

디지털 도서관을 위한 확장된 Dublin Core 기반 메타데이터 관리 시스템

권재길[†]·장원호[†]·김현주^{††}·박인호[†]·배종민^{†††}·강현석^{†††}

요약

최근 WWW의 성장과 더불어 멀티미디어 전자 문서의 활용이 많아지고 있다. 특히, 디지털 도서관은 다양하고 방대한 양의 멀티미디어 전자 문서들을 다루게 되는데, 이들에 대한 효율적인 검색 방법과 체계적인 관리가 필요하다.

본 논문은 디지털 도서관에서 관리되는 멀티미디어 전자 문서들을 효율적으로 저장하고 검색하기 위해 Dublin Core 메타데이터를 확장하여 사용한다. 그리고 이렇게 확장된 Dublin Core 메타데이터에 기반한 멀티미디어 문서 데이터베이스 스키마를 설계하고 이를 관리하는 시스템인 xD-MMS를 구현한다.

A Metadata Management System Based on Extended Dublin Core for Digital Libraries

Jae-Gil Kweon[†] · Won-Ho Jang[†] · Hyun-Ju Kim^{††} · In-Ho Park[†] ·
Jong-Min Bae^{†††} · Hyun-Syug Kang^{†††}

ABSTRACT

With the growth of World Wide Web, multimedia electronic documents have been frequently used in a lot of applications. Especially, digital libraries that deal with huge and various multimedia electronic documents need to support efficient search techniques and systematical management techniques.

In this paper, we suggest an extended Dublin Core metadata to store and search the multimedia electronic documents managed at digital libraries. We also design a document database schema based on the extended Dublin Core metadata and implement a system, called xD-MMS which manages the multimedia electronic documents.

1. 서론

최근 인터넷의 급속한 보급에 따라 이를 통해 교환

되는 전자 정보의 양이 증가하고 있다. 즉, 컴퓨터를 통해 전자화된 정보는 통신망을 이용함으로써 유통 및 재생산이 매우 용이해지고 있다. 그러나 실제로 인터넷을 통하여 수많은 정보가 생산되어 유통되고 있으나 이들은 대개 조직적으로 관리되지 못하고 있다. 따라서 이들을 체계적이고 효율적으로 관리하고 검색하는 방안이 제시되어야 한다. 이러한 요구에 따라 출현한 디지털 도서관은 텍스트뿐만 아니라 비디오, 오디오,

* 이 논문은 1997년도 학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

† 정회원 : 경상대학교 대학원 전자계산학과

†† 준회원 : 경상대학교 대학원 전자계산학과

††† 종신회원 : 경상대학교 컴퓨터과학과/정보통신연구센터 교수

논문접수 : 1998년 2월 6일, 심사완료 : 1998년 7월 11일

이미지 등을 포함하는 각종 멀티미디어 정보를 전자화하여 네트워크를 통해 편리하게 다양한 정보를 사용할 수 있도록 한다[1, 2, 3, 4].

메타데이터란 “데이터에 대한 데이터” 라고 할 수 있다. 즉, 데이터 또는 데이터 집합을 효율적으로 접근하고 관리할 수 있도록 해주는 데이터에 대한 정보를 총칭한다. 메타데이터를 통한 전자 문서 관리는 정보 검색시 사용자가 보다 정확하고 쉽게 전자 문서를 검색할 수 있도록 하며, 대용량 정보를 취급하는데 유지 비용을 절감시킬 수 있다. 특히, 멀티미디어 전자 문서의 메타데이터 모델링은 비정형 디지털 문서의 구조화를 가능케 하는 정형화된 포맷을 제공하기 때문에, 멀티미디어 전자 문서에 대한 메타데이터의 관리는 반드시 필요하다.

현재 전자 문서의 검색 도구로써 많이 사용되고 있는 Lycos[http://lycos.cs.cmu.edu/]나 WebCrawler[http://webcrawler.com/] 등은 자동으로 생성된 인덱스를 통해 문서를 관리하며 사용자에게 위치 정보만을 제공한다. 그런데 이것은 화일명 자체가 메타데이터이므로 문서에 대한 유용한 정보가 부족하여, 효율적으로 사용자가 요구하는 문서를 정확하게 검색하기 힘들 뿐만 아니라 특히, 내용 기반의 검색이 어렵다.

이러한 단점을 보완하기 위해 본 논문에서는 Dublin Core를 이용하며, 특히 멀티미디어 전자 문서의 관리를 위해 이를 확장해 사용한다. 즉, 멀티미디어 전자 문서들의 체계적인 관리와 검색을 위해 Dublin Core 메타데이터[2, 3]를 확장하여 데이터베이스화하고 이를 이용한 시스템을 구현한다.

Dublin Core는 네트워크상에 존재하는 문서들을 기술하기 위한 메타 엘리먼트 집합(Dublin Core Metadata Element Set)으로 구성되어 있으며, SGML DTD로 표현할 수 있다. 따라서 네트워크상에 산재해 있는 다양한 종류의 전자 문서들을 SGML 문서로 관리함으로써 Dublin Core 메타데이터 정보와의 효과적인 연동을 이룰 수 있다[2, 3].

전자 문서에 대한 메타데이터를 데이터베이스로 관리하면 많은 이득을 얻을 수 있다. 즉, 질의 처리 기능을 이용한 효율적인 문서 관리와 검색, 그리고 병행 제어 등 데이터베이스 시스템이 갖는 여러 특성들을 제공받을 수 있다. 그런데 이러한 메타데이터는 새로운 형태의 문서 형식 및 내용에 대한 융통성있는 변경을 위해 확장성 및 상속성 등을 제공받아야 하므로 객체

지향 데이터베이스로 관리하는 것이 효과적이다[9]. 이를 위해서는 Dublin Core 메타데이터에 기반한 전자 문서 관리에 적절한 객체 지향 데이터베이스의 설계가 이루어져야 한다.

본 논문은 멀티미디어 전자 문서와 메타데이터를 디지털 도서관에서 통합적으로 다룰 수 있도록 객체 지향 데이터베이스 스키마를 설계하고, 이를 관리하는 시스템을 다룬다. 특히 멀티미디어 전자 문서내에 산재된 다양한 미디어 정보의 관리와 메타 정보 관리에 중점을 둔다. 2장에서는 확장된 Dublin Core의 메타데이터를 소개하고, 3장에서는 메타데이터 관리를 위한 데이터베이스 스키마를 설계한다. 4장에서는 3장에서 설계한 데이터베이스 스키마를 관리하는 시스템에 대해 기술하며, 5장에서는 메타데이터 등록 시스템 및 검색 시스템을 구현한다. 6장에서는 결론 및 향후 연구 과제에 대해 논한다.

