

해양 전자정보자원 메타 데이터베이스 시스템 설계 및 구현방안에 관한 연구

A Study on Planning & Implementation of the Meta Database System for Ocean Electronic Resources

한 종 엽(Jong-Yup Han)*

〈 목 차 〉

- | | |
|--------------------------|-------------------------------------|
| 1. 서론 | 4. 해양분야 메타 데이터베이스 시스템
설계 및 구현 방안 |
| 2. 웹자원 메타 데이터베이스 연구동향 | |
| 3. 해양조사자료 메타 데이터베이스 연구동향 | 5. 결론 |

초 록

해양 전자정보자원 메타 데이터베이스 시스템 설계 및 구현을 위해 국내의 관련 연구를 조사하고 분석하였다. 연구대상은 해양분야 웹자원에서 해양조사자료까지를 범위로 하였다. 본 연구의 목적은 네트워크 자원의 기술에 적합한 더블린코어를 기반으로 효율적인 해양분야 전자정보자원 정보검색서비스를 제공하는데 있다. 본 논문에서는 해양분야 전자정보원 조사, 메타데이터 기술요소 분석, 메타데이터 분류체계, 시스템 구성 및 검색 구현방안의 연구를 수행하였다.

주제어 : 해양 전자정보자원, 메타 데이터베이스 시스템, 웹자원, 해양조사자료, 더블린코어,
메타데이터 분류체계

Abstract

A literature analysis for the planning and realization of meta database system was carried out to establish the ocean electronic resources, the first in Korea. The study targeted from web resources and to oceanographic survey data. The focus of the analysis lies in the providing practical information retrieval service for ocean electronic resources based on the framework of effective Dublin Core metadata with network resources description. The analyses included ocean electronic resources, metadata descriptive elements, metadata classification, system organization, and retrieval for planning and implementation of meta database system.

Key Words : Ocean Electronic Resources, Meta Database System, Web Resources,
Oceanographic Survey Data, Dublin Core, Metadata Classification

* 한국해양연구원 문헌정보팀장 (jyhan@kordi.re.kr · http://library.kordi.re.kr)
· 접수일 : 2002. 5. 13 · 최초심사일 : 2002. 6. 4 · 최종심사일 : 2002. 6. 8

1. 서론

정보기술 및 네트워크의 발전으로 인해 정보자원의 범위는 기존의 인쇄매체를 포함하여 전자적 자원으로까지 그 범위가 확대되고 있는 실정이다. 또한 전자적 자원의 급격한 증가와 유통은 새로운 정보환경의 패러다임을 형성하고 있으며, 이에 따라 전자정보자원의 효율적인 조직과 검색활용의 필요성이 보다 중요시되고 있다. 특히 자료수집에 많은 비용과 노력이 투입되는 해양분야 정보원에 대한 학술정보공유 시스템의 구현은 국가적으로도 매우 시급한 일이다.

메타데이터는 데이터에 대한 구조화된 데이터로 조직화되지 않은 정보자원을 체계적으로 기술하기 위한 수단이라고 할 수 있다. 현재 웹 상에서의 데이터는 텍스트뿐만 아니라 이미지, 동영상, 음성정보 등에 이르기까지 그 범위가 다양하다. 또한 해양조사자료도 연구해역의 방대함과 자료의 종류별 특성이 매우 복잡하여 정형화의 어려움이 있다. 이러한 다양한 데이터에 대한 요소를 합리적으로 추출하여 데이터에 대한 정보를 체계화하는 기술 요소를 메타데이터라고 할 수 있는데, 메타데이터의 역할은 검색을 용이하게 할 뿐만 아니라 검색된 자원에 대한 효율적인 관리와 이용을 위한 통제수단 및 다양한 프로토콜에서 상호호환을 가능하게 한다. 또한 서로 다른 메타데이터 형식간에는 기술요소의 범위나 종류, 어의, 요소명 등의 차이로 인하여 기술 요소의 식별이나 정보의 공유를 위해서는 형식 상호간 연계성 확보가 필수적이다.¹⁾

본 연구에서는 네트워크자원을 기술하는데 적합한 더블린 코어를 기반으로 해양분야의 전자정보자원을 체계화하여 한 차원 진화된 정보검색서비스의 제공을 목표로 한다. 본 연구의 대상인 해양분야 전자정보자원이란 웹자원과 더불어 해양에서의 조사활동으로 획득하여 전자적으로 저장된 해양조사자료까지를 범주로 한다. 이를 위해 먼저 국내외 관련 연구동향을 살펴보고, 해양 연구분야의 자료 특성에 부합하고 효율적인 전문주제분야 전자정보자원에 대한 메타데이터 시스템 설계 및 구현방안을 제시하고자 한다.

1) 김이경 등, "메타데이터 연계성을 위한 RDF 응용스키마 설계에 관한 연구", 《정보관리학회지》 제17권, 제1호(2000), p. 22

2. 웹자원 메타 데이터베이스 연구동향

2.1 정보통신분야 메타 데이터베이스 구축

정보통신 메타데이터베이스는 1999년도에 한국데이터베이스진흥센터(DPC)에서 국내외 정보통신 분야의 웹자원을 대상으로 메타데이터 요소를 적용하여 구축한 정보검색 시스템이다. 이는 인터넷으로 유통되고 있는 방대한 정보자원을 체계적으로 기술할 수 있는 방안이 모색되고 있는 시점에서 정보통신 분야에서 처음으로 시도된 메타 데이터베이스라고 할 수 있다. 정보통신 메타데이터베이스의 데이터 요소의 내용은 다음의 <표 1>과 같다.

<표 1> 정보통신 메타데이터베이스 데이터 요소²⁾

번호	항 목 (Qualifier)	입력 구분	한정요소 (Sub-element)	작성방식 (Scheme)	타 형식과의 매핑관계	
					Dublin Core	GILS
1	제목 (Title)	M	Main Alternative	Internal AACR2	Title	Title
2	저자 (Author)	O	Person	Internal	Author or Creator	Originator
3	주제 (Subject)	M	Keyword	Controlled List LCSH, UDC, DDC	Subject and Keywords	Uncontrolled Term
4	내용요약 (Description)	M	Brief free text Abstract	Internal	Description	Abstract
5	제공기관 (Provider)	M	Organization Corporate	Internal, USMARC	Publisher	Distributor
6	협력자 (Contributor)	O	Name, Address	Internal, USMARC	Other Contributors	Contributor
7	발행일 (Date)	O	Created, Issued Modified, Valid	ANSI X3.30-1985	Date	Date of Publication
8	자료유형 (Type)	M	Data Type	Controlled List DC Object Types	Resource Type	Medium
9	제공형태 (Format)	M	Medium	Controlled List MIME	Format	Available Linkage Type
10	정보위치 (Locator)	M	Site Directory Document	URL	Resource Identifier	Available Linkage

2) 한국데이터베이스진흥센터, 『정보통신메타DB구축사업 완료보고서』. 서울 : 한국데이터베이스진흥센터, 1999. p. 39.

4 한국도서관·정보학회지 (제 33권 제 2호)

번호	항 목 (Qualifier)	입력 구분	한정요소 (Sub-element)	작성방식 (Scheme)	타 형식과의 매핑관계	
					Dublin Core	GILS
11	정보원 (Source)	O	X	URL, URN, ISBN, ISSN	Source	Sources of Data
12	제공언어 (Language)	M	Main Language Multilingual	Controlled List Z39.53, ISO 639	Language	Language of Resource
13	관련정보 (Relation)	O	X	X	Relation	Cross Reference
14	취급범위 (Coverage)	O	Reference Period, Place	Free text LCSH, ANSI X3.30-1985	Coverage	Supplemental Information
15	권리 (Rights)	O	X	Free text (copyright Info.)	Rights Management	Use Constraints

※ 입력구분 : M(Mandatory) = 필수항목, O(Optional) = 선택항목

정보통신 메타데이터베이스의 데이터 요소는 기본적으로 더블린코어 요소에서 크게 벗어나지 않는 형태로 구성되어 있다. 더블린코어의 15개 요소와 그 숫자가 동일하며 creator를 author로 publisher를 provider로 identifier를 locator로 대치한 것이 다른 점이다. 그러나 의미상으로 데이터 내용에는 차이가 없다. 단 요소 기입시 필수적으로 기입해야 하는 요소와 선택적으로 기입해야 하는 요소가 고정적으로 규정되어 있고 같은 요소를 반복하여 사용할 수 없다. 이것은 선택, 반복적으로 요소를 사용할 수 있는 더블린코어와 다른 점이다.

정보통신 메타데이터베이스의 분류체계는 정보 이용자 관점에서 웹 디렉토리 주제분류, 산업 분류체계 및 도서관자료 분류체계 등을 참고하여 작성되었으며 데이터 양, 형태의 다양성, 가독성, 유지 관리의 용이성 등을 고려하여 설계되었다. 최상위 메뉴는 5개 영역과 8개의 주류(class)로 배열하였으며, 주류 아래에는 중분류와 세부분류가 나타나는 각 영역별 3-4 단계의 메뉴 구조로 구성되어 있다. 또한 주제별 접근방식 뿐만 아니라 제공주체별, 정보유형별 분류방식으로 배열하여 디렉토리 서비스가 가능하도록 배려하였다.³⁾ 메타데이터베이스 분류체계는 다음의 <표 2>와 같다.

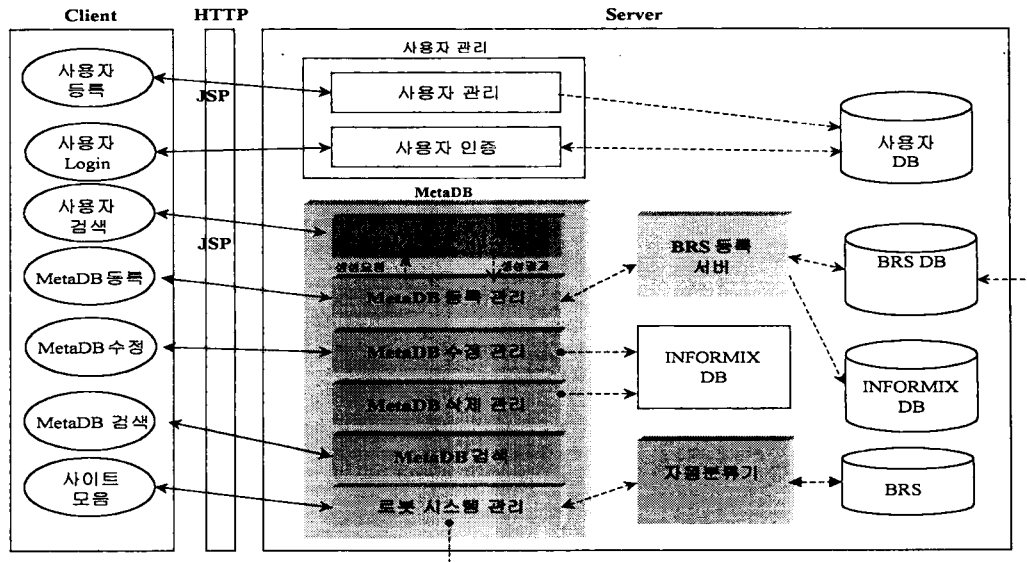
3) "정보통신 메타데이터베이스 분류체계 구성".

