디지털도서관의 문헌 특성 및 관리

The Characteristics and Management of Documents in the Digital Libraries

김 성 혁(Sung-Hyuk Kim)

목 차

1. 서론
2. 디지털 환경하에서 문헌의 특성
   2. 1 전통적인 문헌에 대한 변화
   2. 2 하이퍼미디어 시스템에서 문헌의 특성
3. 디지털도서관의 문헌 관리
   3. 1 디지털도서관의 시스템 환경
   3. 2 디지털도서관에서 문헌 관리시 고려점
   3. 3 디지털도서관에서 표준의 중요성
4. HyTime
   4. 1 등장배경 및 개요
   4. 2 HyTime의 구조와 선언
   4. 3 HyTime에서 제공하는 하이퍼링크 기능
   4. 4 HyTime에서 제공하는 주소 및 위치 부여 기능
5. 결론

초 록

디지털도서관에 관한 연구가 진행됨에 따라 문헌의 디지털화에 관한 관심이 고조되고 있다. 본 논문은 전통적인 인쇄문헌에 비해 디지털 문헌의 특성은 무엇이고, 디지털 문헌을 어떻게 관리하고, 디지털 문헌을 관리할 때 어떤 점을 고려하여야 할 것인가에 관한 연구 결과이다. 문헌의 디지털화에 따른 문헌 구조화를 위하여 SGML과 HyTime의 관계를 설명하고, 구체적으로 HyTime의 적용 예를 제시하였다.

ABSTRACT

This paper is the result of study on what is the characteristics of digital document, how to manage it, and what is the consideration to manage it, compared to the conventional paper document. For the document structuring, I also explain the relationship between SGML and HyTime, and , at the same time, suggest in detail the practical examples of HyTime application.

* 본 연구는 숙명여자대학교 95년도 교비연구비 지원에 의하여 수행되었음.
** 숙명여자대학교 문헌정보학과 조교수
접수일자 1997년 2월 14일

1. 서 론

디지털도서관은 전통적인 도서관과 여러 가지 측면에서 많은 차이점을 가지고 있으나, 가장 근본적인 차이점은 정보가 수록되는 매체의 변화와 정보 표현의 다양화에 기인한다고 볼 수 있다. 전통적인 도서관이 정보 표현 및 저장 매체로 종이와 문자를 사용한 인쇄매체인 문헌을 수집하고, 조직화하여, 이용자에게 제공하는 데 중점을 두었다면, 디지털도서관은 디지털화된 문헌속에 포함된 텍스트, 이미지, 사운드 등 다양한 정보들을 전자적으로 다루어 이용자에게 전자적인 서비스를 제공하는 데 중점을 두고 있다.

그러나 이러한 차이점에도 불구하고, 디지털도서관의 가장 근본적인 목적은 과거 전통적인 도서관이 그러하였던 것처럼, 인류의 지식을 보존하고 이를 이용자에게 제공하여, 지금까지 축적된 지식 및 정보를 후세에 전달하는 것이다. 따라서 과거의 주된 정보전달 매체였던 인쇄된 문헌이 현재 새로운 정보전달 매체인 디지털 문헌으로 변화하고 있다 하더라도 도서관은 이러한 새로운 매체 및 정보 표현방법을 수용함으로써 도서관의 기본적인 의무를 수행하여야 한다. 즉, 디지털도서관에서 수용해야 할 멀티미디어 정보의 특성을 파악하고 그에 기반을 둔 디지털 문헌의 효율적인 관리는 필수적이라고 할 수 있다.

지금까지의 도서관정보시스템은 인쇄된 문헌의 메타데이터적인 요소들을 추출하여 시스템을 구축하였기 때문에 원 문헌(original document)의 구조를 반영시키지 못한 죽은 문헌 (dead document)에서 정보를 찾거나 검색하였다고 볼 수 있다. 따라서 도서관은 메타데이터적인 요소들만을 관리하고, 유지하는 일에 중점을 두었고, 이용자는 원 문헌이 갖고 있는 물리적 및 논리적인 구조가 무시된 메타데이터만을 이용하여 원하는 정보를 검색하는 결과를 낳았다. 그러나 디지털도서관 환경에서는 인쇄된 문헌이 아닌 디지털 문헌을 주 대상으로 하기 때문에 지금의 도서관자동화시스템적인 문헌처리 방법으로는 디지털 문헌의 처리가 불가능하다. 디지털도서관은 수많은 정보 보존소 (repositories)들이 네트워크에 의해 연결되어 다양한 정보서비스를 제공하여야 하는데, 인쇄문헌 중심적인 자료처리 방법으로는 이를 충족시킬 수 없다. 그렇기 때문에 인쇄된 문헌과 구조적으로 동일하게 문헌을 디지털화하여 컴퓨터에 저장하여야 한다. 즉 인쇄된 문헌이 살아있는 문헌 (live document)이라면 디지털화한 문헌도 죽은 문헌이 아닌 살아있는 문헌이어야 한다는 의미이다.

우리보다 앞서 디지털도서관시스템을 구축하고 있는 외국의 경우, 살아 있는 디지털 문헌을 만들기 위한 연구가 활발히 진행되고 있다. 이들은 문헌의 구조화(document structuring)를 통해 원 문헌의 구조를 디지털 문헌에 반영시키고, 문헌속에 포함된 다양한 유형의 정보들을 서로 연결함으로써 살아 있는 디지털 문

헌을 구축하고 있다. 본 연구는 이러한 관점하에 멀티미디어 정보가 하이퍼링크로 연결된 하이퍼미디어 시스템을 구축하는 디지털도서관 환경에서 디지털 문헌의 특성 및 문헌의 관리시 필요한 고려사항을 조사·연구하고, 하이퍼미디어 디지털 문헌을 구축하기 위한 국제 표준인 HyTime을 중점적으로 소개하고자 한다.

2. 디지털환경하에서 문헌의 특성

2. 1 전통적인 문헌에 대한 변화

전통적인 형태의 문헌 (document)은 종이에 문자를 사용하여 저자의 의도를 기술한 것이었다. 문헌의 가장 큰 목적은 의사소통 (communication)을 위한 것으로서, 이 의사소통은 문헌에 수록된 정보의 전달을 통하여 이루진다. 문헌은 동시대 여러 사람들 사이의 의사소통은 물론, 과거와 현재, 현재와 미래의 인류사이에서 의사소통을 담당하고 있다.

