



## 차세대 디지털도서관에서의 콘텐츠 재활용 기술

한국전자통신연구원 김태현  
충남대학교 이용배·맹성현\*

### 1. 서론

기존의 도서관들이 그래왔듯이 전자 도서관의 기본적인 역할은 자료를 보관하고 이를 많은 사람들이 손쉽게 찾아 사용할 수 있도록 하는데 있다. 그러나, 컴퓨팅 시스템의 일반화로 인해 수 많은 전자 자료들이 생성됨에 따라 전자 도서관의 역할은 단순히 자료를 보관하고 이용하도록 하는데 그치지 않고, 만들어진 자료를 효과적으로 구성함으로써 사람들이 쉽게 그 내용을 파악할 수 있도록 하고, 또한 중복된 자료 보관에 따른 여러 가지 문제점을 해소하도록 하는 것으로 까지 확대되고 있다. 이는 전자 도서관의 시스템적인 관점에서 다루기 보다는 정보구성 방법, 즉 문서표현 양식의 관점에서 새로운 표현 방식들을 도입함으로써 해결할 수 있다.

초기의 전자 도서관 시스템은 단순히 텍스트, 이미지, 오디오, 비디오 등의 다양한 자료를 각 자료의 형식에 따라 독립적으로 저장 및 접근하도록 하는 방식을 취하였다. 이는 일반 도서관 시스템의 형식을 그대로 전자화시킨 형태로 전자화에 따라 얻을 수 있는 여러 가지 이점들을 취하지 못하게 된 중요한 원인 중에 하나가 되었다. 이러한 문제에 대한 인식에서 출발하여, 다양한 형식의 전자 자료들을 결합하여 하나의 통일된 뷰로 제공하고자 하는 노력들이 생겨나게 되었는데, 대표적인 예로 다가문서(Multivalent document), 워릭 프레임워크(Warwick framework), 가상문서(Virtual Document), 토픽맵(XML Topic Map) 등의 표현 양식들이 있다. 이러한 표현 양식들은 모든 전자 자료들을 하나의 통일된 표현방식으로 결합함으로써 사용자들이 분산되어 있는 전자 개체들을 하나의 문서를 다루듯 사용할 수 있게 하며, 동

일한 개체를 여러 개의 서로 다른 통합 문서에 포함시켜 사용할 수 있도록 함으로써 자료의 중복 저장에 따른 저장소의 낭비나 자료 불일치 등의 문제를 해결하도록 하고 있다. 최근 들어서는 이러한 단순 자료의 통합에서 한 단계 더 나아가 자료 간의 관계를 정의하고, 이에 의미를 부여함으로써 보다 세련되고 정제된 자료를 제공하도록 하려는 노력이 이루어지고 있다.

이 글에서는 위에서 말한 여러 가지 전자 자료 표현양식에 대해 차례로 살펴보고, 이러한 표현양식들을 사용한 전자 도서관 시스템에 대해 다루어보기로 한다.

### 2. 문서 프레임 관련기술

#### 2.1 다가문서(Multivalent document)

90년대 중반부터 시작한 버클리 대학의 디지털도서관 프로젝트에서는 다가문서(Multivalent Document, 이하 다가문서)라는 개념의 새로운 문서형태를 정의[9]하고 이를 디지털도서관 시스템에 적용 [6,7,8]하려는 시도를 하고 있다.

디지털도서관에 존재하는 같은 내용의 여러 형태의 문서들을 하나의 일관된 인터페이스를 통해서 접근하려면 응용차원의 프로토콜이 필요하다. 예를 들어, 그림 1과 같이 베토벤의 교향곡 5번 운명은 오디오 문서로 저장되어 있을 수 있고 악보 형태의 이미지로 저장되어 있거나 교향곡의 느낌이나 내용을 설명한 텍스트 형태로 저장되어 있을 수 있다. 이들은 모두 같은 문서이지만 서로 다른 형태이므로 디지털도서관 시스템에서 이것들은 통합하고 관리하기 위해서는 일관된 하나의 문서구조가 요구된다. 이러한 이유로 진행된 연구결과물이 다가문서이다.

\* 중신회원

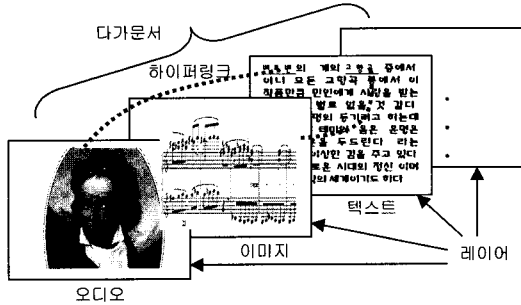


그림 1 다가문서

다가문서는 여러 형식의 문서들을 각각의 레이어(layer)로 나누어 하나의 통합된 문서로 관리한다. 각 레이어에서는 다른 레이어의 데이터로 하이퍼 링크나 주석을 만들 수 있으며 링크나 주석도 독립된 레이어로 관리된다. 다가문서의 구성[9]은 다음과 같다.

- 문서 트리(Document tree) : 다가문서를 편집하고 재현하기 위한 기본 자료구조로서 XML로 기술된 허브 문서(Hub Document)를 파싱하여 돔(DOM : Document Object Model)을 생성한다. 다가문서로 표현되는 HTML이나 TeX DVI, PDF 등 여러 형태의 문서들은 허브 문서에 기록되며 파싱되어 모두 문서 트리로 정보를 나타낼 수 있다.

- 미디어 활성화(Behaviors) : 다가문서에 연결된 미디어 데이터들을 동작시킬 수 있는 단위 응용프로그램이다. 즉, 사용자의 요구시에 텍스트, 비디오, 이미지, 하이퍼링크 등의 미디어들은 몇 개의 미디어 활성화기의 자동 조합으로 사용자에게 보여진다.

