting the mapping between the integrated metadata database and source server, the Integrated Metadata Manager(ISM) for registering and managing the metadata and schema mapping, the Distributed Query Processor (DQP) for processing the user query into the source server query, the Distributed Data Set Integrated Manager(DDSIM) for transforming the total retrieval results by merging to the HTML format, and the integrated retrieval engine for managing the query results. It is assumed that the integrated metadata follows ISO/IEC 11179 metadata registration procedure with the metadata registry system which is a subsystem of the proposed system. There are two kinds of queries for users: the basic query and the detailed query. The users may select the databases or organizations for results by their own choices before giving the queries. The proposed system has been developed over KRISTAL-2002 systems with Visual C++ and C++-CGI and tested and verified with the six database systems.

키워드 : 통합정보검색(Integrated Information Retrieval), 메타데이터 레지스트리(Metadata Registry), 데이터베이스(Databases)

1. 서 론

1990년대 초기부터 각 기관 혹은 분야별로 수많은 정보

검색 시스템이 구축되어 서비스를 제공하고 있다. 인터넷 검색이 보편화 되면서 전문 정보에 대한 사용자의 통합 검색 욕구는 더욱 늘어나고 있는데 비해, 많은 비용과 오랜 시간에 걸쳐서 구축된 전문정보에 대한 정보검색 시스템을 통합하는 속도는 느리게 진행되고 있다. 모든 전문 분야에 대한 통합검색시스템을 구축하기 위해서는 수천만 건에 이르는 서지정보를 관계형 데이터베이스에서와 같이 정규화

† 정 회 원 : 연세대학교 전산학과
 ** 준 회 원 : 연세대학교 전산학과
 *** 정 회 원 : 한국과학기술정보연구원(KISTI) 정보시스템개발실
 **** 정 회 원 : 연세대학교 전산학과 교수
 논문접수 : 2004년 11월 9일, 심사완료 : 2005년 1월 7일

를 통해 분리된 여러 개의 테이블을 이용하여 저장 검색하거나, 아니면 하나의 테이블을 이용하여 저장 검색한다. 전자의 경우는 조인에 의한 부담 때문에 시스템의 속도가 느리고, 하나의 테이블을 이용하여 저장할 경우에는 너무 중복된 자료가 많아 저장 공간을 효율적으로 사용하지 못하는 문제점을 가진다. 후자의 경우는 같은 종류의 데이터를 하나의 시스템만을 이용하여 사용하기에는 시스템의 부하가 많이 걸리는 단점이 있기 때문에 여러 테이블 스키마와 여러 대의 시스템을 이용하여 개발해야만 검색 시스템의 성능 향상을 기대할 수 있다.

비슷한 분야의 정보에 대해 많은 시간과 비용을 들여서 구축된 기존의 검색 시스템을 하나의 통합된 동질의 검색 시스템으로 구축하기에는 위험성이 너무 크고 서비스를 중단해야 하기 때문에, 기 구축된 시스템을 변환하는 것은 현실적이지 못하다. 또한 비슷한 분야의 정보를 각 시스템마다 사용자가 등록하여 각각 검색하는 것도 사용자에게는 매우 불편하다.

이질적인 하드웨어 상에서 이질적인 검색엔진들을 통합 검색하기 위한 프로토콜들로, Z39.50[21, 22], K-protocol[23] 등 분산 통합 검색 프로토콜들이 정의, 연구되고 있다. 분산통합검색 시스템은 정의된 프로토콜들을 구현하여 각 서버에 이식하여 수행하는데, Z39.50 같은 경우는 개발에 대한 부담이 너무 커서 대형 서버를 이용한 대용량 서비스 기관 등에만 적용될 수 있다. KISTI가 개발한 K-protocol은 그러한 부담은 줄였지만 시스템 통합시 전문가에 의한 장시간 프로그래밍 작업이 필요하다. 이 작업은 KRISTAL-2002를 사용하여 전문분야별로 기 구축된 시스템(이하 원천서버라 칭함) 간의 분산통합검색에도 필요하며, 프로그램 전문가만이 할 수 있어 시스템 통합을 어렵게 하는 원인이기도 하였다. KRISTAL-2002는 데몬 형태로 운영되는 검색관리시스템이며, 분산통합검색시스템을 KRISTAL-2002 기반으로 할 경우 데이터베이스들의 통합검색 서비스는 매우 단순하게 구현될 수 있다.

본 논문에서는 현재 구축되어 있는 각 원천스키마를 변경하지 않고 별도의 프로그래밍 작업 없이 스키마 매핑과 서버의 위치정보만을 이용하여 분산된 시스템을 통합 검색할 수 있는 방법을 제안하고 구현하였다. 시스템 통합 및 검색 절차는 다음과 같다. 원천서버관리자는 각 원천서버의 데이터베이스에 대한 스키마 정보와 통합검색 시스템의 대상이 되는 메타정보를 정의한다. 통합서버관리자는 통합절차를 쉽게 하기 위해서 각 원천 서버의 데이터베이스 혹은 테이블에 대한 메타정보를 이용하여 ISO/IEC 11179에서 제안한 방법으로 통합 메타데이터를 구축한다. 원천서버관리자는 원천스키마를 표준화된 절차로 등록 관리되는 통합 메타스키마의 속성과 일치되도록 매핑한다. 이때 구조적 이질성은 온톨로지(ontology) 개념을 적용하여 해결한다. 매핑 정보를 이용하여 사용자는 어떠한 서버 및 데이터베이스에 어떠한 정보가 있는지에 관계없이 현재 분산되어 있는 정보에 대한 정보를 쉽게 검색할 수 있다. 사용자가 제

시한 질의는 각 원천서버의 데이터베이스에 맞게 재생성되어 각 원천서버의 검색엔진에게 보내지며, 각 원천서버에서 얻어진 결과를 통합하여 사용자에게 출력한다. 이 경우 간단한 메타정보와 검색하기 전의 약간의 전처리와 후처리를 통해 시스템의 성능 향상을 가져올 수 있으며, 사용자에게 검색 자료 유형별로 찾아다니며 검색하는 수고를 덜게 해준다.

통합 검색 시스템은 Visual C++과 C++CGI를 이용하여 KRISTAL-2002의 API를 통해 어떠한 시스템이든지 스키마의 의미만 알면 원천서버를 연결하여 통합 검색에 참여시킬 수 있도록 개발하였다. 실험의 결과는 현재 KISTI가 원천서버별로 제공하고 있는 과학기술문헌(LITERATURE), 동향(TREND), 연구보고서(REPORT), 분석(ANALYSIS), 특허(PATENT), 인력(HRST-DB) 등 6개 데이터베이스를 통합 검색하여 검증하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를, 3장에서는 KRISTAL-2002 시스템의 개요를 소개하며, 4장에서 분산 통합 검색기의 각 컴포넌트들에 대한 설명과 구축 결과에 대해서 기술한다.

2. 관련 연구

여러 곳에 분산되어 있는 데이터 통합과 관련 데이터들의 이질성 문제를 해결하기 위한 다양한 방법이 검색엔진 및 정보검색시스템의 개발과 함께 지속적으로 연구되고 있다. 그러나 정보를 보유하고 있는 분산된 원천 시스템들 간의 스키마적인 불일치 문제에 초점을 맞추고 있는 스키마 통합과 스키마 매핑 방법은 의미 상호교환 문제에 있어서 일부분만을 해결하고 있다.

웹 검색엔진 기술 및 디렉토리 기반 정보서비스 포털은 일반 정보 검색 이용자를 위한 대용량 정보 검색 분야를 발전시키는 주요 요인이 되었다. 그러나 매우 짧은 질의에 비해 상대적으로 많은 양의 정보를 대상으로 검색을 해야 하는 어려움이 있으며, 이를 해결하기 위한 방법으로 일반 문서에는 존재하지 않은 링크 및 추가적인 구조 정보의 특성을 반영한 페이지 랭킹과 같은 새로운 기술이 개발되고 있다. 웹 검색과 관련된 새로운 기술 분야인 크롤링(crawling) 방법은, 분산된 여러 시스템에 저장되어 있는 문서를 수집하여 색인하는 기술로써, 어느 분야의 문서를 얼마나 빠른 시간에 검색 대상에 포함시키는가에 초점을 맞추고 있다[19]. 그러나 문서의 수집, 저장 및 서비스 문제에서 발생하는 엔지니어링 문제를 해결해야 한다. 클러스터링(clustering) 방법은 너무 많은 검색 결과가 사용자에게 반환되기 때문에 이를 정리해서 결과를 제시함으로써 정보의 과부하를 줄이고 사용자가 원하는 문서에 신속하게 접근할 수 있도록 하는 기술이지만, 속도의 개선과 정보의 정확성을 유지해야 하는 어려움이 있다[17]. 웹 검색의 일환으로 개발된 메타-검색(meta-searching)은 사용자 질의를 웹 검색엔진으로 보낸

후 검색된 결과를 통합하여 사용자에게 제시하는 기술이다[5, 12]. 검색 서비스를 제공하는데 있어 웹 크롤러나 대용량 문서에 대한 색인을 구축할 필요가 없다는 장점과 효과적인 통합이 이루어질 경우 다양한 원천 시스템으로부터 검색된 결과를 활용한다는 장점이 있다. 그러나 기존 시스템의 협력이 없이는 운영이 불가능하다는 한계를 가지고 있고, 다양한 알고리즘으로 생성된 랭킹과 서로 다른 문서집합을 통합하여 양질의 단일 결과를 생성하는 것은 기술적으로 중요한 문제이기 때문에 아직 해결해야 할 과제로 남아있다.

또한 자료의 의미 문제를 해결해야 하는 부분 즉, 데이터 소스들의 내용 및 표현을 실세계의 개체와 개념에 연관시키는 부분을 고려하여 메타데이터들 간에 존재하는 모든 의미론적 상호운용을 보장하기 위한 연구가 활발히 진행되고 있다. 의미적 상호운용을 해결하기 위한 방법으로 중개 자기반(intermediary-based)[2, 4, 16, 20], 매핑기반(mapping-based)[18], 질의지향(query-oriented)[11] 접근 방법들을 사용한다. 중개자기반 접근법은 에이전트(Agent)[2], 온톨로지(Ontology)[4, 16], 메디에이터(Mediator)[20]와 같은 매개수단이 필요하다. 이러한 매개 수단으로 다양한 정보 자원을 통합하기 위해서는 특정 도메인 지식이나, 매핑 지식 혹은 규칙을 가져야 한다. 대부분의 경우, 표준 어휘를 공유하기 위해 온톨로지를 사용하거나 상호 통신을 위한 프로토콜을 사용한다. 온톨로지 사용의 장점은 서로 다른 데이터베이스가 같은 개념에 대해서 서로 다른 단어나 식별자를 사용할 경우에 이를 해결해주지만, 지식 도메인의 고유한 복잡성으로 독립적이고, 동적이며 이질적인 데이터베이스에서 온톨로지를 유지하기가 어렵다. 그러므로 이 방법은 단지 제한된 응용 도메인에 적용한다. 매핑기반 접근은 의미적으로 관련된 정보 자원을 통합스키마와 원천스키마 간의 매핑을 통해 해결하는 방법이다[18]. 매핑은 속성, 관계, 개체와 같은 스키마 구성요소를 제한하지 않을 뿐만 아니라 도메인과 스키마 구성요소 사이의 매핑도 가능하다. 그러나 특정 스키마나 응용에 독립적으로 설계되지 못하고, 매핑시 상호운용성과 관련된 문제를 해결하기 위해 명시적인 표현이 필요하다.