2. 확장된 Dublin Core 메타데이터

WWW에서 멀티미디어 전자 문서들을 효율적으로 검색하는 시도가 많이 이루어지고 있다. 그러나 기존의 검색 시스템들(Lycos, Webcrawler 등)은 주로 화일명을 통하여 색인을 발생시켜 인덱싱한 후 위치 정보만을 제공하므로 문서에 대한 유용한 정보가 부족할 뿐만 아니라, 내용 기반의 검색이 어려워 네트워크상에 존재하는 대용량의 정보에 대한 사용자의 요구에 정확한 정보를 제공하지 못한다. 또한, 이러한 문제를 해결하려고 기존의 도서관에서는 TEI 혹은 MARC[6]와 같은 분류 시스템을 사용하였으나, 이는 전문 정보 제공자가 아닌 경우 사용이 힘들며, 많은 양의 정보를 기술하는데 어려움이 있다.

이러한 어려움을 극복하기 위해 1995년 3월 OCLC(Online Computer Library Center)와 NCSA(National Center for Supercomputing Applications)에서 지원한 메타데이터 워크샵에서는 네트워크상에 존재하는 정보 객체를 효율적으로 검색하는데 사용할 수 있는 15개의 메타데이터 엘리먼트[2, 3]를 아래와 같이 정의하였다.

- (1) Title : 문서의 이름 또는 화상/동화상의 설명구
- (2) Creator : 문서의 내용에 대해 책임을 지는 개인이나 단체
- (3) Subject : 문서에서 강조된 내용이나 문서 객체가

속하는 지식 분야

- (4) Description : 동영상 문서의 경우는 내용, 텍스트 문서의 경우는 요약
- (5) Publisher : 문서를 전자화하는데 책임을 갖는 개인이나 단체
- (6) Contributor : 문서의 편집자나 기술자
- (7) Date : 문서가 발표된 날짜
- (8) Type : 문서의 장르
- (9) Format : 문서의 화일 형태
- (10) Identifier : 문서를 유일하게 식별하는 숫자나 문자열
- (11) Source : 문서의 기원 혹은 역사
- (12) Language : 문서의 내용을 구성하는 언어
- (13) Relation : 다른 문서와의 관계
- (14) Coverage : 문서의 시공간적인 특징

(15) Rights : 문서를 접근하는 권한에 대한 내용

이는 문서의 저작자와 제공자에게 효율적인 기술 수단을 제공하고, 다양한 탐색 도구들 사이의 상호 운용성을 쉽게 하며, 분류 시스템들간의 사상이 가능하도록 한다. 그리고 사용자가 쉽게 이해할 수 있으며, 유사 문서들(Document-Like Objects, DLO)을 기술하기 위해 확장이 가능하다. 그러나 문서의 버전 관리와 다른 단체의 분류 시스템과 사상을 쉽게 하기 위한 확장성 그리고 메타데이터의 상속성에 대한 문제가 여전히 과제로 남아 있다.

한편 워크샷에서는 15개 메타데이터 엘리먼트들을 기존 HTML의 META 엘리먼트의 속성인HTTP-EQUIV, NAME, CONTENT 등을 이용하여 표현하는 방법을 고려하였으나, META 엘리먼트를 사용하는데 있

```

<!DOCTYPE DublinCore
<!ENTITY %
a.global          'type CDATA #IMPLIED
                  scheme CDATA "uncontrolled"
<!ENTITY %
packageType "package | dublinCore | packageRef"
<!ELEMENT
container - O
(%packageType)* >
<!ELEMENT
package - O
(metadata | metaGroup | %packageType)*>
<!ATTLIST
package
name          CDATA          #REQUIRED
URI           CDATA          #IMPLIED
version       CDATA          #IMPLIED
packageRef - O EMPTY
packageRef
URI           CDATA          #IMPLIED
name          CDATA          #REQUIRED
version       CDATA          #IMPLIED
<!ELEMENT
metaGroup - O
(#PCDATA | metadata | metaGroup)*
<!ELEMENT
metadata - O
(#PCDATA | metadata)*
<!ATTLIST
metadata
type          CDATA          #REQUIRED
sheme         CDATA          'uncontrolled'
show          (show | noshow | inherit) inherit
sortkey       CDATA          #IMPLIED
index         (index | noindex | asparent) asparent
<!ELEMENT
dublinCore - O (
title
| creator
| subject
| description
| publisher
| contributor
| date
| type
| format
| identifier
| source
| language
| relation
| coverage
| rights
| Media
| Keyword
| Location
| metadata)*
<!ATTLIST
dublinCore version CDATA #IMPLIED
<!ELEMENT
title - O
(#PCDATA)
<!ATTLIST
title
%a.global
    
```

<!ELEMENT	creator	- O	(#PCDATA)	>
<!ATTLIST	creator		%a.global	>
<!ELEMENT	subject	- O	(#PCDATA)	>
<!ATTLIST	subject		%a.global	>
<!ELEMENT	description	- O	(#PCDATA)	>
<!ATTLIST	description		%a.global	>
<!ELEMENT	publisher	- O	(#PCDATA)	>
<!ATTLIST	publisher		%a.global	>
<!ELEMENT	contributor	- O	(#PCDATA)	>
<!ATTLIST	contributor		%a.global	>
<!ELEMENT	date	- O	(#PCDATA)	>
<!ATTLIST	date		%a.global	>
<!ELEMENT	form	- O	(#PCDATA)	>
<!ATTLIST	form		%a.global	>
<!ELEMENT	format	- O	(#PCDATA)	>
<!ATTLIST	format		%a.global	>
<!ELEMENT	identifier	- O	(#PCDATA)	>
<!ATTLIST	identifier		%a.global	>
<!ELEMENT	source	- O	(#PCDATA)	>
<!ATTLIST	source		%a.global	>
<!ELEMENT	language	- O	(#PCDATA)	>
<!ATTLIST	language		%a.global	>
<!ELEMENT	relation	- O	(#PCDATA)	>
<!ATTLIST	relation		%a.global	>
<!ELEMENT	coverage	- O	(#PCDATA)	>
<!ATTLIST	coverage		%a.global	>
<!ELEMENT	rights	- O	(#PCDATA)	>
<!ATTLIST	rights		%a.global	>
<!ELEMENT	Media	- O	(Video Image Audio)*	>
<!ELEMENT	Video	- O	(#PCDATA)	>
<!ATTLIST	Video			
	fileLoc		CDATA	#REQUIRED
	fileName		CDATA	#REQUIRED
	frameRate		CDATA	#IMPLIED
	scanline		CDATA	#IMPLIED
	compress		(MPEG1 MPEG2)	MPEG2
<!ELEMENT	Image	- O	(#PCDATA)	>
<!ATTLIST	Image			
	fileLoc		CDATA	#REQUIRED
	fileName		CDATA	#REQUIRED
	resolution		CDATA	#IMPLIED
	height		CDATA	#IMPLIED
	width		CDATA	#IMPLIED
	colordepth		CDATA	#IMPLIED
	mainColor		CDATA	#REQUIRED
<!ELEMENT	Audio	- O	(#PCDATA)	>
<!ATTLIST	Audio			
	fileLoc		CDATA	#REQUIRED
	fileName		CDATA	#REQUIRED
	samplingRate		CDATA	#IMPLIED
	audioSize		(8 16)	16
	chanel		(mono stereo)	stereo
	audioType		(midi wave)	wave
<!ELEMENT	(Keyword Location) - -		(#PCDATA)	>]>

(그림 1) 확장된 Dublin Core DTD
(Fig. 1) Extended Dublin Core DTD

어 표준화 문제와 문서를 풍부하게 기술할 수 없는 문제점이 있어, 이후 이러한 문제를 해결하고 이 기종 시스템들간에 전자 문서 교환이 가능하도록 마크업 (Markup) 언어 구문을 정의하는 메타 언어인 SGML [7, 8]의 DTD로 기술하였다.