- 허브 문서(Hub Document) : 다가문서 안에 들어갈 미디어들과 미디어들을 동작시키는 미디어 활성화기를 XML 형식으로 기술한다. 문서 트리 생성시에는 이 허브 문서를 참조하여 돔을 구성한다.

다가문서는 여러 미디어 형태로 존재하는 문서들을 하나로 통합하여 일관된 형태로 의미를 전달하는데 효율적이다. 특히, 레이어 단위로 문서를 구성하므로 이미지 레이어 위에 주석이나 링크 레이어를 생성하여 학습자들에게 의미전달을 하는데 매우 효과적이다. 또한 다가문서 안에 포함되는 내용을 링크로 연결하여 구성할 수 있으므로 멀티미디어 문서 재사용이나 저장공간을 감소시키는 장점이 있으며 문서를 버전별로 관리하기가 쉽도록 프레임이 구성되어 있다. 이와 같이 문서를 표현하는데 있어 여러 가지 장점이 있지만 몇 가지 보완해야 할 기능들이 남아있다.

- 다양한 미디어 타입이 혼합되어 나타나는 XML로 기술된 문서는 다가문서로 표현하기 어렵다. 현재는 텍스트나 이미지, 오디오 등의 독립된 데이터에 한정하여 다가문서를 구성하고 있다.

- 새로운 종류의 문서가 발생하면 이를 처리할 수 있는 미디어 활성화기를 추가적으로 만들어야 한다.

- 다가문서를 재현하기 위해서는 다가문서 전용 브라우저[5]를 이용해야 한다.

## 2.2 워릭 프레임워크(Warwick framework)

워릭 프레임워크는 1996년 영국의 워릭(Warwick) 지방에서 열린 컨퍼런스에서 제안되었으며 이는 다양한 메타데이터와 자원들을 논리적으로 특정 형식을 가지는 패키지(package)라는 일관된 관점으로 다룸으로써 메타데이터와 자원을 효율적으로 구성 및 관리할 수 있도록 하는데 목적이 있다. 초기 워릭 프레임워크는 다양한 타입을 가지는 패키지들을 컨테이너(container)라는 하나의 구조에 포함시키기만 하는 단순한 형태를 띄고 있었다. 이로 인해 사용자들이 컨테이너에 있는 각 패키지에 접근하여 정보를 얻고자 할 때 컨테이너에 어떤 패키지들이 있는지 모두 훑어보아야 한다는 문제점이 있었다. 이러한 문제점의 인식에서 출발하여 새롭게 제안된 것이 워릭 프레임워크 카탈로그(Warwick Framework Catalog: WFC)인데, 이는 컨테이너에 포함되어 있는 각 패키지들과 이들 간의 관계에 대한 설명적인 정보들을 포함하고 있는 목록으로써, 이 카탈로그 또한 하나의 패키지로써 컨테이너 내에 포함된 형태로 다루어졌다. 다음의 그림 2는 WFC를 포함한 형태의 컨테이너 예이다. 이 컨테이너는 데이터 패키지 P2와 P2에 대한 서지정보를 가지고 있는 메타데이터인 더블린 코어 레코드[10](Dublin Core Record) P1, 그리고 P2에 대한 접근권한 목록을 가지고 있는 패키지인 P3로 이루어지며, P2와 P1, P2와 P3의 관계를 기술하는 패키지인 P0가 컨테이너의 첫 패키지로 포함되어 있다. 따라서, 사용자는 P0를 참조하여 컨테이너 내에 있는 모든 패키지에 대한 기본적인 정보를 얻을 수 있기 때문에 패키지에 대한 정보를 파악하는데 드는 노력을 줄일 수 있게 된다[13].

WFC에서 한 단계 더 확장된 개념으로 DAR (Distributed Active Relationship)이 있다. 이는 메타데이터도 하나의 데이터일 뿐이라는 관점에서 시작

하여 모든 패키지를 동일하게 자원 패키지로 취급하고, 이들 간의 관계를 단순한 이름이 아닌 분산환경에서 디지털 자원의 유일한 식별자인 URI(Uniform Resource Identifier)를 이용하여 표현함으로써 관계까지도 해석 가능한 하나의 자원으로 생각할 수 있게 한 것이다. 다음의 그림 3은 이러한 DAR의 개념을 사용하고 있는 컨테이너의 예이다[15].

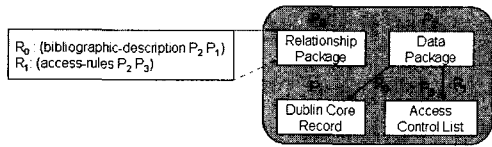


그림 2 WFC를 포함하는 워릭 프레임워크 컨테이너

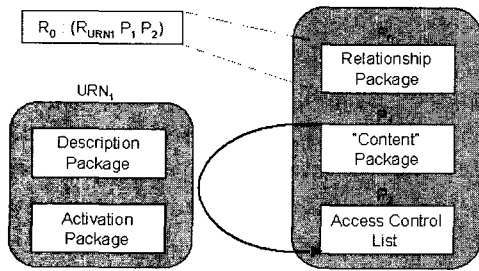


그림 3 권한 관리를 위한 DAR 사용 예

이 예에서는 콘텐츠 패키지 P1과 이에 대한 접근 권한 조정 목록 패키지 P2가 컨테이너 내에 존재하고, P1과 P2의 관계가 P0에 기술되어 있으며, P0에 기술된 관계는 단순한 이름 형식이 아닌 실행가능한 또 다른 패키지들을 포함하는 하나의 URN 형태로 명시되어 있다. 즉 접근권한 제어를 위한 조정 목록과 실제 실행을 위한 메커니즘이 분리되어 있으면서, 이 두 패키지 간의 관계가 DAR에 의해 실행가능한 액티브한 관계로써 명시된 것이다.