질의 지향 방법은 논리 기반 언어 혹은 확장된 SQL과 같은 상호운용 언어에 기반한 방법이다[11]. 여러 데이터베이스에 질의를 표현하는 것이 가능하지만 데이터 구조와 의미상의 불일치를 해결하기 위해 데이터와 메타데이터의 범위를 위한 고수준의 표현 방법이 요구된다. 또한 각 원천 데이터베이스의 하부 구조를 이해하고 있어야 하며, 의미적 불일치를 구별하기 어렵기 때문에, 사용자는 의미적 불일치를 찾고 해결해야만 하는 문제가 있다. 의미적 불일치는 데이터 단계와 스키마 단계에서 발생한다. 데이터 단계의 불일치는 데이터 값(value)과 표현(representation), 정확성(precision)의 불일치에서 발생한다. 분산된 시스템의 통합과 정보공유를 위해서는 시스템적인 통합 메커니즘의 개발 뿐만 아니라, 데이터에 대한 의미(Semantic), 구문(Syntax),

표현(Representation)의 불일치를 해결해야 한다. 정보를 표준화된 업무 절차와 형식으로 공유하기 위한 프레임워크가 제안되고 있다. ISO/IEC JTC1/WG2의 데이터 관리 및 교환(Data Management and Interchange)에서는 정보공유를 위한 데이터 표준화를 지원하기 위한 작업의 일환으로 국제표준 'ISO/IEC 11179 - 메타데이터 레지스트리'에서 데이터의 의미, 구문, 표현을 표준화할 수 있는 프레임워크를 제시하였다. 메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지, 관리하며, 메타데이터의 명세와 의미의 공유를 목적으로 한다[13]. 현재 여러 국가의 주요 프로젝트에서 메타데이터 레지스트리 시스템을 구현하고 있으며, 메타데이터의 등록을 통해 각 시스템에서 데이터의 호환성을 유지하는 방안을 제시하고 있다. 효율적인 정보의 교환은 제공하는 시스템이 의도하는 것과 동일하게 제공받는 기관이 정보를 같은 의미로 받아들이는 것을 보장하는 환경에서 이루어지며, 쉽게 정보의 위치를 지정하고 검색할 수 있어야 한다. 즉, 정보의 의미와 표현방법 등 정보의 공유를 위해서는 정보를 교환하는 상대방과의 일정한 약속이 있어야 가능하다. 이러한 정보의 공유가 가능하도록 표준화된 의미와 형태를 가진 정보의 요소를 표준 데이터 요소라 하며, 이 표준 데이터 요소는 자동화된 정보처리 시스템에서 사용될 수 있다. 스키마 단계의 불일치는 명명(naming) 불일치, 요소 식별자(entity-identifier) 불일치, 스키마 이질동상(schema-isomorphism), 개념적(generalization) 불일치, 집합적(aggregation) 불일치, 스키마적 불일치(schematic discrepancy)와 같이 개별적으로 설계된 데이터베이스의 논리적 구조나 스키마가 다른 시스템을 통합할 경우 발생한다. 그러므로 의미적, 구조적으로 서로 다른 여러 시스템을 통합하기 위해서는 데이터 수준과 스키마 수준의 상호 운용성을 모두 고려하여 개발해야 한다.

RDF(Resource Description Framework)[3], 스키마 통합(Schema Integration)[15], 지능형 정보통합(Intelligent Integration of Information)[6], 지식 공유(Knowledge Sharing Effort)[8] 등의 연구도 진행되고 있다. RDF는 컴퓨터가 처리할 수 있도록 데이터의 의미를 기술하는 표준을 제공한다. 그러나 클래스 정의와 같은 제약조건 및 규칙을 정의하는 방법을 제공하지 않기 때문에 사용자가 직접 이러한 정보들을 기술해 주어야 한다. Ontolingua[1], XOL[7]과 같은 온톨로지 언어들이 개발되었고, 데이터와 지식을 표현하는데 사용되고 있으며 의미론적 단계에서의 데이터 통합을 가능하게 하는 체계적인 형식을 제공하고 있다.

KRISTAL-2002 시스템을 기반으로 하는 통합 검색 시스템 개발을 위해 앞에서 언급한 여러 방법들의 특징을 이용하고 보완하여 적용하였다. ISO/IEC 11179 메타데이터 레지스트리의 표준 절차에 따라 통합 메타스키마를 등록하고, 기 개발되어 제공되고 있는 각 원천서버의 스키마를 매핑한다. 새로 개발될 원천스키마는 표준화된 메타

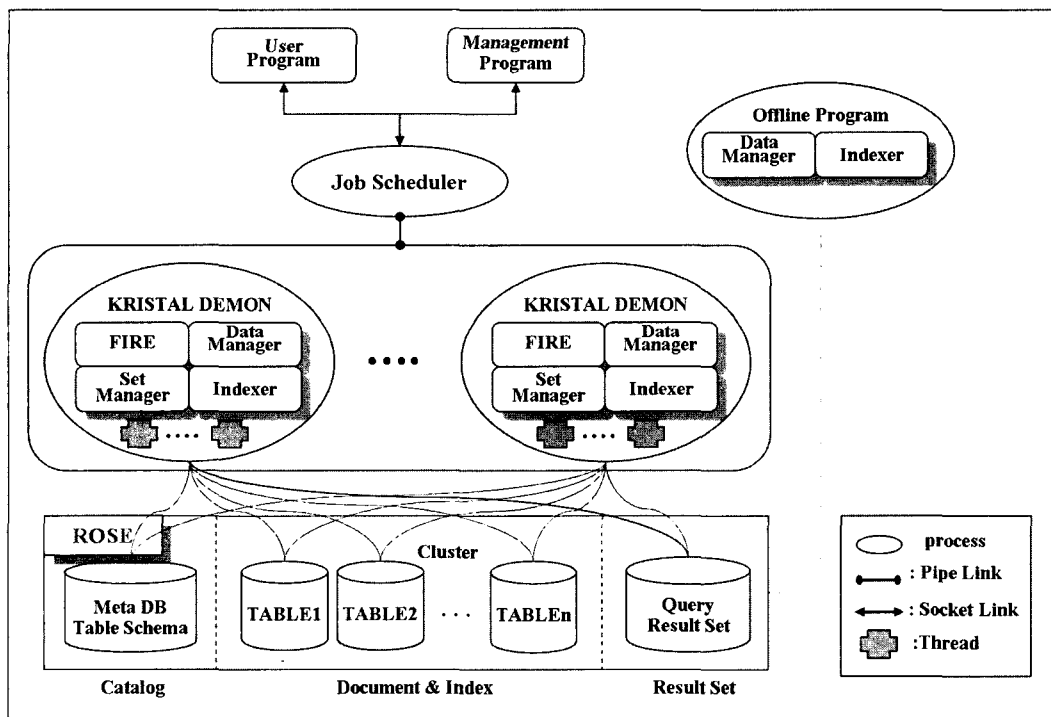
데이터를 적용하도록 유도한다. 매핑시 메타데이터에 의미를 해결할 수 있는 함수정보를 같이 포함시키고, 온톨로지 개념을 적용하여 기존 시스템의 스키마나 구조의 변경 없이 통합시스템에 즉시 반영될 수 있도록 문제를 해결하였다.

3. KRISTAL-2002 개요

KRISTAL-2002는 한국과학기술정보연구원(Korea Institute of Science & Technology Information, KISTI)에서 개발한 정보검색관리 시스템(Information Retrieval & Management System, IRMS)으로, 윈도우 및 유닉스 기반에서 데이터에 대한 저장, 관리, 검색에 관한 처리를 수행한다[10].

KRISTAL-2002는 (그림 1)과 같이 다섯 종류의 프로세스로 구성되어 있으며, 프로세스들은 소켓이나 파이프 통신을 통해 데이터를 전달한다. 사용자 요청 작업을 관리하는 작업관리자(Job Scheduler)가 있으며, 검색관리자(FIRE), 데이터관리자(Data Manager), 셋관리자(Set Manager), 색인

기(Indexer)를 하나의 데몬으로 통합하여 여러 개의 데이터베이스를 관리할 수 있고, 상용 DBMS에서 지원해 왔던 다중 스키마 기반 다중 테이블을 지원한다. 경량의 데몬 구조를 수용함으로써 소규모 최소용량 플랫폼부터 대규모 정보 서비스까지 지원할 수 있다. 일반 문서는 정형 데이터, 비정형 데이터 및 대용량 데이터(BLOB)를 모두 지원하며, 정형색선을 지원함으로써, 데이터베이스의 물리적 공간을 절약하고, 데이터 갱신 시 발생하는 시간을 단축하였다. XML 문서의 구조를 정확하게 파악하여 구조적 정보와 함께 문서를 처리함으로써, XML 데이터 관리 및 검색을 효과적으로 지원한다. 여러 색인 방법을 이용하여, 사용자는 자신이 처리하고자 하는 데이터에 가장 적합한 색인 방식을 선택하고, 색인 결과를 생성함으로써 검색의 정확도와 재현율을 향상하였다. 온라인 데이터 관리 기능 수행시 트랜잭션 처리를 지원하여 데이터관리의 안정성을 보장하며, 핫 백업 방법과 함께 이중으로 데이터의 안정성을 높일 수 있다. 검색에 많은 지식을 갖고 있지 않은 일반 사용자와 전문 검색을 원하는 전문 사용자들을 모두 수용하기 위해서 벡터 검색과 불리언(boolean) 검색을 제공한다.



(그림 1) KRISTAL-2002 구조도

각 모듈별 기능을 살펴보면, 작업관리자(Job Scheduler)는 사용자로부터의 연결 요청을 수락하고 유휴 상태인 검색관리자(FIRE)에게 작업을 분배해주며, 온라인 문서 관리시 데이터관리자(Data Manager)에게 문서 관리를 요청하여 그 결과를 사용자에게 전달한다. 또한, 데이터 변경시 데이터관리자, 검색관리자, 셋 관리자(Set Manager)에게 데

이터베이스의 변경을 알려 각 프로세스가 작업을 수행하게 한다. 검색관리자는 검색을 수행하는 프로세스로, 작업관리자의 서비스 요청에 대한 검색 결과를 사용자에게 전달하며, 셋관리자에게 저장을 요청한다. 셋관리자는 검색관리자의 서비스 요청에 대해 결과 셋 번호와 문서 개수를 넘겨주고, 검색 결과를 저장, 관리한다. 데이터관리자는 온라인

문서 관리를 담당하는 프로세스로서, 작업관리자의 문서 저장 혹은 변경에 대한 요청을 처리한 후 결과를 작업관리자에게 전달하며, 삽입이나 변경이 성공적이면 셋관리자에게 결과 셋을 갱신할 것을 명령한다. 색인기(Indexer)는 색인 시스템의 기능인 한글 형태소 분석, 영어 스테밍, 한자 변환, 불용어 관리 및 사용자 사전 기능을 사용자가 보다 유연하게 활용할 수 있도록 지원한다. 즉, 입력 문서를 분석하여 그 문서를 효과적으로 표현하는 색인어를 추출해 내고, 그 유형에 맞는 분석 방법을 선택하게 한다.

