

XML을 활용한 고문헌의 원문디지털화 방안에 대한 연구*

- 고문헌을 위한 DTD개발을 중심으로 -

A Study on Methods of Digitalization of Older Oriental Books by Using XML

리 상 용(Sang-Yong Lee)**

목 차

- | | |
|--------------------------------|-------------------------------|
| 1. 서론 | 3. 1 고문헌을 위한 XML DTD 개발 |
| 2. XML의 개념 | 3. 2 DTD를 통한 실제 고문헌의 원문 입력 |
| 2. 1 XML이란 무엇인가? | 3. 3 고문헌 XML파일의 활용방안 |
| 2. 2 원문입력방식에서 XML의 위치 | 4. 결 론 |
| 2. 3 XML의 필요성 | 5. 부편: 일반고문헌을 위한 XML DTD 프로그램 |
| 3. XML을 활용한 고문헌의 원문디지털화 및 활용방안 | |

초 록

본 논문은 1996년 W3C에서 웹문서의 표준으로 제안된 차세대 웹언어인 XML(eXtensible Markup Language)을 활용하여 고문헌의 원문을 디지털화하고 웹상에서 이용자들이 고문헌의 원문을 손쉽게 검색할 수 있도록 하기 위한 방안에 대해 연구한 것이다. 고문헌을 디지털화하기 위하여 일반고문헌을 위한 XML DTD(Document Type Definition)를 개발하고, 이 DTD를 토대로 실제로 XML 상에서 고문헌의 원문을 입력해 보고, 그 과정에서 발생하는 문제점과 기술적인 한계를 살펴보고, 향후 본 고문헌 XML파일의 국가적인 활용방안에 대한 시스템 모형을 구축하였다.

ABSTRACTS

This article is a study on methods of digitalization for older oriental books using XML(eXtensible Markup Language) proposed as a standard for Web documents by W3C in 1996. In this paper, to digitalize older books, I designed the XML DTD(Document Type Definition) for general older oriental books and input the texts of the older books on the basis of the XML DTD, examined the problems and limitations in light of the technical level these days, and established the model of the system for using XML files of older oriental books at a national level in the future.

키워드: XML, DTD, 고문헌, 고서, 디지털도서관, Older books, Digital library

* 이 논문은 2000년도 한국학술진흥재단의 지원에 의하여 연구되었음. (KRF-2000-003-C00476)

** 건국대학교 충주캠퍼스 인문학부 문헌정보학전공 부교수(sylee@kku.ac.kr)

논문접수일자 2003년 2월 25일

게재확정일자 2003년 3월 19일

1. 서론

최근 들어 전자도서관에 대한 논의가 활발히 진행되고, 각 도서관마다 이용자들에게 더 나은 서비스를 하기 위해서 전자도서관 구축에 박차를 가하고 있다. 각 도서관마다 초기에는 주로 현대문헌의 디지털화로 그 작업을 시작했으나, 최근에는 현대문헌의 디지털화 작업과 더불어 소장 고문헌에 대한 디지털화 작업을 실시하고 있으며, 일부 대학 및 기관에서는 이미 상당 수 고문헌의 원문을 축적하여 이용자에게 서비스하고 있다.

현재 우리 나라에서 고문헌에 대한 전자도서관 구축기관 중 가장 대표적인 곳으로는 국립중앙도서관, 성균관대학교, 서울대 규장각 등등을 들 수 있다. 이들 기관에서는 일부 소장 고문헌을 대상으로 원문서비스를 해 주고 있는데, 그 원문입력방식을 살펴보면, 이미지 스캐너나 디지털카메라를 사용하여 고문헌의 원문을 입력하는 이미지방식을 사용하고 있다. 이 이미지방식은 텍스트방식에 비해 장점이 있는 반면, 많은 단점을 내포하고 있다. 따라서 이 이미지방식의 단점을 극복하기 위해서는 고문헌을 텍스트로 입력하는 방안을 고려해야 할 것이다.

이 텍스트입력방식은 시스템 구축에 많은 비용은 들지만, 이용자가 손쉽게 고문헌의 원문을 검색할 수 있고 그 원문을 활용할 수 있는 잇점이 있다. 텍스트입력방식에는 여러 가지가 있는데, 그 가운데 최근에 가장 각광을 받는 것이 XML(eXtensible Markup Lan-

guage)이다. 이 XML은 1996년 W3C에서 웹 문서의 표준으로 제안된 차세대 웹언어로서, 유니코드를 채택하고 있어서 주로 한자로 구성된 우리 나라 고문헌의 원문을 디지털화할 경우 상당히 좋은 결과를 가져올 것이다. 또한 익스플로러5.0 이상¹⁾의 웹브라우저에서 XML 브라우저 기능을 지원해 주고 있으므로, 웹상에서 이용자가 고문헌의 원문을 손쉽게 검색할 수 있게 될 것이다.

본 논문에서는 차세대 웹언어인 XML을 활용하여 고문헌의 원문을 디지털화 하여, 웹상에서 이용자들이 고문헌의 원문을 손쉽게 검색할 수 있도록 하기 위한 방안²⁾에 대해 연구하고자 한다. 논의순서는 우선, XML을 활용하여 일반 고문헌을 위한 DTD(Document Type Definition)를 개발하여 고문헌의 원문을 디지털화하기 위한 기초작업을 마친 후, 필자가 개발한 DTD를 토대로 XML상에서 고문헌의 원문을 직접 구현해 보고, 아울러 향후 국가적인 차원에서의 활용방안에 대해 살펴보기로 보기로 한다.

2. XML의 개념

2.1 XML이란 무엇인가?

XML(eXtensible Markup Language)은 1996년 W3C(World Wide Web Consortium)에서 제안되고 1998년 2월 XML 1.0이 발표되면서 웹문서의 표준형식으로 확정되었

1) 익스플로러 5.0 이전 버전의 경우 파서를 받으면 브라우저가 가능하다. 하지만 버전이 높아질수록 XML 문서의 브라우저가 용이하게 된다.

다. 이 XML은 기존의 HTML(HyperText Markup Language)과 SGML(Standard Generalize Markup Language)이 안고 있는 문제점들을 보완해서 만들어진 것이다. 즉, XML은 사용자가 임의로 태그를 만들어 사용할 수 없어서 문헌의 논리적 구조를 표현하지 못하는 HTML의 한계를 극복하는 동시에, 문서의 태그를 정의하여 생성할 수는 있지만 사용법이 복잡한 SGML의 단점을 피하는 등 이들 두 언어의 장점만을 수용하여 쉬우면서도 확장성이 있도록 만들어진 차세대 언어인 것이다. XML은 시스템간에 호환성이 뛰어나고, 어떤 플랫폼에서도 읽을 수 있는 포맷을 지원하기 때문에 특정 회사의 환경에 얽매이지 않아, 향후 웹상에서 원문을 구현하는 데 최적의 도구로 부각되고 있으므로, 차세대 문서의 표준으로 자리잡을 것으로 전망되고 있다.

2. 2 원문입력방식에서 XML의 위치

컴퓨터 상에서 문헌의 원문을 구현하는 방식은 바이너리 방식과 텍스트 방식, 이미지 방식, 텍스트방식과 이미지방식으로 각각 활용할 수 있는 PDF방식으로 구분할 수 있다(김용, 남궁황 1998, 남영준 1998, 박형욱 1998, 유사라 1997, 리상용 2000).

1) 바이너리방식은 <한글>(속칭 아래아한글), <MS워드>, <훈민정음>과 같이 기존의 상용 소프트웨어를 사용하여 투입방식(Key in)으로 입력하는 것을 의미한다.

2) 텍스트방식은 순수텍스트방식(Plain Text)과 마크업텍스트방식(Mark up Text)

으로 구분할 수 있다. 순수텍스트방식이란 완전한 아스키형태로 문서의 내용을 저장할 수 있는 아스키텍스트(ASCII Text)를 의미하며, 마크업언어방식(Markup Language)이란 순수텍스트방식의 단점을 보완해서 나온 것으로, 웹상에서 검색이 용이하도록 문헌을 구조화하여 표현한 HTML, SGML, XML 등이 여기에 해당한다.

3) 이미지방식은 이미지스캐너 또는 디지털카메라를 사용하여 원문을 입력하고 이를 JPG, GIF, TIFF, PDF 등의 파일형식으로 광파일이나 마그네틱테이프 또는 컴퓨터 하드디스크에 저장하는 것이다.

4) PDF는 어도비(ADOBE)사에서 포스트스크립트언어에 기반을 두어 개발한 것으로서 이미지방식과 텍스트방식으로 각각 사용이 가능하며, 윈도, 유닉스, 매킨토시 등 어떤 형태의 컴퓨터 시스템 환경 하에서도 활용될 수 있도록 만들어진 포맷으로서 어떠한 어플리케이션에 의해 생성된 문서도 PDF파일로 변환시켜 준다(장수진 1999, Adobe 1999).

본 논고에서 다루고 있는 XML은 위의 입력 방식 중 마크업언어에 해당하는 것으로 원문입력방식 중 최근에 가장 각광을 받고 있으며 향후 웹을 주도해 나갈 것이다. 필자는 원문입력방식 중 XML과 PDF가 향후 웹을 주도할 것으로 예상했는데, 최근 국내는 물론 전 세계적인 추세로 보아 XML이 주도권을 장악하고 그 뒤를 PDF 등 다른 방식들이 부수적으로 활용될 것으로 보인다.

필자는 이 XML이 처음 소개되었을 때부터 그 유용성에 대해서 충분히 이해하였으나 막상 이를 활용하여 고문헌의 원문을 구현하

는 작업은 시도하지 못하고, 그보다 편리하고 손쉽게 고문헌의 원문을 컴퓨터 상에서 구현할 수 있는 PDF방식을 활용하여 고문헌 원문의 디지털화 방안에 대한 연구를 시도해 본 적이 있다(리상용 2000). 하지만 당시에 웹상에서 통용되는 한자 수의 부족으로 인하여 고문헌에 수록된 수많은 한자들을 컴퓨터 상에서 구현하는데 상당히 많은 어려움을 겪었다.

이 XML을 활용하여 고문헌의 원문을 디지털화 할 경우 유니코드의 지원으로 이전까지 가장 어려운 문제로 대두되었던 컴퓨터 상에서의 한자의 구현 문제가, 비록 완전하다고는 할 수 없지만, 예전보다 상당 부분 해결이 가능하게 되었다. 유니코드가 나오기 전까지 웹상에서 통용되었던 한자는 KSC5601의 4888자에 불과했으나, 이 유니코드의 적용으로 유니코드 3.0에 정의된 한자 27,484자를 구현할 수 있게 된 것이다. 그러나 현실적으로 웹상에서 고문헌의 원문을 구현하는데, 이 27,484자의 한자로는 부족함이 많아서, 고문헌의 원문을 완벽하게 서비스하는 데 여전히 어려움이 남아 있다.

향후 한국, 중국, 일본의 각국 정부당국은 더욱 많은 한자를 웹상에서 공유할 수 있도록 정책적인 노력을 기울여야 할 것이다. 한편 국내의 전자책과 전자문서의 표준 역시 이 XML기반으로 채택되었으므로, 고문헌과 고문서 역시 현대 문헌과 발맞추어 XML기반으로 구현하는 것이 바람직할 것이다.

2.3 XML의 필요성

필자는 고문헌의 활용 단계를 제1 단계 고문

헌의 보관 단계(서고보존단계), 제2단계 고문헌정보의 제공단계(원문의 이미지파일 구현 단계), 제3단계 고문헌정보활용의 극대화 단계(원문의 텍스트파일 구현 단계)로 구분한 바 있다(리상용 2000).