본 논문에서는 워크샵에서 제시한 Dublin Core 메타데이터 DTD를 바탕으로 멀티미디어 형태의 전자 문서를 체계적으로 관리하고 효율적으로 내용 기반 검색에 사용할 수 있도록 새롭게 확장한 DTD를 정의한다(<그림 1> 참조). Dublin Core는 주로 텍스트 문서에 대한 메타데이터를 기술한 것이기에 때문에 다양한 미디어 형태를 가지는 멀티미디어 전자 문서를 기술하기 위해서는 기본 메타데이터에 대한 확장이 필요하다. 즉, 정보 매체에 따라 효율적인 저장 및 검색을 지원하기 위해 아래와 같이 3가지 종류의 엘리먼트들과 관련 속성들을 추가하여 멀티미디어 전자 문서의 메타데이터로 사용한다(기존 Dublin Core에서 추가된 부분은 <그림 1>에서 굵은 글자체로 나타남).

- (1) Media : 문서상에서 지원되는 각종 멀티미디어 자료의 저장 및 검색을 지원하는 엘리먼트로서 다음과 같은 하위 엘리먼트들을 관리한다.
 - 1) Video : 동영상 정보의 저장과 검색 지원 엘리먼트
 - 2) Audio : 동영상 또는 문서와 관련된 소리 정보의 저장 및 검색 지원 엘리먼트
 - 3) Image : 문서상의 정지 화상 또는 사진 정보의 저장과 검색 지원 엘리먼트
- (2) Keyword : 검색의 효율성을 부가한 기능으로 빠른 검색 지원을 위한 엘리먼트
- (3) Location : 문서의 메타데이터 입력시 해당 실제 문서 주소 저장 엘리먼트

전자 문서에 포함된 멀티미디어 객체는 Media 엘리먼트에 의해 관리되며 형태에 따라 각각의 속성을 가진다. 또한 내용 기반 검색을 위해 각 객체에 대한 설명부가 존재한다. Keyword 엘리먼트는 문서의 주요 키워드, Location 엘리먼트는 실제 문서의 주소를 기술한다. 이러한 엘리먼트들을 추가함에 따라 멀티미디어의 종류에 따라 그 특성을 반영할 수 있게 되어 보다

체계적인 관리와 검색이 가능하게 된다.

3. 메타데이터 관리를 위한 데이터베이스 스키마 설계

본 장에서는 2장에서 제안한 멀티미디어 전자 문서 지원을 위한 확장된 Dublin Core 메타데이터의 DTD를 XOMT도로 나타내고, 이에 대한 객체 지향 데이터베이스 스키마를 설계한다.

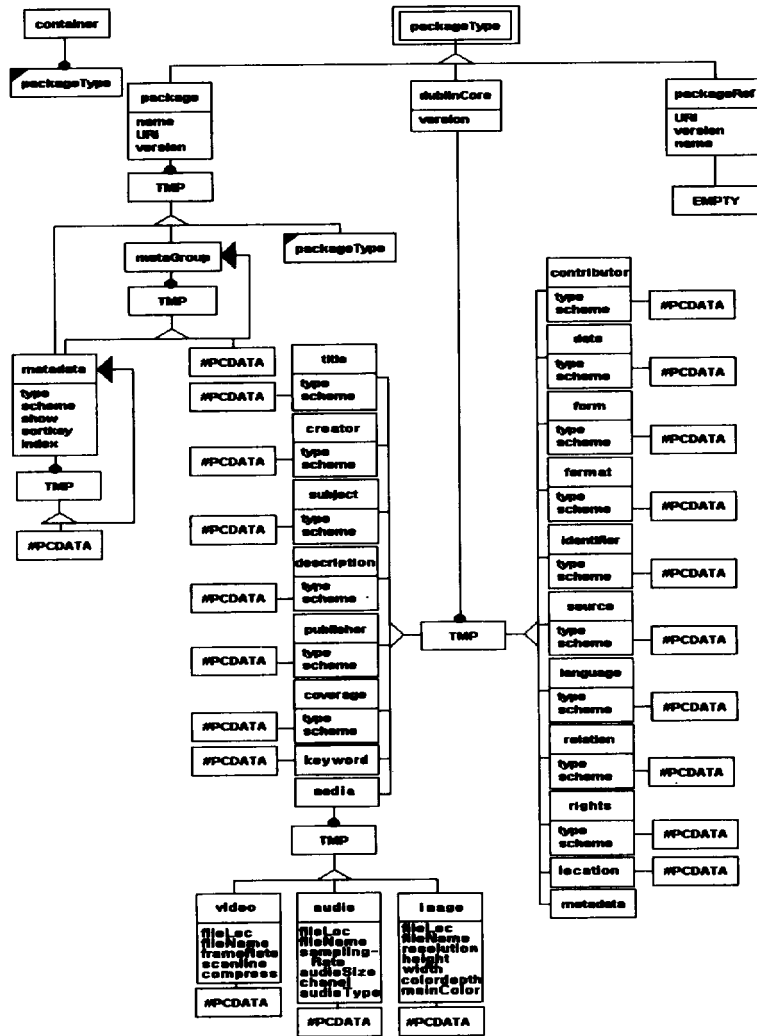
3.1 확장된 Dublin Core 메타데이터 DTD의 XOMT도

SGML DTD를 효율적으로 기술하기 위해서는 체계적인 절차에 따라 이를 작성할 수 있는 다이어그램 기법이 요구된다[7, 8, 10]. 여러 가지 다이어그램 기법들 중 본 논문에서는 XOMT(eXtended Object Modeling Technique)도[10]를 사용하는데, 이는 객체 지향 개발 방법론인 OMT[12]의 객체도를 확장하여 SGML DTD를 다이어그램으로 표현하는 방법으로, SGML DTD 문서의 데이터베이스 스키마 설계에 매우 유용하다[10]. 그리고 XOMT는 객체 다이어그램 기법이므로 메타데이터 데이터베이스 스키마 변경시 DTD의 형식을 그대로 유지하면서 데이터베이스 스키마로의 변환이 아주 용이하다[11]. 그러므로 Dublin Core의 속성 중 사용자의 기술 요구가 달라지거나, 시간이나 기능에 따라 멀티미디어 문서가 변할 때 구성 엘리먼트를 쉽게 변경시킬 수 있는 장점이 있다. 따라서 본 절에서는 확장된 Dublin Core DTD에 대한 객체 지향 데이터베이스를 설계하기 위해 XOMT도를 이용한다.