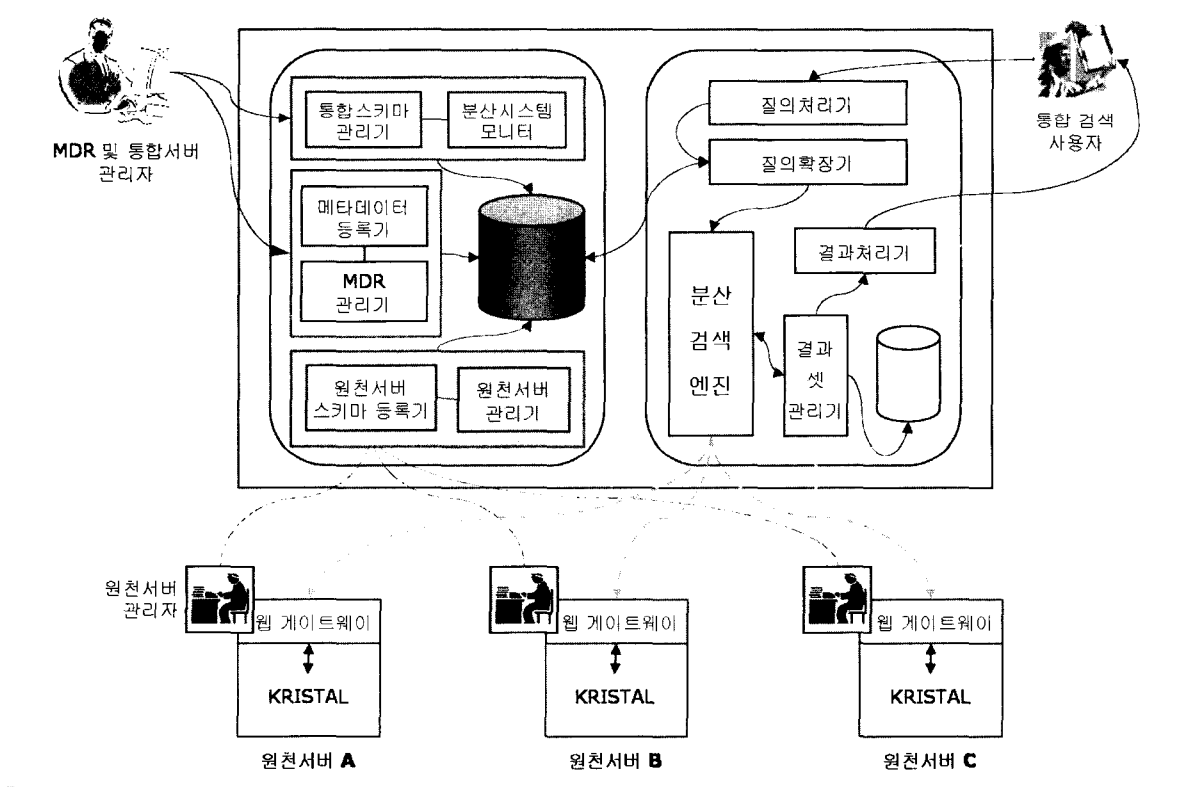
KRISTAL-2002 데이터베이스는 다수 사용자의 동시 검색 및 온라인/오프라인 데이터 관리, 백업 그리고 회복 기능을 안정적으로 수행한다. 또한 검색 엔진의 필수 조건인 대용량 데이터베이스에 대한 빠른 검색속도를 보장하기 위해 멀티쓰래드(multi-thread)를 이용한 분산 검색을 수행한다. (그림 1)의 하단에 도식화된 KRISTAL-2002 데이터베이스의 구조는, 목록, 문서 및 색인 그리고 결과 셋 데이터베이스로 구성하며, 구성된 데이터베이스는 KRISTAL-2002 데이터베이스 시스템이 제공하는 기능을 수행할 수 있는 데이터 관리 서비스의 대상이 된다. 목록 데이터베이스는 테이블 속성에 관한 스키마 정보를 저장한다. 즉, 테이블 구조, 색인 방법, 주키, 불용어 등에 대한 정보가 저장된다. 문서 및 색인 데이터베이스는 하나 또는 여러 개의 클러스터로 구성되며, 각각의 클러스터는 같은 스키마를 사용하는 한 개 이상의 테이블로 구성된다. 클러스터는 검색결과와 랭킹이 적용되는 단위이며, 동시 검색의 대상으로 사용될 수 있다. 문서 및 색

인 데이터베이스의 각 테이블은 문서, 주키, 색인 데이터베이스로 구성되며 목록 데이터베이스에서 정의한 테이블 스키마 구조에 따라 구성된다. 결과 셋 데이터베이스는 검색 질의에 대한 빠른 응답을 위해서 검색된 결과를 질의와 함께 저장하고 관리한다. 그 중에서 문서부분은 KRISTAL-2002 데이터베이스의 최소검색 단위인 색션들의 집합으로 구성되며 원시 데이터에 해당한다. 그리고 주키 부분은 하나 이상의 기본색션으로 구성되며, 구성된 키들은 테이블에서 각 문서들을 구분할 수 있는 키로 사용될 수 있다. 마지막으로 색인 부분은 기본색션과 가상색션에 대한 색인정보, 위치정보 그리고 색인어들에 대한 통계정보를 가지고 있다. 동일한 스키마를 갖는 테이블들은 하나의 클러스터로 관리되며, KRISTAL-2002의 최상위 검색대상이 된다.

4. 분산 통합 정보 검색 시스템의 구조

4.1 시스템 구조

분산 통합 정보검색시스템은 분산된 KRISTAL-2002 기반의 원천시스템을 메타데이터 정보를 이용하여 개발하였다. 원천서버에서 사용되는 모든 스키마는 메타데이터와 반드시 매핑되어야 하며, 메타데이터 레지스트리에 등록되지 않은 스키마는 통합 검색에 사용될 수 없다. 각 분야별 원천서버와 메타데이터 등록, 통합서버와 이를 이용한 분산 통합 검색에 대한 전체 시스템의 구조는 (그림 2)와 같다.

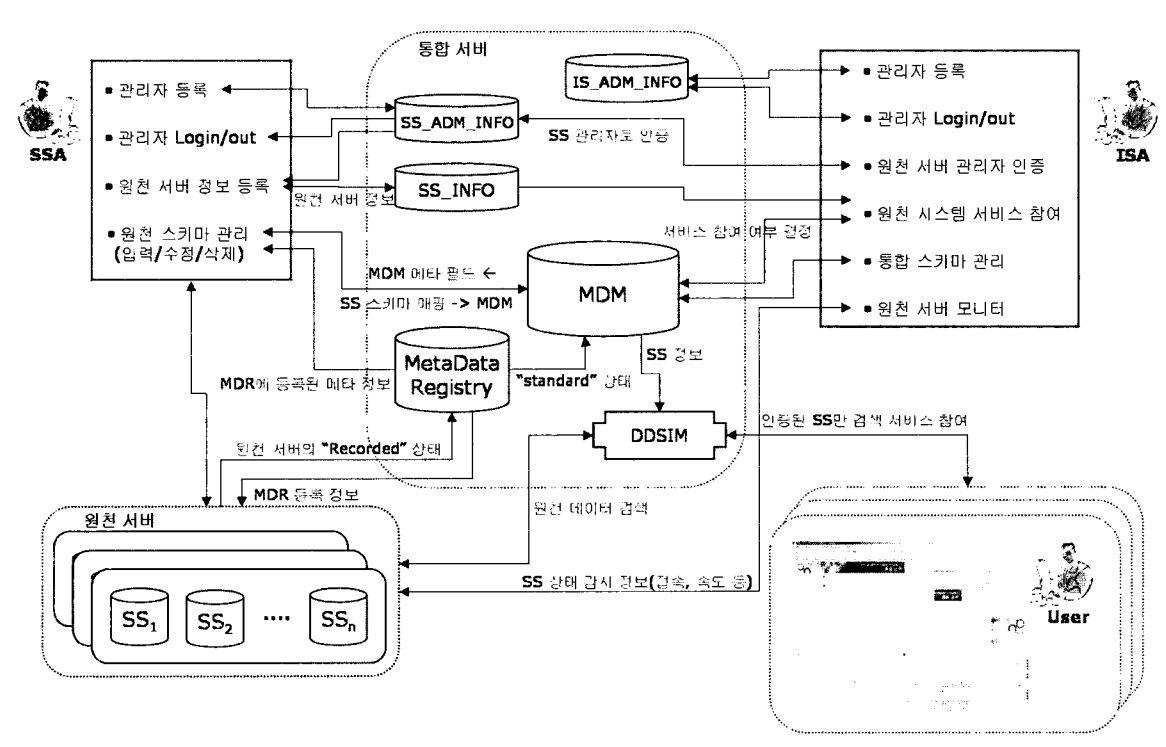


(그림 2) 분산 통합 검색 시스템 흐름도

각 원천서버의 관리자는 원천서버 및 데이터베이스의 구조 정보를 관리한다. 메타데이터 관리 위원은 데이터 요소의 표준화를 유도하고, 통합 관리자는 검색 시스템의 모든 정보를 관리 감독하여 최종 사용자의 검색을 효율적으로 지원한다.

원천서버와 통합서버와의 연계과정은 (그림 3)과 같다. 각 원천서버의 관리자가 관리자정보를 등록하고, 통합서버 관리자가 등록된 원천서버관리자를 인증한다. 인증을 받은 원천서버관리자가 시스템에 접속하여 관리할 원천서버를 등록하고, 등록된 원천서버를 통합서버관리자가 인증한다.

인증된 원천서버에 대해 각 원천서버관리자는 통합스키마와 논리적, 의미적으로 상이한 구조를 일치시키기 위해 통합 메타데이터와 스키마 매핑을 시행한다. 이때 앞서 언급한 메타데이터 레지스트리, 스키마 매핑 및 온톨로지 개념이 적용되어 매핑이 이루어진다. 각 원천서버에서 수행한 스키마 매핑을 통합서버관리자가 인증하여 정보 검색 사용자에게 통합검색 시스템을 제공한다. 원천서버관리자와 통합서버관리자는 관리하고 있는 원천서버에 대해 문제가 있다고 판단되는 경우 서비스 참여를 중지시킬 수 있다. 각 모듈별 기능은 다음 단락에서 상세히 기술한다.