현재 국내 대부분의 도서관에서는 제1 단계에 머물러 있거나 제2단계로 진입을 계획하거나 이미 진입한 상황이다. 향후에 가장 중요한 것은 제3단계인 고문헌정보활용의 극대화 단계로 나아가야 한다는 것이다. 분명 현재의 기술수준 하에서 제3단계에서 추구하는 텍스트 방식으로 고문헌의 원문을 완벽하게 구현하는데 다소간의 어려움에 직면하겠지만, 현 상태에서 기술이 허락하는 선까지 거기에 대한 기초작업이 수행되어야 할 것이다.

현재 국내에서 고문헌에 대한 원문서비스를 해 주는 기관에서 채택한 입력방식을 살펴보면, 대부분의 도서관에서 PDF, JPEG, TIFF 등의 이미지입력방식을 채택하고 있으며, 한국 역사정보통합시스템에서 텍스트방식으로 구축한 한국문집총간, 고전국역총서 등 한국사관련 기초 사료 일부를 서비스 해 주고 있다. 현재 고문헌의 원문에 대한 텍스트입력방식은 이제 겨우 걸음마 단계이고 여전히 국내 대부분의 기관에서는 이미지방식을 채택하고 있다.

이 이미지 방식은 텍스트방식에 비해 개발 비용, 시간, 노력 면에서 간편하며, 고문헌의 원문을 있는 그대로 볼 수 있어서 고문헌의 판본 연구, 활자 또는 서체 연구자에게 나름대로 의의가 있다. 하지만 컴퓨터 내에서의 저장용량을 많이 차지하고, 통신상에서 전송이 용이하지 않으며, 색인어를 별도로 부여해야 하며, 본문에 대한 전문검색(全文檢索)이 되지 않으

며, 뷰어 간에 호환성 문제로 인하여 열람에 어려움이 발생하는 등 많은 문제점을 안고 있다.

그러나 이 XML 방식으로 고문헌의 원문을 디지털화 할 경우, 고문헌의 원문을 입력하기 위해 DTD를 개발해야 하며, 고문헌의 텍스트 파일 구축에 엄청난 인력과 시간과 비용을 필요로 하며, 입력과정에서 오류 발생의 가능성 및 원문 대조 및 정정 작업에 많은 비용이 예상되며, 현 기술 수준에서 유니코드의 한계로 인하여 고문헌의 원문을 완벽하게 구현할 수 없다는 단점이 있다. 그러나 일단 구축해 놓으면 기존의 이미지방식에 비해 컴퓨터 내에 저장용량을 적게 차지하고, 전송속도가 빠르며, 이미 구축해 놓은 데이터를 가공하여 활용할 수 있으며, 무엇보다도 고문헌의 원문에 대한 자연언어 검색 및 요소별 검색이 가능하고, 웹 브라우저를 통해서 브라우징이 가능하며, 이용자가 아주 쉽고 신속하게 필요한 자료를 검색하여 활용할 수 있게 될 것이다.

3. XML을 활용한 고문헌의 원문 디지털화 및 활용방안

이 장에서는 고문헌을 위한 XML DTD를 실제로 개발해 보고 이를 활용하여 실제로 고문헌을 입력해 보고, 이들 고문헌 XML 파일의 활용방안에 대해 살펴보고자 한다.

3. 1 고문헌을 위한 XML DTD 개발

DTD란 Document Type Definition(문서 형식정의)의 약자로서, 문서의 형식 즉 태그

에 의해 표시되는 문서요소 및 그 계층구조의 선언, 요소의 부속 정보인 속성을 선언하여 XML 상에서 태그를 사용하는 방법을 규정한 것으로서, 가급적 정확하고 공식적인 방법을 통해 문서의 구조를 기술하기 위한 것이다.(김영숙, 조성호 2001). 고문헌을 위한 DTD를 개발한다는 의미는 고문헌의 모든 데이터를 입력할 수 있는 일종의 틀을 개발하는 것으로 이해하면 될 것이다.

처음에는 고문헌을 위한 범용 DTD 하나만을 개발하려고 시작했지만, 고문헌의 유형별 특성으로 인하여 그 많은 유형의 고문헌을 하나의 DTD에서 모두 수용하는 데 많은 문제점과 한계점이 발생하였다. 따라서 전 주제분야의 고문헌을 모두 포괄하는 하나의 DTD를 개발해서 여기에 모든 고문헌을 적용시키기보다는 고문헌의 형식, 소장 수량, 문헌의 이용도 및 중요성 등을 고려하여 고문헌의 유형에 따라 DTD를 여러 개 구성하여 각각 그 유형에 맞게 적용시키는 편이 바람직할 것이다. 따라서 고문헌을 그 유형에 따라 1) 일반고문헌류, 2) 문집류, 3) 족보류 등으로 구분하고 이들 문헌의 유형에 맞는 DTD가 각각 개발되어야 할 것으로 보인다.²⁾

가장 좋은 방법은 각각의 DTD 설계 시에 그 유형에 맞는 고문헌의 항목을 모두 조사하여 완벽한 고문헌을 위한 DTD를 개발하는 것이 최상의 방법이나, 문헌에 따라 새로운 항목이 나올 수 있으므로, 이들 새롭게 나오는 항목을 모두 만족시킬 수 있도록 DTD의 지속적인 갱신 및 개정이 필요하게 될 것이다.

이 장에서는 앞에서 구분한 3가지 고문헌의 유형 가운데 일반고문헌을 위한 DTD를 개발

해 보고, 이 DTD에 맞추어 특정 고문헌을 선정하여 고문헌의 원문을 입력해 보고, 그 입력 현황을 살펴보고자 한다.

고문헌은 기본적으로 1910년 이전에 나온 것을 대상으로 하되, 가급적 일제시대 때 간행된 동장본(東裝本)까지 포괄하도록 하였다. 이 작업을 위해 영남대학교 고서열람실에 소장된 고문헌 중 족보류와 문집류를 제외한 나머지 고문헌을 모두 일일이 조사하여 항목을 추출하였으며, 이밖에 사가소장 고문헌 및 국내에서 간행된 영인본을 대상으로 닥치는 대로 조사하여, DTD내에 해당 요소들이 모두 수용될 수 있도록 최선의 노력을 기울였다.

3. 1. 1 일반 고문헌을 위한 DTD 설계

여기에서는 고문헌 가운데 일반 고문헌을 위한 DTD를 설계하고자 한다. 일반 고문헌을 위한 DTD는 크게 다음과 같이 구분할 수 있다(표 1 참조).

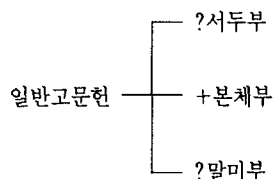
〈표 1〉에서와 같이 일반고문헌을 위한 DTD

에서 고문헌의 구조는 서두부, 본체부, 말미부로 구분한 후에 각각에 알맞는 DTD 항목을 설정하였다. 이 일반고문헌을 위한 DTD는 차후에 개발될 다른 DTD의 모체가 되므로, 가급적 고문헌의 기본적인 요소들을 모두 수용할 수 있도록 설계하였다. 이 DTD는 일반고문헌을 위한 것이므로 차후에 개발될 문집용DTD, 족보용DTD와는 차별성을 부여하였다. 일반고문헌을 위한 DTD를 서두부, 본체부, 말미부 의 3부로 구분하여 각각의 항목들을 살펴보면 다음과 같다.

(1) 서두부(序頭部)

서두부는 기본적으로 고문헌의 서명부터 시작해서 본문이 나오기 전까지, 즉 권수제면(卷首題面)이 나오기 직전까지를 서두부로 하되, 단 본문에 대한 편권목록(篇卷目錄)이 나오는 경우에는 편권목록 직전까지 설정하였다. 서두부에는 일반 고문헌에서 나올 가능성이 있는 항목들을 모두 설정해 주도록 노력하

〈표 1〉 전체 구성도



2) 현재 고문헌을 위한 XML DTD에 관한 연구로는 한국전산원(2002)에서 나온 〈지식정보관리를 위한 XML DTD 표준〉에서 각 문헌의 유형에 따라 학술논문지DTD, 석박사논문DTD, 연구보고서DTD, 편년류 DTD, 사료총서 DTD, 문집총간DTD 등 6종의 DTD를 개발한 것이다. 이들 중 고문헌과 관련된 것은 후자의 3개 DTD이다. 이 중 편년류DTD는 일성록, 승정원일기, 조선왕조실록, 내각일력 등을 입력하기 위해, 사료총서DTD는 국사편찬위원회의 사료총서를 위해, 문집총간DTD는 한국문집총간을 위해 개발하였다. 하지만 이들 DTD는 고문헌 전체를 염두에 두고 DTD를 개발했다기 보다는 특정 기관의 대표적인 문헌을 입력하기 위해 DTD를 개발한 것이다. 현재 일반고문헌을 위한 DTD는 전혀 개발되어 있지 않은 상황이다.

였다. 여기에는 서명군(書名群), 어서군(御書群), 편찬제신군(編纂諸臣群), 범례군(凡例群), 하서군(賀書群), 서문군(序文群), 제사군(題辭群), 발문군(跋文群), 지문군(識文群), 총기군(總記群), 초상화군(肖像畵群), 삽도군(插圖群), 권수(卷首), 연보군(年譜群), 세보군(世譜群), 묵적군(墨蹟群), 목록군(目錄群), 후기군(後記群), 전기군(傳記群), 인명록(人名錄), 기타(其他) 등의 항목을 설정하였다. 서두부에는 필자가 이번에 찾지 못하였으나, 향후 새롭게 나올 항목 및 그 내용이 특이해서 특정 항목군에 속하기 어려운 것들을 위해서 마지막에 기타 항목을 설정하여 DTD의 확장 가능성을 열어 주었다.

서명군(書名群)은 다시 권수제(卷首題), 제첨제(題籤題), 표제(表題), 이제(裏題, 內題), 판심제(版心題), 서근제(書根題) 등으로 구분하여 이용자가 특정 서명만 알 수 있어도 검색할 수 있도록 하였다. 서명에서 특이한 사항은 본문의 권수제면에 있는 권수제(卷首題)를 별도로 서명의 맨 앞에 설정했는데, 그 이유는 권수제가 고문헌의 목록기술상 대표서명으로 채기하는 것이기 때문이다. 어서군(御書群)은 어제어필(御題御筆), 어제서(御製序), 어제소지(御製小識), 운음(綸音), 교서(敎書), 어찰(御札), 조지(詔旨), 기타어서(其他御書) 등으로 구분하였으며, 편찬제신군(編纂諸臣群)은 편집제신직명(編輯諸臣職名), 총재대신(總裁大臣), 교열(校閱), 교정(校正), 서사(書寫), 기타제신(其他諸臣) 등으로 구분하였다. 범례군(凡例群)은 범례(凡例), 교정범례(校正凡例), 개간범례(改刊凡例), 총례(總例), 기타범례(其他凡例) 등으로

구분하였으며, 하서군(賀書群)은 하문(賀文), 축시(祝詩), 축화(祝畵), 축수(祝壽) 등으로 구분하였다. 서문군(序文群)은 서(敍), 원서(原敍), 자서(自序), 자서(自敍), 후서(後序), 표(表), 전(箋), 중편서(重編序), 중간서(重刊序), 중각서(重刻敍), 중수서(重修序), 신간서(新刊序), 신서(新序), 구서(舊序), 총서(總敍), 소서(小敍), 집해서(集解序), 보유서(補遺序), 속지서(續誌序), 합지서(合誌序), 인(引), 서설(序說), 서설(敍說), 서, 자서, 기타서문(其他序文) 등으로 구분하였으며, 제사군(題辭群)은 서제(書題), 제사(題辭) 등으로 구분하였으며, 발문군(跋文群)은 발(跋), 원발(原跋), 속지발(續誌跋), 보유발(補遺跋), 합지발(合誌跋), 기타발(其他跋) 등으로 구분하였다. 지문군(識文群)은 지(識), 소지(小識), 구지(舊識), 중간지(重刊識), 기타지(其他識) 등으로 구분하였으며, 총기군(總記群)은 총론(總論), 총해(總解), 편집대의(編輯大意), 논례(論例), 의례(義例), 총례론(總例論), 집례(集例), 총평(總評), 통론(通論), 기사(紀事), 기타(其他) 등으로 구분하였다. 초상화군(肖像畵群)은 초상(肖像), 유상(遺像), 성상(聖像), 성적도(聖蹟圖), 화상찬(畵像讚), 화상찬(畵像贊), 화상지(畵像識), 화상발(畵像跋) 등으로 구분하였으며, 삽도군(插圖群)은 도표(圖表), 지도(地圖), 사진(寫眞), 변상도(變相圖), 기타삽도(其他插圖) 등으로 구분하였다. 삽도에 사진이 들어간 이유는 석인본(石印本)의 경우 간혹 사진이 들어간 경우가 있어서 그렇게 하였다. 연보군(年譜群)은 연보(年譜), 연표(年表), 연보지(年譜識),