(그림 2)에서 SGML DTD의 엘리먼트는 단선 사각형으로, 엔티티는 이중선 사각형으로, 상속 관계는 작은 세모로, 관계성 대응수는 선 끝의 동그라미로 표현되었다. 그리고 사각형 왼쪽 모서리 검은색 삼각형은 공유 객체를 나타낸다.

3.2 데이터베이스 스키마의 설계

SGML DTD로 정의된 데이터베이스 기술을 이용하여 확장된 Dublin Core 메타데이터는 체계적으로 관리할 필요가 있다. 이때 다양한 종류의 메타데이터를 쉽게 관리하고, 사용자의 요구를 충분히 수용하기 위해서는 객체 지향 데이터베이스를 활용하면 효과적이다 [9]. 질의 처리 등 일반적인 데이터베이스 시스템이 제공하는 특성을 제공받을 수 있으면서, Dublin Core가

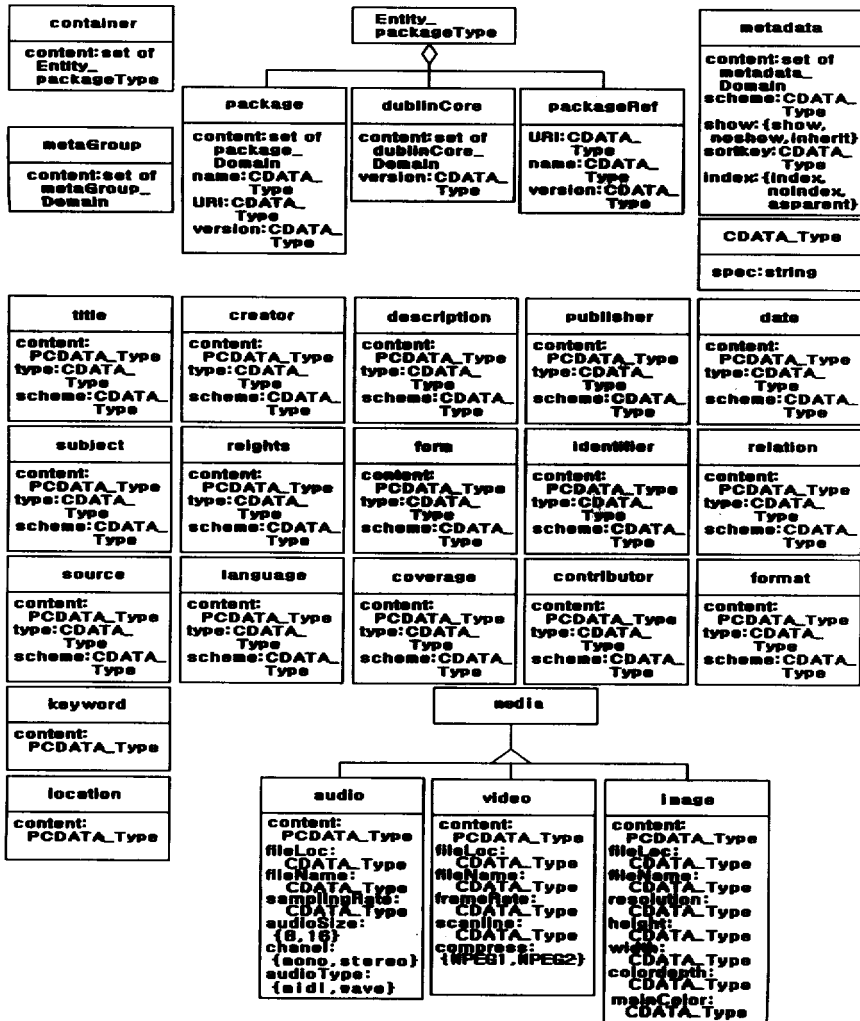


(그림 2) 확장된 Dublin Core DTD의 XOMT도
 (Fig. 2) XOMT diagram for Extended Dublin Core DTD

지향하는 확장성과 상속성을 쉽게 구현할 수 있기 때문이다.

본 절에서는 3.1절의 확장된 Dublin Core 메타데이터를 객체 지향 데이터베이스로 사상하기 위해 설계한 데이터베이스 스키마(그림 3)와 속성 도메인 클래스 타입 계층(그림 4)을 OMT의 객체도로 표현한다[11]. 즉, 클래스는 클래스명과 속성을 갖는 네모로 나타내고 클래스들 간의 상속 관계(is-a)는 작은 세모, 부품 관계(part-of)는 마름모로 표현하였다.

본 절에서 설계한 스키마는 [11]에서 제안한 SGML DTD에 대한 데이터베이스 스키마의 자동 생성 기법을 이용하여 (그림 2)의 DTD에 대해 생성된 데이터베이스 스키마 정보를 디지털 도서관에서 멀티미디어 전자 문서 관리에 사용할 수 있도록 변형시킨 스키마이다. 즉, [11]에서는 SGML DTD에 나타나는 모든 엘리먼트가 동일한 방식으로 단독 클래스로 정의되는데, 여기서는 확장성과 상속성 등을 이용하기 위해 클래스들 사이의 관계성을 보다 체계화시켰다.