(그림 3) 원천서버와 통합서버의 연계 과정

4.2 원천서버 관리기(Source Server Manager)

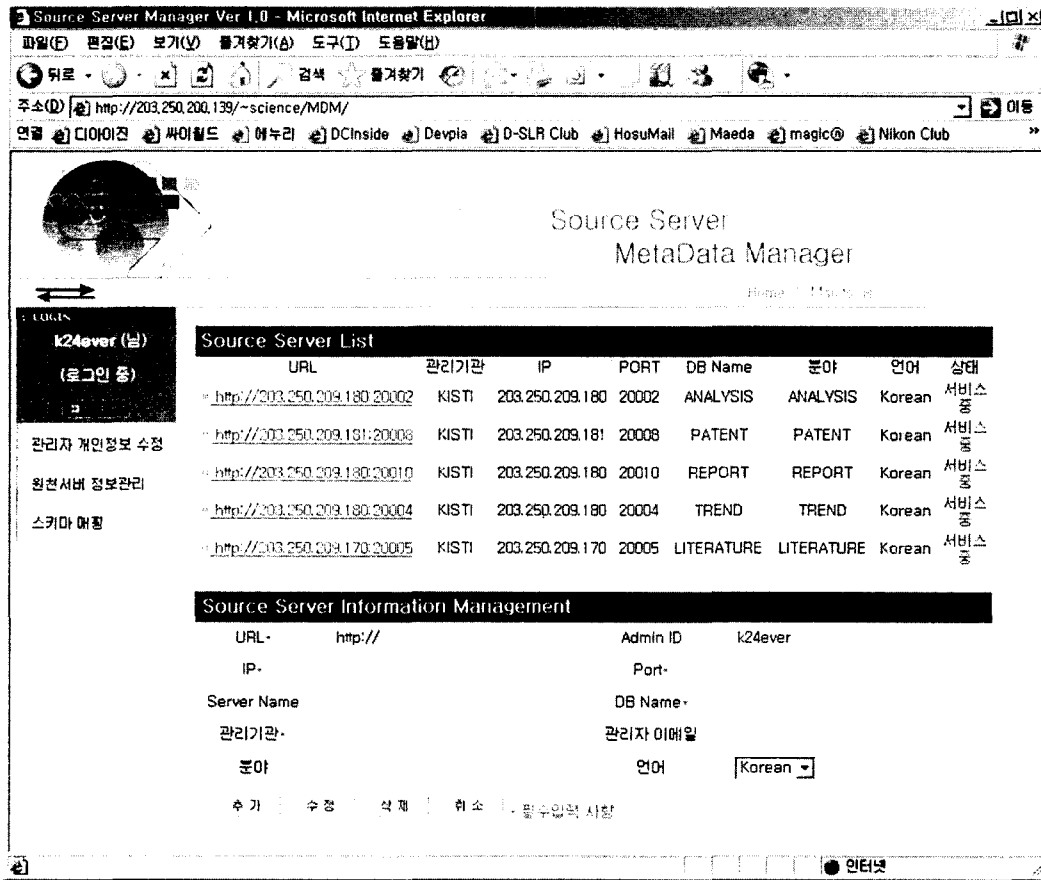
원천서버는 소관 지식정보자원의 수집, 구축 및 관리와 통합서버가 제공하는 검색결과에 대한 원문정보를 제공하는 서버이다. 전문화된 검색서비스 제공을 위해 원천서버 자체적으로 검색시스템을 구축, 운영할 수 있으며, 각 서버를 관리하는 원천서버관리자를 두어 서비스에 참여할 원천서버의 정보, 해당 서버의 스키마 정보를 관리하도록 한다. 즉, 웹 인터페이스를 통해 원천서버에 접근이 가능한 주소와 포트 등 일반 정보를 입력하고, 통합서버의 메타데이터 레지스트리에 접속하여 원천서버의 데이터베이스 검색 필드에 대한 구조 및 정의를 기입하며, 통합서버에 각 원천서버의 스키마와 통합서버의 스키마를 매핑시켜 통합검색이 가능하도록 한다. 원천서버의 정보 변경, 스키마 자체 내용의 변경 혹은 통합서버의 메타데이터와의 매핑에 변경이

발생할 경우에는 수시로 재등록하여, 변경된 내용이 즉시 통합검색에 반영되도록 한다.

4.2.1 원천서버 정보 관리

원천서버 정보 관리 화면인 (그림 4)의 하단에 있는 원천서버 정보 관리 입력 창에 원천서버의 URL, IP, Port, 원천 서버의 이름, 데이터베이스 이름, 정보 분야 등 원천서버를 구별할 수 있는 정보를 입력하여 시스템에 등록한다.

신규로 등록된 원천서버는 통합서버관리자가 인증하고, 인증이 되기 이전에는 원천서버 스키마 매핑을 할 수 없는 '인증대기' 상태가 되며, 통합서버관리자의 인증 후 '서비스중' 상태로 수정된다. '서비스중'인 원천서버에 대해서만이 스키마 매핑 작업을 진행할 수 있다. 원천서



(그림 4) 원천서버 정보 관리

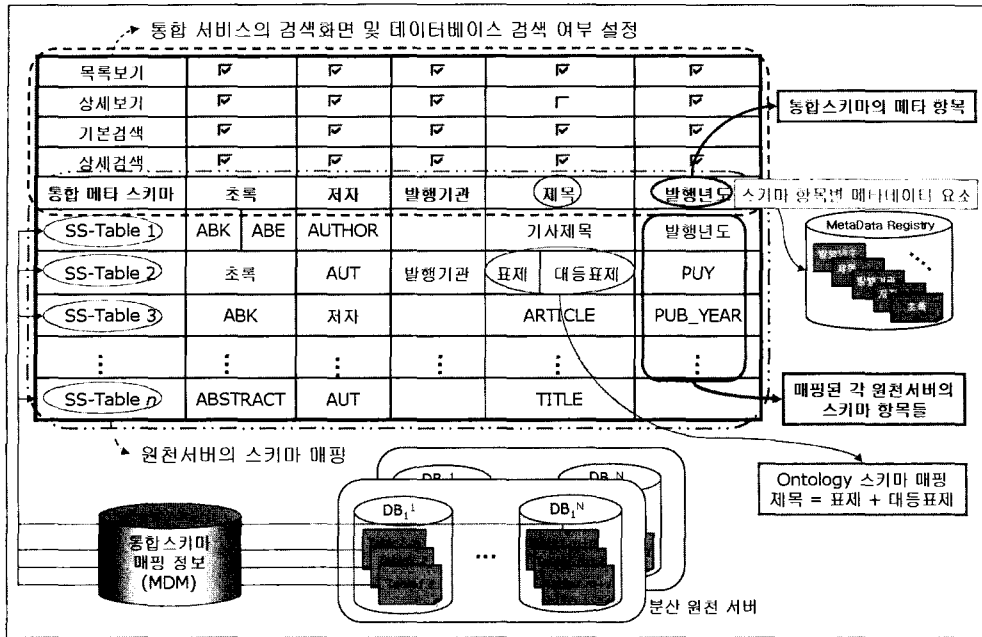
버관리자는 등록된 원천서버의 정보를 수정하거나 더 이상 관리하지 않는 서버의 정보를 수정 또는 삭제 할 수 있다. (그림 4)의 화면에서 이미 등록된 원천 서버 정보 목록으로부터 URL을 클릭하면 해당 서버에 관한 정보 화면이 하단에 나타나고, 원천 서버에 관한 변경된 정보를 수정할 수 있으며, '삭제' 버튼을 눌러 해당 원천 서버를 더 이상 통합 검색 서비스에 참여하지 못하게 할 수 있다.

원천서버의 상태는 통합서버관리자가 서비스에 참여시키기 위해 인증한 '서비스중'인 상태와 서비스에 새로이 참여시키기 위해 원천서버 정보를 인증하지 않은 '인증대기' 상태로 나눈다. 인증한 서버의 경우, '서비스중'과 '일시중지' 상태로 나누어지며, 인증하지 않은 경우 '인증대기' 상태로만 표시된다. '서비스중'인 상태는 원천서버를 통합 서비스에 참여시킨 상태이다. '일시중지' 상태는 자신이 관리하는 원천서버에 이상이 있거나 정보 변경 작업 등으로 인해 통합 서비스에 일시적으로 참여 시키지 않은 상태이며, 필요시 원천서버관리자가 '서비스중'인 서버를 '일시중지' 상태로 변경할 수 있다. 또한, '인증대기' 상태는 통합서버관리자에 의해 통합 서비스에 참여되지 않는 대기 상태를 나타내며, 인증을 거치지 않았으므로 원천서버관리자 임의로 '서비스중'이나 '일시중지' 상태로 변경할 수 없다.

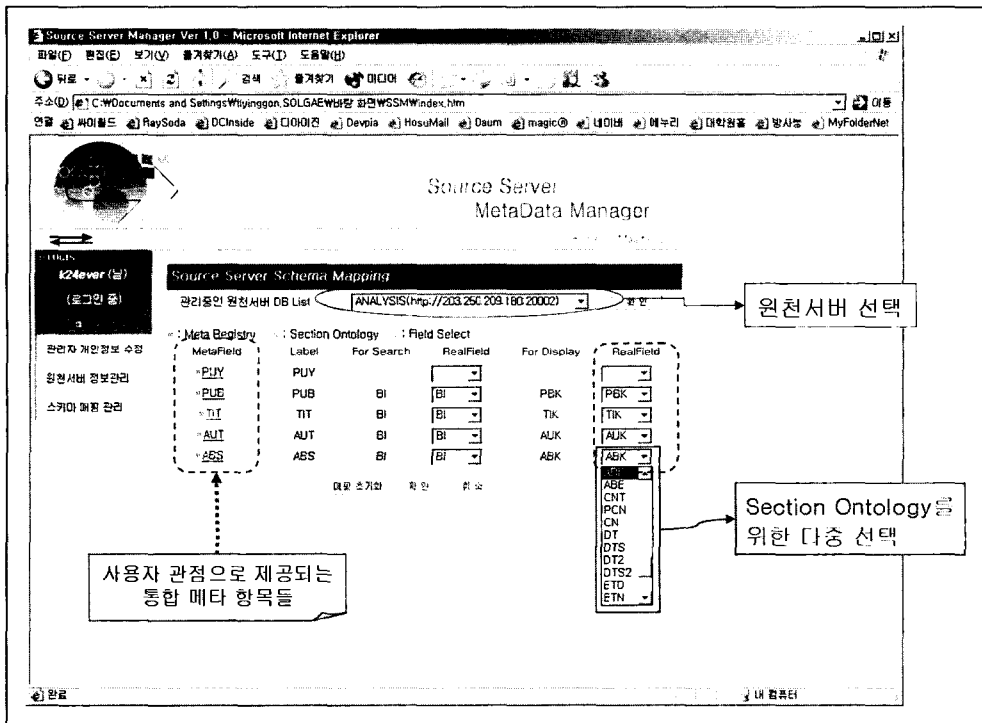
4.2.2 원천서버 스키마 매핑

원천서버의 스키마를 통합 메타 스키마와 매핑하는 개념은 (그림 5)와 같다. 각 원천서버의 구조나 의미의 변경 없이 분산 통합 검색에 참여시키기 위해서는 각 원천스키마를 '통합 메타 스키마'로 매핑시켜야 한다. '통합 메타 스키마'는 메타데이터 레지스트리 등록 절차에 따라 통합 서버가 제공하는 사용자 관점의 스키마로, 사용자에게 각 원천스키마의 정보를 은닉하는 역할을 갖는다. 즉, 사용자는 '통합 메타 스키마'의 항목으로 질의를 발생하며, 시스템 내부적으로 매핑된 원천스키마 항목으로 해당 원천서버에서 검색을 수행하여 결과를 얻어 사용자에게 '통합 메타 스키마'의 항목 이름으로 정보를 제공한다.