연보발(年譜跋), 연고고리(年譜攷異), 연보보유(年譜補遺), 연보부록(年譜附錄) 등으로 구분하였으며, 세보군(世譜群)은 세계요략(世系要略), 세계(世系), 세계도(世系圖), 세보도(世譜圖), 원류도(源流圖), 도통도(道統圖), 세계후록(世係後錄) 등으로 구분하였으며, 묵적군(墨蹟群)은 유묵(遺墨), 필적(筆跡), 유묵도(遺墨圖), 유묵발(遺墨跋) 등으로 구분하였다. 목록군(目錄群)은 목록(目錄), 총목(總目), 총록(總錄), 편목(篇目), 강목(綱目), 부목(部目), 목차(目次), 목례(目例) 등으로 구분하였다. 이들 목차와 목례는 일제시대 때 간행된 서적을 위해 설정한 것이다. 후기군(後記群)은 후기(後記), 후구(後口), 제후(題後), 후론(後論), 서후(書後), 기타후기류(其他後記類) 등으로 구분하였으며, 전기군(傳記群)은 전략(傳略), 본전(本傳), 연기(緣起), 비명(碑銘), 기타전기류(其他傳記類) 등으로 구분하였다. 이들 항목 중 편찬제신군, 초상화군, 삽도군, 연보군, 세보군, 목적군, 후기군 등은 고문헌의 서두부에 나오는 경우도 있고 말미부에 나오는 경우도 있어서 서두부와 말미부에 모두 그 항목을 설정하였다. 이상에서 설명한 서두부의 내용을 구조화해 보면 다음과 같다(표 2 참조).

(2) 본체부(本體部)

본체부의 범위는 기본적으로 권수제면(卷首題面)부터 권말제(卷末題)가 나올 때까지로 하되, 본문의 편권(篇卷)에 대한 목록이 나오는 경우에는 그 편권목록(篇卷目錄)까지 본체부에 포함시켰다. 이 편권목록이 권수제면 앞에 나오는 것이 통례라서 처음에는 이를

서두부의 목록군에 포함시키려고 했다. 하지만 권수제가 나온 다음에 편권목록이 나오는 경우가 간혹 나오는데, 그러한 경우 서두부에 편권목록이 있으면 맨 앞 편권(篇卷)에 대한 편권목록은 서두부로 가고, 본문내에 나오는 부차적인 편권목록은 본체부로 가는 모순이 발생하기 때문에 이 편권목록을 본체부에 수록하고, 실제 DTD상에서 유연하게 활용할 수 있도록 설계 하였다.

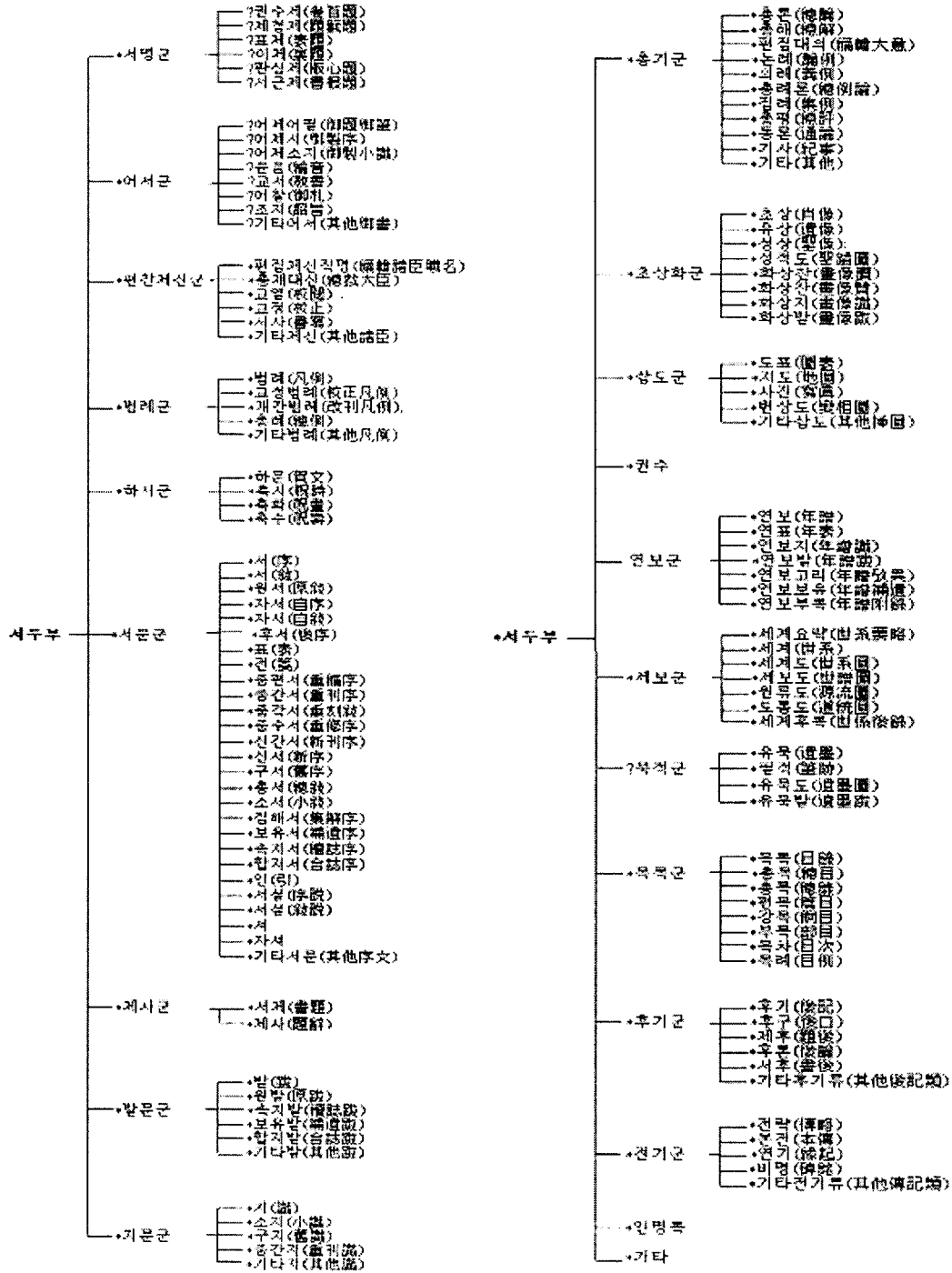
본체부는 본문내용을 설정하여, 이를 편권목록(篇卷目錄), 권차(卷次), 제목(題目), 내용(內容), 도표(圖表), 삽도(插圖), 주석군(註釋群), 본문내부록(本文內附錄), 기타(其他), 권말제(卷末題) 등으로 구분하였다.

본체부의 본문내용에서 특징적인 사항은 다음과 같다. 본문의 내용은 특정 제목에 대해 그의 내용이 나오고, 그 내용에 대해 다시 소제목과 그에 대한 내용이 반복적으로 나올 수 있도록 설계하였다. 주석군(註釋群)은 본문에 나오는 주소자쌍행(註小字雙行)과 두주(頭註 또는 眉批)와 각주(脚註) 등으로 구분하였는데, 실제로 주석군의 경우 본문의 여러 항목에서 출현이 가능하므로, 실제 DTD 프로그램에서는 본문 내용의 각 항목에 적용이 가능하도록 설계하였다. 본문내부록은 권말제가 나오기 전까지 본문 내에 나오는 부록을 의미한다. 이상에서 설명한 본체부의 내용을 구조화하면 다음과 같다(표 3 참조).

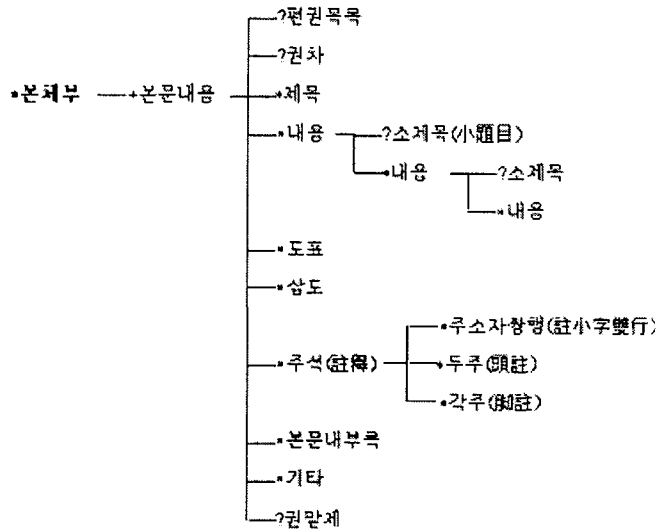
(3) 말미부(末尾部)

말미부의 범위는 권말제(卷末題)가 끝난 다음 페이지부터 책의 마지막까지이다. 말미부는 세보군(世譜群), 후서군(後序群), 말미

〈표 2〉 서두부 구조도



〈표 3〉 본체부 구조도

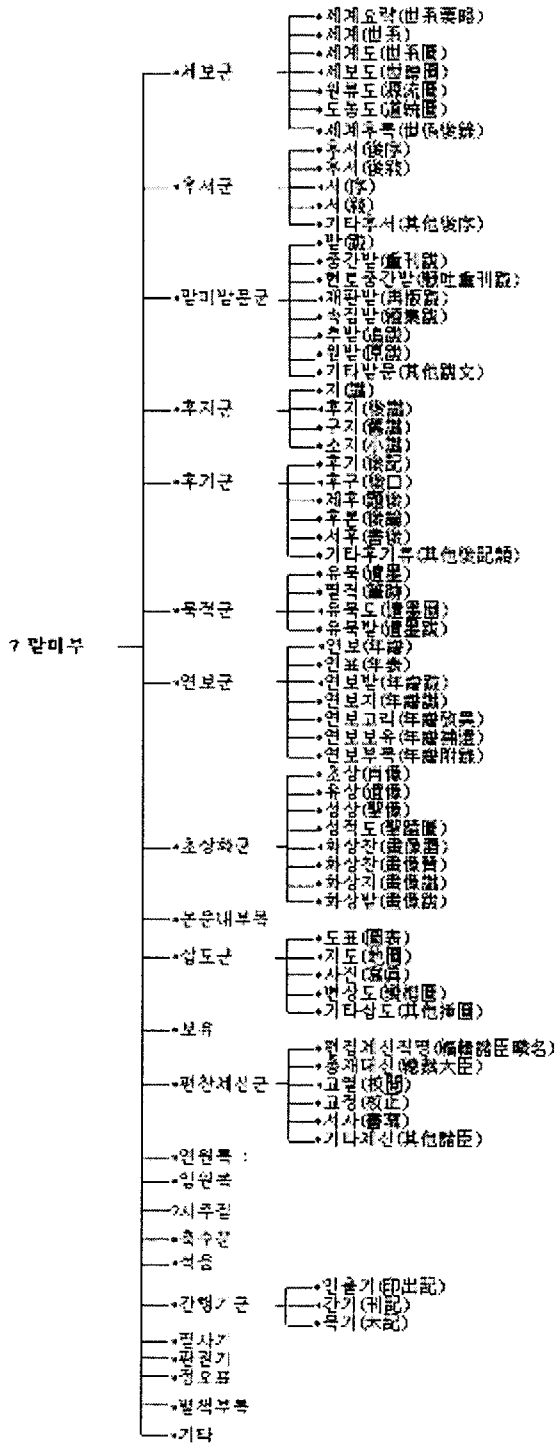


발문군(末尾跋文群), 후지군(後識群), 후기군(後記群), 목적군(墨蹟群), 연보군(年譜群), 초상화군(肖像畵群), 본문내부록(本文內附錄), 삽도군(插圖群), 보유(補遺), 편찬제신군(編纂諸臣群), 연원록(淵源錄), 임원록(任員錄), 시주질(施主秩), 축수문(祝壽文), 석음(釋音), 간행기군(刊行記群), 필사기(筆寫記), 판권기(版卷記), 정오표(正誤表), 별책부록(別冊附錄), 기타(其他) 등으로 구분하여 가급적 고문헌의 말미에 나오는 모든 항목을 수용하도록 설계하였다.

