

국가지식정보통합검색시스템 설계 및 구현

김재원⁰ 최승기 송명원 윤병남
 한국전산원 국가정보화센터
 {kimjw⁰, choe_ski, smw, yoonbnam}@nca.or.kr

Design And Implementaton of Korea National Knowledge-Information Integrated Retrieval System

Kim Jae-Won⁰ Choi Seung-Ki Song Myung-Won Yoon Byeong-Nam
 National Computerization Agency, Center for National Informatization

요 약

세계는 지금 지식정보의 창출 및 활용정도가 국가경쟁력을 좌우하는 지식기반 사회로 급속히 전환 중이며 우리나라는 지식정보의 생산보다는 상대적으로 이를 공유하고 활용하기 위한 기반이 취약한 것으로 평가됨에 따라 Cyber Korea 21(1999. 4)과 지식정보연계활용체제구축사업의 기본계획(1999. 12)에서 전략분야의 DB 확충 및 이들 DB를 연계활용 할 수 있는 통합검색시스템 구축이 제시되었다. 이에 본 논문에서는 국가지식정보통합검색시스템의 추진현황과 기술적인 구현방안 등에 대하여 논하며, 국가지식정보통합검색시스템의 향후 발전방안을 제시하고자 한다.

1. 서 론

한국전산원은 2000년 1월 지식자원관리법의 제정으로 인하여 지식사업의 전담기관으로 지정¹⁾된 이래 지속적으로 지식정보를 구축해왔다. 이에 각각의 사업으로 구축된 지식자원에 대한 통합적인 검색요구가 국가지식정보통합검색시스템 구축을 위한 정보화전략계획수립²⁾을 통해 가시화되었으며, 또한 국가지식정보자원관리센터, 종합정보센터, 전문정보센터 지정을 통해 지식정보자원공유와 유통을 촉진하기 위한 조직체계가 2001년 3월의 2차 지식정보자원관리위원회 의결을 거쳐 확립되었다. 이에 따라 한국전산원이 주관기관이 되어 1단계 국가지식정보통합검색시스템(<http://www.knowledge.go.kr>)을 2001년 8월에 구축·완료하였으며, 이를 통해 정부의 공공 기관에서 운영중인 분산된 국가지식정보 중 331개 기관이 구축·운영하고 있는 9천635만건의 지식정보를 통합 검색할 수 있게 되었다.

통합검색시스템의 구축의 의미는 과학기술, 교육학술, 문화, 역사, 정보통신 등 국가 보유의 5대 지식정보를 하나의 DB처럼 검색 활용할 수 있다는 점과 윈스톱 검색시스템 구현으로 국가 지식정보에 대한 접근성이 강화되었다는 것을 들 수 있다.

2. 제 1단계 국가지식정보통합검색시스템 구축 현황

2.1 국가지식정보자원관리사업 현황

[표 1]은 현재까지 구축을 완료하였거나 구축 중인 32개의 국가지식정보관리사업의 현황을 보여준다.

[표 1] 국가지식정보자원관리사업 현황

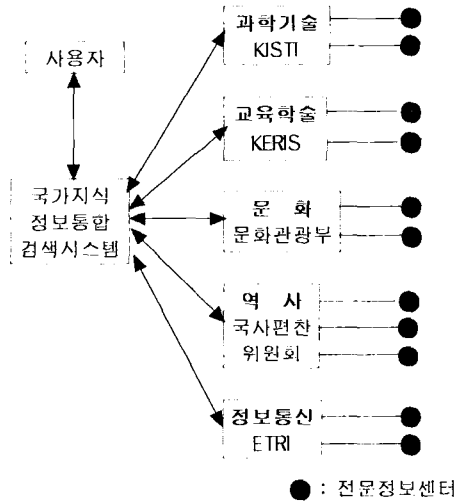
과 제 명	주관기관
한국역사정보통합시스템 구축	서울대학교 규장각
유교문화 종합 정보 DB 구축	한국국학진흥원
학술 논문DB 구축사업	KERS
정보통신통합정보시스템 구축	ETRI
국가문화유산 종합정보시스템	문화관광부
정부출연연구기관 지식정보 DB구축	산업연구원
한국사회과학종합 DB구축	한국사회과학데이터센터
국가통합검색시스템 구축을 위한 ISP 수립	한국전산원
문화예술종합정보시스템	문화관광부
과학기술종합정보시스템 구축	KISTI
해양수산종합정보시스템 구축	해양수산부
한국독립운동사 영문DB구축	독립기념관
한국영화자료 DB구축	한국영상자료원
국가지식정보통합검색시스템	한국전산원
정보통신통합정보시스템 구축	ETRI
유교문화종합정보DB구축	한국국학진흥원
국가문화유산종합정보시스템구축	문화관광부
한국역사정보통합시스템구축사업	국사편찬위원회
한국역사정보통합시스템ISP수립	국사편찬위원회
국가생물종 지식정보 시스템	산림청
해양수산학술연구정보망	해양수산부