원천스키마 매핑을 위해서는 스키마 매핑시에 발생하는 여러 불일치 문제를 해결해야 한다. (그림 5)의 예와 같이, '발행년도' 정보는 각 원천스키마에 '발행년도', 'PUB', 'PUB_YEAR' 등으로 스키마 항목이 정의되어 있으며, 1:1의 매핑 관계가 발생하므로 간단하게 통합스키마의 해당 메타 항목으로 매핑시킨다. 반면 '제목'이 어떤 원천스키마에서는 '표제'와 '대등표제'로 분리되어 있어 1:N의 관계가 성립한다. 이와 같이 원천스키마의 N개의 항목이 통합스키마의 하나의 항목과 매핑되는 경우, 온톨로지 개념을 적용하여 매핑시키고, 해당 원천서버로부터의 '표제'와 '대



(그림 5) 스키마 매핑 개념



(그림 6) 원천서버 스키마 매핑

등표제'의 검색결과는 '제목'으로 합병하여 사용자에게 제공한다.

이러한 개념을 기반으로 원천서버관리자는 (그림 6)의 화면을 통해 관리 중인 원천서버의 스키마 정보를 통합 메타데이터와 매핑하여 통합 서비스에 참여시킨다. 스키마 매핑 화면에서 현재 '서비스중'인 원천서버의 스키마를 선택하면, 각 원천서버에서 이미 매핑된 정보가 있을 경우 매핑 정보

를 보여주며, 변경이 필요할 경우 매핑 정보를 수정한다. '메타필드'에는 메타데이터 레지스트리 등록 절차로 생성된 사용자 관점의 메타 데이터요소, 'Label'은 검색시 제공되는 각 메타 항목의 이름이 표시된다. 각 메타 항목을 선택하여 메타데이터 레지스트리에 저장된 항목별 상세 내역을 조회할 수 있다. 매핑이 끝난 원천 스키마는 통합서버가 관리하는 데이터베이스에 즉시 적재된다.

4.3 통합서버 관리기 (Integrated Server Manager)

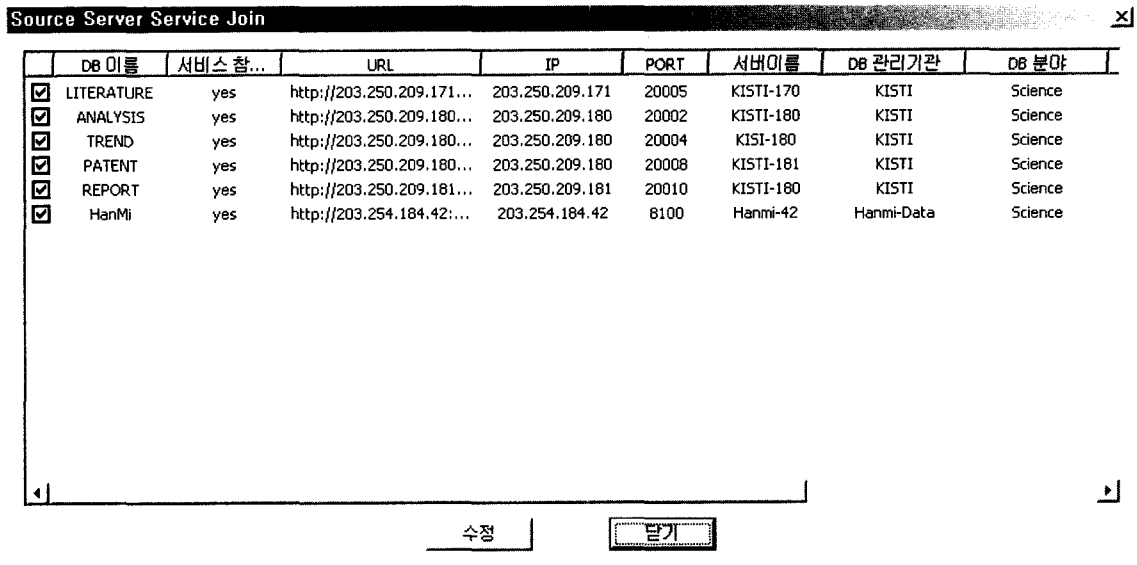
통합서버는 해당 분야별로 특화된 지식정보자원의 효율적 검색과 원활한 유통을 위해 각 원천서버의 정보자원 검색시스템의 통합 및 운영, 메타데이터 데이터베이스 구축 및 관리, 데이터 요소의 공통 표준화 작업을 수행한다. 통합서버는 ISO/IEC 11179에 기반하여 메타데이터를 등록하고 모든 원천서버에서 사용되는 스키마와 매핑하여 분산 통합 정보 검색을 위한 서버 역할을 담당하는 곳으로, 각 원천서버에서 사용되는 스키마와 메타데이터와의 매핑 정보와 각 서버에 대한 정보 및 상태정보를 관리한다. 통합서버에는 시스템을 관리하는 통합서버관리자를 두어 각 원천서버관리자의 인증 및 원천서버를 총괄하여 감독하며 통합검색을 지원한다. 통합서버관리자는 각 원천서버를 통합 서비스에 참여시킬 것인지 여부를 결정하며, 통합 메타데이터 정보 관리 및 통합 서비스에 참여 중인 각 원천서버의 스키마 매핑 상태를 확인한다. 또한 시스템의 상태를 원천서버 모니터를 통해 항상 감시하여 원천서버의 속도가 저하되거나 혹은 멈춰 있을 경우에는 즉시 통합검색시스템의 연결에서 제외시키며,

원천서버관리자가 통합서비스에 참여를 원하지 않을 경우에는 원천서비스 참여를 제거한다.

통합서버 관리기는 원천서버 관리기를 통해 통합 서비스에 참여된 원천서버의 스키마 정보 및 원천서버의 정보를 관리하는 도구이다. 통합서버관리자가 등록된 각 원천서버의 통합 서비스 참여 여부 결정을 할 수 있는 원천서버 서비스 참여 기능, 통합 메타데이터 관리, 통합 서비스에 참여하고 있는 각 원천서버의 매핑 정보를 확인할 수 있는 통합 스키마 관리 기능 및 각 원천서버의 상태를 확인 할 수 있는 분산시스템 감시 기능을 포함한다.

4.3.1 원천서버 서비스 참여

통합서버관리자는 (그림 7)의 화면을 통해 등록된 각 원천서버들의 서비스 참여여부를 결정하여 통합 서비스에 필요한 서버들을 선택한 후 통합 서비스 참여가 이루어지도록 결정한다. 통합 서비스 참여가 승인된 원천 서버는 '서비스 중'인 상태로 원천서버관리자에게 전달되며, 원천서버 스키마 매핑 작업을 수행하여 통합 검색에 참여할 수 있다.



(그림 7) 원천서버 서비스 참여

4.3.2 통합서버 메타데이터 관리기

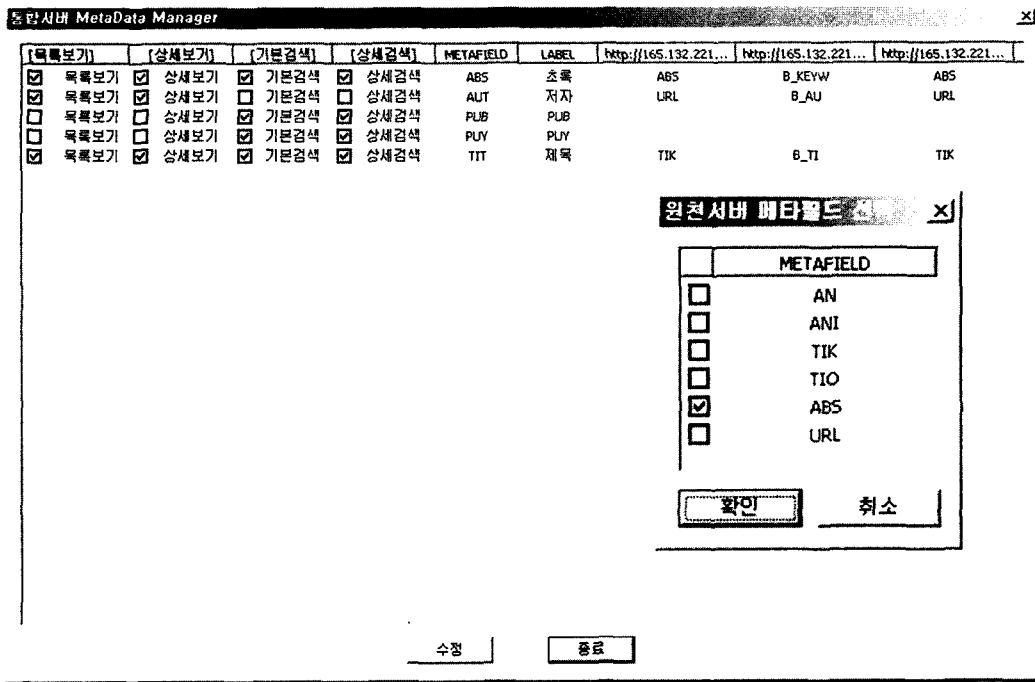
4.2.2절에서 기술된 (그림 5)의 개념은 (그림 8)의 화면으로 제공한다. 사용자의 통합 검색 화면에 나타날 메타데이터 항목의 조건들을 결정하는 요소로는 목록보기, 상세보기, 기본검색, 상세검색이 있으며, 4.4절에서 기술될 분산 통합 정보 검색 관리의 각 검색 화면에 제공되는 항목의 여부를 결정한다.