컴퓨터 기술 및 광학 기술의 발달은 이러한 전통적인 문헌의 물리적인 외형에 큰 변화를 가져왔으나, 문헌의 본질적인 개념은 여전히 '정보의 집합체'이다 (Travis, B. E. & D. Waldt 1995, 66). 다만 정보를 표현하는 방법이 종이에 기록되던 문자 중심의 정보 표현방법에서 컴퓨터 기억매체에 저장되는 컴퓨터 코드, 비트맵으로 표현되는 이미지 등으로 다양해 진 것이다. 이는 컴퓨터를 이용하여 정보를 저장하고 이를 검색하며 네트워크를 통하여 정보를 전달할 수 있게 하는 정보의 디지털화가 가능해졌기 때문이다. 인쇄된 문헌에서 멀티미디어 문헌으로의 전환이라는 문헌의 외형적인 변화는 문헌의 유통과 생명주기에도 많은 변화를 가져오게 되었다. 일반적으로 멀티미디어는 사운드, 그래픽, 애니매이션 (animation), 비디오의 결합체로서 하이퍼미디어의 하부집합으로 간주되고 있다 (HyTime Working Group TR1V1). 하이퍼미디어는 하이퍼텍스트를 동반하는 멀티미디어의 요소들을 결합한 형태로서, 하이퍼링크를 사용하여 서로 관련있는 정보들을 연결한다. 1940년대 중반 Bush에 의하여 소개된 하이퍼링크는 문헌으로의 순차적인 접근방법 외에 임의순서로 문헌에 접근할 수 있도록 허용하였다. 하이퍼링크를 도입한 시스템에서 이용자는 둘 또는 그이상의 대상체간에 존재하는 관계를 표현하고 있는 정보구조체인 링크를 따라 자유롭게 정보공간을 항해 (navigate)하면서 자신이 원하는 정보를 획득할 수 있게 되었다.

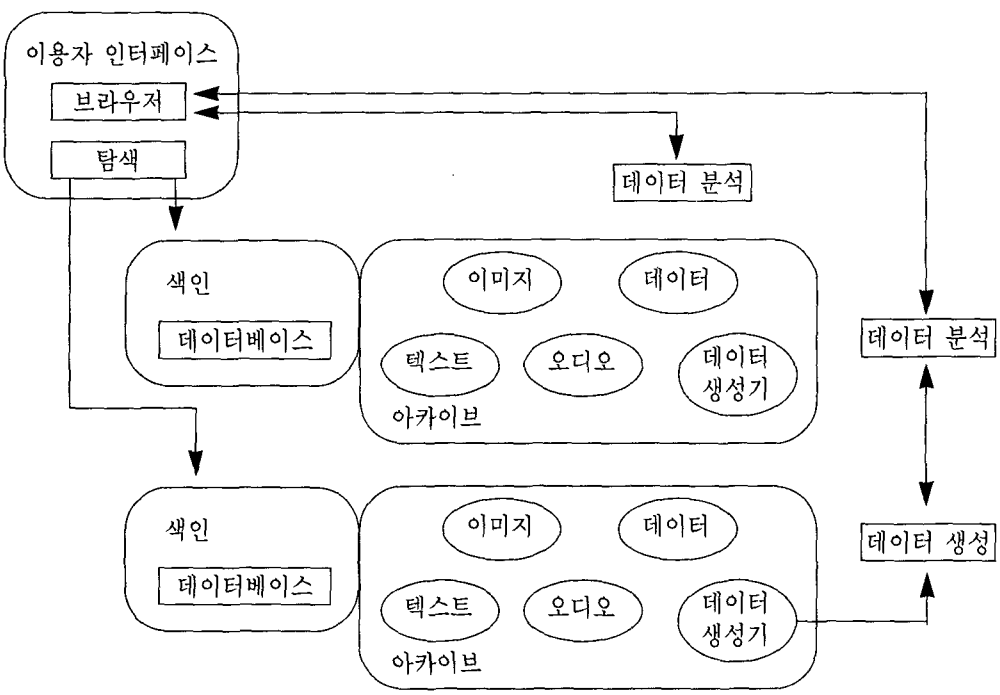
컴퓨터 과학적인 측면에서 본 하이퍼미디어 시스템은 그래픽, 사운드, 비디오가 통합되었거나, 다양한 유형의 미디어가 정보를 저장하고 검색할 수 있게 하는 하나의 시스템으로 통합된 것을 의미한다 (HyTime Working Group TR1V1). 하이퍼미디어 시스템은 이용자 자신의 제어하에 시스템에서 제공하는 하이퍼링크를 이용하여 이용자가 시스템과 상호작용할 수

있게 한다.

이와 같은 사항을 종합해 볼 때, 디지털 환경하에서 문헌의 변화란 첫째, 정보 표현방법의 변화로서 인쇄된 문헌에 수록되던 정보의 디지털화가 가능해졌으며, 둘째, 그에 따른 문헌의 물리적인 변화에 의하여 멀티미디어 문헌이 등장하게 되었으며, 셋째, 문헌의 저자는 자신의 의도에 따라 다양한 표현방법을 사용하여 정보를 전달할 수 있게 되었고, 넷째, 네트워크의 확장과 하이퍼링크 등과 같은 정보기술과 검색기법의 발달로 문헌의 유통 및 이용 방법이 다양해졌다는 것을 문헌에 대한 변화로 요약할 수 있다. 그러나 정보의 집합체라는 문헌 고유의 개념은 문헌의 외형적인 변화에도 불구하고 여전히 지속됨에 따라 디지털도서관은 이러한 문헌의 변화를 수용해야 할 필요가 있다.

2. 2 하이퍼미디어 시스템에서 문헌의 특성

앞서 살펴본 문헌의 변화는 필연적으로 멀티미디어 디지털도서관을 등장시켰다. 〈그림 1〉은 미국의 National Center for Supercomputing Applications(NCSA)의 디지털도서관 구축을 위한 프로토타입



〈그림 1〉 멀티미디어 디지털도서관의 개념도 (Catlett & Terstriep 1992, 91)

으로 전형적인 stand-alone 방식의 멀티미디어 디지털도서관의 개념을 보여주고 있다.

위의 그림에서 하이퍼링크는 표현되지 않았으나 Archive에서 텍스트, 이미지, 사운드 등, 각 유형의 정보매체에 하이퍼링크를 도입하여 관련된 정보들을 서로 연결시키고, 이를 이용자 인터페이스에서 지원해 줄 수 있다면 하이퍼미디어 시스템이 완성된다. 이러한 경우의 하이퍼미디어 시스템은 어떤 한 시스템 내에서 존재하는 하이퍼링크만을 제공한다. 그러나 이러한 시스템이 네트워크 상에 존재하는 여러개의 정보보존소중 하나의 역할을 수행하게 되고 이러한 정보보존소들 간에 하이퍼링크가 구축된다면, 네트워크상의 Client-Server 개념으로 표현되는 하이퍼미디어 시스템으로 확장될 수 있다. 이러한 네트워크 상의 하이퍼미디어 시스템으로 구현되는 디지털도서관은 현재 stand-alone 방식으로 구축되고 있는 하이퍼미디어 디지털도서관의 궁극적인 목표라고 볼 수 있다.