Available from WWW : <<http://www.metadb.net/intro/intro3.htm>> [cited 2001-12-06]

<표 2> 정보통신 메타데이터베이스 분류체계

분류번호	대분류	중분류
1000	인터넷/웹	인터넷/웹-인터넷 정책, e-Business, 인터넷 서비스, 월드와이드웹, 법률, 정치/행정/경영, 교육/취업, 엔터테인먼트, 언론/대중매체, 사이버 커뮤니티, 사이버 예절
2000	통신/네트워크	통신/네트워크-프로토콜, 네트워크 관리, 유선통신, 무선통신, 정보 보호, 정보가전, 서비스제공업체, 장비/기기, 통신이론
3000	컴퓨터과학	컴퓨터과학-컴퓨터일반, 컴퓨터 시스템 구성, 데이터, 전산이론, 정보이론, 전산수학, 정보시스템, 인공지능
4000	소프트웨어	소프트웨어-소프트웨어 일반, 프로그래밍, 소프트웨어공학, DB구조, 프로그래밍 언어, 웹언어, 웹브라우저, 운영체제, 유틸리티, 쉘어웨어
5000	하드웨어	하드웨어-컴퓨터제품, 통제구조/마이크로프로그래밍, 계산/논리구조, 메모리 구조, 투입/산출/데이터커뮤니케이션, 레지스터/트랜스퍼/레벨수행, 논리설계, 집적회로

또한, 정보통신 메타데이터베이스 시스템은 이용자 인증관리 기능이 있으며, Informix DB 와 BRS 검색엔진을 사용하고 있다. 정보통신 메타데이터베이스는 로봇시스템을 통하여 웹정보를 수집하며 자동분류기를 통해 웹자원의 분류가 이루어진다. 시스템 구성은 다음의 <그림 1>과 같다.



<그림 1> 정보통신 메타데이터베이스시스템 구성도

정보통신분야 메타데이터베이스는 메타데이터 형식을 기본으로 적용한 전문적인 데이터베이스로써 국내 최초로 구축되었다는데 의의가 있다. 이로 인해 다른 전문 주제분야에서 메타데이터 요소를 이용한 시스템 구축을 촉진하는 역할을 하고 있다. 해양분야 메타데이터베이스 시스템의 구현방안에 있어서도 이 분야의 메타데이터 요소 적용사례는 많은 시사점을 제시한다.

2.2 연구장비분야 웹자원 메타 데이터베이스 구축

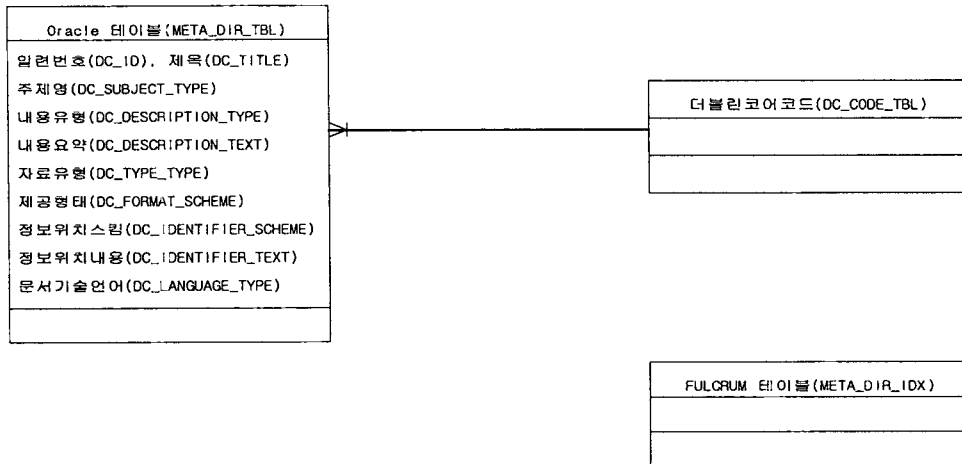
연구장비분야 웹자원 메타데이터베이스 구축은 2001년도에 한국기초과학지원연구원(KBSI)에서 한국과학기술정보연구원(KISTI)과 공동으로 연구사업을 수행하였다.⁴⁾

이 연구는 연구장비분야 웹자원을 기술용 메타데이터(descriptive metadata)로 데이터베이스화하고, 메타데이터베이스의 정보검색 및 관리시스템을 개발하였다. 메타데이터베이스 구축은 국제적 메타데이터 표준안인 더블린코어(Dublin Core)를 준용하여 가능한 상세하게 웹 자원에 대하여 DB화하였다. 메타데이터베이스의 저장과 관리는 Oracle RDBMS를 사용하였으며, 상용의 검색엔진인 Fulcrum Search Server를 활용하였다. 이 시스템에서 관리자 또는 운영자는 웹상에서 정보를 신규등록, 수정, 삭제 등의 정보관리업무를 수행한다.

메타데이터베이스 구축은 먼저 국내외의 연구장비분야 웹자원의 메타데이터는 기술용 메타데이터로 구분하고, 국제적인 표준화가 진행되고 있는 더블린코어(Dublin Core)를 준용하여 원시 정보를 직관적으로 파악하고 활용할 수 있도록 제목, URL, 요약문, 제공기관, 제공주체, 정보유형, 자료형태, 제공언어 등을 데이터 요소로 하여 메타데이터를 작성하였다. 이외의 세부적인 연구장비 메타데이터 항목은 <표 1>의 정보통신 메타데이터베이스의 항목과 동일하다.

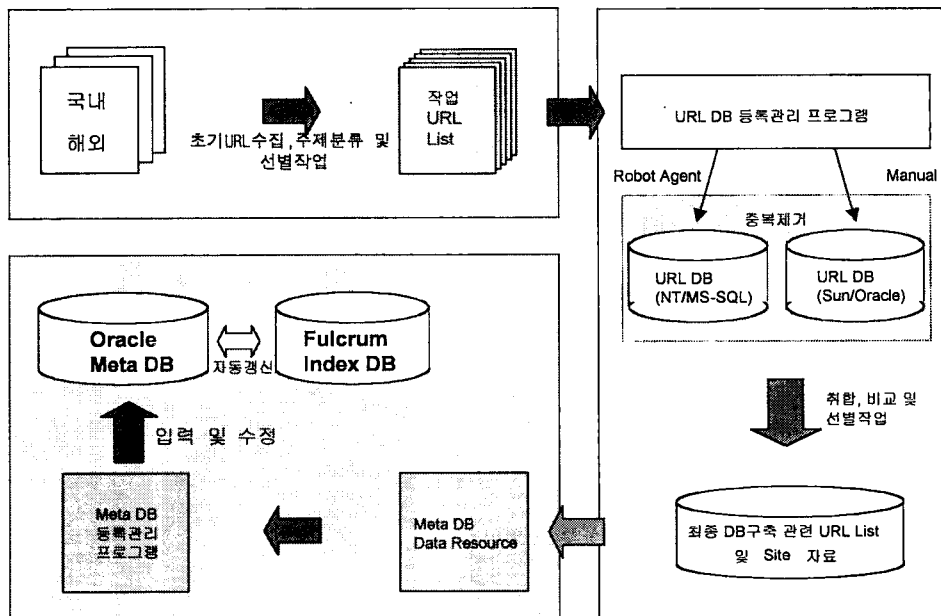
연구장비분야 웹자원 데이터베이스 구축을 위한 엔티티 관계도(ERD)는 다음의 <그림 2>와 같다.

4) 한국기초과학지원연구원, 『연구장비분야 웹자원 메타 DB 구축보고서』. 대전 : 한국기초과학지원연구원, 2001. pp 5-19.



<그림 2> 연구장비 웹자원 데이터베이스 엔티티 관계도

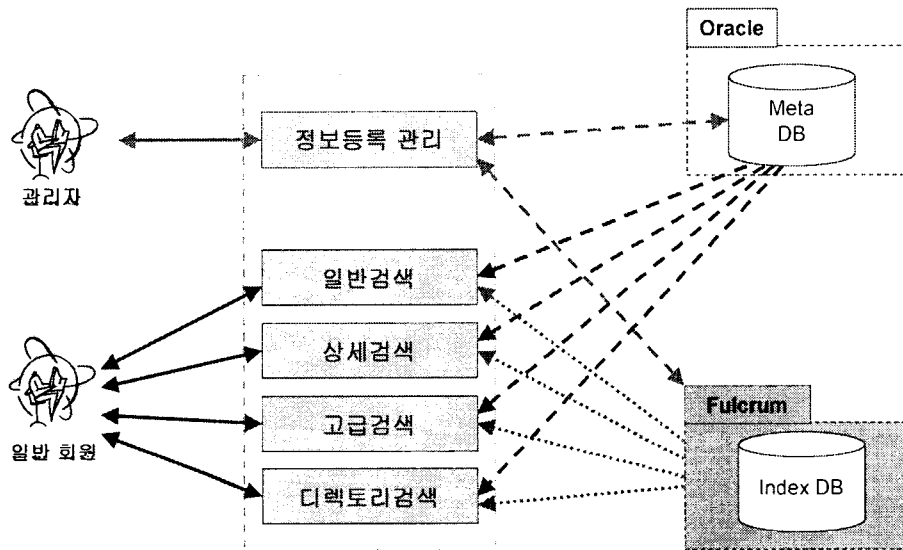
국내 대학, 연구소 등이 보유하고 있는 연구장비 가운데 76%가 외국산인 점을 감안하여 주로 해외 연구장비 제작사의 웹자원을 대상으로 하였고, 다음과 같은 단계를 거쳐 데이터베이스 구축을 수행하였다. <그림 3>은 데이터베이스 구축 과정을 도식화한 것이다.