'목록보기'는 사용자 질의에 적합한 결과를 선택된 각 원천서버에서 검색하여 각 원천 서버의 스키마 필드를 목록 보기 화면으로 제공할 것인지 여부를 정의하는 것으로, 체크 박스에서 선택되지 않은 메타 항목은 검색 목록 보

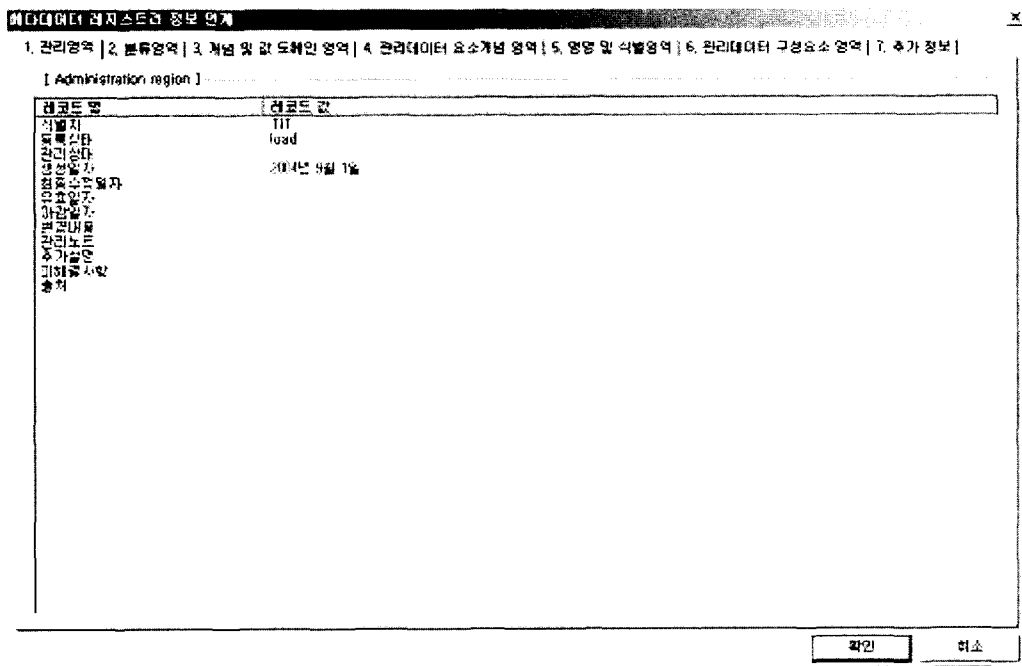
기 항목에 제공되지 않는다는 의미이다. '상세보기'는 목록 보기에 출력된 검색 결과 중 하나를 선택했을 경우 좀더 상세한 정보를 제공하기 위한 항목들을 결정한다. '기본검색'은 간단한 검색 질의어를 입력해서 검색 결과를 얻기 원하는 사용자를 위한 '기본검색 화면'에 제공될 항목을 결정하며, '상세검색'은 여러 검색 요소를 입력할 수 있는 '상세 검색 화면'에 제공될 항목을 결정한다. 'LABEL'에는 검색 화면에서 사용자가 친숙한 이름으로 제공하기 위해 각 메타필드를 재 정의한 레이블 정보를 입력한다. 화면 오른쪽 상단의 URL들은 현재 통합 서비스에 참여하고 있는 원천서버들의 주소이며, 그 아래에 각 원천서버의 메타

필드 매핑 정보를 보여준다. 원천서버 메타필드를 클릭하면 해당 원천서버의 스키마를 읽어 들여 매핑하고자 하는 메타필드가 표시된 팝업창이 나타난다. '원천서버 메타필드 선택' 창에는 해당 원천서버의 데이터베이스 스키마들이 표시되며, 이미 매핑된 스키마 필드는 체크 표시되어 나타난다. 통합서버에서 관리하는 하나의 메타필드에 대해

원천서버의 여러 메타필드가 매핑되는 온톨로지 정보를 위해 원천서버의 여러 메타필드를 동시에 선택할 수 있으며, 다중 선택한 원천서버의 메타필드는 콤마로 구분되어 표시된다. 사용자의 검색 질의어에 대한 결과 검색시 이러한 온톨로지 정보는 합병되어 하나의 항목처럼 검색 질의에 참여한다.



(그림 8) 통합서버 메타데이터 관리 화면



(그림 9) 메타데이터 레지스트리 정보 연계

(그림 8)의 'METAFIELD'를 클릭하면 각 ISO/IEC 11179 메타데이터 레지스트리 등록 절차에 따라 저장된 각 메타항목에 대한 상세 정보를 (그림 9)의 화면에서 검색할 수 있다. 메타데이터 요소는 관리영역, 분류영역, 개념 및 값 도메인영역, 관리데이터요소 개념영역, 명명 및 식별영역, 관리데이터 구성요소영역, 추가정보 등으로 분류하여 상세 정보를 제공한다.

4.3.3 분산 시스템 감시기

분산 시스템 감시기는 (그림 10)과 같이 통합 서비스에 참여하고 있는 각 원천서버 시스템의 상태를 감시할 수 있는 기능으로, 분산된 각 원천 사이트에 대한 정보를 지속적으로 유지하고, 문서 정보에 대한 연속성, 요소 배타 정보에 대한 일관성을 유지한다.

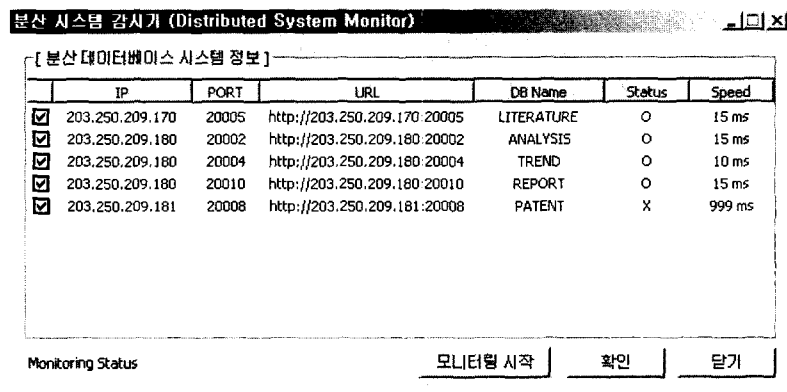
사이트 장애에 대한 신속한 대처를 위해 분산시스템을 감시하고 검색에 참여시킬 서버의 목록들을 체크한다. 'status' 항목에는 정상적으로 시스템이 작동하고 있는 원천서버의 경우 'O'로 표시되며, 원천서버의 검색 엔진이 동작하지 않는 경우나, 해당 원천서버 데이터베이스에 접근 권한이 없어 접근이 불가능한 경우 'X'로 표시가 된다. 'X'

로 표시되는 서버의 경우는 통합 검색에서 사용하지 않도록 좌측의 체크박스를 해제하여 서버를 '일시중지' 상태로 두어 통합 검색에 일시적으로 참여하지 않도록 결정한다.

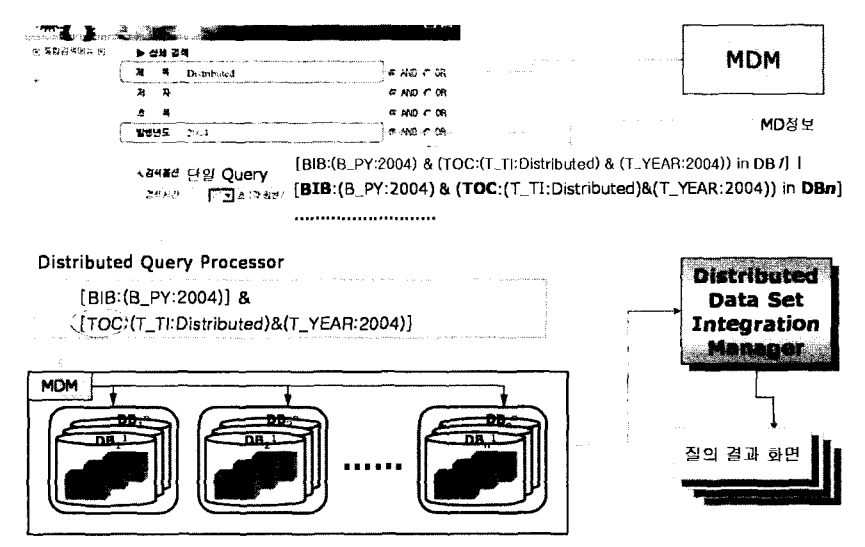
4.4 분산 통합 정보 검색 관리

4.4.1 분산질의 처리기(Distributed Query Processor : DQP)

DQP는 각 원천서버에서 지원하는 질의 형태로 변환하기 위해 통합시스템에 저장된 메타데이터의 스키마 매핑 정보를 참조하여 각 원천서버에 대한 정보 및 데이터베이스 정보를 포함하는 질의를 생성하고, 각 원천서버에 질의를 전달한다. 사용자가 입력한 검색어는 DQP를 통해 불리언/벡터 질의어를 생성하게 된다. 데이터베이스 필드와 검색 화면 항목의 관계는 4.3.2 절의 (그림 8)에서 정의하여 시스템에 저장되며, 각 검색어 항목별로 새로운 질의 생성 모듈을 통해 검색 화면에서 입력받은 항목 값과 데이터베이스상의 검색 항목에 해당하는 필드를 통합스키마로부터 얻은 메타정보를 이용해 대응시킨다. (그림 11)에 DQP와 다른 모듈과의 관계가 표현되어 있다.



(그림 10) 분산 시스템 감시기



(그림 11) 분산질의 처리기와 다른 모듈 사이의 자료 흐름

분산 질의어를 생성하기 위해 원천서버 정보와 테이블 정보를 읽어 들여 검색 화면에 입력된 각각의 항목에 대하여 원천서버의 데이터베이스 구조와 비교한다. 이때 대응하는 항목이 없는 데이터베이스는 검색에서 제외되며, 해당 항목이 있는 데이터베이스에 대해서는 그 검색 값을 필드 이름과 함께 저장한다. 검색 화면의 항목에 해당하는 질의를 각각의 데이터베이스의 필드에 대응시켜 '[데이터베이스 이름 : (필드1:검색어1) op (필드4:검색어4)]'의 형태로 검색 항목을 차례대로 검색하여 질의어를 생성하며, 마지막 항목까지의 질의어 처리를 완료하면 각각의 데이터베이스에 대한 필드에서의 완성된 검색 질의문이 (그림 12)와 같이 생성된다.

4.4.2 분산데이터 셋 통합 관리기(Distributed Data Set Integrated Manger : DDSIM)

DDSIM은 DQP에서 생성된 질의어를 검색 결과 화면에 출력하기 위한 적절한 불리언/벡터 질의어로 생성한다. 즉,

DQP에서 가공된 질의어를 이용해 메타 필드 개수만큼 필드가 존재하는 데이터베이스를 검색하고, 원하는 결과를 추출해 내며, 웹 응용 라이브러리 인터페이스를 이용해 HTML 문서로 변환한다. (그림 13)은 DDSIM과 다른 모듈과의 자료 흐름을 보여준다.

DDSIM에서 원천서버의 주소, 데이터베이스명, 테이블명, 질의어에 부여된 고유 ID를 포함해서 생성된 불리언/벡터 질의어를 HTML 문서로 변환하기 위해 C/C++용 웹 응용 라이브러리 인터페이스인 qDecoder를 이용하며, 생성된 출력력을 모두 HTML 문서로 만들기 위한 초기화를 수행한 후 HTML 헤더를 출력한다. 만약 검색 모드라면 질의어를 생성하기 위해 DQP 모듈을 다시 수행한다. 즉, 데이터베이스 테이블을 초기화하고, 메타필드 개수만큼 필드가 존재하는 원천서버의 데이터베이스에 대해 분산 질의어를 생성한다. 질의어에 해당하는 원천서버의 데이터베이스 개수만큼 각 원천서버의 데이터베이스별로 출력할 항목의 목록을 만들고, 원천서버의 데이터베이스별 분산 질의어로 검색을 수행한다.