### 2.3 가상문서(Virtual Document)

디지털도서관의 유용성이 사용자가 분산환경에서 저장되어 있는 정보를 쉽고 신속하게 찾을 수 있도록 도와줄 수 있는지에 달려있으므로 이를 위해 핵심적으로 연구해야 할 부분은 원격지에 흩어져 있는 정보를 효과적으로 연결시켜서 의미있고 일관된 형태로 보여지도록 하는 것이다. 이 목적을 달성하기 위해

제안된 방법 중 하나가 가상문서이다. 가상문서[1]란 특정 저장장소에 저장되어 존재하던 물리적 문서들의 일부분 또는 전체를 연결하여 하나의 가상적인 문서를 구성하고 재현시에 이들을 동적으로 통합하여 브라우징 할 수 있는 문서를 의미한다. 새롭게 생성된 가상문서에는 실제 데이터는 존재하지 않고 기존에 존재하던 콘텐츠로의 링크들만 갖게 된다.

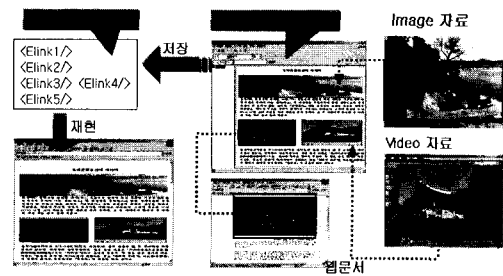


그림 4 가상문서의 개념

그림 4는 디지털 콘텐츠(XML/HTML, 이미지, 동영상)로부터 가상문서를 생성하고 상용 브라우저를 이용하여 가상문서를 재현하는 모습을 보여준다. 그림 2와 같이 가상문서의 생성시에 물리적으로 존재하는 콘텐츠를 복사하여 삽입하는 것이 아니라 필요한 부분에만 링크를 걸어 문서를 구성하므로 멀티미디어 기반의 새로운 문서를 생성하는데 저장공간이 거의 들지 않으며 기존의 제작되어 관리되던 멀티미디어 콘텐츠의 재사용을 활성화[3]시킬 수 있다는 장점이 있다.

그림 4에서는 가상문서에서 주로 이용하는 내포링크와 참조링크를 사용한 예를 보여 주었다. 가상문서에서 사용하는 링크[1]들은 의미와 역할에 따라 다음과 같이 분류될 수 있다.

- 내포링크와 참조링크: 링크의 사용목적에 따라 분류한 것으로 내포링크는 가상문서 재현시에 링크의 목적 콘텐츠가 직접 문서 안으로 삽입되어 나타나는 링크를 의미하고, 참조링크는 콘텐츠가 문서 안으로 삽입되지 않고 앵커로만 남아 사용자의 선택시에 향해나 브라우징을 할 수 있도록 표시하는 링크를 의미한다. 예를 들어, 내포링크는 웹 문서의 이미지 삽입과 유사한 기능을 지원하고 참조링크는 하이퍼링크와 유사한 개념이다.

- 일대일(1:1), 일대다(1:N), 다대일(N:1) 대응 링

크: 링크의 대응관계에 의한 분류로 일대일 대응은 링크의 목적 콘텐츠가 유일한 것을 의미하며, 일대다 대응은 링크의 목적 콘텐츠가 두개 이상인 링크를 의미하며, 다대일 대응은 여러 개의 링크가 같은 하나의 목적 콘텐츠를 갖는 것을 의미한다.

· 특정링크(specific link)와 총칭링크(generic link): 링크의 시작점에서 보는 관점으로 가상문서의 특정위치에 있는 개체가 특정 콘텐츠를 가리킬 경우, 이를 특정(1:1 대응)링크라 하며, 임의의 도메인에 포함된 모든 특정개체들이 모두 하나의 특정 콘텐츠를 가리키는 경우에는 총칭링크(N:1 대응)라 한다.

· 전체링크(total link)와 부분링크(partial link): 링크의 목적지에서 보는 관점으로 전체링크는 링크의 목적 콘텐츠가 콘텐츠 전체인 것을 의미하고 부분링크는 링크의 목적 콘텐츠가 콘텐츠내의 일부분인 것을 의미한다.

가상문서의 구성은 가상문서의 구조적 틀을 설명하는 허브(Hub)와 가상문서별 디스플레이 형식에 관한 기술을 하는 스타일시트(Style Sheet)로 이루어져 있으며, 허브는 다시 참조링크 리스트, 내포링크 리스트, 가상문서의 메타데이터로 구성된다. 메타데이터는 가상문서 전체가 개별적으로 갖는 메타정보들로 더블린코어(Dublin Core)[10]의 15가지 속성을 모두 수용하고 있으며 문서 저장시에 메타데이터를 포함한 허브정보가 XML로 자동 변환되어 하나의 가상문서로 저장된다.

## 2.4 토픽맵(XTM : XML Topic Map)

토픽맵은 TopicMaps.Org를 주축으로 하여 진행되고 있는 XML을 기반으로 한 지식 표현의 표준안 [ISO13250]이다. 이는 추상적이거나 구체적인 대상을 포함하는 어떠한 정보집합을 구조화하여 표현하기 위해, 자원들의 중첩된 레이어나 지도(map)를 통해 자원들에 대한 지식을 전달하고, 구현에 독립적인 방법으로 자원들이 말하고자 하는 주제들과 이러한 주제들 사이의 관계를 나타내는데 그 목적이 있다 [16].

토픽맵은 어떠한 주제(subject)를 컴퓨터 시스템 상에 표현 가능한 대상으로 구체화한 토픽과 이러한 토픽들 사이의 관계(association), 그리고 토픽과 관련된 정보(occurrence)를 주요 개념으로 하여 구성된다. 다음의 그림 5는 이러한 개념들을 이용하여 표현

된 토픽맵의 예를 도식화하여 나타낸 것이다[17].