<그림 3> 연구장비 웹자원 데이터베이스 구축도

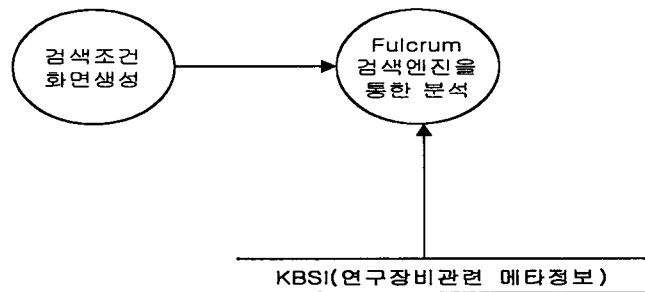
연구장비 웹자원 데이터베이스 구축은 웹자원의 양식 설계(Dublin Core 준용)를 위해 대분류체계 설계 및 가공 지침서 작성을 하고, 웹자원의 조사·선정과 입력·수정 및 품질관리, 웹자원의 정보가공 및 분류를 위해 KDMS(Knowledge Data Management System)를 활용하는 단계로 구성되어 있다.

데이터베이스 운영은 <그림 3>과 같이 관리자에 의해 수집·작성된 메타 데이터베이스 구축을 통하여 일반이용자에게 인터넷 검색서비스를 제공하게 되며, <그림 4>와 같은 시스템 운영이 이루어진다.



<그림 4> 연구장비 웹자원 데이터베이스 운영도

연구장비분야 메타정보의 검색서비스는 상용의 검색엔진인 Fulcrum Search Server를 사용하였으며, 일반검색, 상세검색, 고급검색, 디렉토리검색 등의 검색방법을 지원한다. <그림 5>는 이에 대한 검색서비스의 응용흐름도(DFD)이다.



<그림 5> 연구장비 웹자원 검색서비스 응용흐름도

Fulcrum Search Server 검색엔진을 사용하여 제공하는 메타데이터베이스 정보검색서비스의 주요한 특징은 다음과 같다. 첫째, 검색어 입력시 복잡한 연산기호나 연산식이 전혀 필요 없고, 검색을 원하는 단어나 구절을 입력한 후 검색을 수행한다. 이때 구절 입력 시에는 기본적으로 &(AND)연산이 이루어진다. 둘째, 검색결과내 재검색과 오름차순·내림차순 정렬기능을 지원하고, 검색어의 하이라이팅 기능을 제공한다. 셋째, 검색결과의 개수(5, 10, 15)와 출력 형태(간단히, 자세히) 선택기능을 제공하여 검색에 소요되는 시간과 노력을 절감하도록 한다.

이상에서 살펴본 바와 같이 정보통신분야와 연구장비분야의 웹자원 메타데이터베이스 구축 사례는 국내에서 시도된 웹자원 메타데이터베이스 구축사례로서 상당한 의의를 갖는다.

정보통신 메타데이터베이스의 경우에는 국내 최초의 메타데이터베이스 구축사례로서 첨단 산업인 정보통신산업계의 경쟁력 강화에 많은 도움을 주고 있다. 특히 연구장비 메타데이터베이스 구축은 연구장비의 급증에 비하여 연구장비에 대한 국내 전문운영인력과 운영능력 부족으로 많은 어려움을 겪고 있는 가운데 이루어진 것으로서 향후 연구장비 공동활용의 기반을 형성하였다는 점에서 시의적절한 연구로 평가된다.

그러나, 지금까지의 국내 메타데이터베이스의 연구사례는 일반적인 단순 웹자원 검색에 국한되어 있기 때문에 포털사이트 자체 검색의 어려움이 있었으며, 이중의 관문(gate)을 통한 검색으로 필요정보를 획득해야 하는 불편한 점이 나타난다. 그러므로 앞으로는 메타검색엔진을 사용하여, 한번에 필요정보를 검색할 수 있는 검색시스템의 보완이 필요하다. 또한, 구축된 웹사이트 자원의 경우, 자동수집과 자동분류로 시간단축의 장점을 가지고 있기는 하지만, 검증되지 않은 웹사이트 정보로 인해 불필요한 웹사이트 정보나 내용이 부실한 정보원들이 많기 때문에, 이용자가 메타데이터베이스 사용시 탐색결과에 대한 불만족을 낳을 우려가 있다. 그러므로 자동수집과 분류된 웹사이트의 경우, 해당 주제분야의 전문가로 구성된 웹사이트 검증 위원회를 구성하여 각 분야의 수집된 웹사이트 정보에 대해 주제전문가의 심의를 거쳐 핵심 정보원 중심의 웹자원 데이터베이스 구축이 필요하다.

3. 해양조사자료 메타 데이터베이스 연구동향

3.1 국외의 연구

해양조사자료에 대한 메타 데이터베이스 구축방안은 국제적으로도 많은 연구가 진행되고 있다. 대표적인 프로젝트 중의 하나로는 EUROCORE 및 EUMARSIN(The European Marine Sediment Information Network) 프로젝트 사례이다. EUROCORE는 유럽 내에서 연구기관,

대학, 문서보관소 등에서 수집, 저장하고 있는 해양데이터를 원활하게 유통시키고자 하는 문제의식에서 시작된 프로젝트이다. 이를 위해 유럽지역의 핵심 메타 디렉토리를 개발하여 인터넷을 통해 해당 정보자원에 직접적으로 접근할 수 있게 하고자 한다. EUROCORE는 EUMARSIN과 함께 핵심 메타데이터를 링크된 서버에서 함께 통합 검색 서비스를 제공한다. EUROCORE는 1998년에 시작되었으며 2002년 상반기에 완결되는 프로젝트로써 현재 진행 중에 있다. 이 프로젝트는 독일의 GEOMAR, 스페인의 CSIC, 네델란드의 NITG-TNO, 이탈리아의 IGM, 프랑스의 BRGM 등의 공동 협력으로 진행되고 있으며 책임연구기관은 영국의 Southampton Oceanography Centre이다.⁵⁾

EUROCORE의 목표는 유럽지역 연구소에서 보유하고 있는 해양정보에 대한 메타데이터를 정의하여 해양 저구조 데이터베이스 구축에 기여하고자 하는 것이다. 이로 인해 이용자의 해양 데이터 자료 접근을 유도하고 유럽 및 국제적 협력증진을 목표로 한다. 또한 유럽 기반의 코어 큐레이터 그룹(Core Curator Group)의 중심축을 형성하기 위한 것이다. 코어 큐레이터 그룹은 정기회의를 통하여 내용 및 디자인 측면에서 데이터베이스를 갱신하고 이용자의 요구를 적극 반영한다. 또한 코어 큐레이션에 관한 일반 문제를 논의하고 유럽 및 국제 사회 내에서 광범위한 협력을 이끌어 내고자 한다.⁶⁾

EUROCORE 메타데이터로 구축되고 있는 데이터베이스는 EU-SEASED Meta-database라고 하며 데이터베이스 형식은 GEIXS Project(Catalogue of European Geological Data)와 NOAA의 해양분야 코어 큐레이터즈(Core Curators') 데이터베이스에서의 메타데이터 형식을 기반으로 하였다.⁷⁾ 해양지질시료 데이터베이스 색인(Index to Marine Geological Samples database)으로 알려진 코어 큐레이터즈 데이터베이스는 1976에서 1977년 사이 미국립지구물리데이터센터(National Geophysical Data Center)와 합동으로 해양지질시료 관리자 그룹(Group of Curators of Marine Geological Samples)이 제작한 것이었다. 이것은 1991년에 개정되었고 1996년 새로운 코드 및 필드가 추가되었으며 필드확장도 이루어 졌다.⁸⁾ 코어 큐레이터즈 데이터베이스는 19개 해양학 관련 연구기관의 협력으로 만들어진 시스템으로 데이터 목록(inventory), 암석학 표준 및 해양 연구기관 등에서 소장하고 있는 101,000개 이상의 해저 시료 샘플에 대한 연도별 정보(age information)를 포함하고 있으며, 이에 대한 온라인 검

5) "EUROCORE PROJECT DURATION". Available from WWW : <<http://www.maris.nl/eurocore.htm>> [cited 2002-01-15]

6) "EUROCORE Project Info".

Available from WWW : <<http://www.eu-seased.net/eurocore/welcome.html>> [cited 2002-01-15]

7) "Fields in the EU-SEASED Meta-database".

Available from WWW : <<http://www.eu-seased.net/metadatafields.htm>> [cited 2002-01-15]

8) "SUMMARY in the Index to Marine Geological Samples Database / January 2000 Parameters & Information". Available from WWW : <<http://www.ngdc.noaa.gov/mgg/curator/curatorcoding.html/device>> [cited 2002-01-17]

색서비스를 제공한다.⁹⁾

EUMARSIN & EUROCORE 메타데이터 기술요소는 6개의 필수요소와 26개의 선택요소로 구성되어 있는데, 이는 다음의 <표 3>과 같다.

<표 3> EUMARSIN & EUROCORE 메타데이터 기술요소¹⁰⁾

MANDATORY META-DATA FIELDS	<ul style="list-style-type: none"> · RECORD NUMBER · MEASURING ID · MEASURING AREA TYPE · COORDINATES · SAMPLING DEVICE · DATA SOURCE HOLDER
OPTIONAL META-DATA FIELDS	<ul style="list-style-type: none"> · SAMPLE STATE · SAMPLE STORAGE CONDITION · INTERNAL REFERENCE NUMBER · OBJECTIVE MEASUREMENT · TREATMENT · MEASURED PARAMETERS · SURFACE/SUB-SURFACE SAMPLE · GEOGRAPHICAL AREA · MONITORING SITE · PHYSIOGRAPHIC PROVINCE · NAVIGATION SYSTEM · CORE/SAMPLE LENGTH · WATER DEPTH · CORE/SAMPLE PENETRATION · CORER SAMPLING PENETRATION · CORE/SAMPLE DIAMETER · DATE OF COLLECTION · PROJECT/CRUISE NAME · RESEARCH/SURVEY VESSEL · PROJECT/CRUISE REPORT · BASAL AGE OR PERIOD · PREDOMINANT SEDIMENT TYPE · SAMPLE RECOVERY · LIST OF MAPS · REFERENCE · COMMENTS

3.2 국내의 연구

한국해양연구원(KORDI)에서는 해양조사자료 중에서 가장 일반적이라고 할 수 있는 CTD (Conductivity, Temperature, Depth) 자료에 대한 메타데이터 요소를 정의하여 데이터베이스

9) "The Index to Marine Geological Samples". Available from WWW : <<http://www.ngdc.noaa.gov/mgg/curator/curator.html>> [cited 2001-01-17]

10) "Boundary Conditions Meta-Data Format". Available from WWW : <<http://www.eu-seased.net/metadatabase/metadoc>> [cited 2001-11-15]

를 구축한 바 있다.