원래 후서군은 말미발문군에 포함시키려 했으나, 후서(後序 또는 後敍)란 항목으로 발문 앞에 나오는 사례가 많아서 별도로 설정하였다. 후서군(後序群)은 후서(後序), 후서(後敍), 서(序), 서(敍), 기타후서(其他後序) 등으로 구분하였으며, 말미발문군(末尾跋文群)은 발(跋), 중간발(重刊跋), 현토중간발(懸吐重刊跋), 재판발(再版跋), 속집발(續

集跋), 추발(追跋), 원발(原跋), 기타발문(其他跋文) 등으로 구분하였다. 후지군(後識群)은 지(識), 후지(後識), 구지(舊識), 소지(小識) 등으로 구분하였다. 지(識)의 경우 서두부에도 나오고 있으나, 실제 문헌의 조사 결과 말미부에 나온 것과 그 항목 면에서 차이가 있어서 서두부와 다르게 명칭하였다. 간행기군(刊行記群)은 인출기(印出記), 간기(刊記), 목기(木記) 등으로 구분하였다. 말미부의 부록은 본문의 권말제가 나온 뒤에 나오는 부록을 의미하며, 별책으로 간행된 별책부록과는 구별된다. 시주질(施主秩)은 주로 불교서적의 말미부에 나오는 시주자 명단을 말하며, 석음(釋音)은 불교서적의 말미부에 통상적으로 잘 쓰이지 않는 글자의 발음 또는 의미를 해석해 놓은 것이다. 판권기(版卷記)는 일제시대에 간행된 서적을 위해 설정한 것이다. 이상에서 설명한 말미부의 내용을 구조화하면 다음과 같다(표 4 참조).

<표 4> 말미부 구조도



3. 2 DTD를 통한 실제 고문헌의 원문 입력

여기에서는 앞 절에서 설계한 일반 고문헌을 위한 DTD구조도에 따라 필자가 직접 작성한 일반고문헌을 위한 XML DTD 프로그램(5.부편 참조)에 의거하여 XML상에서 고문헌의 원문을 구현해 보고자 한다. 여기에서는 1) 순한문 문헌, 2) 옛한글 문헌, 3) 구절자 등으로 구분하여 <한글>로 작성한 텍스트 파일을 XML 상에서 구현시켜 보고 그 결과에 대해 살펴보고자 한다. 순한문 문헌으로는 <격몽요결>과 <중용>을 택하였고, 옛한글 문헌으로는 <홍길동전>을 택하였으며, 구절자는 <한글>에 나타난 구절자를 대상으로 하여 그 원문을 입력시켜 보았다. 그 결과는 각각 다음과 같다.

3. 2. 1 순한문 문헌

여기에서는 울곡선생이 저술한 <격몽요결(擊蒙要訣)>과 사서삼경의 하나인 <중용(中庸)>을 앞에서 설계한 DTD 구조에 맞추어서 구현해 보았는데 그 결과는 다음과 같다.

<격몽요결>과 <중용>의 XML 파일화 작업은 TagFree2000 DTD editor, TagFree 2000 XML editor, TagFree2000 XML Builder를 가지고 윈도우2000 및 윈도우XP 메모

장 및 노트패드에서 진행하였다. 작업 순서는 앞 절에서 설계한 일반고문헌용 DTD에 맞추어서 진행하였다. <한글>파일로 구축한 <격몽요결>과 <중용>원문을 XML 에디터에서 불러 들여 해당 엘리먼트에 해당 문장을 끌어다 놓는 식으로 작업하였다.

먼저 <격몽요결>의 파일을 윈도우2000 메모장에서 불러서 작업해 본 결과 <격몽요결>에 수록된 7265자 가운데 犧(犧), 신(齎), 효(曉) 3자를 만을 제외하고 모든 한자가 XML 상에서 구현되었다(표 5 참조).

2000년 필자의 논문 "PDF를 활용한 고문헌의 원문디지털화 방안에 대한 고찰"에서 당시에 <한글>로 작성한 <격몽요결>의 원문 파일을 가지고 PDF 텍스트파일을 만든 적이 있었는데, 그 당시 <격몽요결>의 본문글자 7265자 가운데 28종 32자가 변환이 되지 않은 것에 비하면, 이번 XML 상에서의 한자의 변환율은 대단히 높다고 할 수 있다. 유니코드의 지원으로 인하여 예전보다 훨씬 많은 한자를 웹상에서 구현할 수 있게 된 것이다.

<격몽요결>에서 기대 이상의 좋은 결과를 얻은 데 힘입어, <한글>로 작성한 <중용>파일을 XML 상에서 구현해 보았다. <중용>의 본문 분량이 많고 거기에 수록된 한자 역시 <격몽요결>에서 보다 어려운 한자가 많이 있기

<표 5> 격몽요결

人有發謬我者則必反而自省若我實有可發之行則自責內訟不憚改過若我過甚微而增衍附益則彼言雖過而我實有受謬之苗願亦當刻齒前愆不啻寤寐若我本無過而捏造虛言則此不過妄人而已與妄人何足計較虛言既且彼之虛謬如風之號舞雲之過空於我何與哉夫如是則發謬之來有則改之無則加勉莫非有益於我也若爾過自責?? 本不虛必欲置身於無過之地則其過愈甚而取謬益重矣昔者或問止謬之道文中子曰莫如自修謬益日無辨此言可為學者之法凡稱先生長者當實問吾理難曉處以明其學稱鄉長老當小心恐謬不放言謬有則則改對以實與朋友處當以道義講磨只談文字義理而已世俗鄙俚之說及時政得失守令賢否他人過惡一切不可掛口與鄰人虛聲隨問隨答而終不可發謬妄之言獲莊樂自持而切不可存矜高之色惟當以吾言謬接必欲引而向學與幼者處當謬謬言孝悌忠信使發善心若此不已則邪俗漸可罷也

때문에 그렇게 시도해 본 것이다. 이 <중용> 파일을 XML 상에서 구현해 본 결과 <중용>에 수록된 한자 중 32자가 구현되지 않고 그 글자가 있어야 할 자리에 물음표(?)로 대신 나타났으며(표 6 참조) 브라우징을 해 보아도 역시 같은 결과를 보였다.

<중용> 본문 가운데 XML 상에서 구현되지 않고 물음표로 나타난 32자의 글자들은 다음과 같다.

- 겁(跽), 경(嬰), 고(罟), 구(疚), 달(怛), 벽(辟), 병(並), 봉(丰), 부(鈇), 설(擊), 신(矧), 심(悃), 임(揜), 예(睨), 요(徼), 원(藺), 유(輜), 인(訥), 임(衽), 전(翦), 정(阱), 조(槩), 좌(莖), 지(躋), 체(棧), 체(禘), 치(鷗), 타(鼉), 편(徧), 활(濶), 획(獲), 회(餽)

필자는 위의 글자들이 유니코드 지원 한자 27,484자 외의 글자라서 XML 상에서 구현되

지 않는 것으로 생각했다. 이 구현되지 않는 한자들에 대한 처리를 어떻게 해야하는가 하는 문제에 봉착했는데, 윈도XP에서 돌려본 결과 의외의 좋은 성과를 얻게 되었다.

<격몽요결>과 <중용>에서 구현되지 못한 총 35자를 대상으로 하여 윈도2000보다 업데이트 된 윈도XP의 메모장에서 시험해 보았는데 그 결과는 다음과 같이 나왔다.

윈도XP의 메모장에서 이들 35자를 <훈글> 파일에서 복사해서 붙여 넣기를 해서 저장해 보았더니, 윈도XP 메모장에서 이들 35자 가운데 28자가 깨지지 않고 한자로 구현되고, 겁(跽), 유(輜), 임(衽), 정(阱), 조(槩), 체(禘), 획(獲) 등 7자만이 한자로 구현되지 않은 채 검정색 네모박스가 그 글자 대신 나타났다. 이들 7자가 결국 구현되지 않을 것으로 예상하고, 이 상태에서 인터넷익스플로러 6.0으로 브라우징 해 본 결과, 예기치 않게, 문제가 발생했던 7자를 포함한 35자의 글자 모두가 깨지지 않고 온전하게 브라우징 되었다.

<표 6> 중용

中庸何爲而作 子思子 聖道學之失其傳而作也 蓋自上古 聖神繼天立極而道統之傳 有自天矣 其見於經則尤駭厥中 聖堯之所以授舜也 人心惟危 道心惟微 惟精惟一 允執厥中者 舜之所以授禹也 堯之一言 至矣盡矣 而舜僅爲之以三言者 則所以明夫舜之一言 必如是而後 可原也 蓋聖堯之心 之虛靈知覺 一而已矣 而以爲有人心道心之異者 則以其或在於形氣之私 感源於性命之正 而所以爲知覺者不同 是以感而動而不安 欲應而難見耳 然人豈不有靈形 故雖上智不能無人心 亦莫不有是性 故雖下愚不能無道心 二者雖於方寸之間 而不知所以治之 則危者愈危 險者愈險 而天理之公 罕能以勝夫人欲之私矣 情則察夫二者之間而不雜也 一則守其本心之正而不虧也 從事於斯 無少闕闕 必使通心 常爲一貫之至 而人心 善惡寧重 則危者安 險者寧 而動靜云爲 自無不及之差矣 夫堯舜禹天下之大聖也 而天下相傳 天下之大聖也 以天下之大聖 行天下之大事 而其授受之際 丁寧告祝 不過如此 則天下之理 豈有以加於此哉 自是以來 聖聖相承 若成湯文武之爲君 堯舜伊傅周召之爲臣 雖以此 而推夫道統之傳 若吾夫子 則雖不得其位 而所以繼往聖 開來學 其功 反有賢於堯舜者 然當是時 見而知之者 雖魏氏曾氏之傳 傳其宗 及其曾氏之再傳 而復得夫子之弟子思 則去聖遠而真蹟益矣 子思歷夫悠久而愈失其真也 於是 雖本堯舜以來相傳之意 實以平日所聞父師之言 更互演繹 作爲此書 以昭後之學者 蓋其憂之也深 故其言之也切 其慮之也遠 故其說之也詳 其曰 天命率性 則道心之謂也 其曰 擇善固執 則精一之謂也 其曰 君子時中 則執中之謂也 皆之相法于有餘年 而其言之不異 知言符節 感運前聖之書 所以續道統 開示端矣 未有若其明且盡者也 自是而又再傳 以得孟氏 爲能推明是書 以承先聖之統 及其沒而遂失其傳 則吾道之所寄 亦越乎言說文字之限 而其傳之說 日新月盛 以至於老佛之徒出 則雖近理而大亂矣 然而而幸此書之不泯 故程夫兄弟者出 得有所考 以續夫千載不傳之緒 得有所據 以斥夫二家以盡之非 蓋子思之功 於聖爲大 而堯程夫子 則亦其能因其緒而得其心也 傳乎 其所以爲說者 不傳 而凡石氏之所論 僅出於其門人之所記 是以大善盡明 而微言未折 至其門人所自爲說 則雖詳盡而多所發明 然皆其師說而遷於老佛者亦有之矣 著白深誠 雖存遺而竊疑之 流傳反復 蓋亦有年 一旦慨然 以有以得其善 然後 乃敢會衆而讀其書 既爲字畫盡句 一篇以俟後之君子 而一二同志 復取石氏書 對其疑難 名以歸略 且記所嘗論釋者之意 則爲或問 以附其後 然後此書之旨 支分節解 綴給貫通 詳略相因 巨細畢舉 而凡諸說之同異得失 亦得以曲暢委通 而各復其真 蓋於道統之傳 不敢妄議 然初學之士 或有取焉 則亦原于升高行遠之 助云爾

淳熙己酉春三月戊申 新安朱熹書

윈도2000에서 작업을 했을 때 구현되지 않았던 한자들이 윈도XP에서는 모두 구현된 것이다. 처음에는 윈도2000보다 윈도XP에서 지원하는 한자의 수가 증가하여 그 같은 결과가 나왔다고 생각했는데 꼭 그렇지만은 않았다. 나중에 알게된 사실이지만, 윈도2000에 한자 27,484자를 지원할 수 있는 TTF(True Type Font)를 설치하지 않고 작업하여 좋지 않은 결과가 나왔던 것이다. 따라서 윈도2000에서 오피스2000의 확장판에 있는 NBATANG체를 설치할 경우 윈도XP에서 작업한 것과 같은 결과를 가져올 수 있을 것이다.