하이퍼미디어 시스템의 가장 핵심적인 구성요소는 멀티미디어 문헌과 하이퍼링크이다. 멀티미디어 문헌은 다양한 방법으로 정보를 표현하고 있다. 일차적으로 멀티미디어 정보는 텍스트 정보와 비텍스트 정보로 구분된다. 텍스트 정보는 컴퓨터의 아스키 (ASCII) 코드값으로 표현되거나 이미지로 표현할 수 있으며, 비텍스트 정보는 해당 정보의 특성에 따라 다양한 정보 표현방법을 갖는다. 하이퍼미디어 시스템을 구성하고 있는 멀티미디어 문헌 각각에 대한 구조 및 표현방법에 대한 연구는 각 연구분야에서 지속적으로 수행되고 있는 연구과제이다. 특히, 이미지나 사운드, 동화상 등을 표현하기 위한 표준안에 대한 연구는 인터네트의 WWW (World Wide Web) 발전과 함께 더욱 활발해지고 있다.

그러나 여러 유형의 멀티미디어 각각에 대한 연구가 활발한 것에 반하여 다양한 멀티미디어를 하이퍼시스템으로 통합하고 하이퍼링크를 사용하여 서로 관련있는 멀티미디어 문헌을 연결시켜 주는 하이퍼미디어 시스템 자체에 대한 구조 및 표준안에 대한 연구는 상대적으로 미비하다. 현재 수정작업 중에 있는 1992년에 제정된 ISO/IEC 10744 HyTime (Hypermedia /Time-based Structuring Language) 만이 멀티미디어 문헌의 구조화를 위한 유일한 국제 표준으로 인정되고 있다. 하이퍼미디어 시스템에서 링크는 그 문헌이 표현된 매체에 관계없이 문헌과 분리될 수 있다. 그러나 문헌의 구조는 문헌의 내용에 의존적이기 때문에 문헌과 분리될 수 없다. 따라서 문헌의 구조는 문헌의 내용에 대한 스트림 (stream), 계층구조 (hierarchies), 복합체 (composites) 등을 기준으로 표현 할 수 있다 (DeRose & Durand 1994, 22-25).

스트림은 연속된 데이터의 집합으로, 모든 문헌 구조의 가장 기본적인 구성요소이다. 텍스트 정보를 포함하고 있는 문헌에서 스트림은 아스키 코드일 수 있으며,

비디오 정보를 포함하는 문헌에서 스트림은 비디오 프레임일 수 있다. 그러나 이 모델은 계층구조에 기반을 둔 모델과 비교해 볼 때, 정보단위가 커지면 커질수록 정보의 조직이 어렵다는 단점을 갖는다. 계층구조로 표현된 문헌 구조는 문헌에 포함되어 있는 각 구성요소들 사이의 계층적인 상관관계에 기반을 둔다. 각 구성요소는 이미지나 사운드, 하이퍼링크 등 서로 다른 성질의 대상체를 내포할 수 있다. 이러한 계층구조는 시스템으로 하여금 문헌의 구조에 대한 명확한 지식을 가지게 함으로써 문헌의 직접적인 조작을 용이하게 한다.

한편, 상용되는 것은 아니나 스트림과 계층구조를 결합하는 시스템이 사용되기도 한다. 스트림과 계층구조를 결합하는 시스템은 이용자가 임의로 데이터 스트림을 조작할 수 있게 하고 문헌의 계층적인 구조를 편집할 수 있도록 한다.

일반적으로 인쇄된 문헌에서의 링크란 한 문헌내에서의 상관참조 (cross-reference)를 의미한다. 문헌내 상관참조를 위한 링크의 구조는 다음과 같은 세가지 구성요소를 갖는다.

- 참조를 나타내는 표시 자체 : 독자에게 현 문헌의 다른 곳에 해당 사항과 관련된 정보가 존재하고 있다는 것을 표시한다.
- 이 참조표시에 따른 참조 (reference) : 관련된 정보가 어디에 존재하는 지를 표시한다.
- 참조된 정보 : 참조에 따른 관련 정보를 의미한다.

하이퍼미디어 시스템의 링크에도 위의 세가지 구성요소가 그대로 적용된다. 그러나 링크를 사용하여 참조할 수 있는 정보는 문헌의 공간에 제한을 받지 않는다. 참조된 정보 지역을 앵커 (anchor)라고 하는데 이 앵커는 텍스트, 이미지, 비디오 프레임 등 다양한 멀티미디어 문헌이 될 수 있는 것이다.

HyTime 표준에서 앵커로의 참조는 로케이터 (locator)로 정의된다. 이 로케이터에는 앵커를 찾는 방법과 앵커가 다루어지는 방법에 대한 정보가 제공되어 특수한 포맷팅 기능을 수행할 수 있도록 허용한다. 하이퍼링크는 로케이터들의 집합체로 엄밀히 정의될 수 있다. 링크 내에서 각 로케이터는 하나의 링크엔드 (linkend) 속성을 갖는데, 이 링크엔드 속성값에 의하여 링크의 방향이 결정된다. 이러한 링크에는 앵커의 역할에 대한 정보가 추가될 수 있다.

하이퍼미디어 시스템은 하이퍼미디어 시스템에 수록된 문헌들이 서로 자유롭게 연결될 수 있도록 하며, 새로운 문헌의 추가를 용이하게 하고, 다양한 상호참조를 지원함으로써 독자들에게 글읽기에 대한 유용한 통로를 제공하며, 슬라이드나 오디오, 비디오 등 여러 유형의 미디어를 조합하고 그 순서를 정하는데 융통성을 발휘하는 특징을 갖는다. 또한 이용자가 데이터에 접근할 때, 이용자의 요구나 시스템의 환경에 따라 여러 가지 viewing point

를 제공하며 이에 따른 문헌의 갱신이나 수정시, 문헌의 버전을 관리할 수 있는 방법을 제공한다. 이와 같은 기능을 갖는 하이퍼미디어 시스템에서 유지되는 문헌은 역동적이며, 복잡하고 다양한 체계를 가질 수 있으며, 또한 서로 다른 질 (quality)의 정보가 하나의 문헌에 나타날 수 있고 다양한 포맷으로 표현될 수 있다는 특성을 갖는다.