CTD 자료의 획득은 해양조사 연구사업을 기반으로 이루어지기 때문에 CTD 메타데이터베이스의 테이블은 크게 연구사업정보(Project Information)와 항해정보(Cruise Information) 및 정점정보(Station Information)로 구성하고, 각 테이블에 해당하는 메타데이터를 <표 4>와 같이 설정하여 관계형 구조의 데이터베이스 테이블을 작성하였다.

구축된 데이터베이스 시스템은 이용자가 웹상에서 사용할 수 있으며, 자료 검색항목은 자료종류(CTD, 채수기), 자료출처, 조사기간, 조사해역, 관측수심, 연구사업 및 조사선 등으로 설정하였고, 검색결과는 해당자료의 정점 리스트와 간단한 통계처리 결과 그리고 해당정점에 대한 수온·염분의 수직적 프로파일이 검색구현(display) 되도록 하여, 해양조사자료 이용자에게 부가적인 정보도 함께 제공하도록 한다.

이러한 CTD 메타데이터 요소 정의는 해양분야 메타데이터베이스 시스템 구축에 있어서 각 정보원에 대한 메타데이터 요소를 규정하는 기본 토대가 될 것이다. 메타데이터 사용예시는 다음의 <표 4>와 같다. 또한, 이를 더블링크어 요소형식으로의 통합 가능성을 위해 CTD 데이터의 관련 필드에 대한 더블링크어 요소화를 시도하였다.

<표 4> KORDI의 CTD 메타데이터 사용 예시

종류	항목(필드)	항목 설명	더블링크어요소
Project Information	PROJECT_NAME	연구사업 한글명칭	제목(title)
	PROJECT_NAME_ENG	연구사업 영문명칭	
	ORGANIZATION	한국해양 연구원 고유 ID(Code)	제작자(creator)
	DIVISION	연구사업 수행부서 고유 ID(Code)	
	PROJECT_MANAGER	연구사업 책임자 한글이름	
	PROJECT_MANAGER_ENG	연구사업 책임자 영문이름	
	PROJECT_ACCOUNT_NO	한국해양연구원부여연구사업 계정번호	
	PROJECT_AREA	연구해역(Code)	
	SPECIFIC_AREA_NAME	상세연구해역명:좁은해역의 범위일때	
	START_DATE	연구사업 시작년/월/일	발행일(date)
	END_DATE	연구사업 종료년/월.일	
	SECURITY_GRADE	자료공개대상에 대한 보안등급(Code)	
	ABSTRACT	연구사업의 목적을 요약 기술	주제(subject)

종류	항목(필드)	항목 설명	더블링크어요소
Cruise Information	CRUISE_NAME	Cruise 고유명칭	내용요약 (description)
	PRINCIPAL_INVESTIGATOR	Cruise 책임자 한글이름	
	PRINCIPAL_INVESTIGATOR_E	Cruise 책임자 영문이름	
	PLATFORM_TYPE	Platform Type(Code)	
	PLATFORM_NAME	Platform Name(Code)	
	CRUISE_AREA	연구해역 (Code)	
	SPECIFIC_AREA_NAME	상세연구해역명:좁은해역의 범위일때	
	START_DATE	조사항해 시작년/월/일	
	END_DATE	조사항해 종료년/월/일	
	REMARK	참고사항 기술	
Station Information	STATION_NAME	정점 고유 명칭	내용요약 (description)
	CASTING_TIME	정점에서의 관측년/월/일/시/분	
	CASTING_DIRECTION	CTD Casting의 수직적 방향:U, D	
	POSITION_LAT	정점 위치 위도(도/분/초)	
	LATITUDE_HEMISPHERE	정점 위치 위도(N or S)	
	POSITION_LON	정점 위치 경도(도/분/초)	
	LONGITUDE_HEMISPHERE	정점 위치 경도(E or W)	
	BOTTOM_DEPTH	정점의 수심	
	MAX_OBSERVATION_DEP	최대관측 수심	
	CTE_GEAR	CTD 관측장비(Code)	
	NAVIGATION_GEAR	위치 측정 장비(Code)	
	TEMPERATURE_PRECISION	관측 수온의 유효숫자	
	SALINITY_PRECISION	관측 염분의 유효숫자	
	DENSITY_PRECISION	관측 밀도의 유효숫자	
	DO_PRECISION	관측 용존산소의 유효숫자	
PH_PRECISION	관측 pH의 유효숫자		

CTD 메타데이터는 <표 5>에서 밝힌 해양분야 웹자원의 필수요소와 비교할 때, 필수요소 8개 항목 중에서 제목(title), 주제(subject), 내용요약(description)과 같은 3개의 경우는 각 요소로 변환이 가능하지만, 나머지의 자료유형(type), 제공형태(format), 웹주소(identifier), 제공언어(language)와 같은 4개의 경우는 요소 변환이 어려운 상황이다. 하지만, 이외 필수요소인 제공기관(provider)의 경우는 한국해양연구원으로 국한됨으로써 따로 요소를 정의할 필요가 없다. 그리고, 선택요소 중에서 제작자(creator)와 발행일(data)의 경우도 변환이 가능하다. 이에 대해서는 앞으로도 추가적인 검토가 이루어져야 한다.

4. 해양분야 메타 데이터베이스 시스템 설계 및 구현방안

4.1 해양분야 전자정보원

각 자원의 특성에 맞는 메타데이터의 개발 및 적용은 전자자원을 체계적으로 기술하는데 필연적인 작업으로 인식되고 있다. 특히 해양 주제분야는 정보자원이 방대하고 다양할 뿐만 아니라 학술적으로도 가치가 높은 정보로써, 국가 지식정보자원이라고 할 수 있다. 그러므로 해양분야의 정보자원에 대한 메타데이터 요소를 정의하고 이를 바탕으로 한 검색 시스템의 개발은 전문적인 연구자 및 일반 이용자들의 정보요구를 충족시켜 줄 수 있는 방안이 될 것이다.

해양분야 웹자원 데이터베이스는 더블린코어 메타데이터 기술을 위해 텍스트 형태의 웹자원을 일반적인 정보원으로 삼고 있다.

해양분야의 웹자원에 대한 리스트 정보는 인터넷을 통해서 해양분야 웹자원을 조사하는 것이 기본적인 탐색 방법이 되겠지만, 제1차적으로는 해양과학분야에 있어서의 웹자원 동향을 파악할 수 있는 최신 정보매체로써 IANSLIC(International Association of Marine Science Libraries and Information Centers)의 계간 뉴스레터를 활용하고자 한다. IANSLIC은 세계 40여개국 100여 해양수산정보센터의 협의회로써 각국에서 생산되는 최신 웹자원에 대한 소개 자료를 연재하고 있다. 이러한 검증된 자료는 해양분야 웹자원 메타데이터 시스템의 주요한 정보원이 된다.

그리하여, PDF 형식으로 제공되는 IANSLIC 뉴스레터를 Text로 변환한 후, 필터링을 통하여 필요한 웹정보를 추출하고 이에 대한 분류작업을 거쳐 해양분야 웹자원 메타데이터베이스에 등록을 한다.

또한 IANSLIC 뉴스레터의 정보 뿐만 아니라, 고성능 로봇을 통하여 웹자원 추출을 가능케 하여 여러가지 다양한 정보를 지원하도록 하고, 수집된 정보를 자동분류기와 지식관리시스템을 활용하여 자동적으로 메타데이터 시스템에 저장하는 시스템으로 설계한다. 검색시스템에 있어서도 기존의 일반적인 사이트 검색엔진과 더불어 각 정보센터의 자료를 상세히 검색 가능하게 하는 메타검색엔진을 활용하도록 한다.

4.2 해양분야 메타데이터 기술요소

해양분야 메타데이터 요소는 더블린코어를 기반으로 작성한다. 메타(meta)라는 단어의 어원은 보다 근본적이고 고차원의 특성을 지닌 중요한 것을 상징하는 그리스어에서 비롯되었다.

메타데이터는 데이터를 한 단계 더 가공하고 조직하였다는 의미로 해석할 수 있으며 이용자의 특정 정보요구에 대한 정확성 확인 및 이용의 적합성 등을 평가하기 위해서 사용된다.¹¹⁾

더블린코어 메타데이터의 특성은 단순성에 기초하여 네트워크 자원을 효율적으로 기술하는데 있다. 생성 및 유지의 간략성, 구문의 용이성, 확장성 및 국제적 사용 등은 더블린코어 표준의 전형적인 특징이다.

더블린코어는 HTML과 XML을 사용하는 RDF 구문으로 기술된다. HTML은 더블린코어 형식을 이해하기는 쉽지만 한정어를 사용하는데 적합하지 않은 측면이 있다. 복잡한 한정어 이용은 XML 구문에서 잘 표현될 수 있다. 그러나 어떤 구문을 사용하는 것이 적합한가를 결정하는 문제는 더블린코어의 개념을 검색엔진 및 인간이 모두 해석할 수 있는 형태로 유지하는 것이 중요하다는 사실을 바탕으로 해야 한다.¹²⁾

이런 메타데이터를 기술하는 방법은 HTML 상에서 메타 태그를 생성하는 방법, XML/RDF 구문으로 표현하는 방법, SGML DTD로 표현하는 방법 등이 있다. 여기서는 가장 일반적인 구문 형태인 HTML에서 메타 태그를 생성하여 데이터를 작성하도록 한다. HTML에서의 메타 태그는 <META NAME="DC.Title" CONTENT="Ask an Earth-Scientist">의 형태로 생성되며 <HEAD>...</HEAD> 사이에 위치한다. 메타 태그에서 나타나는 명칭 중에서는 name은 더블린 코어 요소명을 content는 요소의 내용 기술 부분이며 scheme은 외부 표준 및 준거를 의미한다.

