

디지털 도서관의 기술요소 및 문헌구조화와 SGML

울산대학교 의학도서관

최 홍 식

I. 서 론

오늘날 과학기술의 발달과 각종 전자 매체의 급속한 발전으로 인하여 도서관은 새로운 형태의 모습으로 변화되어야 한다는 인식이 팽배하고 있다. 디지털도서관은 그 대안으로 빠르게 대두되었고 이미 미래의 도서관으로서 필연적으로 갖추어야 할 형태로 인식되고 있다.

디지털도서관은 정보자원을 이용하기 위하여 데이터의 디지털화가 필수적으로 선행되어야 하므로 이를 강조하기 위한 용어로 도서관 소장자료와 정보서비스 및 다른 정보자원에 대한 원격 접근을 기본개념으로, 전자적 네트워크를 통하여 전세계적인 정보 및 지식자원에 대한 접근은 물론 그 전문을 받아 이용이 가능하게 하는 것으로 시간과 공간을 초월한 접근과 이용이라고 그 개념을 정의하고 있다.

이러한 변화에 도서관이 효과적으로 적응하고 대처하기 위하여 적극적인 관심과 지속적인 노력을 경주함에 있어 예외 일 수는 없을 것이다.

따라서 본고에서는 디지털도서관에 요구되는 기술과 문헌의 디지털화의 방법론 및 멀티미디어 데이터의 디지털화에 대하여 살펴보고, 문헌의 구조화와 SGML에 대하여 고찰하고자 한다.

II. 요구되는 기술

디지털도서관에 요구되는 기술은 과업을 수행하

기 위한 기본적인 도구와 정보처리 기술, 표준화로 대별하여 살펴보면 다음과 같다.

1. 기본적인 도구

디지털도서관에 필요한 기본적인 도구를 살펴보면,

첫째, 한글 전자 사전 및 시소러스이다. 이는 디지털도서관의 전제사항으로서 자연언어처리 및 검색을 위한 기초도구이다.

둘째, SGML Application tool로서 editor, parser, automatic markup등 이다.

에디터(editor)는 DTD(Document Type Definition)를 기본으로 하여 DTD에 기술된 순서대로 문헌요소의 시작태그를 화면상에 출력하고 데이터 입력을 촉구하는 것으로 저자는 시작태그나 종료태그를 의식하지 않고 TEXT를 입력할 수 있다. 파서(parser)는 검토와 변환, 링크의 기능을 가진다. 즉 DTD를 검토하고, 작성한 TEXT가 DTD에 따라 작성되었는지를 검토하며, 입력한 TEXT로 부터 완전한 SGML TEXT DB로 변환하는 기능, SGML 형식의 DB로 부터 다른 형식(예> DSSSL등)의 파일로 변환하는 링크의 기능을 가진다.

셋째, 멀티미디어와 SGML 데이터를 처리할 수 있는 DBMS이다. 사운드 뿐만 아니라 이미지 및 동화상등의 멀티미디어 데이터를 처리할 수 있는 데이터 베이스 관리 시스템이 필요하다. 그 밖에 Web Server, Web Browser 및 전거통제를 위한 데이터등을 기본도구로 필요로 한다.

2. 정보처리 기술

디지털도서관에 정보처리의 기술을 살펴보면, 첫째, 한글 자동색인 및 검색기술이다. 자동색인의 기본원리는 문헌을 구성하는 단어들을 일정한 기준에 의거하여 주제어와 비주제어로 구분하고 주제어로 평가된 단어로 부터 색인어를 선정하는 것이다. 따라서 색인어의 자동부여, 색인어의 자동추출, 시소러스의 자동 생성등의 기술을 요한다.

둘째, 한글 자연언어 처리 기술이다. 자연언어가 가지고 있는 애매성과 표현의 다양성, 구문분석의 수준, 사용할 문법의 선택 및 의미분석의 수준등이 고려되어야 한다. 구체적으로 내포문의 처리나 복잡한 명사구의 처리, 관계대명사나 생략구문의 처리등을 말할 수 있다.

셋째, 한글 OCR 및 이미지나 비디오의 색인 및 검색기술이다. 광학문자인식 시스템은 현재 속도가 많이 빨라져 과거에 비하여 시간이 상당히 절약되는 장점이 있다. 그러나 정확도에 있어서 99%의 정확도가 있더라도 나머지 1%를 수정하는데에는 많은 인력과 시간이 필요하여 보다 완벽한 기술이 개발되어야 한다. 이미지와 비디오의 색인 문제는 정지화상과 동화상의 색인과 그에 따른 검색기술의 개발이 요구되고 있다.