[표 1] 국가지식정보자원관리사업 현황(계속)

과 제 명	주관기관
국가학술연구 DB구축사업	KERIS
국가학술연구 DB구축 ISP 수립	KERIS
한국학 고전원문 디지털화사업	서울대 규장각
건설교통기술 지식정보DB구축 사업	한국건설기술연구원
문화예술종합정보시스템구축 사업	문화관광부
문화예술종합정보 ISP 수립	문화관광부
한국여성사지식정보시스템 구축	한국여성개발원
한국독립운동사 종합지식정보시스템 구축	독립기념관
과학기술 문헌 및 산업기술기반 전문 정보 DB구축	KISTI
과학기술 종합정보시스템 정보화전략계획 (ISP)수립	KISTI
정부출연 연구기관 지식정보DB 구축	산업연구원

이렇게 다양하고도 방대한 정보에 대한 효과적인 검색 방안 및 관리기법에 대한 요구로 인하여 국가지식정보통합검색시스템의 구축이 추진되었다.

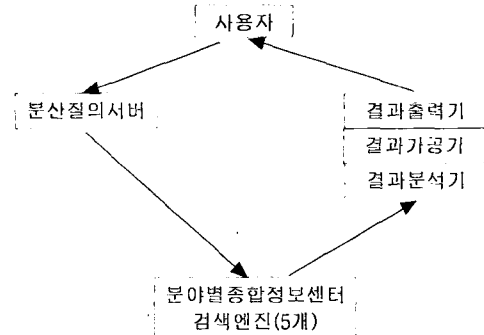
2.2 구축시스템의 개요



[그림 1] 국가지식정보통합검색시스템 개념도

[그림 1]은 국가지식정보통합검색시스템에 연계되어 있는 각 기관을 보여준다. 5개의 분야별 종합정보센터와 300여개의 전문정보센터가 국가지식정보통합검색시스템을 구성하고 있으며, 2단계의 분산검색방법을 이용하여 사용자의 질의를 처리하고 있다. 국가지식정보통합검색시스템의 검색엔진의 특징은 분산검색방식과 메타검색방식을 들 수 있다.

2.1.1 분산검색방식(Distributed Retrieval System)



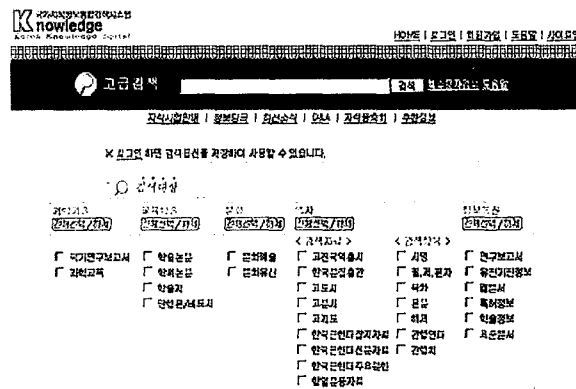
[그림 2] 국가지식정보통합검색시스템 분산검색

국가지식정보통합검색시스템의 검색 메커니즘은 먼저 사용자의 검색요청을 통합검색시스템의 분산질의 서버가 받아서 5개 분야의 종합정보시스템에 있는 검색 어플리케이션에 검색을 요청한다. 검색 어플리케이션은 자신이 가지고 있는 검색엔진을 통하여 검색을 수행하여 검색 결과를 통합검색시스템에 전달하고, 통합검색시스템의 결과분석기, 결과가공기, 결과출력기 등의 처리를 거쳐 최종적으로 사용자에게 그 결과가 전달되게 된다.

2.2.2 메타검색방식(Meta Search Agent)

국가지식정보통합검색시스템은 자체적인 DB와 인덱스 그리고 Robot을 보유하지 않고 5개의 종합정보센터내에 있는 검색엔진을 이용한 체계이다. 따라서 본 시스템에서는 5개의 종합정보센터의 검색엔진의 특성과 Query 스타일에 대한 분석이 실시되었다.