국내 최대의 해양전문 포털사이트인 한국해양연구원 문헌정보서비스 홈페이지에 대한 메타데이터를 영국문헌정보네트워크(UK Office for Library and Information Networking)에서 개발한 더블린코어 메타데이터 편집기인 DC-dot¹³⁾으로 생성한 예시는 다음과 같다.

```
<html>
<head>
<title>한국해양연구원 문헌정보서비스</title>
<link rel="schema.DC" href="http://purl.org/dc">
<meta name="DC.Title" content="한국해양연구원 문헌정보서비스">
<meta name="DC.Subject" content="한국해양연구원">
<meta name="DC.Publisher" content="Korea Ocean Research and Development Institute">
<meta name="DC.Date" scheme="W3CDTF" content="2000-03-30">
```

11) "Diane Hillmann. Using Dublin Core. Working Drafts". Available from WWW : <http://purl.org/DC/documents/wd/usageguide-20000716.htm> [cited 2001-11-21]

12) 문헌정보처리연구회 편, 『메타데이터의 형식과 구조』. 서울 : 문헌정보처리연구회, 1998, pp. 44-58.

13) "DC-dot". Available from WWW : <http://www.ukoln.ac.uk/metadata/dcdot/> [cited 2001-12-13]

```
<meta name="DC.Type" scheme="DCMIType" content="Text">
<meta name="DC.Format" content="text/html">
<meta name="DC.Format" content="1776 bytes">
<meta name="DC.Identifier" content="http://library.kordi.re.kr/">
</head>
<body>
.....
</body>
</html>
```

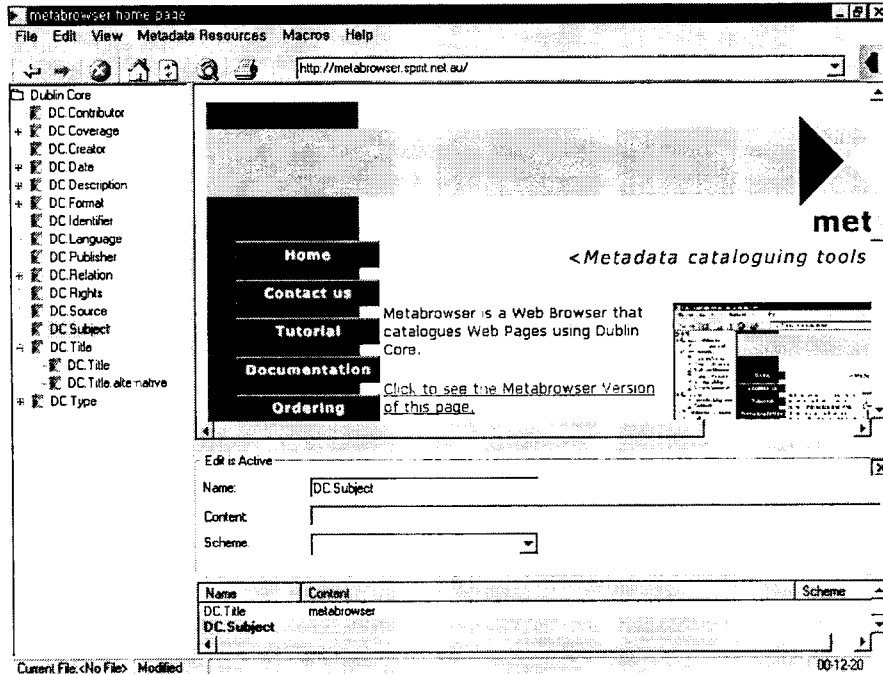
DC-dot은 해당 웹사이트의 URL만 입력하면 자동으로 메타데이터가 생성되는 것이 가장 큰 특징이다. 또한 웹상에서 직접 이용할 수 있기 때문에 프로그램을 따로 설치할 필요가 없다. 그러나 한정어가 지원되지 않는다.

색인 프로그램은 <HEAD> 태그 사이에 메타 레코드가 삽입되어 있는 것을 인식하게 하기 때문에 자동적으로 메타데이터를 추출할 수 있는 것이다. 이런 메타 데이터는 일반 문서의 포맷이나 인쇄에서는 나타나지 않는다. 레코드를 인식하는 웹 브라우저만 메타데이터를 탐색할 수 있을 것이다. 그래서 현재 많은 검색 엔진들에서는 웹문서 내에 HTML <META>를 이용하는 성능이 포함되기 시작했다.

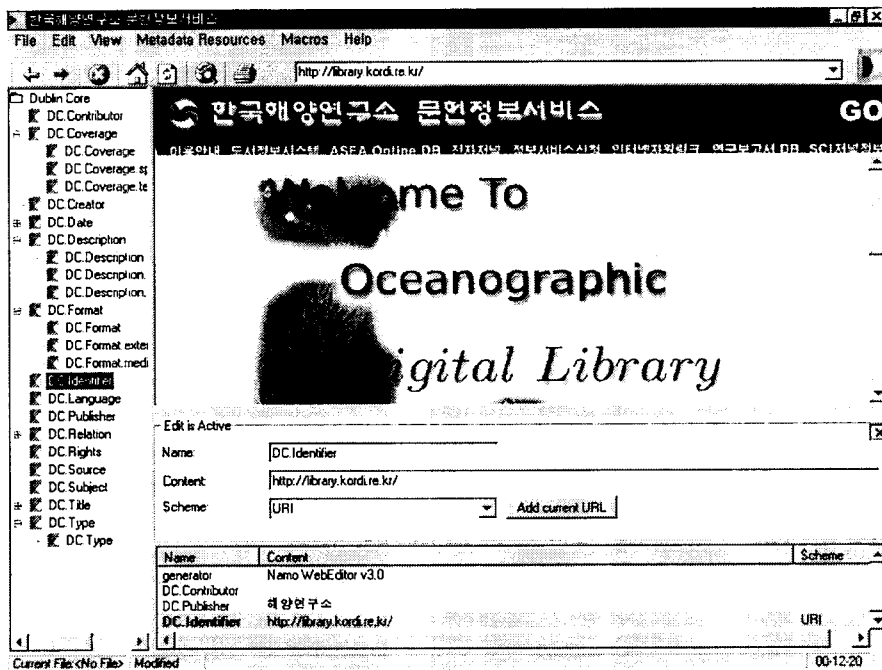
메타데이터를 생성하는 도구로는 DONOR Metadata Generator, 유닉스 및 윈도우즈2000/NT용 Meta Web's Software, DC-dot, Nordic Metadata Project에서 지원하는 Dublin Core Metadata Template, The PrismEd Metadata Editor, Reggie, Hiawatha Island Software Co.에서 제작한 상업용 메타데이터 제작기인 Metabrowser, TagGen(Dublin Core Edition), 소프트웨어 패키지인 What's Hot 등이 존재한다.¹⁴⁾

이러한 편집 프로그램 중에서 최근 개발된 메타브라우저¹⁵⁾는 URL로 웹사이트를 접속하여 메타데이터를 생성할 수 있도록 만들어진 편목도구이다. 메타브라우저는 더블린 코어 요소의 원칙을 바탕으로 COVERAGE, DATE, DESCRIPTION, FORMAT, RELATION, TITLE, TYPE의 7개 요소에서 한정어를 사용할 수 있도록 제작되었다. 이외에도 AGLS 등과 같은 다른 메타데이터 스키마를 제공한다. <그림 6>은 메타브라우저를 통한 데이터 생성창이며, <그림 7>은 해양분야 색인대상 웹사이트의 접속화면 예시이다.

14) "Metadata Related Tools". Available from WWW : <<http://purl.org/DC/tools/index.htm>> [cited 2001-11-24]
15) "Metabrowser Homepage". Available from WWW : <<http://metabrowser.spirit.net.au/>> [cited 2001-12-07]



<그림 6> 메타 브라우저 데이터 생성창



<그림 7> 해양분야 색인대상 웹자원 접속화면

웹자원 메타데이터 요소의 내용과 요소기입방식에 대한 한국데이터베이스진흥센터, 한국기초과학지원연구원, 한국해양연구원 3개 기관의 비교는 다음의 <표 5>와 같다.

<표 5> 3개 기관의 더블링크어 기반 메타데이터 요소 비교

기관명	DB진흥센터(DPC)		기초연(KBSI)		해양연(KORDI)	
	메타데이터 요소	기입	메타데이터 요소	기입	메타데이터 요소	기입
TITLE	제목(title)	필수	제목(title)	필수	제목(title)	필수
CREATOR	저자(author)	선택	저자(author)	선택	제작자(creator)	선택
SUBJECT	주제(subject)	필수	주제(subject)	필수	주제(subject)	필수
DESCRIPTION	내용요약(description)	필수	내용요약(description)	필수	내용요약(description)	필수
PUBLISHER	제공기관(provider)	필수	제공기관(provider)	필수	제공기관(provider)	필수
CONTRIBUTOR	협력자(contributor)	선택	협력자(contributor)	선택	협력자(contributor)	선택
DATE	발행일(date)	선택	발행일(date)	선택	발행일(date)	선택
TYPE	자료유형(type)	필수	자료유형(type)	필수	자료유형(type)	필수
FORMAT	제공형태(format)	필수	제공형태(format)	필수	제공형태(format)	필수
IDENTIFIER	정보위치(locator)	필수	정보위치(locator)	필수	웹주소(identifier)	필수
SOURCE	정보원(source)	선택	정보원(source)	선택	출처(source)	선택
LANGUAGE	제공언어(language)	필수	제공언어(language)	필수	제공언어(language)	필수
RELATION	관련정보(relation)	선택	관련정보(relation)	선택	관련정보(relation)	선택
COVERAGE	취급범위(coverage)	선택	취급범위(coverage)	선택	취급범위(coverage)	선택
RIGHTS	권리(rights)	선택	권리(rights)	선택	권리(rights)	선택

비교대상 3개 기관의 메타데이터 요소는 더블링크어 기반으로 요소명에서만 약간의 상이 점이 있을 뿐, 기입방식에 대해서는 동일하다. CREATOR를 한국데이터베이스진흥센터와 한국기초과학지원연구원은 저자(author)로 하고, 한국해양연구원은 제작자(creator)로 명명하였다. PUBLISHER는 모두 제공기관(provider)으로 표시하였으며, IDENTIFIER를 한국데이터베이스진흥센터와 한국기초과학지원연구원은 정보위치(locator)라고 하였고 한국해양연구원은 웹주소(identifier)라고 명명한 점이 다를 뿐이다. 기입방식은 3기관 모두 동일한데, 한국해양연구원의 경우를 보면, 필수요소는 제목(title), 주제(subject), 내용요약(description), 제공기관(provider), 자료유형(type), 제공형태(format), 웹주소(identifier), 제공언어(language)이며, 선택 요소로는 제작자(creator), 협력자(contributor), 발행일(date), 출처(source), 관련정보(relation), 취급범위(coverage), 권리(rights)이다.