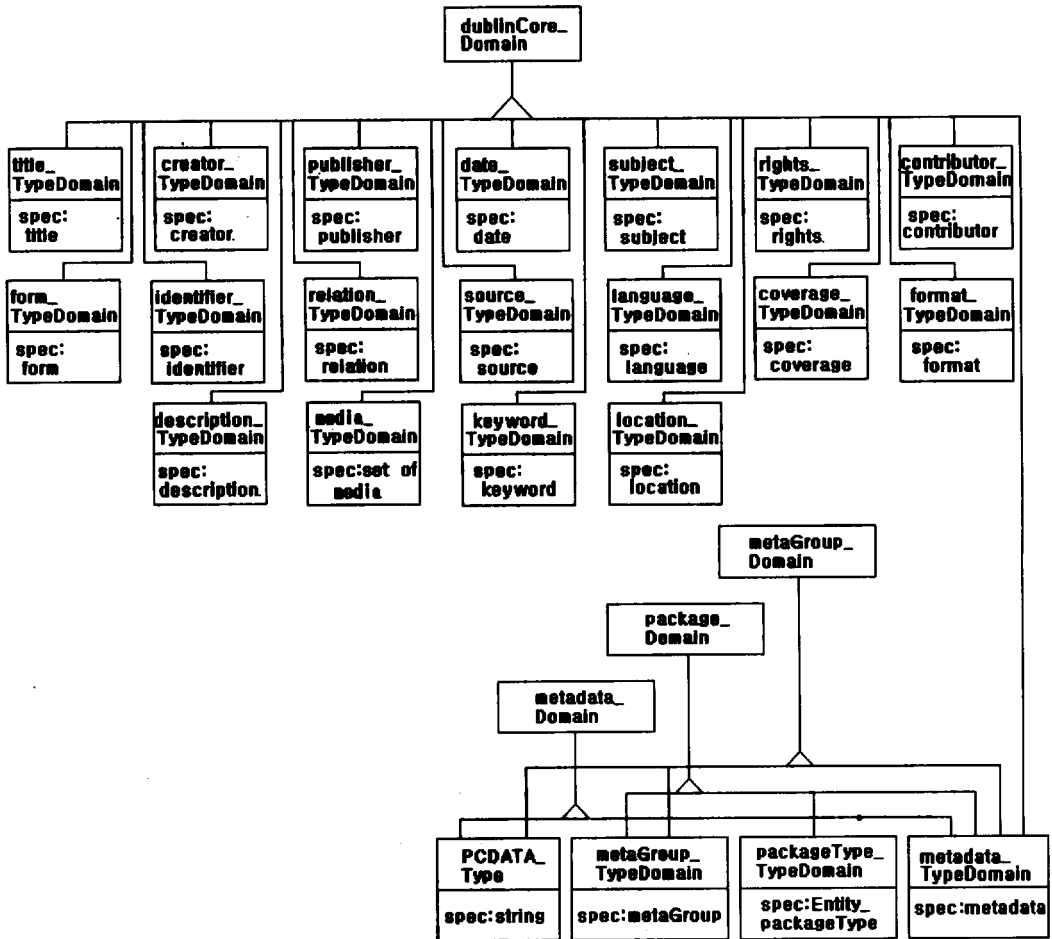


(그림 3) 확장된 Dublin Core 메타데이터를 위한 데이터베이스 스키마
(Fig. 3) Database schema for Extended Dublin Core metadata

(그림 3)은 DTD의 엔티티와 엘리먼트들을 각각의 클래스들로 정의한 것이다. 최상위 클래스인 container 클래스는 Entity_packageType 클래스의 집합으로 정의되며, Entity_packageType 클래스는 package, dublinCore, packageRef 클래스 중 하나의 클래스를 가진다.

확장된 Dublin Core를 구성하는 엘리먼트들 중 15개의 기존 엘리먼트 클래스들은 content, type 그리고 scheme을 속성으로 가지며, 확장된 엘리먼트중 Media 클래스는 Image, Audio 그리고 Video 클래스를 하위

클래스로 가진다. Image 클래스는 문서에 포함된 이미지 데이터를 관리할 수 있는 7개의 속성, Video 클래스는 비디오 데이터를 관리할 수 있는 5개의 속성, 그리고 Audio 클래스는 소리 정보를 관리하기 위한 6개의 속성을 가진다. Keyword 클래스와 Location 클래스는 PCDATA_Type 클래스, meta 클래스는 metadata_Domain의 집합과 type, schema, show, sortkey와 index를 속성으로 가지며, metaGroup 클래스는 metaGroup_Domain의 집합으로 구성되어 있다.



(그림 4) 확장된 Dublin Core 메타데이터를 위한 속성 도메인 클래스 타입 계층
 (Fig. 4) Attribute domain class type hierarchy for Extended Dublin Core metadata

(그림 4)는 확장된 Dublin Core 메타데이터 데이터베이스 스키마의 속성 도메인 클래스 타입 계층이다. 이는 Dublin Core DTD의 내용 모델부에 나타나는 자기 자신 참조 등 복잡한 형식의 도메인 클래스를 체계적으로 다루기 위해 타입 계층을 이용하여 정의한다.

(그림 3)에 나타난 메타 클래스에는 멀티미디어 전자 문서의 실제 메타 정보가 저장되고, (그림 4)의 속성 도메인 타입 계층에는 메타 정보를 참조하기 위한 객체 식별자와 클래스의 속성들 사이에 나타나는 타입 계층을 참조하기 위한 구조가 저장된다.

4. 확장된 Dublin Core 메타데이터 관리 시스템(xD-MMS)

이 장에서는 멀티미디어 문서 정보 검색을 지원하는 확장된 Dublin Core 메타데이터를 관리하는 시스템인 xD-MMS(eXtended Dublin Core Metadata Management System)를 제안한다.

xD-MMS의 설계에는 객체 지향 개발 방법론인 OMT [12]를 사용한다. 여기서는 전체 구조도, 객체도, 그리고 기능도만을 다룬다.

4.1 xD-MMS 구조도

xD-MMS는 전자 문서를 효율적으로 관리하고 검색하기 위해 확장된 Dublin Core 메타데이터의 정의를 기반으로 설계된 데이터베이스 스키마를 관리하는 시스템이다.

(그림 5)는 xD-MMS 전체 구조도이다.

xD-MMS는 편집 부시스템, 확장된 Dublin Core 메타데이터 관리 부시스템, 문서 검색 부시스템 등 크게 3개의 부시스템으로 이루어진다.

편집 부시스템은 Web상에 산재되어 있는 멀티미디어 전자 문서들의 효율적인 검색과 관리를 위해 본 논문에서 제안한 확장된 Dublin Core 메타데이터로의 편집을 위한 확장된 Dublin Core 메타데이터 편집기, 사용자가 전자 문서 검색시 SQL 질의문을 통해 검색할 수 있도록 질의문을 생성, 편집하는 문서 질의 편집기로 이루어진다.

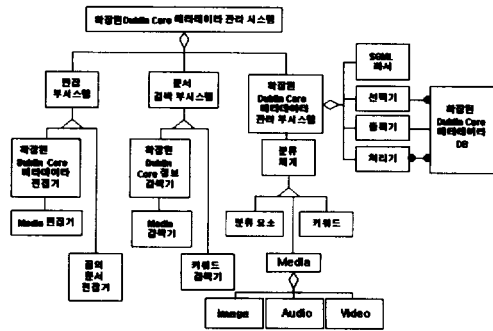
확장된 Dublin Core 메타데이터 관리 부시스템은 편집기에서 작성된 메타데이터를 설계된 데이터베이스 스키마에 저장시키는 등록기, 스키마에서 해당되는 메타데이터를 선택하는 선택기, 메타데이터의 관리를 위한 처리기, 그리고 SGML 파서로 구성된다.

문서 검색 부시스템은 크게 Dublin 워크샵에서 제시한 15개 Dublin Core 메타데이터와 키워드를 통한 문서 검색을 지원하는 키워드 검색기, 전자 문서에 포함된 멀티미디어 정보 검색을 지원하는 Media 검색기로 이루어진다.

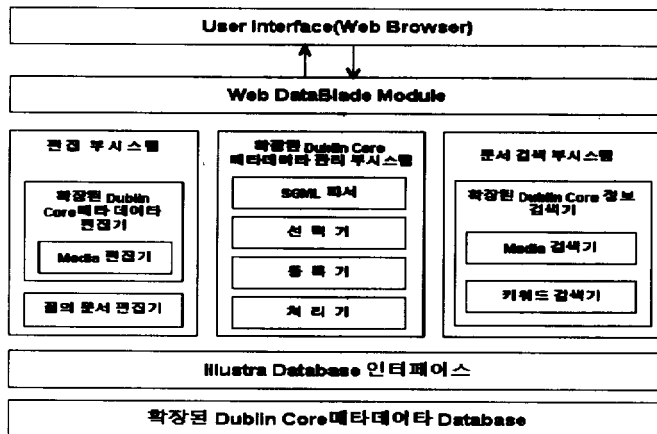
4.2 xD-MMS 객체도

(그림 6)은 xD-MMS의 전체 구조를 OMT의 객체도(Object Diagram)로 나타낸 것이다.