비록 완벽하다고는 할 수 없지만 XML 상에서 <격몽요결>과 <중용>에 나타난 원문 전체를 구현할 수 있게 되었다. 종전에 KSC5601의 지원 가능 한자 4888자로는 순한문 고문헌의 원문을 구현하기가 사실상 힘들었는데, 이제는 고문헌의 원문 디지털화 작업을 하는데 상당히 좋은 환경으로 변화된 셈이다.

이들 XML상에 입력된 한자는 모두 자연언어 검색이 가능하게 되어서 이 파일에서 원하는

글자를 쳐 넣으면 아주 손쉽게 검색이 가능하게 되므로, 이용자들이 고문헌의 원문을 검색하는데 상당히 편리함을 줄 수 있게 되었다.

3. 2. 2 옛한글 문헌

여기에서는 <한글>로 작성한 <홍길동전>³⁾ 파일을 가지고 XML상에서 구현해 보았는데 그 과정 및 결과는 다음과 같다.

윈도XP의 디폴트 상태에서 <한글>로 작성한 <홍길동전>파일을 XML 상에서 구현시켜본 결과, <홍길동전> 본문에 수록된 옛한글이 제대로 구현되지 못하고 그 글자가 공란이 되거나, 유사한 다른 현대어로 표현되었다(표 7 참조).

이 XML 파일을 MS사에서 만든 익스플로러로 브라우징할 수 있으므로, <한글>파일보다는 <MS워드>파일이 XML파일을 구현하기 쉽지 않을까 하는 생각으로, <한글>로 입력한 <홍길동전>파일을 <MS워드(2002)>로 변환하여 다시 XML상에서 구현해 보기로 하였다. <MS워드>상에서 <한글>파일을 불러보았다

<표 7> 홍길동전

• 명문거족으로 소년 등과 여 비 이 니췌판서의 니르 , 물망이 도
 아의 웃들이오, 흥효 경비 기로 일홈이 일국의 진동 더라. 일족 두 아
 들을 두어시니, 일 일홈이 인형이니 영실 누시 쇼 미오, 일 일
 홈이 길동이니 시비 춘섬의 쇼 이라. 선시의 공미 길동을 나췌 의 일
 물을 어드니, 문득 뇌정벽벽이 진동 며 청풍이 슈염을 거 리고 공의게
 향 며 다라들거 , 놀나 다르니 일장춘풍이라. 심중의 희 며
 각 되, ' 미제 등풍을 어더시니 반 시 귀 식을 나흐리라' 고 즉
 시 말으로 드러가니, 부인 누시 니러 맞거 , 공이 으면이 그 록슈
 닛그려 절이 친알고져 거 , 부인이 경 왈, '상공이 제위 존장 시거
 , 년쇼 경박 의 비루 물 코져 시니 형은 몸 치 아니 리로쇼
 이다.' 고, 언파의 손을 치고 나가거 , 공이 가장 무류 며 분기
 자 못 고 외당의 나와 부인의 지식어 업스물 한탄 더니, 맛 시비
 춘섬이 을니거 그 고요 을인 며 춘섬을 밋글고 형실의 드러가
 췌이 친알 니, 미 춘섬의 나히 심꺼이라. 번 몸을 허 후로 문외
 의 나지 아니 고 타인을 이 업스니, 공이 과득이 녀며 인 며

3) 이윤석. 1997. 홍길동전 연구, 대구 : 계명대학교. 부록 II 에 수록된 <홍길동전> <한글>파일을 이용함.

나, <한글>에서 입력이 되었던 옛한글이 <MS워드>에서 그대로 살아나지 않고, 이들 옛글자의 대부분이 깨지거나, 글자가 빠지거나, 다른 유사한 현대어로 바뀌어서, <MS워드> 파일과 <한글>파일을 함께 열어 놓고, <한글>로 입력한 <홍길동전>의 원문을 복사해서 <MS워드>상에 붙인 후에 저장해 보았더니, 앞서 파일 불러오기를 했을 때 문제가 발생하였던 옛한글이 그대로 살아서 복구되었다. 따라서 이 후자의 <MS워드> <홍길동전>파일을 가지고 XML태그를 붙여서 다시 구현해 보았는데, <한글>로 작성한 <홍길동전>파일에서와 마찬가지로 <MS워드>문서로 만든 <홍길동전>역시 XML 상에서 구현되지 않았다. 옛한글이 유니코드에 표준화되지 않아서 시스템 폰트 상에서 브라우징 할 경우 옛한글이 깨지고 화면상에 구현되지 못한 것이다.

그러나 XML 상에서 옛한글의 구현이 반드시 불가능한 것은 아니었다. 비록 유니코드에서 옛한글을 지원해 주지는 않았지만, 별도의 방법을 사용하여 옛한글을 XML 상에서 구현할 수 있었다.

윈도XP에서 이들 옛한글을 구현하려면, 먼저 새바탕체, 한컴바탕체, 한양신명 등의 폰트를 미리 설치한 후에, 이 상태에서 옛한글로 작성한 <한글>문서를 유니코드로 저장하여, 그 문서를 웹브라우저 옵션에서 이미 설치해 놓은 새바탕체 또는 한컴바탕체로 글꼴을 지정한 후에 브라우징 하면, 앞서 깨졌던 옛한글의 원문을 온전하게 볼 수 있게 된다. 이처럼 화면상에서 옛한글문서가 브라우징된 것은 새바탕체가 사용자영역에 임시로 배정해 놓은 옛한글의 코드값을 지원해 주어서 가능하게 된 것이다. 이 방식대로 하여 <한글>로 작성한 <홍길동전>을 XML 상에서 구현해 보았는데, 다음과 같이 <홍길동전>의 원문을 그대로 볼 수 있게 되었다(표 8 참조).

현재 유니코드에서 현대한글은 음절형 11,172자와 초·중·종성 자모 238자(초성, 중성의 채움 코드를 포함하면 240자)만이 사용되어 지원이 가능하지만(김홍규 2002), 옛한글은 유니코드에서 표준화되지 못했기 때문에, 임시적인 방법으로, 한글과 컴퓨터사와 마이크로소프트사에서 현재까지 고문헌에서 발견된

<표 8> 홍길동전

일죽논 일흠이 인형이니 명실 뉴시 쇼침이오, 일죽논 일흠이 길동이니 시비 춘섭의 쇼침이라. 선시의 공이 길동을 나흘 세의 일똥을 어드니, 문득 뇌정벽벽이 진동하며 청동이 슈업을 거스리고 공의게 향하여 다라들거늘, 놀나 췌다르니 일장춘풍이라. 심즁의 뒤회하여 침각하되, '너 이제 똥똥을 어드시니 반드시 귀한 조식을 나흐리라'하고 즉시 니당으로 드러가니, 부인 뉴시 니리 맞거늘, 공이 혼연이 그 옥슈를 닛그라 정이 친압고져 하거늘, 부인이 경척왈, "상공이 채워 존중하시거늘, 년쇼 경박주의 비무하물 형교져 하시니 첩은 농형치 아니하리로쇼이다." 하고, 언파의 손을 쥔치고 나가거늘, 공이 가장 무류하여 분기돌 출지 못하고 외당의 나와 부인의 지식이 업스를 한탄하더니, 맛춤 시비 춘섭이 추를 울니거늘 그 고요하물 인하여 춘섭을 잇금고 협실의 드러가 정이 친압하니, 이 적 춘섭의 나히 심필이라. 홀 번 몸을 허출 후로 문외의 나지 아니하고 타인을 취출 쫓이 업스니, 공이 괴득이 녀어 인하여 임첩을 삼아더니, 과연 그 달부터 락과 잇셔 십 삭 만의 일기 옥동을 칭하니, 괴골이 비범하여 진짓 영웅 호걸의 괴상이라.

옛한글 약 5200여자를 자모순으로 하여 유니코드의 사용자 영역에 배정하여 사용할 수 있게 된 것이다. 이 5200여자로 옛한글 문헌을 완벽하게 구현하기는 어렵겠지만, 예전에 옛한글 처리가 거의 불가능했던 KSC5601에 비하면 환경이 아주 좋아졌다고 할 수 있다.

3. 2. 3 구결자

구결자파일은 <한글>의 문자표에 수록된 구결자 269종을 대상으로 <한글>파일을 작성해 본 후에 이를 XML 상에서 구현해 보았다. 그 과정 및 결과는 다음과 같다.

윈도XP의 디폴트 상태에서 작업해 본 결과, <한글> 구결자 역시 XML 상에서 구현되지 않고 그 해당 구결자들이 물음표(?)로만 나타났다(표 9 참조).

<표 9> 구결자

```

????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????
????????????????????????????????????

```

구결자파일을 <한글>문서가 아닌 <MS워드> 문서로 만들어 보면 혹시 XML 상에서 구현될 수 있지 않을까 하는 기대감에서, <한글>로 작성한 구결자를 <MS워드2002>에서 불러

보았는데, <MS워드>상에서 <한글>의 구결자가 깨지지 않고 화면에 브라우징 되었다. 일말의 기대감을 가지고 XML 상에서 구현해 보았으나, 그 구결자에 해당하는 문자들 역시 물음표로 나타나고, 익스플로러에서 브라우징을 해 보았더니 거기에 이상한 기호만 화면에 뜨고 구결자가 전혀 구현되지 않았다. 구결자 역시 유니코드에서 표준화 되지 않았기 때문에 이같은 결과가 나온 것이다.

따라서 이들 구결자를 XML 상에서 구현하려면, 먼저 새바탕체, 한컴바탕체, 한양신명 등의 폰트를 미리 설치한 후에, 이 상태에서 구결자 <한글>문서를 유니코드로 저장하여, 그 문서를 웹브라우저 옵션에서 이미 설치해 놓은 한컴바탕체⁴⁾로 글꼴을 지정한 후에 브라우징 하면, 이전에 물음표로만 나타났던 구결자 역시 원문 그대로 볼 수 있게 된다(표 10 참조).