4.3 해양분야 메타데이터의 분류체계

해양분야 메타데이터의 분류체계는 국제식량농업기구(FAO)의 해양수산정보시스템(Aquatic Sciences and Fisheries Information System : ASFIS)의 분류표를 참고하여 재구성하였다. FAO/ASFIS는 해양수산분야의 대표적인 분류표로서 국제적인 해양수산정보 학술공유를 위한 협력프로그램으로 작성되었으므로 여기에서의 분류체계는 부분적으로 한반도 연구해역의 실정에 적합하게 재조정후 본 연구의 기본적인 분류체계로 사용가능하다. 이의 분류체계는 크게 3개 영역으로 구분하고 11개의 주류와 200여개의 하위류로 세분전개하였다. 각 하위류에는 주석과 상호참조 기능이 지원된다.

분류작업은 검증된 자원에 대해서는 1차 단계로 해양분야 전문가들에 의한 수작업 분류를 수행하고, 2차로는 로봇을 통한 웹정보 수집과 더불어 자동분류기를 이용한 자동분류가 이루어지도록 하고, 분류된 정보에 대해서는 지식정보 관리 시스템(KMS)을 활용해 저장·관리하도록 한다.

<표 6> 해양분야 분류체계

[주요분류]

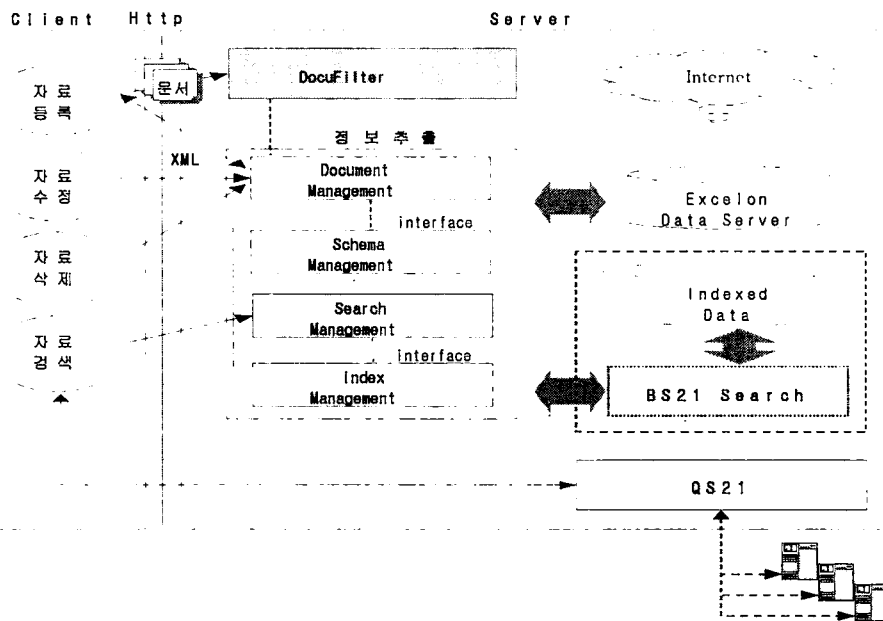
분류코드	분류항목	분류코드	분류항목
1000	생물과학, 생명자원	1600	어획량
1100	종류	1620	수산물 이용
1180	생물학	2000	해양기술, 미생명자원
1200	미생물학	2120	해양법, 정책
1220	식물학	2140	기술적 해양학
1240	무척추동물학	2160	해양역학 및 육수학
1260	연체동물학	2170	연안역학
1280	갑각류학	2180	화학 및 지구과학
1300	곤충학	2200	수중음향학
1320	척삭동물학	2220	수중광학
1340	어류학	2240	해양기상학 및 기후학
1360	조류학	2260	지질학 및 지구물리학
1370	포유류학	2280	해양기술
1380	수중생태학	2300	조사선, 잠수정
1400	생태학	2320	해양구조물
1420	개체생태학	2340	잠수학
1440	개체군연구	2380	해양 기장비
1460	수중군집	2390	해난구조
1480	생산력, 생태계, 개체상호작용	2400	해양자원
1500	수산	2420	해상 교역
1540	부착과 천공	3000	해양환경
1560	어업	3500	해양오염
1580	양식	3520	환경적 특성

[세부분류 예시]

분류코드	주제내용	분류코드	주제내용
2180	화학 및 지구화학	2188	대기의 화학적 특성
2181	일반		해수 관련 대기의 화학적 구성과 물리화학적 특성
	해수와 육수의 화학, 지구화학에 관한 일반적인 학술자료		해수 대기 상호작용. 습도와 열의 교환 use 2163 기상학 use 2241-2245
	생화학 연구 use 1206 기초 생산물 use 1481 생화학적 순환 use 1482	2200	수중음향학
2182	연구방법 및 분석	2201	일반
	현장시료 채취와 분석기술. 데이터 가공기술		해수와 대기, 해수와 퇴적층의 접촉 및 음향
	오염물질의 화학적 분석 use 3502		수중음향학의 특정 연구 use 2202-2205
2183	물리학 및 화학	2202	연구방법 및 분석
	물리적이고 화학적인 특성		음향측정, 수중 청음기, 어군탐지 등
	특정 해역의 화학적 특성 use 2184-2186 압력파 (수중음향) use 2201-2205 발광 (수중광학) use 2221-2225 부유물질의 화학적 특성 use 2186 부식 및 방식 use 2282		지질학상의 음향기술 use 2262 음향흐름 측정과 측정기기 use 2162
2184	해수 성분	2203	음향 전파
	해수의 화학적 성분		수중음향의 발생과 전파의 특성
	해수의 물리적 화학적 특성 use 2183 유기화학 use 2185 부유 입자의 화학적 특성 use 2186 퇴적물의 화학적 특성 use 2187	2204	잔향
	용존, 부유 유기물의 합성		해수 표면, 어류, 기포에 의한 반사와 산란의 특성
	부유 유기물의 화학적 특성 use 2186 퇴적물의 유기화학 use 2187	2205	소음 및 생체음향
2185	유기 합성		과량, 퇴적물의 이동. 선박으로 인한 소음의 특성
	용존, 부유 유기물의 합성	2220	수중광학
	부유 유기물의 화학적 특성 use 2186 퇴적물의 유기화학 use 2187	2221	일반
2186	부유물질의 화학적 특성		해수면과 대기, 수중의 광학적인 현상
	부유 유기물과 무기물의 화학적 특성		특수한 광학적인 특성 use 2223 광학적인 작용 use 2222-2225
	퇴적물의 화학적 특성 use 2187 탁도의 광학적 측량 use 2222	2222	연구방법 및 분석
2187	퇴적물의 지화학적 특성		수중광학 측정기술
	퇴적물의 화학적이고 광물학적인 연구		혼탁도 관찰과 화학적 특성 use 2186
	퇴적물의 유기화학적 특성 use 2185 부유물질의 화학적 특성 use 2186 비화학적으로 취급한 퇴적학 use 2272 광물자원의 시추개발 use 2404	2223	광학적 특성
			광학적 특성, 관찰, 모델링
			열수광상의 구성. 태양과 지구상의 방사선 use 2146
		2224	수중 시계(視界)
			조명과 빛의 근원. 바다의 미래상. 수중카메라

4.4 해양분야 메타데이터베이스 시스템 구성

해양분야 메타데이터베이스 시스템의 H/W 주전산기는 NT-Server를 기반으로 한다. S/W의 선정은 RDBMS로 Oracle 8i(NT용)를 한국해양연구원의 기존 정보관리시스템(Maestro)과의 호환과 함께 장기적인 운영을 고려하여 선택하고, 개발언어로서 ASP(Active Server Page)는 서버 측면 애플리케이션의 기본 골격을 제공하여 보다 용이한 동적 웹응용 개발을 지원하도록 한다. 이외에 일반검색엔진으로는 BS21, 메타검색엔진은 QS21, 웹자원 자동 수집을 위해 Kware Filter 등으로 구성하고자 한다. 해양분야 메타데이터베이스 시스템 구성에 따른 운영도는 다음의 <그림 8>과 같다.

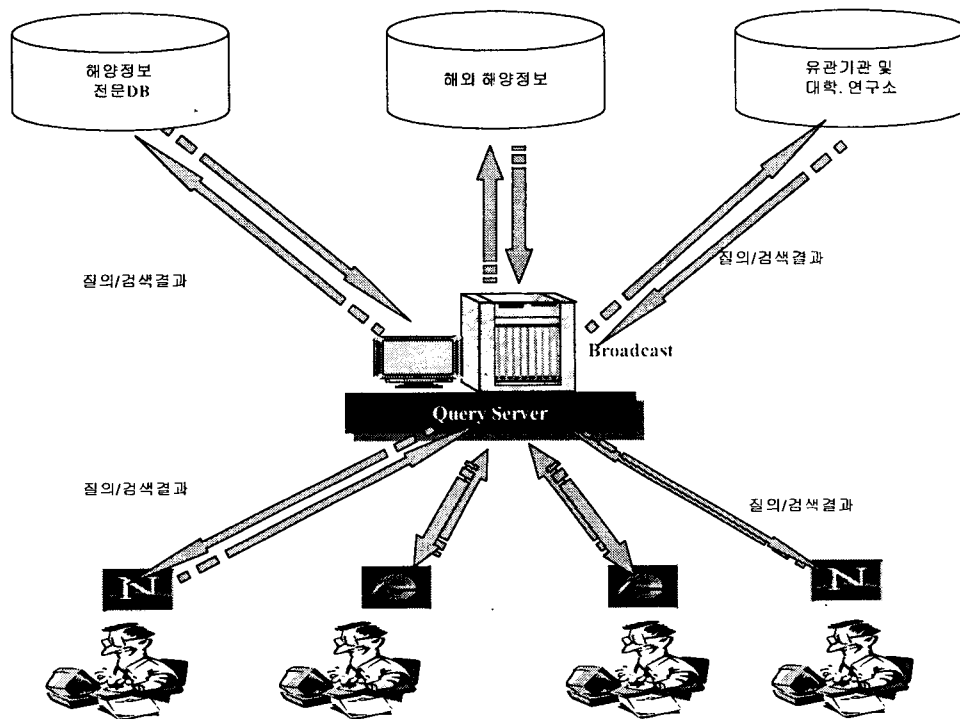


<그림 8> 해양분야 메타데이터베이스 시스템 구성도

해양분야 메타데이터베이스는 기본정보원으로서 PDF 형식의 IAMSLIC 뉴스레터를 Text로 변환시킨 후에 필터(Kware Filter)를 통해 웹정보를 수집하여 자료등록을 하고, RDBMS 형식의 오라클 데이터베이스에 자료를 관리하도록 하며, 일반적인 검색엔진(BS21)과 메타검색엔진(QS21)을 사용한 검색이 이루어지게 설계된다.