3. 디지털도서관의 문헌 관리

3. 1 디지털도서관의 시스템 환경

디지털도서관은 도서관 자체적으로 수집하는 정보의 디지털화와 디지털정보의 효율적인 조직화, 이용자 우호적인 인터페이스의 개발 및 검색도구의 개발 등 다양한 기능을 수행해야 한다. 한편, 각 디지털도서관이 전세계로 확장되어 있는 네트워크 상의 한 정보보존소의 역할을 담당해야 한다는 측면을 고려해 보면, 디지털도서관은 자관이 소장하고 있지 못한 정보를 네트워크를 통해 전달받을 수 있어야 하고 다른 도서관에 자관의 정보를 제공해 줄 수 있어야 한다.

또한 각 정보보존소가 소장하고 있는 정보의 범위를 통합한 막대한 범위의 정보공간에서 효율적으로 정보를 검색하고 이를 이용자 개개인의 정보요구에 적합하게 가공하여 제공해 줄 수 있는 기능을 가져야 한다. 즉, 디지털도서관 환경에서는 정보공간이 네트워크를 따라 산재해 있는 광범위한 정보공간으로 확장된다.

도서관의 환경 변화는 디지털도서관의 이용자에게도 영향을 미친다. 디지털도서관의 이용자는 디지털도서관에서 제공하는 이용자 우호적인 인터페이스를 통하여 다양한 검색도구를 사용함으로써, 광범위한 정보공간에서 정보가 수록된 매체의 포맷이나 시스템의 상이한 환경에 관계없이 자신이 원하는 정보를 자신이 원하는 포맷에 맞추어 제공받게 되길 기대하게 되었다. 따라서 디지털도서관은 디지털도서관이 처한 환경과 이용자의 정보요구를 만족시킬 수 있는 디지털 정보의 효율적인 저장과 관리, 제공을 위한 제반 사항들을 고려해야 할 것이다.

3. 2 디지털도서관에서 문헌 관리시 고려점

앞서 살펴본 디지털도서관의 환경 변화는 디지털도서관의 시스템 구축시 많은 고려사항을 초래하였다. 막대한 양의 정보는 디지털도서관에 있어 문헌 조직화에 대한 중요성을 더욱 증가시켰으며 새로운 매체의 등장과 컴퓨터 환경의 발달은 디지털 정보의 저작권 문제 등 기존의 인쇄매체를 중심으로 하던 도서관 환경에서 크게 고려되지 않았던 사항들을 해결해야 할 사항으로 대두시켰다.

이러한 문제점들은 크게 사회, 경제적인 고려사항과 기술적인 고려사항으로 구분할 수 있다. 저작권이나 인간의 지적 산물에 대한 특성의 분석, 디지털 정보의 사용

에 따른 경제적 지불 문제, 디지털도서관의 정보 수집 범위 선정 등은 사회, 경제적인 측면에서의 대표적인 고려사항들이다. 이외에 기술적인 측면에서 디지털화된 문헌의 관리시 고려해야 할 사항들은 다음과 같다.

첫째, 디지털화된 문헌의 구성요소는 자연적인 형태로 표현되어야 한다. 문헌을 구성하고 있는 각 구성요소는 원 문헌의 구성요소에 따른 특징을 그대로 유지하고 있어 이 문헌을 조작하고 사용하는 이용자가 디지털화된 문헌에 익숙해 질 수 있어야 한다.

둘째, 이용자 우호적인 인터페이스가 지원되어야 한다. 이 사항은 먼저 디지털도서관이 개방 시스템(open system)으로 구축되어야 함을 의미한다. 디지털도서관 구축시, 이용자들의 요구사항을 조사하고 이를 분석하여 디지털도서관의 검색시스템에 그 결과를 반영시킬 수 있는 환경이 마련되어야 한다. 진정한 의미의 이용자 우호적인 인터페이스란 서로 이질적인 시스템들 사이에서의 정보 검색이나 정보 전달, 또는 정보를 표현할 때, 이용자들이 시스템의 사용에 어려움을 느끼지 않도록 함을 의미하기도 한다. 이는 서로 이질적인 시스템에 수록된 서로 다른 데이터 포맷을 갖는 정보들도 이용자가 정보의 결절을 느끼지 않고 자유롭게 검색하여 이용자의 편의에 따른 포맷에 맞추어 이용자에게 디스플레이 되거나 제공되어야 함을 뜻한다.

셋째, 디지털도서관 시스템은 진보된 검색방법을 지원해 줄 수 있어야 한다. 텍스트 정보에서 비텍스트 정보로 정보의 유형이 다양화되고 멀티미디어화 되면서, 각 정보의 특성에 맞는 정보검색기법이 개발되고 있다. 이러한 정보검색기법은 그에 따른 정보의 조직화가 선행되어야 한다. 과거 전통적인 도서관이 인쇄매체인 문헌을 대상으로 독자적인 분류 및 색인체제를 유지하고 있었던 것처럼 새로운 매체의 도입에 따른 효율적인 정보의 구조화 및 색인 시스템이 유지되어야 한다. 이러한 기능은 무한한 정보공간에서 이용자가 자신이 원하는 정보를 효율적으로 검색할 수 있도록 지원하는 가장 강력한 수단으로 작용할 것이다.

넷째, 막대한 정보가 효율적으로 저장되고 유지되어야 한다. 이 사항은 도서관의 'electronic storage repository'라는 측면이 부각된 것으로 (Graham 1996), 다양한 세부적인 고려사항이 부차적으로 존재한다. 빠르게 증가하고 있는 디지털화된 문헌 모두를 디지털도서관 시스템과 온라인으로 연결된 저장매체에 수록할 필요는 없다. 정보가 수록된 매체의 특성과 정보의 이용횟수, 정보의 중요도 등을 고려하여 정보에 대한 엑세스의 우선순위를 결정해야 한다. 정보를 저장하는 기억매체의 종류를 선택할 때, 이 우선순위가 적용될 수 있다. 디지털도서관에서 유지되는 정보들은 다른 응용시스템에서와 동일하게 자동적이고 효과적인 백업시스템을 갖추고 있어야 한다.

다섯째, 인증 (authentification)과 권한

부여 (authorization) 프로세스를 제공해야 한다. 인증(authentification)은 어떤 사람의 신분을 확인하여 시스템에 접근할 수 있도록 허용하는 처리과정이며, 권한부여(authorization)는 어떤 오퍼레이션을 허용할 것인지 허용하지 않을 것인지를 결정하는 과정이다 (Masinter 1995). 이 두가지 처리과정은 모든 컴퓨터 응용시스템에서 요구되는 공통적인 사항이라고 할 수 있다. 디지털도서관 시스템의 메타데이터나 하이퍼링크 등에 대한 추가나 갱신, 수정작업을 위하여, 또는 온라인 저작물을 유지하고 있는 디지털도서관에서는 이와 같은 프로세스가 반드시 제공되어야 한다.