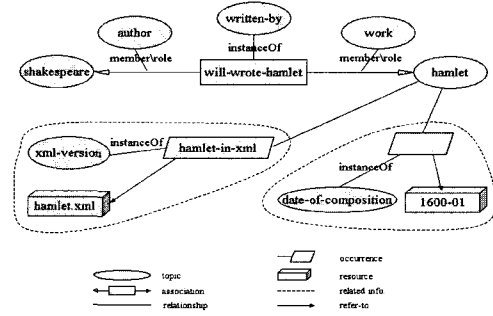


그림 5 XTM 활용 예

위 그림에서 'hamlet'이라는 토픽과 관련된 토픽인 'shakespeare'가 관계 'will-wrote-hamlet'에 의해 상호 연관 지어짐으로써, 'shakespeare가 hamlet이라는 작품을 썼다'는 지식을 표현하고 있다. 또한 토픽 'hamlet'에 대한 정보로써 'hamlet.xml'이라는 자원과 '1600-01'이라는 자원을 명시함으로써, 작품이 쓰여진 시기와 그 작품을 XML이라는 표현 형식을 빌어 기술한 물리적인 문서 자체를 'hamlet'에 대한 부수적인 정보로 표현하고 있다.

XTM은 정보 자원들에 포함되어 있는 지식 구조를 모델링하는 표준화된 방법을 제공한다. 또한 XTM은 XML을 기반으로 기술되기 때문에 정보를 구조적으로 기술할 수 있을 뿐 아니라, 기술된 각 정보는 XTM의 DTD가 가지는 특수화된 구조와 스키마에 의해 상호 간에 다양한 의미의 관계와 계층을 가질 수 있다. XTM의 이러한 특성은 XTM 문서를 이용하는 다양한 응용에서 정보를 검색하고 네비게이션하고 재구성하는데 있어 기술적·의미적으로 유용한 '지식표현' 및 '정보관리'를 가능하게 한다. 따라서, XTM을 동의어 사전이나 의미 네트워크, 또는 지식베이스 구축 및 관리 등과 관련된 응용 분야에서 활용하기 위한 연구가 활발히 진행되고 있다.

## 3. 디지털 도서관 시스템에의 응용

### 3.1 FEDORA(Flexible and Extensible Digital Object Repository Architecture)

#### 3.1.1 FEDORA의 기본 컴포넌트 및 특징

DAR과 워터 프레임워크에 기반한 전자 도서관

시스템으로, 전자 도서관의 내용을 저장 및 분배하기 위해 코바(CORBA)를 이용하여 구현된 저장소 아키텍처(architecture)이다. FEDORA는 전자객체(DigitalObject), 데이터흐름(DataStream), 인터페이스(Interface), 엔포서(Enforcer)의 네 가지 컴포넌트를 중심으로 구성된다. 다음은 컴포넌트에 대한 설명이다[14,15].

- 데이터흐름 : 바이트 스트림으로 이루어지는 실질 데이터를 포함하는 패키지이다.
- 전자객체: 데이터흐름을 캡슐화하기 위한 컨테이너 추상화를 제공하는 개체로써, 데이터흐름들을 생성 및 접근하고, 자신의 내부에 있는 구조들을 조작할 수 있게 하는 일반적인 활성화기(behavior)들의 집합을 제공한다.
- 인터페이스: 전자객체에 캡슐화되어 있는 데이터에 대해 문맥적으로 의미가 있는 뷰(view)나 배포판(dissemination)을 제공하는 기능을 담당한다.
- 엔포서: 권한 관리(right management)를 위해, 자원에 대한 접근 제어를 담당한다.

데이터흐름과 전자객체는 앞선 2.2절에서 다루었던 일반적인 패키지에 해당되고, 인터페이스와 엔포서는 전자객체 내에 있는 데이터흐름이라는 패키지에 접근하기 위한 특별한 관계를 정의하고 있는 일종의 DAR에 해당된다. 이러한 컴포넌트를 기반으로 하는 FEDORA의 주요 특징들로는 다음과 같은 것들이 있다[14].

- 지역적으로 분산된 서로 다른 타입의 데이터를 모아놓은 복합 객체들의 생성이 가능하다.
- 객체의 기본 구조에 관계없이, 콘텐츠의 전역적 혹은 도메인 한정적인 표기를 연관시키는 활성화기(behavior)들의 집합을 이용해 객체에 능력을 부여한다.
- 사용자는 이러한 콘텐츠 타입 행위자들을 실행하는 외부 메커니즘들을 실행함으로써 해당 객체의 다양한 뷰나 배포판들을 얻을 수 있다.
- 어떠한 객체 콘텐츠에 대한 어떠한 배포판에 대해서도 외부 권한 관리 스킴 및 관계(association)들을 형성할 수 있다.

### 3.1.2 FEDORA 전자 객체 구조

FEDORA에서 사용되는 전자객체의 기본 구조는 다음의 그림 6과 같다. 전자객체는 확장가능한 활성화기 계층(Extensible Behavior Layer: EBL) 내부에

커널 구조를 가지고 있게 된다. 커널에는 포스트스크립트(postscript) 타입의 데이터흐름과 메타데이터 정보를 포함하는 마크(MARC) 레코드 타입의 데이터흐름이 기본 데이터 패키지로 놓이게 되고, 이들에 대한 기본적인 접근 인터페이스로 사용되는 원시 배포자(Primitive Disseminator)가 제공된다. 이러한 기본 구조에 추가적으로 새로운 타입의 배포자를 추가함으로써 사용자에게 원하는 타입으로 변환된 데이터를 제공할 수 있게 된다. 새로운 타입의 배포자는 EBL에 추가된다[14].