넷째, SGML데이터 처리기술과 SGML Application 도구 개발 기술이다. SGML에 관련된 제품은 현재 개발이 많이 되어 있다. 구미의 MARK-IT, Author/Editor, The Publisher, DynaText등이 있고, 일본의 SGML Plus 등이 있다. 그 밖에 정보처리의 기술을 보면, Full Text/Image database modelling, 다양한 유형의 정보 디스플레이 포맷, Meta data의 표준화, VOD 관련 기술 및 오디오의 색인, 저장 및 검색 기술, 멀티미디어의 압축 및 저장과 관련된 기술 등으로 개발을 요한다.

3. 표준화

표준화 기술은 디지털도서관의 핵심기술의 하나로서 여러가지 관련기술등과 통합적으로 연구되어

져야 하고, 상호연결을 위하여 절대적으로 필요한 기술이다. 이와 관련된 기술을 간략하게 살펴보면 다음과 같다. 먼저 데이터의 표준으로서 문자의 코딩은 검색이 가능한 ASCII코드로, 문헌은 SGML로 표준화 하고, 메시지처리의 표준은 전자메일 X.400과 디렉토리 X.500, 정보검색 관련 표준은 정보검색 프로토콜로서 전문검색, 질의언어, 데이터 포맷 개선 등을 포함하는 Z39.50, 범용 통신 명령언어의 Z39.58, 마크업 표준의 Z39.59, 전자 데이터의 교환 표준의 X.12, 멀티미디어 관련 표준으로서 사진, 필름자료등의 정지화상의 표준화인 JPEG(Joint Photographic Experts Group)와 실시간으로 애니메이션과 음성의 압축 전개를 위한 알고리즘을 의미하는 동화상을 위한 MPEG(Moving Picture Experts Group), 통신을 위한 프로토콜로서 TCP/IP(Transmission Control Protocol/Internet Protocol)와 OSI(Open Systems Interconnection), 이미지 포맷에 관한 표준화등이다. 그 밖에 오디오 정보 포맷을 위한 UNIX AU(Unix-audio-film Format)등이 있다.

이러한 표준은 미국의 ANSI, NISO와 국제적 단체인 ISO, IEC, CCITT등과 같은 표준관련 기관에서 제기되고 인정되고 있는 실정이다.

III. 디지털 도서관 관련 기술

1. 전자출판 분야

전자출판이란 종래의 단순한 종이매체를 이용한 인쇄출판이 아니다. 전자화된 정보를 기반으로 하여 종이매체 뿐만 아니라 CD-ROM등 그 밖의 전자매체로 출판하는 것을 전자 출판이라고 한다. 정보의 축적매체의 발전과정을 크게 살펴보면 종이매체에서 마이크로 매체로, 마이크로 매체에서 자기매체로, 자기매체에서 광매체로 발전하여 왔다. 또한 정보의 출판형식은 인쇄출판(종이)에서 전자출판(컴퓨터)으로 변화하여 발전하였다.

즉 오늘날의 출판은 광매체로 축적되는 전자출판이라 할 수 있다. 이러한 전자출판의 범위는 크

게 3가지로 나눌 수 있다.

첫째, 컴퓨터와 주변기기를 이용하여 종이책을 편집해 내는 편집의 전산화이다.

여기에는 DTP(Desk Top Publishing)나 CTS(Computer Typesetting System)등이 그 범주에 속한다.

둘째, 디지털 신호로 문자와 그림 등을 기록, 보존하는 전자 출판물이다.

여기에는 여러가지 장점으로 많이 이용되고 있는 CD-ROM이나 LD등이 포함된다.

셋째, 통신방법을 활용하여 대량보급을 가능케 하는 전자통신 출판이다.

비디오 디스크, 전자저널, 전자우편, 전자신문, 전자회의등이 있다.

이러한 전자출판은 디지털도서관의 장서개발과 밀접한 관계를 가지고 있으며, 이들을 네트워크 환경하에서 적절하게 운영하는 것이 기술적인 과제이다.

2. 하이퍼미디어 분야

디지털도서관에서 하이퍼미디어 분야는 매우 중요하다. 하이퍼미디어란 멀티미디어와 하이퍼텍스트가 조합되어진 개념으로 이해되고 있다. 하이퍼텍스트란 텍스트 정보내에서의 노드 단위간의 연관관계에 의하여 어떤 곳에서 다른 어떤 곳으로라도 입체적으로 이용자들이 이동할 수 있는 그물망의 구성으로 설계되어진 정보접근 방식이다. 하이퍼미디어에서는 시스템에서 정보를 조직함에 있어서 노드와 링크를 사용하는 하이퍼텍스트의 개념에 기초하고 있으며, 이때 노드는 멀티미디어에서 다루는 여러 유형의 미디어가 결합된다. 하이퍼미디어의 기법은 이용자들에게 일직선상에서 정보를 읽는 것이 아니라 같은 주제를 개념적으로 연결하여 원하는 정보를 이쪽 저쪽으로 마치 항해하듯이 찾을 수 있다.