<표 10> 구결자

ハ力力加可去去臣命ナロ古具本考：スハ本光四果曾巴ハナ平既
 乃尹男明女及叔トヒ尾色色斤丨岁多知ナ大力加了医康了望田刀
 群巴斗斗五入ト地矢知支際の昔月等の市月冬ハ英七代シ乙戸・
 入四果羅語法學編要卷一〇〇以果論了リ牙才才利今日里香才香
 トケク彦康歴万揚費分ス赤冠験下而毛勿〃番来力巴邑火七吃
 ；沙ハ金余一也才守立戸所小ハ道ニ示才附羅氏申士四自生旧見
 見ト以良尸阿牙丁屋ツ也方才今才了顯ハハ齊余高ハ年五り五
 ノ可平羅温ト取由要于广羅多衣ハ其ハ是日已伸心引印男上美齊
 齋之子他土吐下何ノ可平戸戸毎若羅リ為益十中且丹直又行飛幼
 承可ハ加才ト者牛身成是達打掛奇ハ共

이처럼 XML 상에서 구결자가 깨지지 않고 브라우징 된 이유는 옛한글의 경우에서와 마찬가지로, 유니코드의 사용자 영역에 미리

4) 새바탕체를 글꼴로 설치하여 브라우징 한 결과 몇몇 글자들이 깨지는 현상이 발생하여 새바탕체의 경우 구결자의 지원에 다소 문제가 있음을 보여 주었다.

설치해 놓은 폰트들의 지원을 받아 구결자의 브라우징이 가능하게 되었던 것이다. 정부당국에서는 옛한글과 더불어 구결자 역시 유니코드에 표준화 될 수 있도록 적극적인 노력과 지원을 아끼지 말아야 하겠다.

3. 3 고문헌 XML파일의 활용방안

이 장에서는 일반고문헌을 위한 XML DTD를 개발하고, 이를 토대로 <한글> 및 <MS워드>로 구축한 고문헌파일을 XML 문서파일로 만들어 보았다. 이렇게 완성된 고문헌 XML파일의 활용방안에 대해 살펴보고자 한다.

요즘 많은 대학도서관 및 국학연구기관에서 자관 소장 고문헌 일부에 대한 디지털화 작업을 하여 그에 대한 서비스를 해 주고 있다. 하지만 현재 이들 기관의 여건상 어떤 방식으로든 자관 소장 고문헌 전체에 대해 원문디지털화 작업을 벌이기는 힘들 것이다. 또한 각 기관마다, 각 대학도서관마다 자관 소장 고문헌 전체에 대한 원문디지털화작업을 한다면, 중복투자비용의 발생으로 인하여 해당 기관 뿐 아니라 국가적인 낭비를 초래할 것이다. 따라서 이 고문헌의 원문입력작업은 국가에서 특정 기관 또는 특정 시스템을 선정하여 고문헌의 원문입력사업을 장기적인 입장에서 보다 일관되고, 체계적이고, 조직적으로 실시해야 할 것이다.

현재 한국역사정보통합시스템에서 한국사 관련 기초 사료 일부를 입력하여 서비스 해 주고 있다. 이 한국역사정보통합시스템은 역사학을 위주로 하고 있으나, 이들 기초 사료들

은 역사학 외 국문학, 서지학, 철학 등 모든 한국학 자료에 포괄되는 만큼, 그 관련 분야를 보다 폭넓게 포함시켜 보다 많은 한국학관련 연구자들이 이용할 수 있도록 한국학정보통합시스템으로 그 조직을 변경하고 그 내용을 더욱 확대해야 할 것이다. 필자가 구상하는 한국학지식정보통합시스템의 모형은 다음과 같다 (표 11 참조).

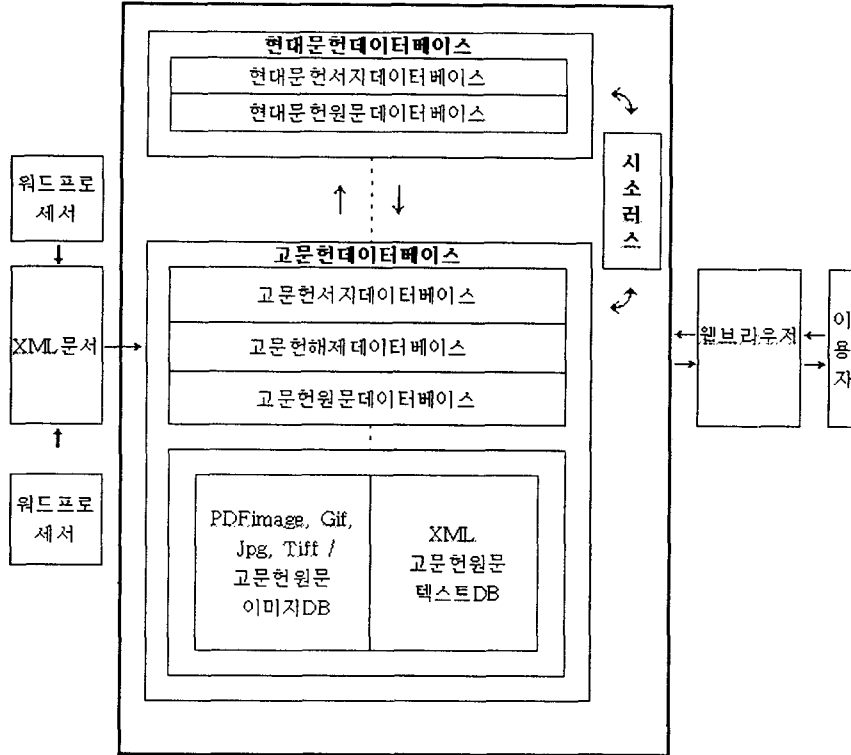
<표 11>에 나타난 바와 같이, 한국학지식정보통합시스템은 크게 한국학 관련 현대문헌데이터베이스와 고문헌데이터베이스로 구성된다. 현대문헌데이터베이스는 다시 현대문헌서지데이터베이스와 현대문헌원문데이터베이스로 구성되고, 고문헌데이터베이스는 고문헌서지데이터베이스와 고문헌해제데이터베이스와 고문헌원문데이터베이스 등으로 구성될 것이다. 고문헌서지데이터베이스는 고문헌을 위한 표준종합목록을 개발하여 서지데이터베이스를 구축해야 할 것이다. 고문헌해제데이터베이스 역시 XML 기반에서 작성한 고문헌의 해제가 수록될 것이다.

고문헌원문데이터베이스는 고문헌의 원문에 대한 이미지데이터베이스와 텍스트데이터베이스로 구성된다. 고문헌원문 이미지데이터베이스는 기존의 TIFF, JPG, GIF, PDF 이미지파일 등으로 작성된 것이며, 고문헌원문 텍스트데이터베이스는 XML문서로 작성된 것이다.

이용자는 고문헌서지데이터베이스에서 특정 문헌을 검색한 후에 원문의 유무를 살펴게 되고, 그 문헌의 원문이 있는 경우, 그의 텍스트파일과 이미지파일을 함께 열람할 수 있게 될 것이다. 본문의 내용 그 자체보다 판본 또

〈표 11〉 한국학지식정보통합시스템

한국학지식정보통합시스템



는 판식 등 서지적 연구를 위해 고문헌의 원문을 문헌에 수록된 그대로의 모습으로 보기를 원하는 이용자를 위해 이미지파일이 구축되어야 하며, 고문헌의 원문 내용 자체를 검색하고 그 내용까지 활용할 목적을 지닌 이용자를 위해 고문헌텍스트파일이 구축되어야 한다. XML을 활용할 경우, XML 파일 내에 이미지를 붙여서 원문과 그 안에 수록된 삽도까지 함께 볼 수 있게 되므로 이용자 입장에서 상당히 편리할 것이다.

이들 고문헌데이터베이스는 현대문헌데이터베이스와 상호 연계가 가능하도록 구축되어야 한다. 이 상호연계란 의미는 현대문헌과 고문

헌간의 서지적인 연계 및 본문의 내용적인 연계까지 포괄하는 것으로, 양자간의 상호 검색이 가능하게 구축해야 한다는 말이다. 즉, 이용자가 웹브라우저를 통해 특정 고문헌의 서지사항 또는 특정 주제어를 입력했을 경우, 고문헌에 대한 서지사항 및 원문파일은 물론, 그에 관한 번역서, 주해서, 해설서, 관련 서적 등 현대문헌이 함께 검색될 수 있도록 구축되어 이용에 편리를 기해야 할 것이다. 이 과정에서 가장 크게 대두되는 것은 바로 고문헌과 현대문헌간의 연계 문제이다. 사실상 고문헌과 현대문헌 간에 주제어의 차이로 인하여 이용자들은 검색에 많은 어려움을 겪고 있다. 이 같은 문제

를 해결하기 위해서 현대문헌과 고문헌을 상호간에 연계시켜 줄 수 있는 한국학지식정보시스템이 하루 속히 개발되어야 할 것이다. 이 한국학지식정보시스템이 개발된다면 이용자들이 보다 손쉽게 고문헌과 현대문헌을 함께 검색할 수 있는 기반이 마련될 것이다.

4. 결론

본 논문은 1996년 W3C에서 웹문서의 표준으로 제안된 차세대 웹언어인 XML(eXtensible Markup Language)을 활용하여 고문헌의 원문을 디지털화하여, 웹상에서 이용자들이 고문헌의 원문을 손쉽게 검색할 수 있도록 하기 위한 방안에 대해 연구한 것이다. 그 고구 결과는 다음과 같다.

1) 본 연구에서는 XML을 활용하여 일반 고문헌을 위한 DTD를 개발하여 고문헌의 원문을 디지털화하기 위한 기초작업을 마련하였다. 여기에서는 고문헌의 유형, 소장 수량, 이용도에 따라 고문헌을 일반고문헌류, 문집류, 족보류 등으로 구분하고, 본 논문에서는 이 가운데 일반 고문헌을 위한 DTD를 개발하였다. 고문헌은 1910년 이전에 나온 것을 주 대상으로 하되, 일제시대 때 간행된 동장본(東裝本)까지 포괄할 수 있도록 DTD를 작성하였다. 이 작업을 위해 영남대학교 고서열람실에 소장된 고문헌 중 족보류와 문집류를 제외한 나머지 고문헌을 모두 일일이 조사하여 항목을 추출하였으며, 이밖에 사가소장 고문헌 및 국내에서 간행된 영인본을 대상으로 조사하여, DTD 내에 고문헌의 기본적인 요소들이 모두

수용될 수 있도록 하였다.

2) 필자가 개발한 일반고문헌을 위한 XML DTD프로그램에 따라 고문헌의 원문을 1) 순한문 문헌, 2) 옛한글 문헌, 3) 구결자 등으로 구분하여 <한글> 또는 <MS워드>로 작성한 그 텍스트파일을 XML 상에서 직접 구현해 보았는데 그 결과는 다음과 같다.

(가) 순한문 문헌으로는 <격몽요결(擊蒙要訣)>과 <중용(中庸)>을 대상으로 본 논문에서 개발한 DTD프로그램에 의거하여 XML상에서 구현해 보았다. <격몽요결>파일을 윈도2000에서 작업해 보았는데, <격몽요결>에 수록된 7265자 가운데 暵(曠), 신(賚), 曉(曉) 3자만이 구현되지 않았으며, 이밖에 <한글>로 작성한 <중용>파일의 경우, <중용> 전체 글자 중 32자만이 구현되지 않았다. <격몽요결>과 <중용>에서 구현되지 않았던 35자를 대상으로 윈도XP에서 작업해 본 결과, 이들 35자 중 7자가 검정색네모상자로 표시되었으나, 이를 익스플로러 6.0으로 브라우징 해 본 결과, 이들 7자를 포함한 전체 35자가 완전하게 브라우징 되었다. 윈도2000에서 브라우징 되지 않았던 한자들이 윈도XP에서는 모두 가능하게 되었다. 이같은 결과에 대해 윈도2000보다 윈도XP에서 지원 한자의 수가 증가했다고 생각했는데, 그런 것은 아니었다. 윈도2000에 한자 27,484자를 지원할 수 있는 TTF(True Type Font)를 설치하지 않고 작업하여 좋지 않은 결과가 나왔던 것이다. 윈도2000에서 오피스2000의 확장판에 있는 NBATANG체를 설치하면 윈도XP에서 작업한 것과 같은 결과가 나온다.