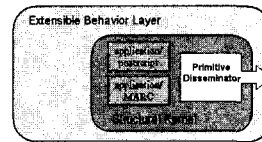


그림 6 전자객체 기본 구조

다음의 그림 7은 새로운 타입의 배포자를 이용해 확장된 전자객체의 예를 보여주고 있다. 이 객체에 대해 사용자는 그림에서와 같은 순서대로 배포자를 이용함으로써, 전자객체에 포함된 콘텐츠에서 원하는 부분만을 얻을 수 있게 된다. 순차적으로 살펴보면, 우선 사용자는 전자객체가 제공하는 배포자들의 타입에 관한 리스트를 얻는다(ListDisseminatorTypes). 사용자는 현재 객체는 Book과 Dublin Core 타입의 배포자를 가지고 있다는 정보를 얻게 된다. 이 중에서 Book 타입의 배포자를 이용하기 위해 Book 배포자가 가지고 있는 메소드 목록을 요청한다. 결과로 Book 타입의 배포자는 GetChapter(n), GetPage(n), GetTOC()의 세 가지 메소드를 가진다는 정보가 반환된다. 사용자는 최종적으로 GetPage() 메소드를 이용해 전자객체의 1페이지에 해당되는 내용을 얻어올 수 있게 된다[14].

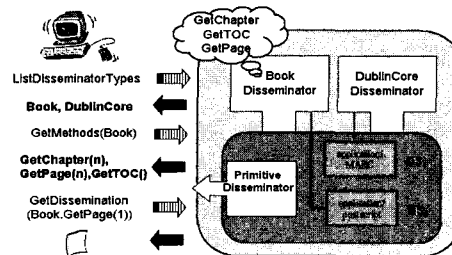


그림 7 확장된 전자객체의 예

FEDORA의 이러한 전자객체 구조는 동일 내용을 포함하고 있는 전자자료를 사용자의 요구에 맞추기 위해 포맷별로 유지해야만 했던 기존의 불필요한 저장소 낭비 및 관리 측면에서의 부담을 줄여주는 역할을 한다. 전자객체에 포함된 콘텐츠에 다양한 타입의 배포자를 붙여주는 것 만으로도 충분히 사용자들의 다양한 요구를 충족시켜줄 수 있기 때문이다.

### 3.2 MIRAGE-X 디지털도서관 시스템

#### 3.2.1 MIRAGE-X 디지털도서관의 구조

그림 8에서와 같이 가상문서를 기반으로 한 MIRAGE 계열 디지털도서관 시스템[2]은 용도에 따라 개인 디지털도서관(MIRAGE-Lite)과 공용 디지털도서관(MIRAGE-Regular)으로 구분된다. 두 시스템의 전반적인 구조는 같지만 개인 디지털도서관은 개개인이 사용하기 편리하도록 사용자에 의존적으로 설계되었고 공용 디지털도서관은 개인 디지털도서관보다 사용자에이전트 기능을 확장시켜 같은 MIRAGE 계열의 디지털도서관 시스템이나 분산환경하의 다른 디지털도서관 시스템들과 서비스를 교환할 수 있으며 메타검색을 할 수 있는 기능이 추가된다.

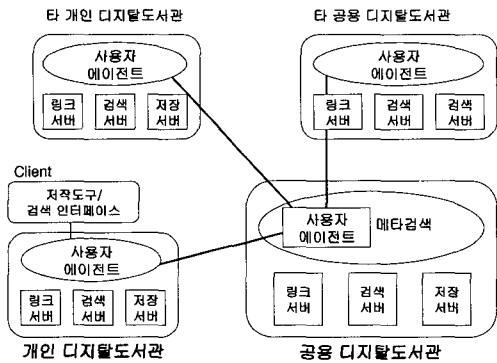


그림 8 MIRAGE-X 디지털도서관

그림 8의 서버부분에서 사용자 에이전트는 MIRAGE계열 디지털도서관 시스템 사이의 메시지 교환 및 디지털도서관으로 입력되는 사용자 서비스를 분석하여 이에 해당하는 처리결과를 사용자에게 전달하는 기능[12]을 수행한다. 검색서버[2]는 기존의 디지털도서관에 존재하던 콘텐츠와 이를 재활용하여 생성한 가상문서를 대상으로 XML문서 검색기

법을 적용한 검색을 수행하며 사용자에게 다양한 단위의 검색결과를 제공한다. 링크서버[11]는 검색서버의 링크정보 요청시에 문서의 일부분이나 문서 전체에서 나가고 들어오는 링크정보를 검색하여 검색서버에 전달한다. 또한 저장서버는 기존의 디지털 콘텐츠와 가상문서를 저장관리하며 가상문서를 재현시에 URI에 해당하는 문서를 잘라내어 사용자 에이전트에 전달하는 기능을 수행한다.

클라이언트[4]의 저작도구와 검색 인터페이스에서는 디지털 콘텐츠를 검색한 후 필요한 콘텐츠의 일부 또는 전체에 링크를 생성하여 XML로 기술된 가상문서를 쉽게 만들 수 있으며 일반 웹 브라우저를 통한 브라우저링을 가능하게 해준다.

#### 3.2.2 MIRAGE-X 디지털도서관의 특징

링크정보만으로 구성된 가상문서를 기반으로 설계된 MIRAGE-X 디지털도서관 시스템의 특징은 다음과 같다.

- 가상문서 제작을 위해 새로운 멀티미디어 콘텐츠를 구축할 필요없이 기 구축된 디지털 콘텐츠를 활용하므로 데이터베이스 구축비용을 절약할 수 있고, 멀티미디어 콘텐츠의 재활용 및 유통의 활성화를 촉진시킬 수 있다.

- 가상문서는 원격지에 흩어져 있는 교육용 콘텐츠를 복사하지 않고 링크만으로 이루어진 이들을 포함하는 가상문서를 저장하므로 대용량의 멀티미디어 콘텐츠를 저장하기 위해 필요한 대규모 저장공간을 절약할 수 있다.