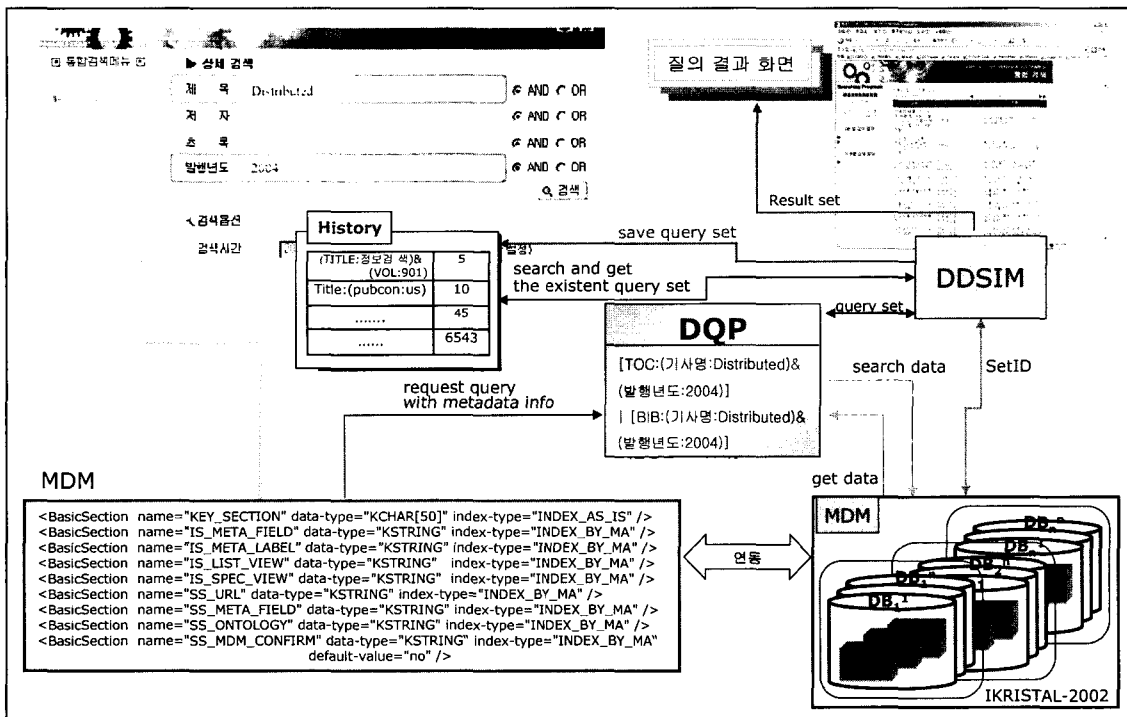
1) AND인 경우

→ [BIB:(B_PY:2004)] & [TOC:(T_TI:Distributed)&(T_YEAR:2004)]

2) OR인 경우

→ [BIB:(B_PY:2004)] | [TOC:(T_TI:Distributed)&(T_YEAR:2004)]

(그림 12) 분산질의 처리기에 의해 생성된 질의문



(그림 13) 분산 데이터 셋 관리기와 각 모듈 사이의 자료 흐름

검색 결과 화면을 생성하기 위한 HTML 코드를 출력한 후, 검색결과 값의 개수를 이용해 페이지 목차를 출력한다. 만약 결과가 생성되었다면 결과 화면의 행 별로 배경색을 출력하며, 페이지에 해당하는 문서 개수와 필드 개수만큼 다음과 같은 절차를 수행 한다. 즉, 하나의 레코드를 읽어서 만약 레코드가 특정 필드라면 상세보기로 이동할 HTML 코드를 출력하지만, 토큰을 처리해야할 레코드일 경우, 구분자를 기준으로 토큰을 처리한 후 토큰 처리된 코드를 출력한다. 반면, 상세보기 모드에서는 문서번호와 테이블번호를 얻어 테이블 번호에 따른 데이터베이스 테이블을 초기화한다. 검색 화면을 생성하기 위한 HTML 코드와 계층구조에 따른 DIV 메뉴용 텍스트를 출력하고, 레코드의 필드 상태에 따른 처리 후 검색 결과를 출력한다.

4.4.3 분산 통합 정보 검색 관리

분산 통합 정보검색 서비스는 (그림 8)에서 정의한 항목을 기반으로 기본검색과 상세검색으로 분류하여 검색 결과를 출력한다. 기본검색은 기본 질의를 모든 메타 항목의 결과와 비교하여 검색을 수행하며, 상세검색은 상세검색 항목으로 정의된 메타 항목으로 좀 더 세분화하여 질의를 수행할 수 있는 기능이다.

4.4.3.1 검색 기능

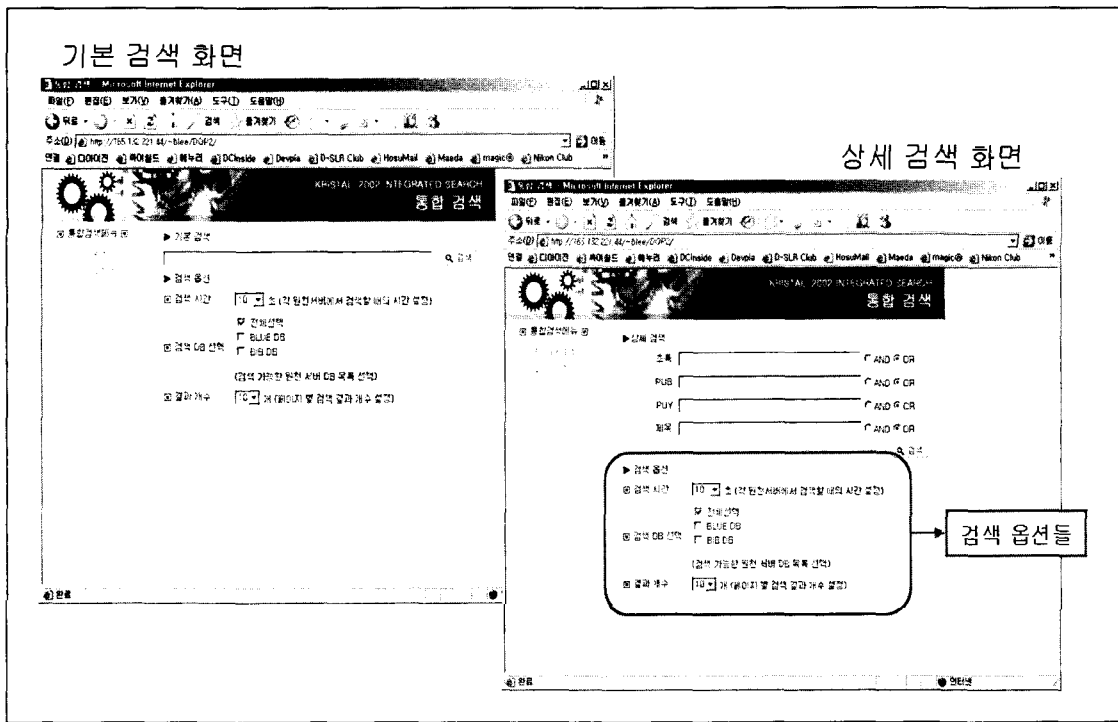
기본검색은 (그림 14)의 좌측과 같이 질의를 입력하는 단일창에 질의어를 입력하여 검색을 수행한다. 사용자는 검색 질의어와 검색 대상이 되는 원천서버의 데이터베이스, 화면당 검색 결과 개수, 원천서버로부터의 검색 응답 시간을 설정한다. 사용자가 입력한 질의는 (그림 8)의 통합서버 메타

데이터 관리 화면에서 설정한 '기본검색'에서 선택하여 정의한 메타항목과 1:1 대응하여 검색을 실시할 수 있도록 각 원천서버의 데이터베이스의 형태에 알맞게 생성되어 각 원천서버로 보내진다. '검색 시간'은 결과를 얻을 때까지의 최대 소요 시간을 제한할 수 있는 기능으로, 사용자 임의로 각 원천서버에 질의를 보내 검색 응답을 받는 시간을 설정할 수 있다. '검색 DB 선택'은 검색 가능한 원천서버 데이터베이스 목록을 체크 박스에서 선택하고, '결과 개수'는 한 페이지 당 검색 결과를 몇 개 출력할 것인지 설정하는 항목이다.

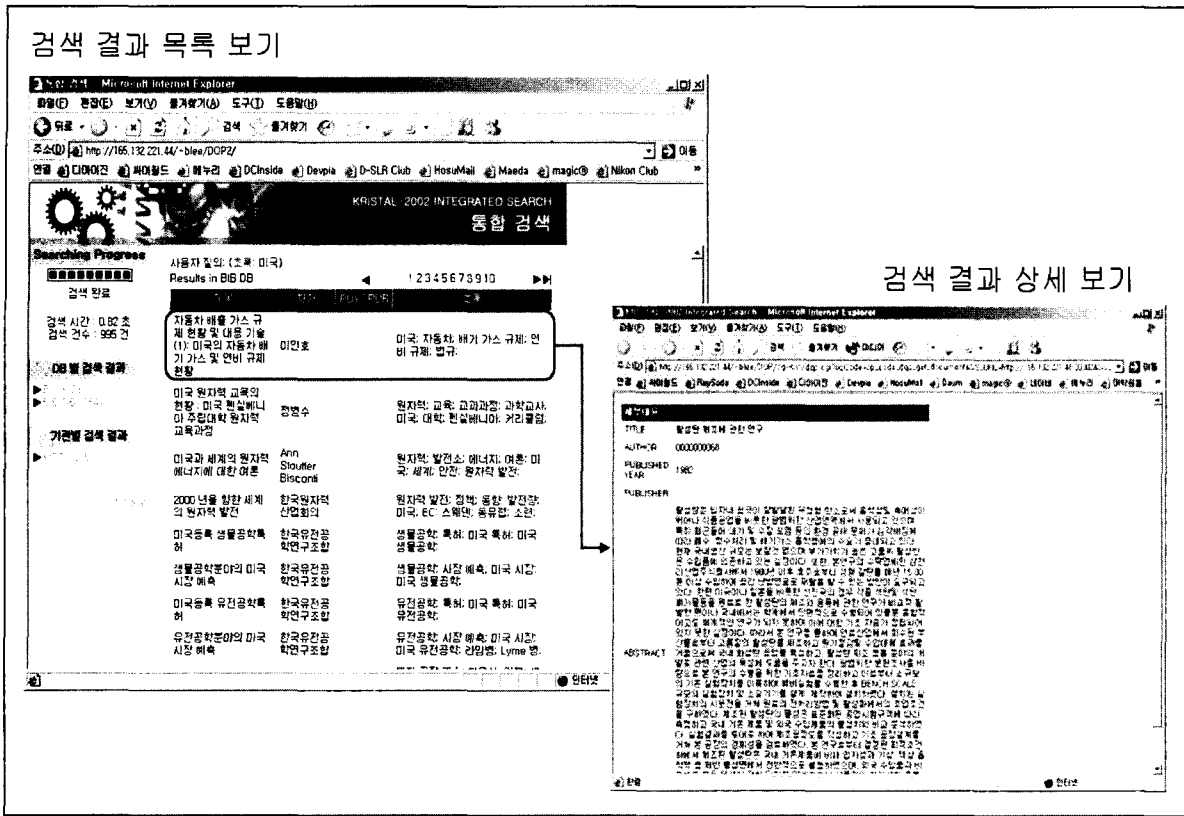
상세검색은 기본검색보다 검색 항목을 좀 더 세분화하여 상세하게 검색 질의를 생성할 수 있는 기능을 제공하며, (그림 14)의 우측과 같다. 상세검색 항목들은 통합 관리자가 필요할 때마다 (그림 8)의 통합 메타 스키마 관리를 통해 설정 및 해제할 수 있어, 동적으로 검색 항목이 제공된다. 또한 검색의 효율을 높이기 위해 AND, OR 연산자를 추가하여 질의에 포함시킬 수 있다. 사용자가 입력한 질의는 (그림 8)의 '상세검색'에서 선택된 각 원천서버의 메타 필드와 1:1 대응하여 검색을 실시할 수 있도록 질의어가 생성된다.