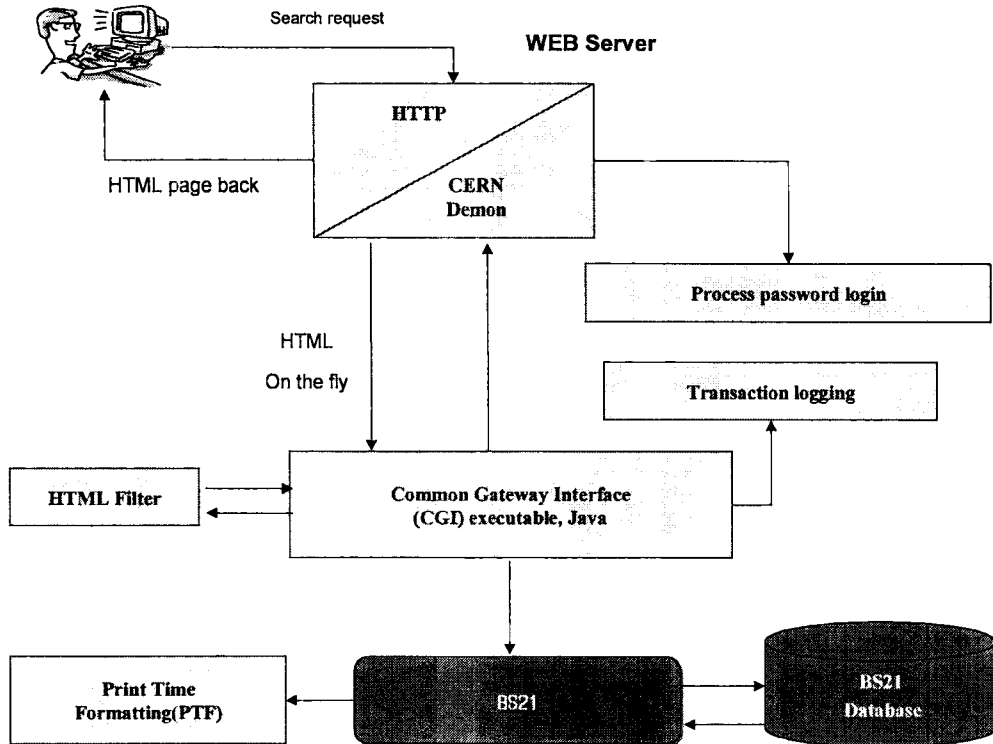
4.5 해양분야 메타데이터베이스 검색시스템 구현 방안

해양분야 메타데이터베이스는 다음의 <그림 9>와 같은 메타통합검색시스템으로 구성한다. 일반적으로 사용되는 특정주제분야 사이트 및 정보저장소에서만 정보를 검색하는 웹 포털 사이트의 검색 뿐만이 아니라 여러 개의 정보저장소에 접근하여 한번에 필요한 정보의 검색결과를 가져오는 메타검색엔진을 사용하는 것이 특징이 된다.



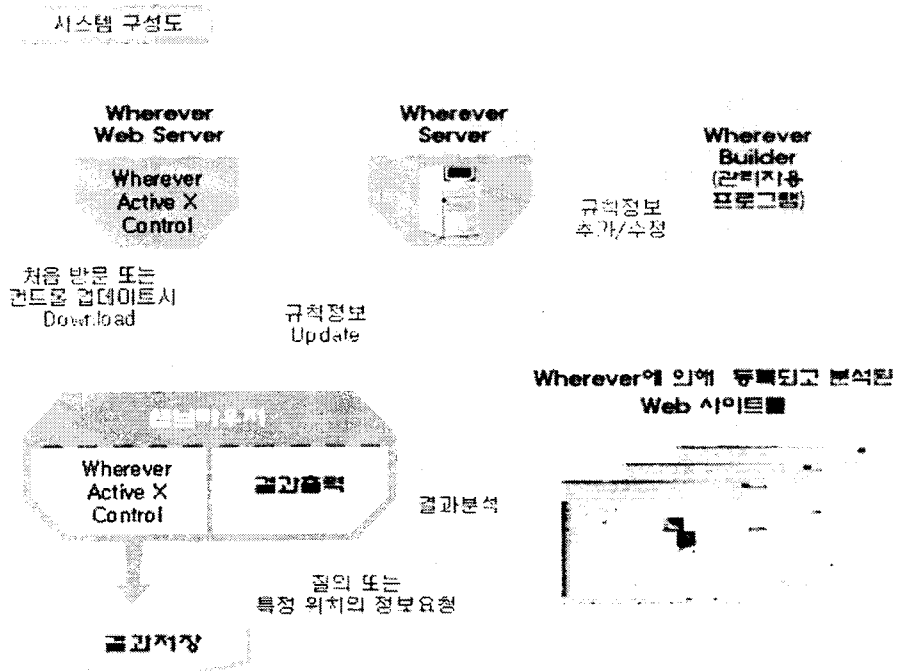
<그림 9> 해양분야 메타 통합검색 구성도

따라서 이에 대한 검색시스템은 일반적인 웹사이트 검색엔진과 메타검색엔진으로 함께 구성된다. 먼저 웹포털사이트 검색엔진은 구축시스템의 저장 및 검색을 지원하기 위해 대용량의 데이터 지원기능과 고유의 압축기능을 사용하여 디스크 사용효율을 극대화해야 한다. 또한 도치 색인파일을 사용하여 검색의 신속성을 보장하고 다양한 검색조건을 지원한다. 그리고, 각 시스템과 연계하여 운영되기 때문에 실시간 삽입, 갱신, 삭제기능까지 수반하도록 한다. <그림 10>은 웹포털사이트 검색엔진의 구성도이다.



<그림 10> 웹포탈사이트 검색엔진의 구성도

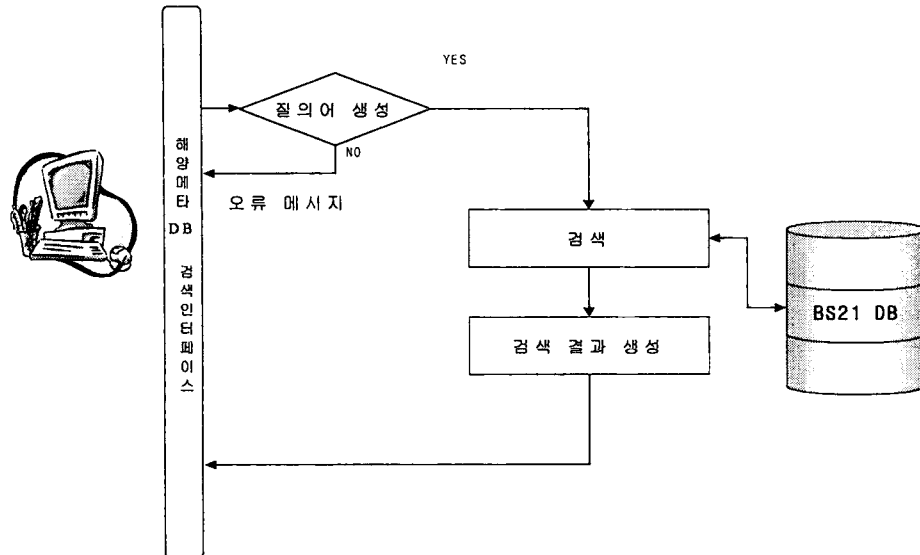
또한, 메타 검색엔진을 사용한 정보검색은 여러 개의 분산된 웹 사이트를 하나의 화면에서 동시에 검색하고, 그 검색결과를 취합하여 하나의 통합목록으로 제공함으로써 이용자가 각각의 사이트를 서핑하지 않아도 원하는 정보를 보다 신속하게 검색할 수 있는 장점이 있다. 그러므로 이용자가 원하는 검색 사이트를 GUI(Graphical User Interface) 환경에서 보다 간편하게 구축할 수 있도록 도와주며, 이용자 질의를 각자의 검색엔진이 이해할 수 있는 형식으로 변환시킨 다음, 검색된 결과를 즉각적으로 데이터베이스에 삽입하거나 수집된 HTML 형태의 문서를 분석하여 제공한다. 메타 검색엔진의 구성도는 다음의 <그림 11>과 같다.



<그림 11> 해양분야 메타검색엔진 구성도

한편 한국기초과학지원연구원에서는 Fulcrum Search Server를 사용하였는데, 이것은 복잡한 연산식이 필요없이 원하는 단어나 구절을 입력하여 검색할 수 있기 때문에 광범위한 검색이 가능하다.

해양분야 메타데이터베이스는 키워드 검색 기능을 보유하여 다음의 <그림 12>와 같은 키워드 검색서비스를 지원하도록 한다. 이를 통하여 간단한 키워드 입력만으로 원문과 메타데이터를 검색하여 빠른 검색을 가능케 하고 결과내 재검색 기능을 지원하여 보다 정확하고 빠르게 접근할 수 있는 시스템으로 구현한다.



<그림 12> 해양분야 키워드 검색서비스

5. 결론

해양분야 전자정보자원이란 웹자원과 더불어 해양에서의 조사활동으로 획득하여 전자적으로 저장된 해양조사자료까지를 범주로 하고 있는데, 본 연구는 웹자원의 메타데이터베이스 시스템 구축방안 설계와 해양조사자료데이터베이스의 더블린코어 연계가능성에 대하여 살펴 보았다. 국내외 연구동향 조사를 통하여 나타난 내용은 다음과 같다. 또한 선행연구의 해양 분야에의 적용 가능성도 부분적으로 제시하였다.

첫째, 국내에서의 웹자원 메타데이터베이스 구축사례는 한국데이터베이스진흥센터의 정보통신 메타데이터베이스와 한국기초과학지원연구원의 연구장비 메타데이터베이스가 조사대상이 되었으며, 해외의 해양분야 웹자원 메타데이터베이스 구축사례는 아직까지 보고된 바가 없기 때문에 본 연구에서는 해양조사자료 메타데이터베이스 연구동향만을 조사하였다.