- 링크만을 사용하여 크기가 작아진 가상문서는 네트워크를 통한 전송시에 네트워크 트래픽을 현저히 감소시킬 수 있으므로 네트워크를 효율적으로 사용할 수 있다.

- 일대다 링크를 만들 수 있으므로 문서를 표현할 때 여러 개의 혹은 같은 내용의 서로 다른 미디어 형식으로 문서를 표현할 수 있다.

- XML언어로 저장되는 가상문서는 XML구조검색 기법을 이용하여 문서의 일부분 또는 문서전체를 검색할 수 있으며 링크정보에 대한 검색도 가능하다.

- 가상문서를 저장할 때 내용을 저장하지 않고 링크정보만을 저장하는 것과 문서 재현시에 내포된 내용의 출처를 표시해주는 것은 현재 사회적으로 커다란 이슈가 되고 있는 저작권 보호 문제를 근원적으로 해결해 줄 수 있다.

### 3.3 XTM 응용 시스템

XTM을 기반으로 한 지식구축 및 관리와 관련된 연구 중 대표적인 예로 Ontopia에서 진행중인 'Ontopia topic map'과 TheBrain Technologies Corporation에서 진행중인 'TheBrain'이 있다. 각 시스템에 대해 살펴보기로 한다.

#### 3.3.1 온토피아 토픽맵(Ontopia topic map)

온토피아 토픽맵은 일반 웹 환경에서 주로 사용되는 네비게이션 방식을 이용하여 XTM으로 표현된 정보집합을 브라우저할 수 있게 하는 시스템이다. 이 시스템은 크게 토픽맵 엔진, 네비게이터 프레임워크, 클라이언트 에디터 프레임워크의 세 부분으로 이루어진다. 토픽맵 엔진(Topic map Engine)은 XTM 표준을 따르는 문서를 저장 및 보관, 검색할 수 있도록 하는 기능을 담당하고, 네비게이터 프레임워크(Navigator Framework)는 사용자들이 XTM 문서를 로딩 및 브라우저할 수 있게 하는 응용 프로그램을 위한 프레임워크이다. 클라이언트 에디터 프레임워크(Client Editor Framework)는 협업으로 토픽맵문서를 편집할 수 있도록 하는 웹 응용 프로그램을 위한 프레임워크로, 네비게이터 프레임워크를 확장하여 사용자 인터페이스 엘리먼트들을 이용해 토픽맵을 손쉽게 수정할 수 있도록 하는 기능을 지원하고 있다[18].

이 시스템에서 XTM 문서는 토픽, 정보자원의 유형, 관계 유형, 메타데이터 속성 등의 기준에 따라 다양한 관점에서 접근할 수 있도록 처리된다. 다음의 그림 9는 오페라와 관련된 각종 정보를 기술하고 있는 XTM 문서를 네비게이션하는 한 예를 보이고 있다.

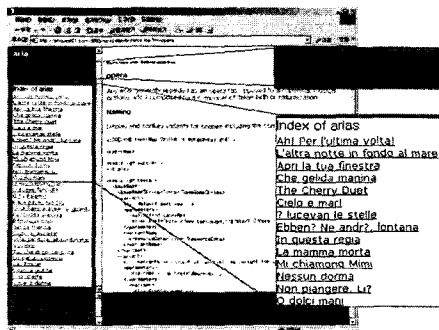


그림 9 Ontopia 네비게이션 인터페이스  
온토피아 토픽맵은 XTM 문서에 기술된 정보를

각 관점에 따라 나열식으로 사용자에게 제시하고, 이를 사용자가 순차적으로 선택하여 원하는 정보에 접근하도록 한다. 위 그림은 XTM 문서에 기술된 정보를 association role type이나 topic type에 따라 볼 수 있도록 선택하는 하이퍼 링크와 그러한 구분 없이 전체 정보를 나열식으로 제시하여 이로부터 사용자가 네비게이션을 시작할 수 있도록 하는 인터페이스를 제공하고 있다. 또한 사용자 인터페이스의 우측에는 사용자가 선택한 정보자원을 디스플레이 하는 부분이 있다. 이러한 인터페이스는 사용자가 원하는 정보를 찾기 까지 여러 단계의 선택과정이 필요하고, 전체 정보집합의 구조와 실제적인 정보자원이 구분되어 제시되기 때문에 사용자가 종합적인 정보를 받아들이는데 있어 어려움이 따른다.

#### 3.3.2 TheBrain

TheBrain은 정보집합에 기술된 정보들을 각 정보가 유기적으로 결합된 신경망의 형태로 표현한다. TheBrain 시스템은 크게 4개의 서브 시스템으로 구성된다. 범용 데이터 접근(Universal Data Access) 시스템은 웹 사이트, 관계형 데이터베이스, 파일 시스템 등의 다양한 자원들에 있는 정보들을 통합할 수 있도록 하는 데이터 접근 계층을 제공하고, 통합 협업(Integrated Collaboration) 시스템은 정보가 생성되거나 저장된 위치에 관계없이 이들에 접근할 수 있도록 하고, 여러 사람이 동시에 이러한 자원들을 사용 및 편집할 수 있도록 하기 위한 커뮤니케이션 기능을 제공한다. 지식 모델(Knowledge Model) 시스템은 정보들이 어떻게 생성 및 연결되었는지에 대한 지식을 학습할 수 있도록 하며, 다양한 경로를 통해 중요 개념들이나 정보들에 접근할 수 있도록 한다. 마지막으로 비주얼 사용자 인터페이스(Visual User Interface) 시스템은 사용자가 손쉽게 시스템을 사용하고 네비게이션할 수 있도록 하는 그래픽 사용자 인터페이스를 제공한다[19].