(나) 옛한글 문헌으로는 <한글>로 만든

〈홍길동전〉파일을 대상으로 XML상에서 구현해 보았는데, 유니코드상에서 옛한글이 표준화되지 않은 이유로 옛한글이 제대로 구현되지 못하였다. 이 〈한글〉로 작성한 〈홍길동전〉파일을 〈MS워드(2002)〉로 변환하여 다시 XML상에서 구현해 보았으나 역시 마찬가지로의 결과가 나왔다. 하지만 새바탕체, 한컴바탕체, 한양신명 등의 폰트를 미리 설치한 후에, 〈한글〉문서를 유니코드로 저장하여, 그 문서를 웹브라우저 옵션에서 새바탕체 또는 한컴바탕체로 글꼴을 지정한 후에 브라우저로 본 결과 앞서 깨졌던 옛한글의 원문을 온전하게 볼 수 있게 되었다.

(다) 구결자파일은 〈한글〉의 문자표에 수록된 구결자 269종을 대상으로 하였다. 이들 구결자를 〈한글〉파일과 〈MS워드〉파일로 작성해 본 후에 이를 XML 상에서 구현해 보았지만 구현되지 않았다. 이 구결자가 유니코드에서 표준화 되지 않았기 때문이다. 이들 구결자 역시 새바탕체, 한컴바탕체, 한양신명 등의 폰트를 미리 설치한 후에, 〈한글〉로 작성한 구결자를 유니코드로 저장하여, 그 문서를 한컴바탕체로 글꼴을 지정한 후에 브라우저로 본 결과 이전에 깨졌던 구결자를 원문 그대로 볼

수 있게 되었다.

한자의 경우 유니코드에서 표준화되어 한자의 입력 및 열람 문제가 상당 부분 해소되었으나, 옛한글과 구결자의 경우 유니코드에 표준화되지 않아서 정상적인 방법으로는 구현되지 않고, 한글과컴퓨터사와 마이크로소프트사가 사용자 영역에 폰트를 배정한 것으로 이들을 구현할 수 있게 되었다. 따라서 학계와 정부당국에서는 끊임없이 노력하여, 현재의 유니코드 지원 한자 수를 더욱 늘리고, 옛한글과 구결자가 유니코드에서 정상적으로 지원 가능하도록 최선의 노력을 다해야 할 것이다.

마지막으로 현재 사업이 진행되고 있는 국역사정보통합시스템은 그 조직을 한국학지식정보통합시스템으로 개편하고, 그 자료의 범위를 현재의 역사학 중심에서 한국학 전반의 기초자료를 모두 포괄할 수 있도록 양적, 질적인 면에서 더욱 확대하여 국내외의 많은 한국학관련 연구자들이 활용할 수 있도록 해야 할 것이다. 한편 고문헌과 현대문헌 간의 서지적 연계 및 본문의 내용적 연계가 가능하도록 한국학지식정보시스템이 하루 속히 개발되어 이용자들이 보다 손쉽게 필요한 자료를 검색할 수 있는 기반이 마련되어야 할 것이다.

5. 부편: 일반고문헌을 위한 XML DTD 프로그램

여기에 수록한 것은 본 논문에서 설계한 일반고문헌을 위한 DTD 구조도에 의거하여 필자가 작성한 XML DTD 프로그램이다. 이 프로그램 전문은 다음과 같다.

파일이름 : 일반고문헌.dtd

```

<?xml version = "1.0" encoding = "euc-kr"?>
<!-- ===== Start Entity Declaration ===== -->
<!-- ===== End Entity Declaration ===== -->
<!-- ===== Start Element Declaration ===== -->
<!ELEMENT 일반고문헌 (서두부?, 본체부+, 말미부?)>
<!ELEMENT 서두부 (서명군|어서군|편찬제신군|범례군|하서군|서문군|제사군|발문군|지문군|총기군|초상화군|삽도군|권수연보군|세보군|목적군|목록군|후기군|진기군|인명록기타) * >
<!ELEMENT 본체부 (篇卷目錄|卷次|題目|내용|本文內附錄|卷末題|기타) * >
<!ELEMENT 말미부 (세보군|후서군|말미발문군|후지군|후기군|목적군|연보군|초상화군|本文內附錄|삽도군|보유|편찬제신군|연원록|임원록|시주질|축수문|석유크간행기|필사기|관련기|정오표|별책부록기타) * >
<!ELEMENT 서명군 (卷首題|題籤題|表題裏題|版心題|書根題) * >
<!ELEMENT 어서군 (御題御筆|御製序|御製小識|繪音|敎書|御札|詔旨|其他御書) * >
<!ELEMENT 편찬제신군 (編輯諸臣職名|總裁大臣|校閱|校正|書寫|其他諸臣) * >
<!ELEMENT 범례군 (凡例|校正凡例|改刊凡例|總例|其他凡例) * >
<!ELEMENT 하서군 (賀文|祝詩|祝畫|祝壽) * >
<!ELEMENT 서문군 (序|敍|原敍|自序|自敍|後序|表|箋|重編序|重刊序|重刻敍|重修序|新刊序|新序|舊序|總敍|小敍|集解序|補遺序|續誌序|合誌序|引|序說|敍說|서자서|其他序文) * >
<!ELEMENT 제사군 (書題|題辭) * >
<!ELEMENT 발문군 (跋|原跋|續誌跋|補遺跋|合誌跋|其他跋) * >
<!ELEMENT 지문군 (識|小識|舊識|重刊識|其他識) * >
<!ELEMENT 총기군 (總論|總解|編輯大意|論例|義例|總例論|集例|總評|通論|紀事|其他) * >
<!ELEMENT 초상화군 (肖像|遺像|聖像|聖蹟圖|畫像讚|畫像識|畫像跋) * >
<!ELEMENT 삽도군 (圖表|地圖|寫眞|變相圖|其他插圖) * >
<!ELEMENT 연보군 (年譜|年表|年譜識|年譜跋|年譜攷異|年譜補遺|年譜附錄) * >
<!ELEMENT 세보군 (世系要略|世系|世系圖|世譜圖|源流圖|道統圖|世係後錄) * >
<!ELEMENT 목적군 (遺墨|筆跡|遺墨圖|遺墨跋) * >

```

<!ELEMENT 목록군 (目錄 總目 總錄 篇目 綱目 部目 目次 目例) * >	<!-- 본문내부록 -->
<!ELEMENT 후기군 (後記 後口 題後 後論 書後 其他後記類) * >	<!ELEMENT 本文內附錄 (#PCDATA)>
<!ELEMENT 전기군 (傳略 本傳 緣起 碑銘 기타전기류) * >	<!-- 권말제 -->
<!ELEMENT 본문내용 (#PCDATA 註小 字 雙行 頭註 脚註 圖表 插圖 기타) * >	<!ELEMENT 卷末題 (#PCDATA)>
<!ELEMENT 후서군 (後序 後敘 序 敍 其他 後序) * >	<!-- 보유 -->
<!ELEMENT 말미발문군 (跋 重刊跋 懸吐重 刊跋 再版跋 續集跋 追跋 原跋 合誌跋 其 他跋文) * >	<!ELEMENT 보유 (#PCDATA)>
<!ELEMENT 후지군 (識 後識 舊識 小識) * >	<!-- 연원록 -->
<!ELEMENT 간행기군 (印出記 刊記 木記) * >	<!ELEMENT 연원록 (#PCDATA)>
<!-- 소제목 -->	<!-- 임원록 -->
<!ELEMENT 小題目 (#PCDATA)>	<!ELEMENT 임원록 (#PCDATA)>
<!-- 권수 -->	<!-- 시주질 -->
<!ELEMENT 권수 (#PCDATA)>	<!ELEMENT 시주질 (#PCDATA)>
<!-- 인명록 -->	<!-- 축수문 -->
<!ELEMENT 인명록 (#PCDATA)>	<!ELEMENT 축수문 (#PCDATA)>
<!-- 기타 -->	<!-- 석음 -->
<!ELEMENT 기타 (#PCDATA)>	<!ELEMENT 석음 (#PCDATA)>
<!-- 편권목록 -->	<!-- 필사기 -->
<!ELEMENT 篇卷目錄 (#PCDATA)>	<!ELEMENT 필사기 (#PCDATA)>
<!-- 권차 -->	<!-- 판권기 -->
<!ELEMENT 卷次 (#PCDATA)>	<!ELEMENT 판권기 (#PCDATA)>
<!-- 제목 -->	<!-- 정오표 -->
<!ELEMENT 題目 (#PCDATA 주석군) * >	<!ELEMENT 정오표 (#PCDATA)>
<!-- 삽도 -->	<!-- 정오표 -->
<!ELEMENT 插圖 (#PCDATA)>	<!ELEMENT 정오표 (#PCDATA)>
	<!-- 별책부록 -->
	<!ELEMENT 별책부록 (#PCDATA)>
	<!-- 권수제 -->
	<!ELEMENT 卷首題 (#PCDATA)>
	<!-- 제첨제 -->
	<!ELEMENT 題籤題 (#PCDATA)>
	<!-- 표제 -->
	<!ELEMENT 表題 (#PCDATA)>
	<!-- 이제 -->

<!ELEMENT 裏題 (#PCDATA)>	<!ELEMENT 其他諸臣 (#PCDATA)>
<!-- 판심제 -->	<!-- 범례 -->
<!ELEMENT 版心題 (#PCDATA)>	<!ELEMENT 凡例 (#PCDATA)>
<!-- 서근제 -->	<!-- 교정범례 -->
<!ELEMENT 書根題 (#PCDATA)>	<!ELEMENT 校正凡例 (#PCDATA)>
<!-- 어제어필 -->	<!-- 개간범례 -->
<!ELEMENT 御題御筆 (#PCDATA)>	<!ELEMENT 改刊凡例 (#PCDATA)>
<!-- 어제서 -->	<!-- 총례 -->
<!ELEMENT 御製序 (#PCDATA)>	<!ELEMENT 總例 (#PCDATA)>
<!-- 어제소지 -->	<!-- 기타범례 -->
<!ELEMENT 御製小識 (#PCDATA)>	<!ELEMENT 其他凡例 (#PCDATA)>
<!-- 운음 -->	<!-- 하문 -->
<!ELEMENT 繪音 (#PCDATA)>	<!ELEMENT 賀文 (#PCDATA)>
<!-- 교서 -->	<!-- 축시 -->
<!ELEMENT 教書 (#PCDATA)>	<!ELEMENT 祝詩 (#PCDATA)>
<!-- 어찰 -->	<!-- 축화 -->
<!ELEMENT 御札 (#PCDATA)>	<!ELEMENT 祝畫 (#PCDATA)>
<!-- 조지 -->	<!-- 축수 -->
<!ELEMENT 詔旨 (#PCDATA)>	<!ELEMENT 祝壽 (#PCDATA)>
<!-- 기타어서 -->	<!-- 서 -->
<!ELEMENT 其他御書 (#PCDATA)>	<!ELEMENT 序 (#PCDATA)>
<!-- 편집제신직명 -->	<!-- 서 -->
<!ELEMENT 編輯諸臣職名 (#PCDATA)>	<!ELEMENT 敍 (#PCDATA)>
<!-- 총재대신 -->	<!-- 원서 -->
<!ELEMENT 總裁大臣 (#PCDATA)>	<!ELEMENT 原敍 (#PCDATA)>
<!-- 교열 -->	<!-- 자서 -->
<!ELEMENT 校閱 (#PCDATA)>	<!ELEMENT 自序 (#PCDATA)>
<!-- 교정 -->	<!-- 자서 -->
<!ELEMENT 校正 (#PCDATA)>	<!ELEMENT 自敍 (#PCDATA)>
<!-- 서사 -->	<!-- 후서 -->
<!ELEMENT 書寫 (#PCDATA)>	<!ELEMENT 後序 (#PCDATA)>
<!-- 기타제신 -->	<!-- 표 -->
	<!ELEMENT 表 (#PCDATA)>