디지털도서관에서 하이퍼 미디어 분야의 기술은 중요하다. 왜냐하면 저자들이 출판과정에서 하이퍼텍스트 시스템의 도움으로 저술작업을 수행하고 기사들을 통합하는 방법을 알게되고, 학문적인 활

동이 많은 하이퍼 베이스와 지식베이스를 생산함에 따라 적당한 형태로 정보의 개인적 조직화와 멀티미디어 책체장서의 구성이 가능하기 때문이다. 현재 하이퍼베이스와 지식베이스들은 많은 연구자들에 의하여 개발되고 있다.

3. 정보관리 분야

디지털도서관에서 정보관리분야는 특수한 기술을 필요로 한다. 즉 공간과 지리정보, 데이터의 압축에 관한 기술, 멀티미디어시스템 등에 전문화된 기술이 요구된다. 데이터 베이스의 관리방법은 관계형이건 오브젝트지향형이건 디지털도서관에서 문헌데이터베이스의 이용 뿐만 아니라 목록, 사용료의 관리, 사용 로그, 보안 통제, 다른 서비스의 관리와 직접적인 관련이 있어 매우 중요하다. 또한 텍스트 분석과 정보검색기법은 변환, 색인, 표현, 탐색, 결과물의 제공과 관련이 있고, 이용자 인터페이스나 가이드시스템은 이용자가 요구한 정보를 찾을 수 있도록 도와주는 시스템으로서 꼭 필요하다.

이러한 시스템은 전적으로 이용자 중심으로 설계되어야 하고 그 방법이 채택되었을 때 그 효과는 더욱 크게 나타날 것이다.

IV. 문헌의 디지털화

1. 문헌 디지털화 방법론

1) 이미지 기반 시스템(Image-based system)

문헌의 내용을 이미지로 인식하는 시스템으로서 검색유형은 키워드 검색이다. 즉, 전문가가 일정한 정의에 의하여 일관되게 키워드를 부여하고, 이용자는 시소러스나 검색요어집등을 이용하여 원하는 정보에 접근,검색할 수 있도록 하는 시스템이다. 다시 말하면 이미지로 표현된 각 문헌들에 전문가가 문헌의 내용을 나타내는 키워드를 부여하고, 이용자는 이 키워드를 이용하여 자신이 원하는 문헌에 접근하게 하는 것이다. 이 시스템은 문헌의 내용이 이미지로 저장되어 있어 검색이 불가능하므로 서지사항을 이용하여 문헌에 접근하여 이용

되므로 정보검색시 서지 데이터베이스를 이용하여야만 한다.

이 시스템은 구축이 편리하고, 키워드로서 검색하기 때문에 검색속도가 빠르며, 보조기억장치등에 의하여 문헌의 이미지만을 따로 보관하기 때문에 저장매체의 안정성과 관리의 편리성등을 장점으로 들 수 있다.

그러나 문자가 이미지로 표현되어 있기 때문에 전문검색이 불가능하고 키워드의 선정에 있어서 전문가의 선입관이나 주관이 개입될 여지가 있고, 동일한 개념을 표현하는데 있어서도 이용자와 다른 용어를 사용할 수 있기 때문에 검색의 정확율이 저하된다는 단점이 있다. 또한 대규모의 기억공간을 필요로하며, 이용자가 이미지 정보의 특성에 맞는 하드웨어와 소프트웨어를 갖추어야만 해당 정보를 사용할 수 있어 기기의 호환성 문제도 단점으로 지적되고 있다.

2) 아스키 기반 시스템(ASCII-based system)

문헌의 내용을 기계가독형의 아스키 코드로 변환시킴으로서 컴퓨터가 문헌에 나타나는 문자를 검색할 수 있게 하는 시스템이다.

이 시스템의 저장방법은 재입력하거나 스캐너, OCR 소프트웨어등을 이용하여 저장한다. 또한 이 시스템은 색인 파일이 자동적으로 생성되는데, 이것은 일반적으로 아스키 기반 시스템은 문헌의 색인 작성을 위하여 색인 생성 프로그램을 제공하기 때문에 새로운 문헌이 입력되면 문헌에 나타나는 단어를 기반으로 색인파일이 자동적으로 생성되기 때문이다.

이 시스템의 장점은 모든 문헌의 내용이 아스키 코드로 저장되기 때문에 전문검색과 주제검색이 가능하고, 입력에서 검색까지 다른 시스템에 비하여 시간이 단축되며, 이미지에 비하여 저장공간이 절약되고, 전송속도가 빠르다. 또한 그래픽이나 오디오, 비디오 정보와의 공존이 가능하기 때문에 응용성이 보증된다. 아스키 기반 시스템은 정보의 재사용이라는 측면에서 효율적이고, 그 밖에 현존하는 컴퓨터가 아스키 코드를 지원하고 있기 때문에 심리적인 편안함도 장점으로 들 수 있다. 그러

나 아스키 코드로 변환하는데 어려움이 따르고, 본문 이외의 테이블이나 도표등은 검색이 불가능하다는 단점이 있다.