여섯째, 디지털화된 문헌을 식별할 수 있는 식별기호 체제가 마련되어야 한다. 문헌의 식별기호는 정보 공간에서 해당 문헌으로의 일관성있는 참조를 가능하게 한다. 식별기호 체제는 식별기호를 부여하는 문헌의 단위나 정보공간에서 문헌의 위치 등을 고려하여 문헌의 식별기호에서 해당 문헌에 대한 정보를 추측할 수 있는 체계를 유지해야 한다. 또한 수시로 변경되는 역동적인 정보공간에서 문헌식별기호의 고유성을 유지할 수 있는 체제를 가져야 한다.

일곱째, 디지털화된 문헌에 대한 메타데이터가 유지되어야 한다. 메타데이터란 문헌의 일부분인 데이터가 아닌, 문헌에 관한 데이터를 뜻하며 네트워크로 연결된 정보 자원들의 기술 (description)과 조직, 검색, 접근에 대한 핵심적인 역할을 담당한다. 메타데이터를 구성할 때에는 문헌의 특성을 표시하는 표준적인 속성을 사용함으로써 서로 이질적인 여러 시스템을 대상으로 정보를 검색할 때, 오류가 발생하지 않도록 해야 한다.

여덟째, 문헌 내부와 문헌들 사이에 존재하는 링크들은 체계적으로 기록되고 유지되어 조직화되어야 한다. 또한 필요시 링크생성 프로세서에 의하여 자동적으로 링크가 생성될 수 있어야 한다.

아홉째, 디지털 정보의 refreshing이 지속적으로 이루어져야 한다. 이는 컴퓨터 기술의 진보에 따른것으로서, 이에 관한 포괄적인 연구가 1994년부터 미국의 Research Library Group과 Commission on Preservation and Access에 의하여 이루어져 왔다. 발전하는 컴퓨터 관련 기술은 기존의 시스템에 대한 변화를 요구하고 있다. 이 변화는 지속적으로 시스템에 수용되어야 한다. 만일 refreshing 작업이 수행되지 않는다면, 새로운 기술의 도입시 기존의 시스템에 대한 기술적인 차이에 의하여 이전 시스템을 새로운 시스템에서 사용할 수 없게 될 수 있다. refreshing 작업은 기존 시스템에 수록되어 있던 데이터의 포맷을 현재의 시스템에서도 활성화시킴으로써, 이전 포맷을 현재 시스템의 데이터 포맷으로 전환시켜 앞으로 등장하게 될 새로운 기술에 대비한다.

3. 3 디지털도서관에서 표준의 중요성

디지털도서관에서 다루어지는 문헌 관

리시 고려해야 할 사항들을 살펴보면 디지털도서관에서 표준이 중요한 역할을 담당하고 있음을 알 수 있다. 이질적인 소프트웨어와 하드웨어 환경 사이에서 정보흐름의 결절 없이 정보를 검색하고 교환하기 위한 고려사항이 대부분을 차지하기 때문이다.

지금까지 이러한 문제에 대한 강력한 해결책으로 손꼽혀 온 것은 각 분야에 대한 표준안을 제정하고 이를 준수한 어플리케이션을 개발하는 것이었다. 정보를 처리하기 위한 각 부분에서 이미 많은 표준들이 국제표준화기구와 그 외 여러 표준화기구를 통해 제정되었다. 멀티미디어를 위한 표준들은 JPEG, MPEG 등과 같이 각 매체마다 독립적으로 제정된 표준들이 존재한다.

앞서 언급한 것처럼 하이퍼미디어의 조직과 링크의 생성을 위한 국제 표준은 ISO 10744 HyTime이 유일하다. HyTime은 텍스트 문헌의 구조화를 위한 국제 표준 언어인 ISO 8879 SGML (Standard Generalized Markup Language)의 응용으로서 SGML과 상호 호환될 수 있다. SGML의 장점을 그대로 지니고 있는 HyTime의 가장 강력한 특징은 하이퍼미디어 시스템을 생성하기 위하여 멀티미디어 문헌을 조직화 및 구조화할 수 있는 방법과 강력한 링크기능을 제공하는 것이다.

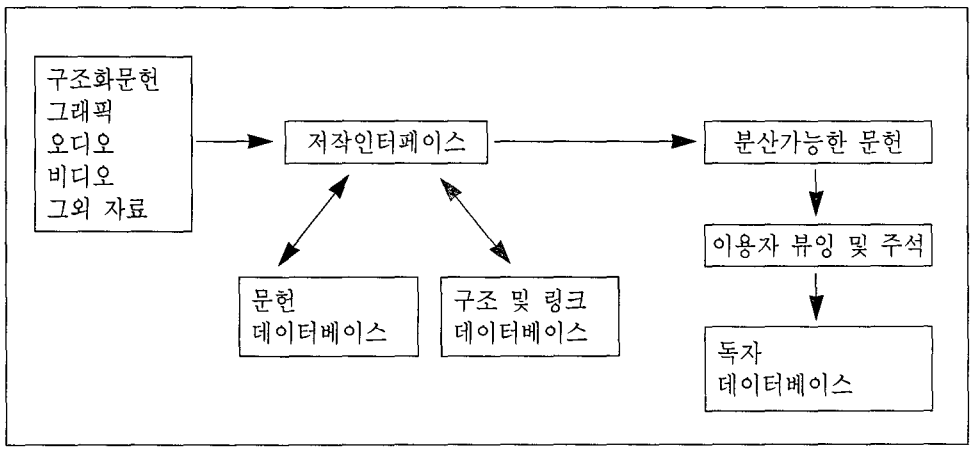
4. HyTime

4. 1 등장배경 및 개요

HyTime은 SGML을 개발한 Charles F. Goldfarb에 의하여 주도적으로 개발되었다. 1985년 음악분야의 어플리케이션을 위한 표준안이 ANSI X3 SD3-542D로 제안되었다. 이 표준안을 기반으로 음악의 상호참조 기능을 확장하고 음악적 구조에서 나타나는 복잡한 분기와 반복을 표현하기 위하여 1989년 Goldfarb는 hyperlinking과 location addressing architecture를 개발하였다. 이 사항들은 Standard Music Description Language (SMDL)에 반영되었고, 이 SMDL에서 'time model' 부분만을 분리하여 HyTime 표준안이 제정되었다. 이후 SMDL과 HyTime은 분리되어 ISO 10743과 ISO 10744로 표준화되었다.