4.4.3.2 검색 결과 출력

검색결과는 목록 보기와 상세 보기로 구분한다. 검색 결과 목록 보기는 사용자가 기본검색 또는 상세검색에서 생성한 질의를 검색 옵션에서 선택한 원천서버 데이터베이스로부터 검색하여 생성된 검색 결과를 (그림 15)의 좌측과 같이 간결한 목록으로 보여주는 화면이다.



(그림 14) 검색 질의 입력 화면



(그림 15) 검색 결과 화면

(그림 15)의 화면에 나타난 검색 결과 또한 (그림 8)의 '목록보기'와 '상세보기'가 선택된 메타필드만 검색 대상이 되며, 사용자는 검색 프로세스 바를 통해 현재 검색이 진행되는 상황을 확인할 수 있다. 검색이 완료되면 총 검색 시간 및 원천서버의 데이터베이스 별 총 검색 건수와 관리 기관별 총 검색 건수를 보여준다. 'DB별 검색 결과' 목록에 출력된 원천서버의 데이터베이스 이름을 선택하면 해당 원천서버의 데이터베이스별 검색 결과의 목록을, '기관별 검색 결과' 목록에 출력된 기관명을 선택하면 해당 기관별 검색 결과의 목록을 출력한다.

검색 결과 상세 보기는 (그림 15)의 좌측 목록 보기 화면에 출력된 검색 결과 중, 사용자가 선택한 특정 검색 결과를 상세하게 보여주는 결과 화면으로 (그림 15)의 우측과 같다.

5. 결 론

본 연구에서는 KRISTAL-2002 기반에서 기개발되어 서비스를 제공하고 있는 원천서버 스키마를 메타정보를 이용하여 통합 메타데이터와 매핑하는 방법으로 분산 통합검색시스템을 개발하였다. 이 시스템은 현재 구축되어 있는 각 원천스키마를 변경하지 않을 뿐만 아니라 별도의 프로그래밍 작업 없이도 분산된 시스템을 통합할 수 있다. 정보가 제공되고 있는 실시간 중에 간단한 스키마

매핑 작업만으로 시스템 통합이 이루어지며, 시스템 분리 작업 또한 화면상의 단순한 선택 여부로 결정할 수 있다. 사용자는 데이터의 내용 및 시스템의 위치에 관계없이 원하는 데이터를 통합하거나 부분적으로 검색할 수 있다.

KRISTAL-2002 시스템에 추가로 원천서버 관리기, 통합서버 관리기, 메타데이터 관리기, 분산질의 처리기, 분산데이터 셋 통합 관리기 등을 개발하였으며, 통합적인 정보 검색을 지원하기 위한 분산 통합 정보검색시스템을 개발하였다.

향후 과제로서는 본 연구를 바탕으로 KRISTAL-2002 기반에서 개발되지 않은, 데이터베이스의 종류에 관계없이 모든 종류의 데이터베이스 시스템을 통합 검색에 참여시키는 연구가 필요하다. 본 논문에서는 이러한 확장될 시스템 개발을 KRISTAL-2002 기반 시스템을 통합하여 제공함으로써 가능성을 제시하였다.

참 고 문 헌

[1] A. Farquhar, R. Fikes and J. Rice. The Ontolingua Server : A Tool for Collaborative Ontology Construction. International Journal of Human-Computer Studies, 46, pp.707-728, 1997.
 [2] A.M Ouksel, A framework for a scalable agent archi-

ecture of cooperating heterogeneous knowledge sources. In *Intelligent Information Agents : Agent-Based Information Discovery and Management on the Internet*, M. Klusch, Ed. Springer, Berlin, Germany, pp.100-124, 1999.

[3] D. Brickley, R. Guha, eds. Resource Description Framework(RDF) Schema Specification. W3C Proposed Recommendation. March, 1999, <http://www.w3.org/1999/TR/PR-rdf-schema>

[4] D. Calvanese, G. Giacomo, and M. Lenzerini. Ontology of Integration and Integration of Ontologies. Proc. of the 2001 Description Logic Workshop(DL 2001), 2001.

[5] Hsinchun Chen, Haiyan Fan, Michael Chau, and Daniel Zeng, MetaSpider : Meta-Searching and Categorization on the Web, *Journal of the American Society for Information Science and Technology*, 52(13), pp.1134-1147, 2001.

[6] Intelligent Integration of Information. <http://mole.dcs.isx.com/13>.

[7] Jonathan Robie, Joe Lapp, David Schach, XML Query Language(XQL), <http://www.w3.org/TandS/QL/QL98/pp/xql.html>.

[8] Knowledge Sharing Effort. <http://www-ksl.stanford.edu/knowledge-sharing>.

[9] Kobayashi M. and Takeda, K. Information retrieval on the web, IBM Research Report, RT0347, April, 2000.

[10] KRISTAL-2002 User's Manual V.1[1].1.5.pdf, Group for Intelligent Information System, KISTI, 2002.

[11] Lakshmanan L. V. S., Sadri, F., Subramanian, I. N., Logic and algebraic languages for interoperability in multidatabase systems. *J. Logic Programm.* 33, 2, pp.101-149, 1997.

[12] Luis Gravano, Chen-Chuan K. Chang, Hector Garcia Molina, STARTS : Stanford Proposal for Internet Meta Searching, In *Proceedings of the 1997, ACM SIGMOD Conference*, 1997.

[13] Metadata Registry, <http://metadata-stds.org/11179/>

[14] M. Robert, Losee, Lewis Church Jr., *Information Retrieval with Distributed Databases : Analytic Models of Performance*, IEEE Transactions on Parallel and Distributed Systems, Vol.14, No.12, December, 2003.

[15] Nishizawa Itaru, Takasu Atsuhiko, Adachi Jun. A Schema Integration and Query Processing Information Retrieval. *IPSJ SIGNotes DataBase System Abstract*, No.094-009, 1993.

[16] Ontology, <http://www.w3.org/2001/sw/WebOnt/>

[17] Oren Zamir and Oren Etzioni, Web Document Clustering : A Feasibility Demonstration, SIGIR 1998,

Melbourne, Australia, 1998.

[18] Prasad, B. E., Reddy, P. G., Gupta, A., A methodology for integration of heterogeneous databases. *IEEE Trans. Knowl. Data Eng.* 6, 6, pp.920-933, 1994.

[20] Soumen Chakrabarti, Martin van den Berg and Byron Dom. Focused crawling : a new approach to topic-specific Web resource discovery, Elsevier Science B. V. 1999.

[21] Yannis Papanikolaou, Hector Garcia-Molina, Jeffrey Ullman, Medmaker : A mediation system based on declarative specifications. In *Proceedings of the 12th IEEE International Conference on Data Engineering(New Orleans, LA, Feb. 26-March 1)*, S. Y. W. Su, Ed. pp.132-141, 1996.

[22] Z39.50 Gateway, <http://www.loc.gov/z3950/gateway.html>

[23] ANSI/NISO Z39.50 Revision, January 2002, <http://www.loc.gov/z3950/agency/revision/revision.html>

[24] 분산 통합 프로토콜 (k-protocol v.1) : 프로토콜스펙 및 스키마 문서 정의, <http://giis.kisti.re.kr/download/k-protocol-03-02-14.pdf>

최 귀 자



e-mail : gjchoe@hosu.yonsei.ac.kr
 1991년 서울산업대학교(학사)
 2005년 연세대학교 대학원 전산학과(석사)
 2005년~현재 연세대학교 대학원 전산학과 박사과정
 관심분야 : 프로그래밍언어, 소프트웨어공학, XML, 메타데이터, 정보검색

김 재 곤



e-mail : k24ever@hosu.yonsei.ac.kr
 2002년 연세대학교(학사)
 2005년 연세대학교 대학원 전산학과(석사)
 2005년~현재 연세대학교 대학원 전산학과 석사과정
 관심분야 : XML, 메타데이터, S/W 컴포넌트 기술, 웹 서비스 기술

서 정 현



e-mail : jerry@kisti.re.kr
 1987년 한양대학교 수학과(학사)
 1987년~1993년 시스템공학연구소
 1990년 IBM Scientific Center 초청연구원
 1993년~2000년 연구개발정보센터 실장
 1998년 North Western 대학 초청연구원
 2001년~현재 한국과학기술정보연구원 정보시스템부
 관심분야 : 정보검색



조 한 형

e-mail : eosludis@hotmail.com
2004년 연세대학교(학사)
2005년~현재 연세대학교 대학원 전산학과 석사과정
관심분야 : XML, 메타데이터, 시멘틱 웹



이 민 호

e-mail : cokeman@kisti.re.kr
1998년 충남대학교 컴퓨터과학과(학사)
2000년 충남대학교 컴퓨터과학과 전산학전공(석사)
2000년~2001년 데이콤 중앙연구소 연구원
2001년~현재 한국과학기술정보연구원 정보시스템부

관심분야 : 분산 통합검색, 디지털 콘텐츠 보안, 메타데이터 레지스트리(MDR),



정 창 후

e-mail : chjeong@kisti.re.kr
1999년 충남대학교 컴퓨터과학과(학사)
2002년 충남대학교 컴퓨터과학과(석사)
2003년~현재 한국과학기술정보연구원 정보시스템부
관심분야 : 정보검색 및 추출, 디지털 도서관, 메타데이터 레지스트리(MDR)



박 동 인

e-mail : dipark@kisti.re.kr
1979년 서강대학교 전자공학과(공학사)
1979년~1994년 시스템공학연구소 인공지능연구부 실장
1992년~1994년 폰트개발보급센터 센터장 (겸임)

1994년~1996년 국어공학센터 센터장
1996년~1998년 시스템공학연구소 자연어처리연구부 부장
1999년 에스엔엘(주) 대표이사
2000년 L&H 음성언어기술연구소 소장
2002년 나눔정보테크(주) 기술담당사장
2002년~2003년 리눅스원(주) 대표이사
2004년~현재 한국과학기술정보연구원 정보시스템부 부장
2002년~현재 한국어 정보처리학회 이사
관심분야 : 인공지능, 자연어처리



남 영 광

e-mail : yknam@dragon.yonsei.ac.kr
1978년 연세대학교 수학과(학사)
1985년 한국과학기술원 전산학과(석사)
1992년 Northwestern University 전산학과 (박사)
1993년~1994년 시스템공학연구소 선임연구원

1995년~현재 연세대학교 전산학과 교수
관심분야 : 프로그래밍언어, 소프트웨어공학, 정보검색, XML