3) 마크업 기반 시스템(Markup-based system)

마크업에는 고유(Specific) Markup과 범용(Generalized) Markup이라는 두 가지의 종류가 있다. Specific Markup은 문헌의 레이아웃을 기술하기 위한 명령체계로서 저자가 기술한 원고에 교정을 보거나 식자를 위한 폰트를 지정하는데 사용되고, 또한 주석을 기술하거나 편집자, 타이피스트가 한 문헌 안에서 어떠한 특정한 부분이 인쇄되고 편집되는 방법을 지시하기 위하여 사용한 마크업이다. 따라서 과거 출판계에서 널리 사용되었던 것으로서 절차적 마크업(procedural markup)이라고도 한다. 이 마크업의 단점은 특정 편집시스템이나 응용 환경에 제한적이라는 것이 단점으로 지적되고 있다.

Generalized Markup은 문헌의 레이아웃이 아닌 구조를 기술하는 마크업으로, 색인과 주, 인용텍스트 등 문헌을 구성하는 요소의 시작 부분과 종료 부분을 표시하는 마크업이다. 즉 마크업되는 사항이 무엇인지를 표시하는 태그로 마크업되는 부분이 문헌에서 어떠한 위치에 있는지를 나타내는 마크업이다. 범용마크업은 문헌의 레이아웃이나 폰트 등의 기술에 중점을 두지 않는다. 예를들어 어떤 문자열이 그 문헌의 표제라면 그 문자열에 <표제>라는 마크업을 부가하여 해당 문자열이 그 문헌의 내용을 대표하여 나타나는 열임을 나타내게 한다. 이 범용마크업의 특징은 특정 시스템에 제한되어 있지 않아 어떤 시스템에서도 이용이 가능하고, 문헌의 레이아웃정보와 내용정보를 서로 분리시킴으로서 고유마크업의 단점을 제거하였다는 것이다. 즉 고유마크업에서 발생하였던 각 시스템 고유의 마크업에 대한 의미와 코드 체계로 인한 문헌 교환의 문제점을 해소하였다. 이러한 마크업을 이용한 시스템은 수 많은 정보를 다양한 수준의 이용자에게 정보교환시 발생하는 여러 문제점을 해소할 수 있고, 정보의 재사용과 이용자의 입장에서 정보의 처리를 가능하게 할 것

이다.

V. 문헌의 구조화와 SGML (Standard Generalized Markup Language)

1. 문헌구조화의 필요성

디지털도서관에서의 모든 정보는 디지털화 되어 있기 때문에 과거 인쇄된 책자 형태에서 중요시한 문헌의 레이아웃은 그리 큰 의미를 갖지 못한다. 즉 각 편집시스템에서 원문에 부여한 편집정보들은 다른 시스템으로 전달되면 편집정보가 유실되거나 더 이상의 의미가 없는 정보가 되기 때문이다. 따라서 앞에서 살펴본 마크업으로 문헌의 포맷을 결정 하는 것이 정보의 교환과 유통적 측면에서 유리할 것이다. 문헌은 저자가 의도하고자 하는 내용을 표현하고 있고, 그 내용들은 각자 서로 유기적인 관계를 형성하고 있어 어떠한 특정 구조를 지니고 있다. 이러한 특성을 구조화하여 표현하고, 시스템에서 이러한 문헌의 구조를 인식할 수 있다면 정보 검색시 이용자에게 강력한 검색수단을 제공하게 될 것이다. 다시 말하면 같은 키워드라도 일반적으로 제목에 내포되어 있는 키워드가 절이나 구에 내포되어 있는 키워드보다 더 유용하기 때문에 강력한 검색수단을 제공하게 된다. 또한 모든 문헌을 구조화하고, 어떠한 시스템 하에서도 같은 내용의 형태로 표현된다면 무한의 정보공간이 확장될 뿐 만 아니라, 저자의 표출 내용과 의도가 정확하게 인지되어 전달될 수 있을 것이다. 즉 저자는 자기의 의도를 강조함에 있어 일반적으로 서체의 크기, 굵기, 위치등으로 표현하기 때문에 이러한 특징을 문헌의 구조화를 시킴으로서 그대로 디스플레이 됨에 따라 저자의 의도가 그대로 인지될 것이다.