HyTime은 HyTime meta DTD를 통하여 SGML의 하이퍼텍스트와 멀티미디어 지원기능을 확장시켰으며 강력한 하이퍼링크를 생성할 수 있도록 한다.

HyTime은 SGML에서 지원하는 직접적인 링크 기능외에 간접적인 링크기능과 ID가 부여되지 않은 임의의 장소를 링크할 수 있게 하는 기능을 갖는다. HyTime meta DTD는 architectural form에 의하여 표현되는 규칙들의 집합이다. architectural form에는 element type architectural form과 attribute list architectural form, 두 가지가 있다. element type architectural form은 기본 엘리먼트의 유형과 엘리먼트에 부가할 수



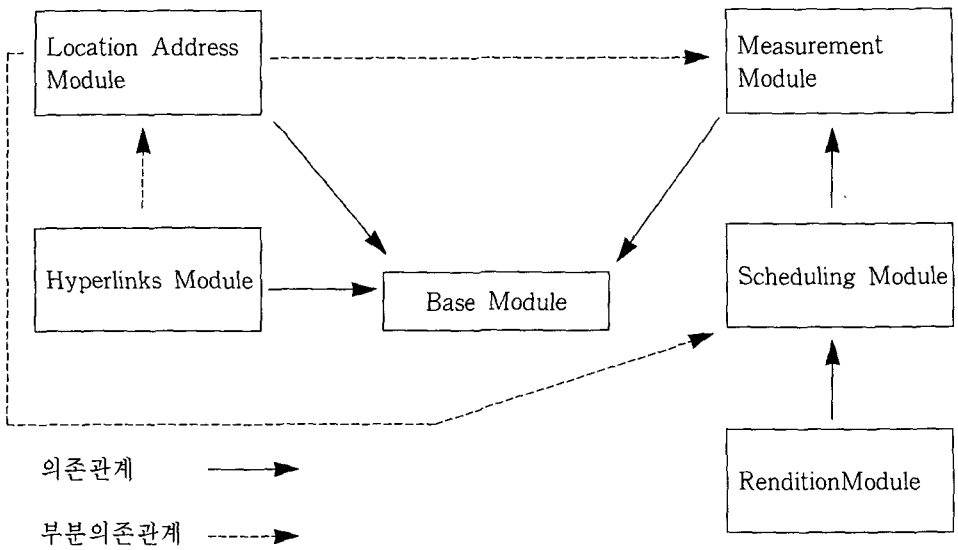
〈그림 2〉 HyTime을 적용한 하이퍼미디어 시스템 환경 (DeRose & Durand 1994, 95)

있는 속성들 및 속성에 허용되는 속성값이 선언되어 있다. attribute list architectural form은 하나의 속성 리스트 선언으로 구성된다. attribute list architectural form 중 일부는 특정 element type architectural form에만 적용되는 속성들을 정의하기도 한다. 〈그림 2〉는 HyTime을 적용한 하이퍼미디어 시스템 환경을 표현한 그림이다.

4. 2 HyTime의 구조와 선언

HyTime은 6개의 모듈로 구성되어 있다. 각각의 모듈은 architectural form의 한 세트를 포함하고 있으며 이 architectural form의 세트가 각 모듈이 제공하는 기능들을 결정한다. 이용자는 자신의 필요에따라 6가지 모듈들을 취사선택하여 하이퍼미디어를 구축할 수 있다. 다음은 6개 모듈들이 가지고 있는 각각의 기능들이며 〈그림 3〉은 6개 HyTime 모듈 상호간의 의존성을 표시하고 있다.

- Base Module : HyTime 문헌이 작동하기 위한 기본적인 최소한의 기능을 제공한다. Base Module의 일부분은 HyTime 문헌에서 항상 요구된다.
- Measurement Module : 오브젝트의 위치와 범위를 기술하기 위한 기능들을 제공한다.
- Location Address Module : 오브젝트에 부여된 ID에 상응하는 기능을 제공한다. 오브젝트의 위치는 name과 coordinate position, semantic construct를 사용하여 명시할 수 있다.
- Hyperlinks Module : 문헌의 계층구조와 독립적으로, 정보 엘리먼트들 사이의 관계를 명확히 기술하는 기능을 제공한다. Location Address



〈그림 3〉 HyTime 모듈의 상호 의존성 (Ferris 1996)

Module과 함께 사용하여 ID가 부여되지 않은 임의의 장소들 사이에 하이퍼링크를 설정할 수 있다.

- Scheduling Module : 한정된 대등 공간 (finite coordinate spaces)의 상관축에 따라 정보 대상체의 scheduling을 허용한다.
- Rendition Module : Scheduling Module이 사용될 때 rendition process에 적용되는 파라미터를 표현할 수 있는 수단을 제공한다.

HyTime 문헌은 기본적인 SGML 선언문에서 "FORMAL YES, APPINFO "HyTime""을 기존의 SGML 선언문과 대치함으로써 선언된다. 프롤로그(prolog)는 HyTime 버전 선언과 HyTime 문헌에서 사용되는 모듈의 종류를 기술함으로써 완성된다. 다음은 HyTime 선언문의 예이다.

```
FORMAL YES
APPINFO "HyTime"

〈?HyTime VERSION "ISO/IEC 10744:1992" HYQCNT=32 〉
〈?HyTime MODULE base context 〉
〈?HyTime MODULE locs pathloc 〉
〈?HyTime MODULE links 〉

〈!DOCTYPE manual 〉
〈!ENTITY % loc "nameloc | pathloc" 〉
〈!ENTITY % link "clink | ilink" 〉
```

```
〈!ELEMENT manual - O
 (front, body, append*)+ +(%loc; |
 %link;) 〉
〈!ATTLIST manual HyTime
 NAME HyDoc 〉
        ...
```

위의 예에서는 base, location address, hyperlinks 모듈이 사용되었다. base모듈과 *location address*모듈 옆에 나타난 context와 pathloc는 각 모듈에서 제공하는 선택사항으로서, 이 HyTime 문헌에서 해당 기능들을 이용하겠다는 선언적 의미를 갖는다. 위의 예에서 선언된 manual 엘리먼트의 속성 'HyTime'은 HyTime 문헌에서 사용되는 모든 엘리먼트에 공통적인 속성으로서, manual 엘리먼트가 HyTime DTD에서 제공하는 HyDoc element type architectural form의 특성을 상속받고 있다는 것을 이용자나 응용프로그램에게 알려준다.

manual 엘리먼트에 부여되는 속성이 HyTime 하나만이라면 manual 엘리먼트는 HyDoc element type architectural form에서 제공하는 모든 속성들을 상속하는 것으로 간주한다. 이용자는 필요에 따라 해당 element type architectural form에서 제공하는 속성들 중 일부를 엘리먼트의 HyTime 속성 다음에 기술함으로써, 엘리먼트에 부가되는 속성들을 제한할 수 있다. 그러나 architectural form에서 제공하지 않는 속성을 이용자 임의로 추가할 수는 없다.