편집 부시스템에서는 확장된 Dublin Core 메타데이터 편집기와 문서 질의 편집기를 통해 전자 문서의 메타데이터를 편집하며, 문서내의 멀티미디어 정보에 대한 편집은 Media 편집기에 의해 관리된다. 확장된 Dublin Core 메타데이터 관리 부시스템의 선택기는 분류 체계에 따라 메타데이터를 관리하며 처리기를 통해 처리된다. 분류 체계는 제시된 15개 Dublin Core 메타데이터 엘리먼트, 키워드 그리고 멀티미디어 정보를 위한 Image, Audio, Video로 구분한다. 문서 검색기는 분류 체계에 따른 검색을 제공한다.



(그림 6) xD-MMS의 객체도
(Fig. 6) Object diagram for xD-MMS



(그림 5) xD-MMS의 구조
(Fig. 5) Architecture of xD-MMS

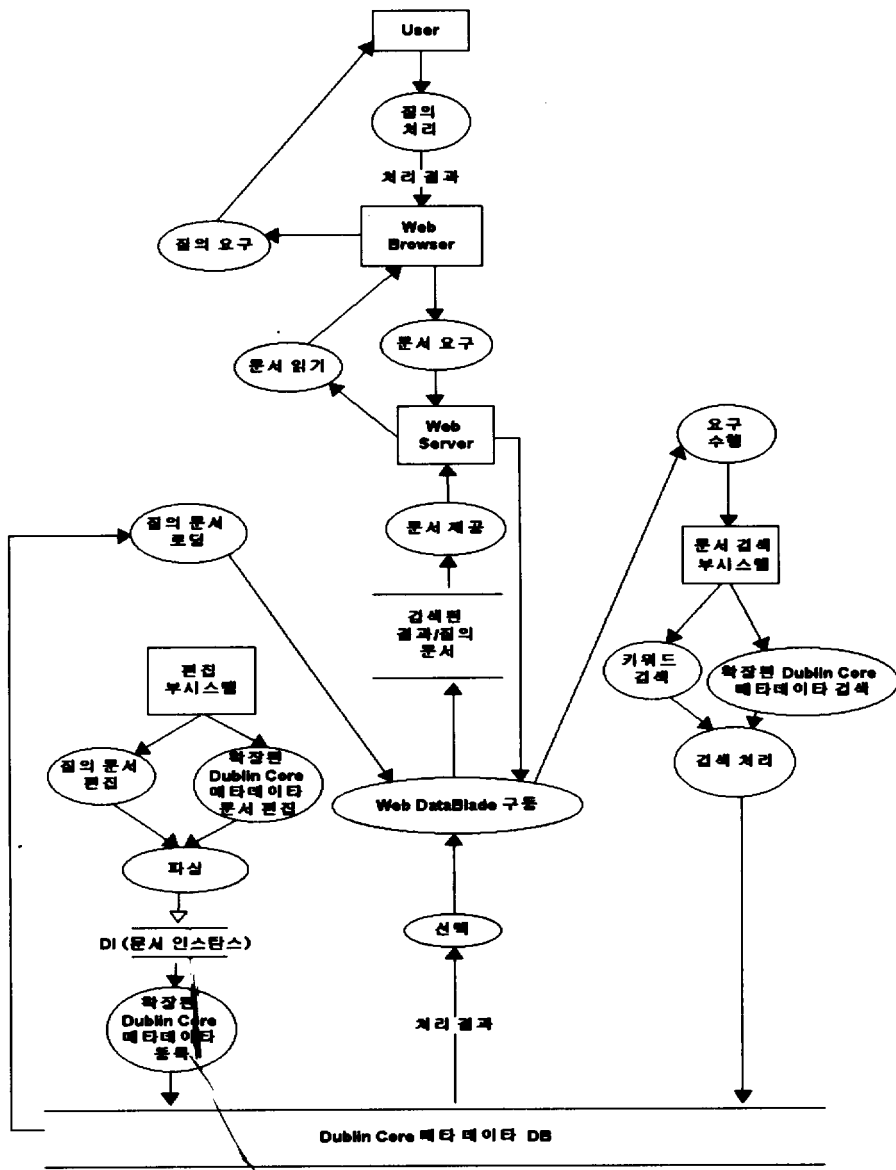
4.3 xD-MMS 기능도

(그림 7)은 xD-MMS의 기능적인 관점을 모아 OMT의 기능도인 자료 흐름도(Data Flow Diagram)로 표현한 것이다.

Web 검색을 통해 검색된 전자 문서의 메타데이터는 편집 부시스템을 통해 편집되어 데이터베이스에 저

장된다. 사용자는 검색을 위한 질의문에 응답하여 원하는 문서의 정보를 획득할 수 있으며, 질의 처리는 DBMS의 기능을 이용한다.

데이터의 흐름은 TCP/IP 프로토콜을 이용한 Web 기반으로 Illustra가 제공하는 Web DataBlade 모듈[14]을 사용한다.



(그림 7) xD-MMS의 기능도
(Fig. 7) Data flow diagram for xD-MMS

5. xD-MMS의 구현

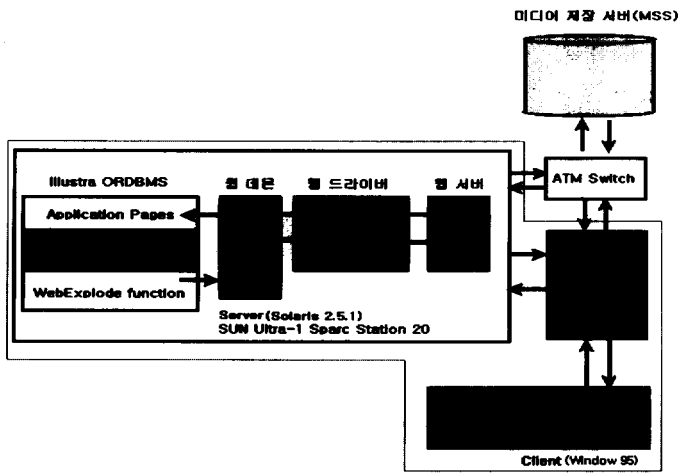
이 장에서는 4장에서 설계한 xD-MMS의 구현 내용을 기술한다. (그림 8)은 xD-MMS의 구현 환경을 도시한 것으로서 사용자는 Web 브라우저를 통하여 서버측의 xD-MMS에 접속하며, xD-MMS는 Illustra DBMS위에서 구축되었다.