2. SGML의 개요

SGML(Standard Generalized Markup Language)은 국제표준화기구에 의하여 표준으로 제정된 기술적 마크업 언어이다. SGML의 단어 하나 하나의 의미를 살펴보면 전체적으로 이해

가 될 것이다. 먼저 'Standard'라고 부르는 이유는 IBM사의 사내 문헌작성에 이용되고 있던 GML(Generalized Markup Language)을 표준화 하였기 때문이다. 'Markup'은 사전에 나타나지 않는 단어로 'Mark up'으로 만든 조어이다. 'Mark up'은 편집자와 인쇄 디자이너가 색인 및 텍스트 그 외의 부분에 대해서 서체나 글자크기 등 레이아웃에 관한 지정을 수작업으로 원고에 기입하는 것이다. 이와 같은 각종 지시사항을 수작업이 아닌 전자적 작업을 통하여 부여 하는 것을 'Markup'이라는 한 단어로 표기한 것이다. 'Language'는 일반적으로 언어를 의미하지만 컴퓨터 과학적인 측면에서 보면 문법(Syntax)을 가지고 있는 것을 말한다. SGML도 엄연히 문법을 채용하고 있기 때문에 Language라고 하는 것이다.

SGML은 문헌의 구조를 기술하는 메타언어이다. SGML로 기술할 수 있는 사항은 문헌의 구조와 한 문헌에서 전달되는 문자들, 문헌에서 한번 이상 사용되는 텍스트, 문헌의 외부에서 생성된 정보를 텍스트에 병합시키는 방법, 텍스트를 구성하기 위하여 사용되는 특수한 기술 텍스트가 처리되는 방법등이다. ISO 8879에서 규정하고 있는 SGML의 특징을 살펴보면 마크업된 문헌은 문헌처리와 워드프로세싱 시스템에서 광범위하게 처리되고, 다양한 기종의 텍스트 입력기를 지원할 수 있고, 문자세트에 독립적이고, 처리시스템이나 기기에 독립적이고, 어떤 한 국가의 언어에 기반을 두지 않고, 마크업 언어는 타이프라이터나 워드프로세 서에서 사용될 수 있고, 특정 데이터 스트림이나 물리적인 파일 조직에 독립적이고, 마크업 된 데이터는 마크업되지 않은 데이터와 공존할 수 있고, 마크업은 인간과 프로그램 모두 해독이 가능하고, 모든 유형의 데이터를 처리할 수 있다는 것을 그 특징으로 하고 있다.

SGML의 효용으로는 전자출판, 문헌 데이터베이스의 작성 및 유지 보수, 문헌의 교환 유통 등을 들 수 있다. 또한 개개의 시스템으로부터 독립되어 있으며, 구조가 엄밀히 정의 되어 있기 때문에 개개의 시스템에 적합하도록 변환시킬 수 있

다. 따라서 다양한 유형의 정보가 상호 교환되며, 다양한 환경에서 정보를 수집하고, 제공, 관리해야 하는 디지털도서 관의 요구사항에 현재로서는 가장 적합한 메타언어로 평가되고 있다.

3. SGML의 문헌의 구조

1) SGML 선언부(SGML declaration)

SGML 선언부는 문헌을 작성하고 응용 시스템을 실행시키기 위한 환경을 설정하는 부분으로 이 문헌을 특정 시스템하에서 처리할 수 있는지를 알 수 있게 한다. 즉 문자코드의 종류, 용량, 구문, 특별한 기능, 애플리케이션 정보 등 하드웨어에 관한 사항을 규정한다. 다시 말하면 SGML 선언은 사용하는 문자집합, 처리에 필요한 기억 용량, 구체구문의 유효범위, 사용하고 있는 구체구문, 사용하는 SGML기능, 처리에 필요한 어프리케이션 고유기능 등을 선언한다. SGML 선언은 SGML 문헌을 처리할 때 필요한 기능을 받는 측에 확실하게 전달하기 위한 것으로, 컴퓨터보다 오히려 받는 측의 사람이 읽는 부분으로 문헌을 특정 시스템하에서 처리할 수 있는지를 알 수 있게 하는 것이다.

SGML 선언은 SGML 문헌의 기본적인 문자셋인 ISO 646 ASCII 코드를 사용한다.

선언의 시작은 코드번호 60과 33인 <!로 시작하고 선언의 종료는 코드번호 62인 >로 끝을 낸다. 선언부의 제일 첫 번째 줄에는 선언부의 작성할 때 근거로 한 SGML의 버전을 나타내기 위하여 사용된 SGML의 표준번호와 날짜를 기록한다. SGML 선언부는 크게 6부분으로 구분할 수 있다. 이 6가지의 구분과 그 실례를 간략하게 살펴보면 아래와 같다.