둘째, 정보통신 메타데이터베이스는 메타데이터 형식을 기본으로 적용한 전문적인 데이터베이스로써 국내 최초로 구축되었다는데 의의가 있다. 또한 연구장비 메타데이터베이스 구축은 해외에서 도입한 고가장비의 공동활용 기반을 형성하였다는 점에서 그 우수성이 인정된다. 그러나, 지금까지의 국내 메타데이터베이스의 연구사례는 일반적인 웹자원 단순검색에

국한되었기 때문에 포털사이트 전체 검색의 어려움과 이중의 관문(gate) 검색으로 필요정보를 획득해야 하는 불편함이 나타난다. 해양분야 메타데이터베이스는 메타검색엔진을 사용하여 이러한 문제점을 해소하고, 자동수집·분류된 웹사이트는 해당 주제분야의 전문가로 하여금 검증을 거쳐 핵심 정보원 중심의 웹자원 데이터베이스 구축을 설계함으로써 진일보한 메타데이터베이스를 구축하도록 하였다.

셋째, 해양조사자료 메타데이터베이스 연구동향에서 유로코어의 경우는 해양분야에서 이루어지고 있는 국제적인 연구프로젝트로써 국내의 해양조사자료관리에 많은 시사점을 제시한다. EUROCORE의 목표는 유럽지역 연구소에서 보유하고 있는 해양정보에 대한 메타데이터를 정의하여 해양 저구조 데이터베이스 구축에 기여하고자 하는 것으로 매년 정기회의를 통하여 데이터베이스를 갱신하고 있다. 이의 메타데이터 기술요소는 6개의 필수요소와 26개의 선택요소로 구성되어 있다.

넷째, 현재 국내에서는 해양조사자료 중에서 가장 일반적이라고 할수 있는 CTD 자료에 대한 메타데이터 요소를 정의하여 데이터베이스구축이 이루어지고 있으며, 구축된 데이터베이스시스템은 이용자가 웹상에서 검색할 수 있으며, 검색항목은 자료종류(CTD, 채수기), 자료출처, 조사기간, 조사해역, 관측수심, 연구사업 및 조사선 등으로 설정하였고, 검색결과는 해당자료의 정점 리스트와 간단한 통계처리 결과 그리고 해당정점에 대한 수온·염분의 수직적 프로파일이 검색구현(display)되도록 하여, 해양조사자료 이용자에게 부가적인 정보도 함께 제공하도록 한다. 또한 CTD 메타데이터를 더블링크어 요소형식으로의 통합 가능성을 위해 CTD 데이터의 관련 필드에 대한 더블링크어 요소화를 시도하였으며, 이에 대해서는 향후 추가적인 연구를 수행하고자 한다.

해양분야 메타데이터베이스 시스템 설계 및 구현방안을 요약하면 다음과 같다.

첫째, 해양분야 전자정보원 중에서 IAMSLIC 뉴스레터 정보는 검증된 핵심적인 자원이 된다. 또한 고성능 로봇을 통하여 여타의 웹자원을 추출하고, 수집된 정보를 자동분류기와 지식관리시스템을 활용하여 자동적으로 메타데이터 시스템에 저장되도록 설계한다. 검색시스템에 있어서도 기존의 일반적인 사이트 검색엔진과 더불어 각 정보센터의 자료를 상세히 검색 가능하게 하는 메타검색엔진을 활용한다.

둘째, 해양분야 메타데이터 기술요소 선정을 위해 정보통신분야와 연구장비분야를 비교하였는데, 더블링크어 기반으로 요소명에서만 약간의 상이점이 있을 뿐, 기입방식에 대해서는 동일하다. 해양분야의 경우 필수요소는 제목(title), 주제(subject), 내용요약(description), 제공기관(provider), 자료유형(type), 제공형태(format), 웹주소(identifier), 제공언어(language)이며, 선택요소로는 제작자(creator), 협력자(contributor), 발행일(date), 출처(source), 관련정보(relation), 취급범위(coverage), 권리(rights)이다.

셋째, 해양분야 메타데이터의 분류체계는 국제식량농업기구(FAO)의 ASFIS 분류표를 참고

하여 재구성하였다. 분류체계는 크게 3개 영역으로 구분하고 11개의 주류와 200여개의 하위류로 세분전개하였다. 각 하위류에는 주석과 상호참조 기능이 지원된다. 분류작업은 검증된 자원에 대해서는 1차 단계로 해양분야 전문가들에 의한 수작업 분류를 수행하고, 2차로는 로봇을 통한 웹정보 수집과 더불어 자동분류기를 이용한 자동분류가 이루어지도록 하고, 분류된 정보에 대해서는 지식정보 관리 시스템(KMS)을 통해 저장·관리하도록 한다.

넷째, 해양분야 메타데이터베이스 시스템의 H/W 주전산기는 NT-Server를 기반으로 한다. S/W의 선정은 RDBMS로 Oracle 8i(NT용), 개발언어는 ASP(Active Server Page), 일반검색엔진으로는 BS21, 메타검색엔진은 QS21, 웹자원 자동 수집을 위해 Kware Filter 등으로 구성한다. 해양분야 메타데이터베이스는 기본정보원으로서 PDF 형식의 IAMSLIC 뉴스레터를 Text로 변환시킨 후에 필터를 통해 웹정보를 수집하여 자료등록을 하고, RDBMS 형식의 오라클 데이터베이스에 자료를 관리하도록 하며, 일반적인 검색엔진과 메타검색엔진을 사용한 검색이 이루어지도록 한다.

다섯째, 해양분야 메타데이터베이스는 메타통합검색시스템으로 구성한다. 일반적으로 사용되는 특정주제분야 사이트 및 정보저장소에서만 정보를 검색하는 웹 포털사이트의 검색 뿐만 아니라 여러 개의 정보저장소에 접근하여 한번에 필요한 정보의 검색결과를 가져오도록 메타검색엔진을 사용하는 것이 특징이 된다. 또한 각 시스템과 연계하여 운영되기 때문에 실시간 삽입, 갱신, 삭제기능까지 수반한다. 메타 검색엔진을 사용한 정보검색은 여러 개의 분산된 웹 사이트를 하나의 화면에서 동시검색한 결과를 하나의 통합목록으로 제공한다. 또한 해양분야 메타데이터베이스는 키워드 검색 기능을 보유하여 간단한 키워드 입력만으로 원문과 메타데이터를 검색하여 빠른 검색을 가능케 하고 결과내 재검색 기능을 지원하여 보다 정확하고 빠르게 접근할 수 있도록 한다.

끝으로 본 연구는 해양분야 웹자원과 해양조사자료에 대한 메타데이터베이스 구축에 관한 선행연구가 국내외적으로 빈약하고, 이를 최초로 웹자원과 조사자료를 망라하여 해양분야 전자정보원 메타데이터베이스 시스템을 시범적으로 구축하고자 하는 설계 및 구현방안에 대한 기초적인 연구이기 때문에, 본 연구의 기반이 되는 해양분야 전자정보원, 메타데이터 기술, 분류체계, 시스템구성에 대한 보다 심도있는 연구는 향후의 연구과제로 수행하고자 한다.

또한 본 연구는 국가 해양전문정보센터 역할을 수행하고 있는 한국해양연구원을 기반으로 하여 해양분야 전자정보자원 메타데이터베이스 시스템 설계 및 구현방안을 제시하였는 바, 향후 국내의 해양수산 관련 정부기관 및 연구기관을 연계한 통합시스템 구축을 촉진하는 기초자료로 활용되기를 기대한다.

참 고 문 헌

- 김이겸 등. “메타데이터 연계성을 위한 RDF 응용스키마 설계에 관한 연구”, 《정보관리학회지》 제17권, 제1호(2000). pp. 21-49.
- 문헌정보처리연구회 편. 『메타데이터의 형식과 구조』. 서울 : 문헌정보처리연구회, 1998. p.359.
- 이재진. “네트워크 환경에서 대표적인 전자자원의 기술 수단”, 《데이터베이스월드》 제65권 (1998.10.). pp. 50-55.
- 최한석. 『정보통신 메타 DB 시스템』. 서울 : 한국지식웨어, 2001.
- 한국기초과학지원연구원. 『연구장비분야 웹자원 메타 DB 구축보고서』. 2001. p. 21.
- 한국데이터베이스진흥센터. 『정보통신메타DB 구축사업 완료보고서』. 서울 : 한국데이터베이스진흥센터, 1999. p. 155.
- 한국해양연구원, 『해양자료서비스 기술 개발』. 안산 : 한국해양연구원, 1999. p. 338.
- “정보통신 메타DB 분류체계 구성”. Available from WWW : <<http://www.metadb.net/intro/intro3.htm>> [cited 2001-12-27]
- “Boundary Conditions Meta-Data Format”. Available from WWW : <<http://www.eu-seased.net/metadatabase/metadatabase.doc>> [cited 2000-11-15]
- “Dublin Core Metadata Initiative”. Available from WWW : <<http://purl.org/DC/>> [cited 2001-11-21]
- “DC-dot”. Available from WWW : <<http://www.ukoln.ac.uk/metadata/dcdot/>> [cited 2001-12-13]
- Diane Hillmann. “Using Dublin Core”, *Working Drafts*. Available from WWW : <<http://purl.org/DC/documents/wd/usaguide-20000716.htm>> [cited 2001-12-15]
- “EUROCORE PROJECT DURATION”. Available from WWW : <<http://www.maris.nl/eurocore.htm>> [cited 2002-01-15]
- “EUROCORE Project Info”. Available from WWW : <<http://www.eu-seased.net/eurocore/welcome.html>> [cited 2002-01-15]
- “Fields in the EU-SEASED meta-database”. Available from WWW : <<http://www.eu-seased.net/metadatabase/metadatabasefields.htm>> [cited 2002-01-15]
- “Metadata Related Tools”. Available from WWW : <<http://purl.org/DC/tools/index.htm>> [cited 2001-12-21]
- “Metabrowser Homepage”. Available from WWW : <<http://metabrowser.spirit.net.au/>> [cited 2001-12-21]
- “Metabrowser Download”. Available from WWW : <<http://metabrowser.spirit.net.au/download/>> [cited 2001-12-21]

- 'SUMMARY in the Index to Marine Geological Samples Database / January 2000 Parameters & Information". Available from WWW : <<http://www.ngdc.noaa.gov/mgg/curator/curatorcoding.html#device>> [cited 2002-01-17]
- 'The Index to Marine Geological Samples". Available from WWW : <<http://www.ngdc.noaa.gov/mgg/curator/curator.html>> [cited 2002-01-17]
- 'What is metadata? in the DC FAQ". Available from WWW : <<http://purl.org/DC/education/index.htm>> [cited 2001-12-13]