다음의 그림 10은 TheBrain의 웹에서 운영되는 버전인 WebBrain의 한 가지 예이다. 이 예제에서 사용자는 'Computers' 분야의 'Software'와 관련된 정보들로 어떠한 것들이 있는지 볼 수 있고 이러한 정보들 중에서 원하는 것을 선택함으로써 현재 'Software'가 중심을 이루고 있는 뷰를 자신이 원하는 정보를 중심으로 한 뷰로 변경할 수 있다. 이는 사용자 생각의 흐름에 따라 전체 망을 옮겨가는 형태로

매우 자연스러운 정보이용 형태라 할 수 있다. 상위 개념 및 하위 개념, 그리고 현재 중심이 되고 있는 개념과 동급의 개념들이 한 화면에 집약되어 표현된다.

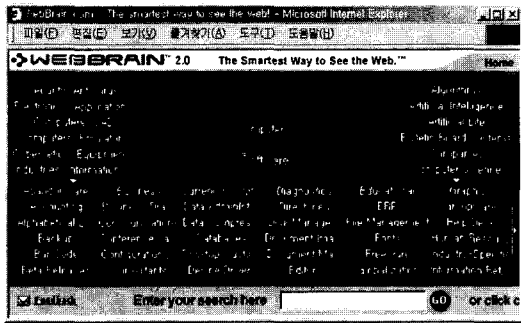


그림 10 WebBrain 이용 예

#### 4. 결 론

전자 도서관의 역할이 증대되면서, 다양한 형식의 전자 자료들을 결합하여 하나의 통일된 뷰로 제공하고자 하는 연구가 활발해 졌다. 본 글에서는 이러한 노력의 일환으로 진행되고 있는 문서 표현 양식과 관련한 연구들인 다가문서, 워릭 프레임워크, 가상문서, 그리고 XTM에 대해 살펴보았다. 이들 연구들은 공통적으로 다양한 형태를 갖는 분산된 멀티미디어 데이터들을 하나의 문서 표현 양식을 이용해 표현함으로써 사용자에게 복합객체에 대한 단일 뷰를 제공할 수 있도록 하며, 새로운 문서를 작성하고자 할 때 기존의 데이터들을 다양한 기법들을 이용해 재사용하도록 함으로써 문서 저장이나 관리 및 배포 시에 발생할 수 있는 문제점들을 최소화하고 있다.

다가문서의 경우 다양한 물리적인 문서들을 통합하기 위해 레이어 개념을 이용하고 있다. 이는 각 미디어 데이터를 하나의 레이어에 두고 이를 여러겹 겹친 형태로 사용자에게 단일 뷰를 제공한다. 워릭 프레임워크의 경우는 자원과 자원에 대한 메타데이터를 모두 특정 타입을 갖는 패키지라는 관점에서 동일시 하고, 이를 컨테이너라는 구조에 담아 사용한다. 이러한 개념을 사용함으로써 사용자는 전자자료를 포함하는 하나의 전자객체를 대상으로 원하는 형태의 자료를 요청하고 이에 대한 서비스 받을 수 있다. 가상문서의 경우는 물리적으로 존재하는 여러 자원들을 다양한 링크들을 이용해 연결하는 방식으로 통합된 하나의 새로운 문서가 생성된다. 따라서, 가상

문서를 구성하고 있는 내용들은 실제로 그 문서 형식대로 물리적으로 저장되는 것이 아니라 재현시에 동적으로 구성되어 사용자에게 보여지게 된다. XTM 문서는 자원들을 실제로 문서에 포함하는 개념보다는 자원들을 참조하고 이에 대한 설명을 기술하는 개념으로 주로 사용된다. XML의 일반적인 형식에 자원에 대한 의미를 부여할 수 있는 형식을 부가적으로 제공함으로써 단순한 자원간의 연결이 아닌 의미를 내포한 새로운 개념의 통합을 꾀하고 있다.

다가문서와 워릭 프레임워크의 경우는 물리적으로 자원들을 통합하는데 중점을 두고 진행된 연구인데 반해, 가상문서의 경우는 물리적인 자원 통합과 더불어 자원 재활용 및 자원간의 관계에 대한 의미부여의 측면도 고려되어 설계되었다. 따라서, 의미 웹 (Semantic web)으로 발전하고자 하는 현재의 웹 분야의 노력과 부합되는 측면이 있다. XTM의 경우는 자원을 통합하는데 있어 표준화된 의미부여 방식으로 자원들 간의 관계를 기술하고자 하고 있다. 그러나, 실제로 사용자에게 보여지는 XTM 문서는 자원에 대한 기술 및 관계와 실제 물리적인 자원이 분리된 형태로 제공되므로 이를 효과적으로 결합하기 위한 연구가 진행되어야 할 것이다.

#### 참고문헌

- [1] Sung Hyon Myaeng, Mann-Ho Lee, Ji-Hoon Kang, Eun-Il Cho, Yong-Bae Lee, Dong-Soo Lim, Jeong-Mook Lim, Hyo-Jung Oh & Jung Shik Yang, "A Digital Library System for Easy Creation/Manipulation of New Documents over Existing Resources", Proceedings of RIAO 2000, Paris, April.
- [2] 이용배, 맹성현, "MIRAGE-III 디지털도서관에서 가상문서 검색서버의 설계 및 구현", 한국정보과학회논문지:컴퓨팅의 실제, 8권2호, pp.219-230, 2002년 4월.
- [3] 정창후, 이용배, 맹성현, "멀티미디어 콘텐츠 유통을 위한 지식서버 시스템의 설계", 한국정보과학회 추계학술발표논문집(I), pp.316-318, 2001년 10월.
- [4] 이상봉, 조은일, 이용배, 맹성현, "디지털도서관에서 이용하는 가상문서 저작도구의 설계 및 구현", 한국정보처리학회 가을학술발표논문집,



pp.171-173, 2000년 10월.