-구분: CHARSET

-document character set details

-사용하는 문자 집합

: CAPACITY

-capacity set details

-처리에 필요한 기억용량

: SCOPE

-the scope of the following concrete syntax

-구체적 구문의 유효범위

: SYNTAX

-the concrete syntax to be used within the document

-사용되고 있는 구체 구문

: FEATURES

-special SGML features used within the document

-사용되고 있는 SGML기능

: APPINFO

-application-specific information entered within the declaration

-처리에 필요한 어프리케이션의 고유기능

2) 문헌유형정의부(Document Type Definition: DTD)

문헌유형정의부는 문헌분석 후 나타나는 문헌구조를 SGML 문법을 사용하여 기술하는 부분이다. 문헌은 책, 논문, 리포트 등 문헌의 유형에 따라 그 구조가 달라지기 때문에 각 문헌 유형에 따라 DTD는 서로 다르다. DTD는 SGML 구문에 대한 지식의 기반하에서 스스로 작성할 수도 있지만 각 문헌의 유형에 따라 DTD가 이미 준비되어 있는 경우도 있어 스스로 DTD를 작성할 필요 없이 준비된 DTD를 사용할 수도 있다. 문헌유형정의부는 3부분으로 구분된다.

첫째, 문헌요소선언부이다. 이는 각 문헌요소의 개괄적 식별기호(Generic identifier 일종의 태그명)를 정의하고 각 문헌요소가 문헌에 나타나는 순서를 정의한다.

둘째, 속성정의 리스트 선언부이다. 이는 각 문헌요소에 명시되는 속성을 정의하고, 이러한 속성값의 범위, 속성값의 초기값을 명시한다.

셋째, 엔티티선언부이다. 이는 긴 문자열을 대치하는 짧은 문자열을 선언하거나 자판기에 의하여 입력할 수 없는 문자를 선언하는 데 사용된다.

<예 1> 기본적인 SGML 선언부

<! SGML "ISO 8879-1986"

-Declaration for typical Basic SGML document-

```
CHARSET BASESET "ISO 646-1983//CHARSET International
Reference Version (IRS)//ESC 2/5 4/0"
DESCSET 0 9 UNUSED
          9 2 9
          11 2 UNUSED
          13 1 13
          14 18 UNUSED
          32 95 32
          127 1 UNUSED
CAPACITY PUBLIC "ISO 8879-1986//CAPACITY Reference//EN"
SCOPE DOCUMENT
SYNTAX SHUNCHAR CONTROL 0 1 2 3 4 5 6 7 8 9 10 11 12 13
                      14 15 16 17 18 19 20 21 22 23 24 25 26
                      27 28 29 30 31 127 255
BASESET "ISO 646-1983//CHARSET International
Reference Version (IRS)//ESC 2/5 4/0"
DESCSET 0 128 0
          FUCCTION RE 13
          RS 10
          SPACE 32
          TAB SEPCHAR 9
          NAMING LCNMSTRT ""
                  UCNMSTRT ""
                  LCNMCHAR "-."
                  UCNMCHAR "-."
                  NAMECASE GENERAL YES
                  ENTITY NO
          DELIM GENERAL SGMLREF
          SHORTREF SGMLREF
          NAMES SGMLREF
          QUANTITY SGMLREF
FEATURES MINIMIZE DATATAG NO OMITTAG YES RANK NO
          SHORTREF YES
          LINK SLMPLE NO IMPLICIT NO EXPLICIT NO
          OTHER CONCUR NO SUBDOC NO FORMAL NO
APPINFO NINE
>
```

이러한 DTD는 TEI(Text Encoding Initiative)와 밀접한 관계를 가진다. TEI는 공개와 공유를 목적으로한 텍스트 데이터베이스의 표준적인 코딩스키마를 작성하기위한 국제적 프로젝트이다. 그 목적은 텍스트 데이터베이스의 공통교환, 포맷의 작성 및 인코딩의 문제점을 해결하고, 텍스트를 인코딩할 때 어떤 요소가 코딩되어야 하며

또 그것을 어떻게 표현하는지에 관해서 구체적인 가이드라인을 작성하며, 주된 인코딩 방법을 조사하여 문서를 작성하는 동시에 그에따른 메타언어를 개발하는데 그 목적을 두고 있다. 현재 TEI의 태그수는 약 400여개이고, 헤더에서 이용되는 기본태그가 약 60여개이다. 결국 DTD는 TEI가 메타언어로 SGML을 선택하였으므로 TEI에 정

의된 태그를 DTD에서 이용하면 별 무리가 없을 것이다.

문헌유형정의부의 실예는 다음과 같다.

```

<예 2> 문헌유형정의부
<-문서요소선언->
<!ELEMENT textbook (front, body, rear)>
<-속성정의리스트선언->
<!ATTLIST fig id ID #IMPLIED>
<-엔티티선언->
<!ENTITY SGML "Standard Generalized Markup Language">
<-공개된 문서종류를 사용하는 선언(manual)->
<!DOCTYPE manual PUBLIC "-//Cave Press//DTD Manual//EN">

```

3) 실제문헌부(document instance)