2.2.3 비 표준화된 확장검색 방식



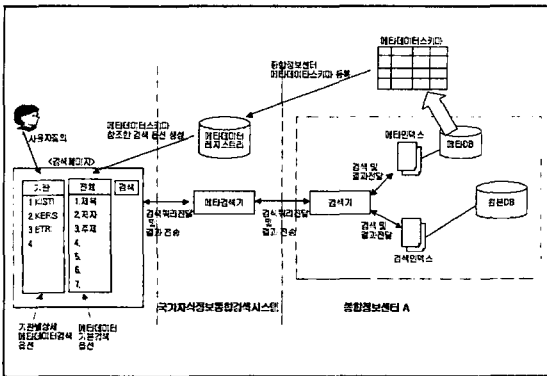
[그림 3] 국가지식정보통합검색시스템 확장검색

국가지식정보통합검색시스템은 각 종합정보센터에서 보유중인 DB의 필드의 특성에 맞는 확장검색방식을 지원하여 제작되었다. 최초 본 시스템이 구축될 시점에서

는 국가적인 차원의 메타데이터에 대한 표준안이 제시되지 않았기 때문에 각 기관에서도 자체적으로 필요한 필드나 메타데이터 자체적으로 구축하였기 때문에 표준화된 인터페이스를 제공하기는 불가능했다. [그림 3]는 현재 운영중인 국가지식정보통합검색시스템의 확장검색 인터페이스를 보여준다.

3. 제 2단계 국가지식정보통합검색시스템 구축 계획

3.1 메타데이터 레지스트리 시스템(Metadata Registry System) 구축



[그림4] 메타데이터 레지스트리를 이용한 검색 흐름도

2000년도 이전의 지식사업에서는 메타데이터에 대한 국가 표준안이 존재하지 않아 각 기관별로 필요에 따라 임의적으로 구축하였으나, 이후 한국전산원이 주축이 되어 마련된 표준안에 따라 7개의 핵심 메타데이터 요소 (Title, Creator, Subject, Publisher, Date, Type, Identifier)는 DB 구축시 필수적으로 동시에 구축하게 되어있으며, 각 기관의 특성에 따라 필수적으로 구축해야 하는 메타데이터 항목에 대해서는 확장자(Qualifier)를 정의할 수 있도록 허용하였다. 이에 국가지식정보통합검색시스템 2단계사업에서는 각 기관에서 임의적으로 구축한 메타데이터의 스키마에 대한 각종 정보의 관리를 통해 효율적인 검색과 정보유통을 지원하기 위하여 메타데이터 레지스트리 시스템을 구축하고자 한다.

3.2 연계기관 추가 연계

현재 연계·운영중인 기관 이외에 해양수산부, 한국건설기술연구원 등 17개 기관의 20개 과제의 산출물에 대한 연계를 추진하고자 한다.

3.3 네트워크 연결 유지 에이전트 개발

분산검색 및 메타검색 엔진의 특성상 사용자의 요구사항에 대하여 실시간으로 반응하여야 하기 때문에 사용자의 검색요구가 발생할 때마다 각 연계기관 시스템과의 연결을 시도하는 네트워크상의 load가 발생한다. 이의 overhead를 제거하기 위하여 가상 전용망 개념의 연결 유지 에이전트를 개발하여 적용하고자 한다.

3.4 단일 로그인 에이전트 개발

국가지식정보통합검색시스템은 현재 5개 종합정보센터가 연계·운영중이지만, 2002년도에 5개의 기관이 추가로 연결될 예정이다. 이에 국가지식정보통합검색시스템에서는 사용자의 편의를 위하여 단일 로그인 에이전트를 개발하고자 한다. 이는 한 시스템의 로그인 정보가 연계된 모든 시스템에 동일하게 적용되어 별도의 로그 프로세스 없이 다른 시스템에 로그인 할 수 있으므로 이를 통하여 각 연계 시스템의 연계성 및 통합성을 높일 수 있다.

4. 결론 및 향후 연구과제

국가지식정보통합검색시스템은 메타데이터 레지스트리 시스템의 구축으로 일관성있는 메타데이터를 이용한 검색이 가능할 수 있으며, 네트워크 연결유지 에이전트로 인하여 검색속도의 향상을 기대할 수도 있게 되었다. 또한 지식자원의 추가적인 연계를 통하여 지식정보 제공의 확대를 도모할 수 있게 되었다. 향후의 과제로는 모바일 환경에 대한 적극적인 지원과 프리서치 에이전트 등의 도입을 통한 시스템의 안정화를 이루어야 하겠다.

5. 참고 문헌

- [1] 지식정보자원관리법(2000. 1. 28), 시행령(2000. 8. 5), 시행규칙(2000. 8. 19)
- [2] 한국전산원, 국가지식정보 통합검색시스템을 위한 정보화전략계획(ISP) 수립, 2000. 8. 3
- [3] 한국전산원, 2000년도 지식정보자원관리사업 "국가지식정보통합검색시스템" 완료 보고서, 2001. 9. 29
- [4] 김재원. 사용자지향 문서검색 에이전트에 관한 연구, 1998. 6
- [5] William B. Franke, Richardo Baeza-Yates, Information retrieval : Data Structure & Algorithms, Prentice Hall