[5] Thomas A. Phelps, Robert Wilensky, "The Multivalent Browser: A Platform for New Ideas", Proceedings of Document Engineering 2001, November 2001, Atlanta, Georgia.

[6] Thomas A. Phelps, Robert Wilensky, "Multivalent Documents: Anywhere, Anytime, Any Type, Every Way User-Improvable Digital Documents", Communications of the ACM, June 2000.

[7] Thomas A. Phelps, Robert Wilensky, "Multivalent Annotations", Proceedings of First European Conference on Research and Advanced Technology for Digital Libraries, 1997.

[8] Thomas A. Phelps, Robert Wilensky, "Toward Active, Extensible, Networked Documents: Multivalent Architecture and Applications", Proceedings of Digital Libraries, 1997.

[9] Thomas A. Phelps, Robert Wilensky, "Multivalent Documents: Inducing Structure and Behaviors in Online Digital Documents", Proceedings of Hawaii International Conference on System Sciences, 1996.

[10] Dublin Core Community, "Dublin Core Metadata Initiative", recommendation 1999. (available at <http://purl.org/DC/documents/>)

[11] 김영승, 박인석, 현순주, "가상 문서의 효율적인 관리를 지원하는 대용량 링크정보 관리서버의 개발", 한국정보과학회 봄 학술발표논문집, 2002년 4월.

[12] 양중식, 임동수, 강지훈, "XML 기반 가상 문서의 브라우저를 위한 문서처리", 한국정보처리학회 추계 학술발표 논문집, 2000년 4월.

[13] Lagoze, Carl, Clifford A. Lynch, and Ron Daniel Jr., The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata, Cornell Computer Science Technical Report TR96-1593, July 1996, <http://cs-tr.cs.cornell.edu:80/Dienst/UI/2.0/Describe/ncstr1.cornell/TR96-1593>.

[14] FEDORA Home Page, <http://www.cs.cornell.edu/cdlrg/fedora.html>

[15] Daniel Jr., Ron and Carl Lagoze, Distributed Active Relationships in the Warwick Framework, Proceedings of the 1997 IEEE Metadata Conference, September, 1997, <http://computer.org/conferen/proceed/meta97/papers/rdaniel/rdaniel.pdf>

[16] XML Topic Maps (XTM) 1.0 TopicMaps.Org Specification "<http://www.topicmaps.org>".

[17] Steve Pepper, "Navigating haystacks and discovering needles", Markup Languages: Theory & Practice 1.4 (1999): 41-68.

[18] Ontopia, "<http://www.ontopia.net>".

[19] The Brain, "<http://www.thebrain.com>".

[20] XML Schema Part 0: Primer, W3C Recommendation, 2 May 2001, "<http://www.w3.org/TR/xmlschema-0/>".

[21] XML 1.1, W3C Working Draft 13 December 2001, "<http://www.w3.org/TR/xml11/>".

[22] 윤여준, 조숙현, 김태현, 맹성현(2001). "가상문서를 XTM 문서로 변환해 주는 자동변환시스템 설계 및 구현", 2001년도 한국정보처리학회 추계학술발표대회, pp1131-1134.

[23] 류궁호, 김운, 김태현, 맹성현(2001), "XTM문서의 가상문서로의 변환 시스템 설계 및 구현", 2001년도 한국 정보과학회 가을 학술발표논문집( I ) pp 172-174.

김 태 현



1999 충남대학교 컴퓨터과학과 학사  
 2001 충남대학교 컴퓨터과학과 전산학 석사  
 현재 한국전자통신연구원 지식처리연구팀 연구원  
 관심분야: 정보검색, 자연어처리, 전자도서관, 정보추출, 문서요약, 지식관리 시스템 등  
 E-mail: heemang@etri.re.kr

**이 용 배**



1999 충남대학교 컴퓨터과학과(학사)  
2001 충남대학교 컴퓨터과학과 대학원(석사)  
현재 충남대학교 컴퓨터과학과 대학원(박사수료)  
관심분야: 정보검색, 자연어처리, 디지털 도서관, 장르분류, 지식관리시스템, 하이퍼미디어시스템  
E-mail: yblee@enquest.co.kr

**맹 성 현**



1983 캘리포니아 주립대학 학사  
1985 Southern Methodist University (SMU) 석사  
1987 Southern Methodist University (SMU) 박사  
1987~88 Temple University 교수  
1988~94 Syracuse University 교수  
1994~현재 충남대학교 정보통신공학부 교수  
관심분야: 정보검색, 자연어처리, 디지털 도서관, 자동요약, 자동분류, 지식 관리 시스템  
E-mail: shmyaeng@cs.cnu.ac.kr

**• 소프트웨어공학 기술 합동 워크샵 2002 •**

- 일 자 : 2002년 8월 22~24일
- 장 소 : 경주 하일라 콘도
- 주 최 : 한국정보과학회 · 한국정보처리학회
- 문 의 처 : 대전대학교 황선명 교수

E-mail: sunhwang@dragon.taejon.ac.kr

**• 제29회 정기총회 및 추계학술발표회 •**

- 일 자 : 2002년 10월 25 ~ 26일
- 장 소 : 수원대학교
- 논문모집 및 발표일정
  - 1) 접수기간 : 2002년 8월 1 ~ 26일
  - 2) 심사결과통보 : 2002년 9월 16일
  - 3) 수정논문 접수마감 : 2002년 9월 25일
  - 4) 사전등록 : 2002년 10월 1 ~ 21일
  - 5) 논문발표 : 2002년 10월 25 ~ 26일
- 문 의 처 : 한국정보과학회 사무국 한영진 과장

Tel. 02-588-9246/7

http://www.kiss.or.kr E-mail: yjhan@kiss.or.kr