실제문헌부는 ISO 8879에서 정의한 SGML 문헌에 대한 마크업 기술 방법과 문헌유형 정의부에 따라서 직접 문서를 기술하고 그 요소들을 식별하는 곳으로 각각의 문헌요소는 시작과 끝에 태그를 표시함으로써 구별된다. 실제문헌부의 실예는 다음과 같다.

```

<예 3> 실제문헌부
<textbook>
<front>
<title> 소아과학 </title>
<author> 홍창의 </author>
<content> 서론
    1. 성장과 발달
    2. 심리적 문제
.....
</content>
</front>
<body> <chap> 서론
<p> 소아과학은 임신부터 청소년에 이르기까지 .....
<p> “소아는 작은 어른이 아니다”라는 .....
</chap>
<chap>

```

```

.....
.....
</chap>
</body>
</textbook>

```

VI. 멀티미디어 데이터의 디지털화

멀티미디어란 광범위한 정보망을 토대로 이루어지는 새로운 정보전달매체로 둘 이상의 미디어를 통합시켜 문자, 도형, 음성, 정지화상, 동화상 및 각종 컴퓨터 프로그램 수치 데이터 등의 통합정보가 하나의 시스템을 통하여 동시에 전달되고 통제될 수 있는 매체를 말한다. 다시말하면 디지털화된 정보가 멀티미디어 정보로 통합되고 사용자가 대화식으로 데이터를 조작하거나 편집할 수 있는 기능의 시스템을 의미한다. 이러한 멀티미디어의 데이터 유형은 크게 사운드, 이미지, 동화상으로 나누기도 한다. 멀티미디어의 특성은 문자, 음성, 영상 등 다양한 정보가 동시에 표현되는 동시성과 다양성, 그리고 사용자와 기기가 상호 작용할 수 있어 사용자가 원하는 시간에 원하는 형식으로 보고 듣고 느낄 수 있는 상호작용성을 들 수 있다.

멀티미디어 데이터의 디지털화시 고려사항 및 연구과제는 데이터 유형의 특성 파악 및 데이터의 구조화를 위한 인코딩의 국제 표준 사항을 고려하여야 하고, 특히 데이터의 압축기법에 대한 개발이 핵심기술로서 연구되어야하며, 또한 멀티미디어 데이터의 색인기법 및 검색기법에 대한 연구가 지속적으로 수행되어야 한다. 멀티미디어 데이터의 구조화를 위한 인코딩 국제표준으로는 Hy-Time(Hypermedia/Time-based Structuring Language)이 있다. HyTime의 기능들을 살펴보면 SGML 문헌이나 멀티미디어 문헌에서 어떤 지점, 또는 문헌 일부분을 서로 연결시키는 하이퍼텍스트 링크를 생성하고, 간단한 편집이나 구조적 변경을 추가함으로써 새로운 하이퍼텍스트 문헌들을 생성하고 열람할 수 있고, 상호참조 등으로 사용자들에게 유용한 글을 읽는 길을 전달하고

정보의 필터링, 전문탐색을 가능하게 하여 전자적인 형태로 문헌을 열람할 수 있게 하고, 성경이나 고문서, 법률서, 의학 서적과 같이 다양한 글을 읽는 방법론이 존재하고 세부적인 주석들을 연구하고 논의할 필요가 있는 유형의 문헌들을 지원하고, 슬라이드나 오디오, 비디오 등 다양한 유형의 미디어를 조합할 수 있게 하며 그 순서를 정하는데 융통성을 발휘하고, 데이터 접근시 다양한 관점을 제공하며 문헌의 갱신이나 수정시 버전을 관리할 수 있게 한다는 기능이 있다. HyTime이 SGML을 공식적으로 채택한 이유는 문헌의 구조와 인스턴스를 모순없이 기술할 수 있고, 오브젝트를 분할하여 기억할 수 있으며, 요소에 속성을 정의할 수 있어 채택하였다. 그러나 HyTime은 동화상 데이터나 음성 데이터와 같은 콘텐츠(Content)의 표현에 관계없이 시변 콘텐츠 등의 시작점과 계속 시간을 다른 정보와 동기시 키기 위한 수단을 정의할 뿐이라는 것이고, 또한 문헌 전체의 체계나 유형의 정의도 규정하고 있지 않다는 점을 염두해야 한다는 지적이 있다. 한편 Carnegie Mellon University(<http://www.informedia.cs.cmu.edu/>)에서는 멀티미디어와 관련하여 멀티미디어 디지털 도서관의 설립과 이용, 데이터의 인코딩, 자연언어처리, 자동색인, 지식기반검색등의 문제를 중점적으로 연구중이다.

디지털도서관은 텍스트의 접근뿐 만 아니라 이미지, 소리, 비디오와 같은 다른 형태의 접근도 제공하는 멀티미디어 도서관으로 이해된다. 과거 이러한 자료들은 유용하지만 검색의 어려움이 있어왔다. 그러나 네트워크상에서 비텍스트 자료들에 대한 전거버전을 만듦으로써 물리적이고 논리적인 접근을 통한 검색 서비스를 향상시킬 수 있게 되었고, 멀티미디어 데이터베이스의 설계와 기술에 관련된 문제들을 광범위하게 논의되고 심도있는 연구가 지속되고 있다.