4. 3 HyTime에서 제공하는 하이퍼링크 기능

HyTime은 정보 대상체에 간접적으로 주소를 부여하는 기능을 제공한다. 이 기능은 링크 엘리먼트와 링크가 지정하는 정보영역을 분리시킨다. 이는 다른 문헌으로의 링크 엘리먼트만을 포함하고 있는 외부 문헌을 따로 작성하여 유지할 수 있게 하며, 저작과 링크 부여작업을 분리시켜 업무의 효율성을 증대시키고 링크의 관리를 용이하게 하는 장점을 제공한다. 간접적으로 주소를 부여할 때에는 name, position, semantic property를 이용하고, HyTime의 name location address (*nameloc*) *element type architectural* form이 기본적으로 사용된다.

HyTime에서 제공하는 링크는 contextual link인 *clink*와 independent link인 ilink가 있다. clink는 한 문헌 내부에서 DTD에 의거하여 clink가 허용되는 모든 장소로의 링크를 제공한다. 예를 들어, 어떤 문헌의 한 구절이 그림을 참조하면 그림에 딸린 표가 동시에 나타나야 한다고 가정해 보자. 다음은 이러한 상황을 허용하는 DTD에 따라 문헌을 태깅한 예이다.

```
〈book〉〈title〉… 〈/title〉
〈chapter〉
〈para〉이러한 현상은 〈figref linkend =
 figloc〉 그림1 〈/figref〉에서 확연히
 나타난다.
```

```
〈/para〉
〈figure id=fig1〉
〈graphic entityref="fig1.gif"〉
〈/figure〉
〈table id=table1〉
〈title〉...〈/title〉
...
〈/table〉

〈nameloc id=figloc〉
〈nmlist nametype=element
 obnames= nobnames〉fig1
 table1〈/nmlist〉
〈/nameloc〉
〈/chapter〉
〈/book〉
```

figref 엘리먼트는 clink 엘리먼트로 이미 선언되었다고 간주한다. *nameloc* 엘리먼트는 ID가 fig1과 table1인 엘리먼트로의 링크를 설정한다. 이때 fig1과 table1은 서로 분리되어 있는 정보영역이 아닌, 동일한 정보영역으로 다루어 진다. 따라서 그림은 실제로 *fig1.gif*이라는 파일에 저장되어 있으나 항상 표와 함께 디스플레이 된다.

ilink는 링크 엘리먼트가 나타나는 위치와 엘리먼트의 linkends 속성에 의하여 기술되는 앵커의 위치가 서로 독립적이다. 즉, clink의 앵커가 링크 엘리먼트가 나타나는 동일한 문헌에 기술된 것에 반하여 ilink는 링크를 설정하는 링크 엘리먼트가 첫번째 문헌에 존재한다면, 앵커인 표1이 두번째 문헌의 내용중에 나타날 수도 있다는 것을 뜻한다. ilink의 linkends 속성은 속성값을 여러개 가질 수 있기 때문에 ilink 엘리먼트는 한번 기술함으로써 여러 문헌에 산재해 있는 앵커를 동시에 연결시킬 수 있다. 앵커들은 각각 독립적인 앵커로 다루어지거나 하나의 집합체로 다루진다.

4. 4 HyTime에서 제공하는 주소 및 위치 부여 기능

HyTime의 대등주소 부여 기능(coordinate addressing)은 mesurement 모듈에서 제공한다. 대등 주소를 부여할 수 있다는 것은 어떤 축이나 범위에 따라 대상체를 식별할 수 있다는 것을 의미한다. 축에서 위치를 식별하려면 축을 따라 형성되는 인정된 측정법이 필요하다. HyTime에서 모든 측정법은 표준 측정단위 (Standard Measurement Unit : SMU)를 사용하여 표현된다. 한 축 위에서 dimension (position과 extent)을 표현하기 위하여 사용되는 값은 한 쌍의 axis markers에 의하여 결정된다. 한정된 범위내에서 각 marker의 음수, 양수 표시는 시작하는 위치와 계속하는 위치를 결정하기 위하여 사용한다. 〈표 1〉은 axis marker의 계수 방향을 설명하고 있다. 예를 들어 1에서 7까지로 한정된 범위에서 axis marker 2와 -3은 위치 2와 5를 표현하고 있다.

location address module은 대등 위치

〈표 1〉 Axis Marker의 계수 방향 (Ferris 1996)

Marker 1	계수 방향
양수	한정된 범위의 처음부터 앞으로 진행
음수	Marker 2가 양수일 때 : 범위의 끝에서부터 뒤로 진행 Marker 2가 음수일 때 : Marker 2로부터 뒤로 진행
Marker 2	계수 방향
양수	Marker 1으로부터 앞으로 진행
음수	범위의 끝에서부터 뒤로 진행

(coordinate position)에 의한 주소부여 기능을 지원하는 다음과같은 architectural form을 제공한다.

- tree location address (treeloc) : 문헌의 인스턴스 파일은 계층적으로 구조화 되어 있다. 이러한 구성요소들 간의 관계는 트리 구조로 표현할 수 있는데, 가장 상위의 엘리먼트를 root로 간주하여 하부 구성요소들과 root와의 관계에 따라 문헌의 각 구성요소들을 표시한다.
- path location address (pathloc) : 트리 구조로 표현된 문헌의 구조가 부모 노드와 자식 노드의 관계를 통하여 하나의 엘리먼트만을 표시하는 것에 대응하여 여러개의 엘리먼트를 식별하고자 할 때 사용한다.
- relative location address (relloc) : path location address와 유사하나 root와 relation 속성을 사용하여 엘리먼트를 식별한다. relation 속성의 속성값은 esib(elder siblings), ysib (younger siblings), anc(ancestors), des(descendants) 중 하나로 결정된다.
- data location address (dataloc) : 비트 (bit) 조합이나 스페이스로 구분되는 토큰 (token)을 quantum 속성의 속성값으로 사용하여 데이터의 위치를 지정한다. 예를 들어 다음과 같은 문헌 인스턴스의 일부분이 있다고 가정해 보자.