향후 ATM 기반에서 멀티미디어 문서를 효율적으로 저장하기 위해 미디어 저장 장비(Media Storage

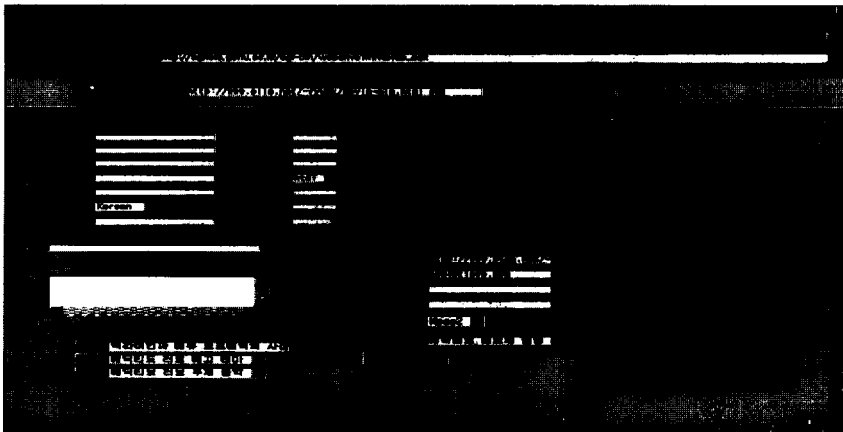
Server, MSS)가 서버와 연결되어 있으나, 본 연구에서는 그림상의 가는 실선 부분만을 구현하였다.

5.1 편집 부시스템

편집 부시스템은 크게 세부분으로 구성된다. 첫 번째, 상위 창에서는 Web에 존재하는 문서의 실제 URL을 입력한다, 두 번째, 우측 창은 상위 창에서 입력한 URL에 의해 연결된 문서의 내용을 보여준다. 이는 메타데이터를 작성하는 사용자가 보다 정확한 메타데이



(그림 8) xD-MMS의 구현 환경
(Fig. 8) Implementation environment for xD-MMS



(그림 9) xD-MMS의 편집 부시스템 인터페이스
(Fig. 9) Edition subsystem interface of xD-MMS

타를 편집할 수 있도록 도와 준다. 세 번째, 좌측 창은 메타데이터 편집 창으로서 우측 창에 나타난 문서를 이용하여 메타데이터를 입력, 저장할 수 있게 한다. 이는 크게 4가지 형태로 나뉘어지는데, (1) Dublin Core 에서 제시된 15개 메타데이터, (2) 키워드, (3) 문서 요약 정보, 그리고 (4) 멀티미디어 관련 정보로 구성되어 있다. 멀티미디어 관련 정보에서는 편집하고자 하는 항목에서 "Add" 버튼을 이용하여 문서내에 포함된 각 미디어에 대한 상세 속성 정보를 입력할 수 있는 스크립트(scriptor)가 나타난다. 이들은 확장된 Dublin Core 메타데이터를 기반으로 전자 문서를 관리, 검색할 수 있는 자료로 이용된다.

5.2 검색 부시스템

(그림 10)은 검색 부시스템의 사용자 인터페이스 화면이다.

검색 부시스템은 크게 두 부분으로 구성된다. 좌측의 검색 창과 우측의 검색 결과를 목록으로 나타내어 주는 창으로 구성되어 있다. 검색 창에서는 검색하고

자 하는 질의 항목을 사용자가 입력할 수 있으며, 멀티미디어 전자 문서를 크게 3가지로 검색할 수 있다. 첫 번째는, Dublin Core 메타데이터를 사용하는 방법이다. 이는 편집 부시스템에서 해당 멀티미디어 전자 문서에 대해 Dublin Core의 15개 메타데이터로 멀티미디어 전자 문서를 정의하여 입력하고, 이를 검색시 사용하는 방법이다. 두 번째는, 키워드를 사용하는 방법이다. 이는 해당 멀티미디어 전자 문서를 대표하는 단어를 편집 부시스템에서 키워드 메타데이터에 하나 이상의 단어를 입력하여 검색시 사용하는 방법이다. 앞의 두 방법으로 검색을 수행하면 해당 메타데이터와 관련된 텍스트 데이터, 멀티미디어 데이터를 모두 검색하여 그 결과를 보여준다. 마지막으로, 멀티미디어 메타데이터를 사용하는 방법이다. 이는 멀티미디어 전자 문서에서 나타난 멀티미디어 데이터를 화일명과 요약 정보 메타데이터로 정의하고, 이를 검색시 사용하는 방법이다. 그리고 우측의 창은 검색된 결과를 문서들의 목록으로 나타낸다. 목록에는 등록된 문서의 제목, 문서 요약 정보의 일부분, 해당 URL과 문서내에 포함된



(그림 10) 검색 부시스템 인터페이스의 초기 화면
(Fig. 10) Initial screen of retrieval subsystem interface

멀티미디어 관련 정보를 아이콘화하여 보여 준다.

DC 아이콘은 검색된 문서의 전체 메타데이터, 메모지 형태의 아이콘은 문서의 요약 정보, 카메라 형태의 아이콘은 이미지 정보, 릴 테이프 모양의 아이콘은 비디오 정보, 그리고 스피크 모양의 아이콘은 오디오 정보의 포함 여부를 나타낸 것으로서, 각 아이콘을 클릭하면 해당 정보들이 JAVA 스크립트나 새로운 브라우저에 의해 나타난다.

(그림 10)은 검색 부시스템 인터페이스 화면의 초기 화면이며, 검색된 결과는 (그림 11)과 같다.

6. 결론 및 향후 연구 과제

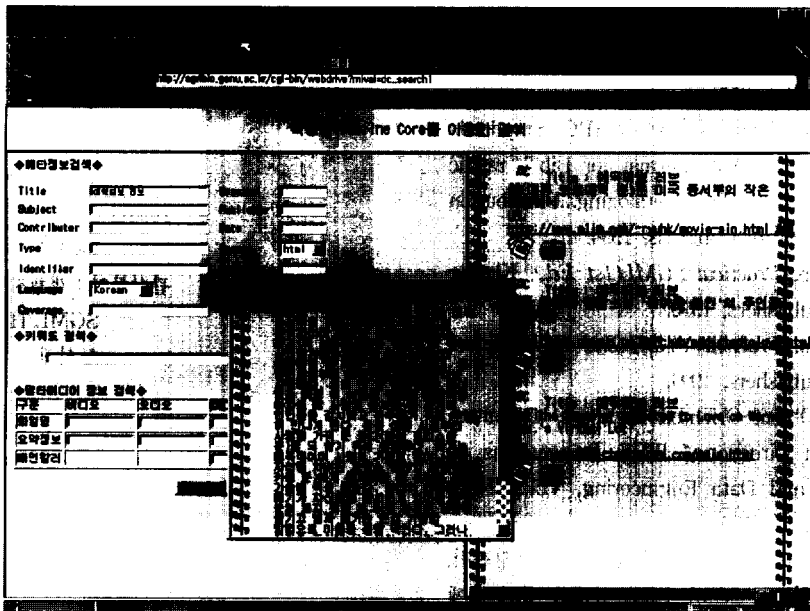
최근 디지털 도서관에서는 텍스트 뿐만 아니라 이미지, 비디오, 오디오를 포함하는 각종 멀티미디어 정보의 사용 요구가 증가되고 있다.