VII. 결 론

컴퓨터와 정보통신망의 발달은 정보화 사회로의

변화에 커다란 영향을 미치고 있다.

도서관의 패러다임을 변화시키는 데에 결정적인 영향을 미친것도 컴퓨팅기술과 통신기술, 전자출판의 급속한 발전이라고 할 수 있다. 디지털도서관은 이러한 정보기술과 통신기술 등의 관련된 첨단 기술의 활용을 기반으로 하고 있다. 디지털도서관의 과제는 원문정보의 디지털화 및 조직화와 이용자들이 디지털정보에 보다 쉽게 접근하여 이용할 수 있도록 정보기술의 기반을 구축하는 것이다. 현재 추진되고 있는 디지털도서관은 자료의 보존적인 측면을 강조하는 데이터의 디지털화를 중심으로 추진되고 있고, 그 활용적인 측면도 연구가 진행되고 있다. 또한 많은 기관들이 다양한 기술을 집약시켜 디지털도서관을 만들고 있지만 어느정도 형태를 갖춘 디지털도서관이 출현할 때까지는 많은 문제를 해결하여야 하기 때문에 상당한 시간이 걸릴 것으로 예상된다. 그러나 명확한것은 전세계적으로 다가올 21세기의 정보화사회에 대비하는 일환으로 디지털도서관 구축에 지대한 관심이 집중되고 있고, 빠르게 추진되고 있다는 점이다.

급속하게 변화하고 있는 과학기술과 사회환경에 도서관이 적극적으로 대처하는 일고로 미래의 도서관으로 주목받고 있는 디지털도서관에 대한 기술적인 요소와 문헌의 구조화 및 SGML에 대하여 살펴본 결과 기술적인 문제는 어느 한 분야의 전문적인 기술이 요구되는 것이 아니라 총망라적인 정보기술을 요구하고 있어 이와 관련된 여러 분야와의 긴밀한 협력과 부단한 연구가 수행되어야 하고, 선진 각국에서 현재 진행중인 여러 프로젝트들을 면밀히 분석, 검토하여 기술적인 시행착오를 최소화하여야 하며, 우리나라의 실정에 맞는 새로운 모형을 보다 구체적으로 개발하여야 한다. 또한 실행과정의 중요 역할에 다른 전문직이 아닌 사서가 중심이 되어 디지털도서관을 발전시켜 나가야 할 것이다.

참 고 문 헌

- 1) 김두식(1993): 전자출판론 I, 서울, 타래. 42.

- 2) 김성혁(1996): *SGML의 기본과 이해*, 서울, 성안당.
- 3) 유사라(1996): “전자/디지털/가상도서관에서이 네트워크”. *정보화 사회와 도서관 정보 네트 워크*. 서울, 나남출판, 441-518.
- 4) 최석두(1995): “디지털도서관의 영향”. 21세기에 있어서의 국립중앙도서관의 기능과 책임: 국립중앙도서관 개관 50주년 기념 논문집, 서울, 국립중앙도서관, 79-100.
- 5) 최석두, 오영화(1995): “디지털도서관의 구성요소에 관한 연구”. 한국정보관리학회학술대회 논문집, 제 2 회, 67-70.
- 6) 최원태(1995): “전자도서관에 관한 연구”. *도서관*, 50(4), 94-124.
- 7) 한국정보관리학회(1996): *국가 디지털도서관 구축 계획에 관한 연구*, 서울, 국립중앙도서관.
- 8) Carnegie Mellon University(<http://www.informedia.cs.cmu.edu/>).
- 9) Edward A. Fox et al(1995): “Introduction”. *Communications of the ACM*, 38(4), 26-28.
- 10) Gary Marchionini and Hermann Maurer (1995): “The roles of digital libraries in teaching and learning”. *Communications of the ACM*, 38(4), 67-75.
- 11) Gio Wiederhold(1995): “Digital libraries, value, and productivity”. *Communications of the ACM*, 38(4), 85-96.
- 12) Lucier RE(1995): “building a digital library for the health sciences: information space complementing information place”. *BMLA*, 83(3), 346-50.
- 13) Robert M et al(1995): “The organization of the digital library”. *Academic Medicine* 70 (4), 286-291.
- 14) SGML Bibliography at Sunsite(<http://sunsite.berkeley.edu/SGML/>)
- 15) Stephen P. Harter(1996): “What is a digital library? Definitions, Content, and Issues”. *Proceedings of the international conference on digital libraries and information services for the 21st century*, 8-17.
- 16) Terry R. Noreault(1996): “Use of SGML in digital libraries”. *Proceedings of the international conference on digital libraries and information services for the 21st century*, 81-87.