〈p id=para1〉
본 문단은 예제를 위한 것입니다.
〈/p〉

위의 문장에서 '예제를 위한 것입니다'의 위치를 지정하기 위한 dataloc 엘리먼트는 〈dataloclocsrc = para1 quantum = word〉 〈dimlist〉-3 3〈/〉〈/〉으로 나타낼 수 있다.

이와 같은 location address 엘리먼트들을 서로 연결시켜 특정 대상체의 위치를 표시할 수 있게 하는기능을 HyTime의 위치사다리 (location ladder)라 한다. 위의 예를 아래와 같이 확장시켜 간단한

위치사다리 기능을 이용하여 동일한 문장의 동일한 위치를 지정하는 과정은 다음과 같다.

〈p id=para1〉
본 문단은 예제를 위한 것입니다.
〈emp〉이 예제와 연관된 문단은 세번째 문단입니다.〈/p〉

〈/p〉 dataloc에 treeloc 엘리먼트를 추가하여 다음과 같은 엔트리를 작성하였다. treeloc의 marklist에 의한 dimension 기술인 1 1 은 locsrc의 첫번째 자식 노드인 문단 내 첫번째 문장을 가리킨다. 이 두 엘리먼트는 locsrc 속성의 속성값을 통하여 서로 연결된다.

〈treeloc id = treeloc1 locsrc = para1〉〈marklist〉1 1〈/〉〈/〉
〈dataloc locsrc = treeloc1 quantum = word〉〈dimlist〉-3 3〈/〉〈/〉

이밖에도 의미적인 위치 주소 형태인 property location address (proploc)는 QPN (Qualified Property Name)을 사용하여 대상체의 속성이나 이용자가 정의한 property의 값을 획득 할 수 있는 방법을 제공한다. PQN은 하나 또는 그 이상의 property name으로 구성되어 있다. 각 property name은 property set name (psn) 속성과 specifier for property (spec) 유형의 엘리먼트를 동반할 수 있다.

5. 결 론

인터네트의 확산과 더불어 전자문헌이 일반화되고 있는 요즈음, 전통적으로 인쇄문헌을 다루어왔던 도서관에서 전자문헌 또는 디지털 문헌의 특성을 조사·연구하고, 특성에 따른 문헌의 관리 및 이용방법을 모색하는 일은 매우 중요한 문제이다. 인쇄된 문헌과 동일한 이미지를 줄 수 있는 디지털 문헌의 개발은 아직 초보적인 단계이지만 문헌구조화 기술 및 관련 표준에 대한 연구가 활발히 진행되고 있기 때문에 가까운 장래에 실현될 것으로 예측된다. 정보기술과 네트워크의 발전은 항상 인간의 예측을 뛰어넘어 왔으며, 디지털 문헌의 확산 및 보급이 이를 가속화시킬 것이다.

본 연구는 디지털도서관 환경하에서 문헌의 특성은 어떻게 변화할 것이며, 그에 따른 문헌의 관리는 어떻게 하여야 할 것인가에 관한 기초적인 연구라 할 수 있다. 지금까지 없었던 디지털 문헌의 등장에 따라 이들 문헌들을 어떻게 조직하여 관리하고 이용시키는가 하는 문제는 문헌정보학적인 측면에서 매우 중요한 연구과제이다. 전통적으로 인쇄된 문헌을 수집하여 정리하고, 이를 이용시키는데 관련된 연구를 수행하는 분야가 문헌정보학의 일부분이었다는 것을 아무도 부정할 수 없다면 당연히 디지털 문헌의 수집, 정리 및 이용 또한 문헌정보학의 한 영역이다. 아직까지 이에 관한 기초적인 연구가 없었다는 것은 매우 유감스러운 일이다. 그러나 최근

들어 문헌의 디지털화, 텍스트인코딩, 문헌구조화, 메타데이터 등 관련 연구의 움직임이 싹트기 시작하였다는 것은 매우 고무적인 현상이라 할 수 있다.

이런 측면에서 아직까지 국내에 잘 알려지지 않은 HyTime 표준을 소개하고, 이 표준이 멀티미디어문헌의 구조화에 어떻게 이용되는지를 개략적으로 제시하는 것도 의미가 있다고 판단된다. 따라서 본 연구는 우리말로 된 인쇄된 문헌을 멀티미디어 디지털 문헌으로 만들기 위한 구조화 연구 및 우리말의 특성이 반영된 표준에 관한 연구 등의 환경조성을 위한 기초연구라고 할 수 있다.

컴퓨터의 등장으로 정보시스템 또는 정보처리라는 미명아래 살아있는 문헌을 죽은 문헌으로 만들어 이용하여 온 것이 오늘날의 현실이다. 문헌이 갖고 있는 본래의 구조를 디지털화에 반영시킨다면 디지털 문헌은 더이상 죽은 문헌이 아니다. 이러한 문헌이 가능할 때 디지털도서관은 이용자 중심적이 될 것이다. 따라서 살아있는 디지털 문헌을 만드는 것이 문헌정보학의 미래 과제중 하나이다.

참고문헌

한국문헌정보학회, 1996. 「국가디지털도서관 구축계획에 관한 연구 최종 보고서」 서울 : 국립중앙도서관

DeRose & Durand 1994, *Making Hypermedia Work*, Boston : Kluwer Academic Publishers

Catlett & Terstrief 1992 "The Use and Effect of Multimedia Digital Libraries in a National Network." 84-98 *Newwork, Open Access, and Virtual Libraries*, Illinois : Graduate School of Library and Information Science

Travis, Brian E. & Dale C. Waldt. 1995. *The SGML Implementation Guide*. New York : Springerverlag.

ISO/IEC 10744: *Hypermedia/Time-based Structuring Language*.

A Brief History of the Development of SMDL and HyTime 1996 (http://www.techno.com/history.html)

Cleveland 1995 Overview of Document Management Technoogy (http://www.nlc-bnc.ca/ifla/VI/5/op/udtopr. htm)

Ferris, R. 1996 HyTime Application Development Guide Version 1.2.4 (http: //www.techno.com/hadg/)

Graham 1996 Requirement for the Digital Research Library (http: //aultnis. rutgers.edu/texts/DRC.html)

HyTime Working Group, Quick Guide to HyTime Basics TR1V1 (http://info.admin.kth.se/SGML/Anvandarforening/Arbetsgruppe r/HyTime/Reports/tr1v1.html)

Masinter 1995 Document Management, Digital Libraries and the Web (http://www.isoc.org/HMP/PAPER/243/ html/paper.html)

Preserving Digital Information, Report of the Task Force on Archiving of Digital Information (http://www-rlg.stanford.edu/ArchTF/tfadi.index.htm)