이러한 멀티미디어 전자 문서의 검색 도구로서 현재 많이 사용되고 있는 것은 Lycos[http://lycos.cs.cmu.edu/]나 WebCrawler[http://webcrawler.com/] 등이다. 이들은 인터넷에 있는 검색 정보를 주제별, 단어별로 분류하여 위치 정보만을 데이터베이스화하여 링크하였다. 그런데 이것은 화일명 자체가 메타데이터이므로

문서에 대한 유용한 정보가 부족하여, 검색어를 올바르게 선정하지 않으면 효율적으로 사용자가 요구하는 문서를 정확하게 검색하기 힘들 뿐만 아니라 특히, 내용기반의 검색이 어렵다.

본 논문에서는 디지털 도서관의 구축에 사용될 수 있는 멀티미디어 전자 문서의 메타데이터를 관리하기 위해 Dublin Core 메타데이터를 확장하고 이를 위한 객체 지향 데이터베이스 스키마를 설계하였다. 그리고 이를 관리하는 xD-MMS 시스템을 설계 및 구현하였다. 멀티미디어 전자 문서의 검색을 위해 텍스트 전자 문서를 주로 고려한 Dublin Core 확장하고 확장된 Dublin Core DTD를 객체 지향 데이터베이스 스키마로 변환함으로써 메타데이터의 확장성, 수정성, 선택성, 반복성을 지원하며, 이를 관리하는 시스템 xD-MMS는 멀티미디어 전자 문서에 대한 사용자의 검색을 용이하게 하고 메타데이터를 효율적으로 관리한다는 특징이 있다.

그러나 설계된 데이터베이스 스키마와 xD-MMS를 실제 응용에 사용하기 위해서는 아직도 몇가지 보완할 점이 있다. 첫째, 제한한 스키마의 확장이다. 새로운 메타데이터의 필요시 이를 위한 확장이 필요하다. 둘째, 메타데이터 뿐만 아니라 실제 전자 문서와 여기에



(그림 11) 검색 결과
(Fig. 11) Retrieval results

포함된 멀티미디어 자료들을 보다 통합적으로 관리할 수 있는 시스템의 개발이 요구된다. 마지막으로, 최근 시간 관계성을 지원하는 HyTime[13]이 제안되었는데, 하이퍼미디어 응용을 위해서는 이의 지원이 반드시 요구된다. 따라서 하이퍼미디어 전자 문서의 관리를 위해 SGML로 설계된 확장된 Dublin Core 메타데이터가 HyTime도 지원할 수 있도록 확장되어야 한다.

참 고 문 헌

[1] F. M. Kappe, "Aspects of the Modern Multimedia Information System," Ph. D Dissertation, Graz University, Austria, 1991.

[2] S. Weibel, J. Godby, and E. Miller, "OCLC/NCSA Metadata Workshop Report," Office of Research, OCLC Online Computer Library Center, Inc. 1995.

[3] L. Burnard, E. Miller, L. Quin, and C. Sperberg-McQueen, "A Syntax for Dublin Core Metadata," Office of Research, OCLC Online Computer Library Center, Inc. 1996. 4.

[4] Stephen P. Harter, "What is a Digital Library? Definitions, Content, and Issues," 1996년 디지털 도서관 국제학술회의 논문집, pp.8-17, 1996. 9.

[5] Makota Nagao, "Multimedia Digital Library: ARIADNE," 1996년 디지털도서관 국제학술회의 논문집, pp.26-32, 1996. 9.

[6] Network Development and MARC Standards Office (Ed.), USMARC Format for Bibliographic Data, Washington, DC, Cataloging Distribution Service, Library of Congress, 1994.

[7] E. Herwijnen, *Practical SGML(1st Ed.)*, Kluwer Academic Publishers, 1990.

[8] E. Herwijnen, *Practical SGML(2nd Ed.)*, Kluwer Academic Publishers, 1994.

[9] W. Kim, "Object-Oriented Database: Definition and Reserch Directions," IEEE Transactions on Knowledge and Data Engineering, Vol.3, No.1, pp.327-341, Sept. 1990.

[10] 박인호, 한에노, 정은주, 김은정, 배종민, 강현석, 김완석, "XOMT : SGML DTD 설계를 위한 객체 다이어그램 기법," 한국정보과학회 논문지, 제3권, 제3호, pp.228-237, 1997. 6.

[11] 한에노, 박인호, 강현석, 김완석, "SGML 문서의 관리를 위한 객체지향 데이터베이스 설계," 한국정보처리학회 논문지, 제4권, 제3호, pp.670-684, 1995. 3.

[12] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorensen, *Object-Oriented Modeling and Design*, Prentice-Hall, 1991.

[13] ISO, *Hypermedia/Time-based Structuring Language : HyTime(ISO 10744)*, 1992.

[14] Illustra Information Technologies Inc. 편집부, *Illustra Web DataBlade Module User's Guide*, Illustra Information Technologies Inc., Release 2.2, July 1996.



권 재 길

1988년 경상대학교 전산통계학과 (이학사)
 1993년 경상대학교 전자계산학과 (공학석사)
 1997년 경상대학교 전자계산학과 박사과정 수료

1995년~현재 창원전문대학 전자계산과 전임강사
 관심분야 : 프로그래밍언어, 하이퍼미디어, 디지털 도서관, 멀티미디어 정보검색



장 원 호

1996년 경상대학교 컴퓨터과학과 졸업(이학사)
 1998년 경상대학교 전자계산학과 (공학석사)
 관심분야 : 객체지향 데이터베이스, SGML/Hy-Time, 디지털 도서관



김 현 주

1988년 경상대학교 전산통계학과 (이학사)
1990년 숭실대학교 전자계산학과 (공학석사)
1994년~1997년 제일정밀 공업 (주) 프린터 연구실 근무

1997년~현재 경상대학교 전자계산학과 박사과정 재학 중
관심분야 : 컴파일러, 정보 검색, 디지털 도서관



배 증 민

1980년 서울대학교 사범대학 수학과(학사)
1983년 서울대학교 대학원 계산통계학과(이학석사)
1995년 서울대학교 계산통계학과 (이학박사)

1982년~1984년 한국전자통신연구원 연구원
1984년~현재 경상대학교 컴퓨터과학과 교수
관심분야 : 병렬 프로그래밍 언어, 디지털 도서관



박 인 호

1990년 경상대학교 전산통계학과 (이학사)
1997년 경상대학교 전자계산학과 (공학석사)
1997년~현재 경상대학교 전자계산학과 박사과정 재학중

관심분야 : 객체지향 데이터베이스, SGML/HyTime, 멀티미디어



강 현 석

1981년 동국대학교 전자계산학과 (학사)
1983년 서울대학교 계산통계학과 (이학석사)
1989년 서울대학교 계산통계학과 (이학박사)

1981년~1984년 한국전자통신연구원 연구원
1984년~1993년 전북대학교 전자계산학과 부교수
1993년~현재 경상대학교 컴퓨터과학과 교수
관심분야 : 객체지향 데이터베이스, 컴퓨터 그래픽스, 멀티미디어 전자문서관리