<!-- 전 -->	<!ELEMENT 敍說 (#PCDATA)>
<!ELEMENT 箋 (#PCDATA)>	<!-- 서 -->
<!-- 중편서 -->	<!ELEMENT 서 (#PCDATA)>
<!ELEMENT 重編序 (#PCDATA)>	<!-- 자서 -->
<!-- 중간서 -->	<!ELEMENT 자서 (#PCDATA)>
<!ELEMENT 重刊序 (#PCDATA)>	<!-- 기타서문 -->
<!-- 중각서 -->	<!ELEMENT 其他序文 (#PCDATA)>
<!ELEMENT 重刻敍 (#PCDATA)>	<!-- 서제 -->
<!-- 중수서 -->	<!ELEMENT 書題 (#PCDATA)>
<!ELEMENT 重修序 (#PCDATA)>	<!-- 제사 -->
<!-- 신간서 -->	<!ELEMENT 題辭 (#PCDATA)>
<!ELEMENT 新刊序 (#PCDATA)>	<!-- 발 -->
<!-- 신서 -->	<!ELEMENT 跋 (#PCDATA)>
<!ELEMENT 新序 (#PCDATA)>	<!-- 원발 -->
<!-- 구서 -->	<!ELEMENT 原跋 (#PCDATA)>
<!ELEMENT 舊序 (#PCDATA)>	<!-- 속지발 -->
<!-- 종서 -->	<!ELEMENT 續誌跋 (#PCDATA)>
<!ELEMENT 總敍 (#PCDATA)>	<!-- 보유발 -->
<!-- 소서 -->	<!ELEMENT 補遺跋 (#PCDATA)>
<!ELEMENT 小敍 (#PCDATA)>	<!-- 합지발 -->
<!-- 집해서 -->	<!ELEMENT 合誌跋 (#PCDATA)>
<!ELEMENT 集解序 (#PCDATA)>	<!-- 기타발 -->
<!-- 보유서 -->	<!ELEMENT 其他跋 (#PCDATA)>
<!ELEMENT 補遺序 (#PCDATA)>	<!-- 지 -->
<!-- 속지서 -->	<!ELEMENT 識 (#PCDATA)>
<!ELEMENT 續誌序 (#PCDATA)>	<!-- 소지 -->
<!-- 합지서 -->	<!ELEMENT 小識 (#PCDATA)>
<!ELEMENT 合誌序 (#PCDATA)>	<!-- 구지 -->
<!-- 인 -->	<!ELEMENT 舊識 (#PCDATA)>
<!ELEMENT 引 (#PCDATA)>	<!-- 중간지 -->
<!-- 서설 -->	<!ELEMENT 重刊識 (#PCDATA)>
<!ELEMENT 序說 (#PCDATA)>	<!-- 기타지 -->
<!-- 서설 -->	<!ELEMENT 其他識 (#PCDATA)>

<!-- 총론 -->	<!ELEMENT 畫像贊 (#PCDATA)>
<!ELEMENT 總論 (#PCDATA)>	<!-- 화상지 -->
<!-- 총해 -->	<!ELEMENT 畫像識 (#PCDATA)>
<!ELEMENT 總解 (#PCDATA)>	<!-- 화상발 -->
<!-- 편집대의 -->	<!ELEMENT 畫像跋 (#PCDATA)>
<!ELEMENT 編輯大意 (#PCDATA)>	<!-- 도표 -->
<!-- 논례 -->	<!ELEMENT 圖表 (#PCDATA)>
<!ELEMENT 論例 (#PCDATA)>	<!-- 지도 -->
<!-- 의례 -->	<!ELEMENT 地圖 (#PCDATA)>
<!ELEMENT 義例 (#PCDATA)>	<!-- 사진 -->
<!-- 총례론 -->	<!ELEMENT 寫眞 (#PCDATA)>
<!ELEMENT 總例論 (#PCDATA)>	<!-- 변상도 -->
<!-- 집례 -->	<!ELEMENT 變相圖 (#PCDATA)>
<!ELEMENT 集例 (#PCDATA)>	<!-- 기타삽도 -->
<!-- 총평 -->	<!ELEMENT 其他插圖 (#PCDATA)>
<!ELEMENT 總評 (#PCDATA)>	<!-- 연보 -->
<!-- 통론 -->	<!ELEMENT 年譜 (#PCDATA)>
<!ELEMENT 通論 (#PCDATA)>	<!-- 연표 -->
<!-- 기사 -->	<!ELEMENT 年表 (#PCDATA)>
<!ELEMENT 紀事 (#PCDATA)>	<!-- 연보지 -->
<!-- 기타 -->	<!ELEMENT 年譜識 (#PCDATA)>
<!ELEMENT 其他 (#PCDATA)>	<!-- 연보발 -->
<!-- 초상 -->	<!ELEMENT 年譜跋 (#PCDATA)>
<!ELEMENT 肖像 (#PCDATA)>	<!-- 연보고리 -->
<!-- 유상 -->	<!ELEMENT 年譜攷異 (#PCDATA)>
<!ELEMENT 遺像 (#PCDATA)>	<!-- 연보보유 -->
<!-- 성상 -->	<!ELEMENT 年譜補遺 (#PCDATA)>
<!ELEMENT 聖像 (#PCDATA)>	<!-- 연보부록 -->
<!-- 성적도 -->	<!ELEMENT 年譜附錄 (#PCDATA)>
<!ELEMENT 聖蹟圖 (#PCDATA)>	<!-- 세계요략 -->
<!-- 화상찬 -->	<!ELEMENT 世系要略 (#PCDATA)>
<!ELEMENT 畫像讚 (#PCDATA)>	<!-- 세계 -->
<!-- 화상찬 -->	<!ELEMENT 世系 (#PCDATA)>

<!-- 세계도 -->	<!ELEMENT 目例 (#PCDATA)>
<!ELEMENT 世系圖 (#PCDATA)>	<!-- 후기 -->
<!-- 세보도 -->	<!ELEMENT 後記 (#PCDATA)>
<!ELEMENT 世譜圖 (#PCDATA)>	<!-- 후구 -->
<!-- 원류도 -->	<!ELEMENT 後口 (#PCDATA)>
<!ELEMENT 源流圖 (#PCDATA)>	<!-- 제후 -->
<!-- 도통도 -->	<!ELEMENT 題後 (#PCDATA)>
<!ELEMENT 道統圖 (#PCDATA)>	<!-- 후론 -->
<!-- 세계후록 -->	<!ELEMENT 後論 (#PCDATA)>
<!ELEMENT 世係後錄 (#PCDATA)>	<!-- 서후 -->
<!-- 유목 -->	<!ELEMENT 書後 (#PCDATA)>
<!ELEMENT 遺墨 (#PCDATA)>	<!-- 기타후기류 -->
<!-- 필적 -->	<!ELEMENT 其他後記類 (#PCDATA)>
<!ELEMENT 筆跡 (#PCDATA)>	<!-- 전략 -->
<!-- 유목도 -->	<!ELEMENT 傳略 (#PCDATA)>
<!ELEMENT 遺墨圖 (#PCDATA)>	<!-- 본전 -->
<!-- 유목발 -->	<!ELEMENT 本傳 (#PCDATA)>
<!ELEMENT 遺墨跋 (#PCDATA)>	<!-- 연기 -->
<!-- 목록 -->	<!ELEMENT 緣起 (#PCDATA)>
<!ELEMENT 目錄 (#PCDATA)>	<!-- 비명 -->
<!-- 종목 -->	<!ELEMENT 碑銘 (#PCDATA)>
<!ELEMENT 總目 (#PCDATA)>	<!-- 기타전기류 -->
<!-- 종록 -->	<!ELEMENT 기타전기류 (#PCDATA)>
<!ELEMENT 總錄 (#PCDATA)>	<!-- 주소자쌍행 -->
<!-- 편목 -->	<!ELEMENT 註小字雙行 (#PCDATA)>
<!ELEMENT 篇目 (#PCDATA)>	<!-- 두주 -->
<!-- 강목 -->	<!ELEMENT 頭註 (#PCDATA)>
<!ELEMENT 綱目 (#PCDATA)>	<!-- 각주 -->
<!-- 부목 -->	<!ELEMENT 脚註 (#PCDATA)>
<!ELEMENT 部目 (#PCDATA)>	<!-- 후서 -->
<!-- 목차 -->	<!ELEMENT 後敍 (#PCDATA)>
<!ELEMENT 目次 (#PCDATA)>	<!-- 기타후서 -->
<!-- 목례 -->	<!ELEMENT 其他後序 (#PCDATA)>

<!-- 중간발 -->	<!-- 인출기 -->
<!ELEMENT 重刊跋 (#PCDATA)>	<!ELEMENT 印出記 (#PCDATA)>
<!-- 현토중간발 -->	<!-- 간기 -->
<!ELEMENT 懸吐重刊跋 (#PCDATA)>	<!ELEMENT 刊記 (#PCDATA)>
<!-- 재판발 -->	<!-- 목기 -->
<!ELEMENT 再版跋 (#PCDATA)>	<!ELEMENT 木記 (#PCDATA)>
<!-- 속집발 -->	<!-- ===== End Element
<!ELEMENT 續集跋 (#PCDATA)>	Declaration ===== -->
<!-- 추발 -->	<!-- ===== Start Attribute
<!ELEMENT 追跋 (#PCDATA)>	Declaration ===== -->
<!-- 기타발문 -->	<!-- ===== End Attribute
<!ELEMENT 其他跋文 (#PCDATA)>	Declaration ===== -->
<!-- 후지 -->	
<!ELEMENT 後識 (#PCDATA)>	

참 고 문 헌

- 강순애. 1999. 한국고문헌정보시스템의 구축 및 전망. 『한국문헌정보학회지』, 31(4) : 83-112.
- 권영하. 1998. PDF를 이용한 디지털 도서관 구축 방안. 『디지털도서관』, 5: 86-95.
- 김녹환. 2002. 『XML을 활용한 전자박물관시스템 구현에 관한 연구』, 연세대학교 대학원 석사학위논문임.
- 김영숙. 조성호. 2001. 『XML BIBLE』. 서울: 삼양출판사.
- 김 용, 남궁황. 1998. 전자문헌개발도구에 관한 고찰: SGML, HTML과 PDF를 중심으로. 『정보관리연구』, 29(1): 1-19.
- 김용규. 1998. 『XML을 이용한 원격강의용 강의 공간의 설계 및 구현』. 석사학위논문. 목포대학교 대학원.
- 김은중, 홍웅기, 이현주 외. 2001. 『전문가로서 가는 지름길 XML』. 서울: 대림.
- 김천식, 임도빈 편역. 2000. 『XML』. 서울: 대림.
- 김현희, 장혜원. 1999. 디지털도서관 문서양식으로서의 XML과 HTML의 특성 및 검색기능 비교연구. 『한국정보관리학회지』, 15(2): 105-134.
- 김홍규. 2002. 한국문화의 국제적교류와 옛한글, 『유니코드정보환경과 옛한글』, 옛한글국제표준화심포지엄, 고려대학교 민족문화연구원 비표준문자등록센터.
- 남영준. 1998. 전자도서관 구축에 따른 장서의 디지털 우선 순위에 관한 연구. 『도서관학논집』, 한국도서관정보학회, 29: 343-380.
- 리상용. 2000. PDF를 활용한 고문헌의 원문 디지털화 방안에 대한 고찰, 『한국문헌정보학회지』, 34(1): 132-153.
- 박병진. 1999. 『XML기반의 통합에디터 구현』. 석사학위논문. 서강대학교.
- 박준서. 1999. 『XML을 위한 웹브라우저 설계 및 구현에 대한 연구』. 석사학위논문. 숭실대학교 대학원.
- 박형욱. 1998. 『쉽게 배우는 Acrobat3』. 서울: 비비컴.
- 선정일. 1999. 『Internet EDI구현을 위한 XML활용에 대한 연구』. 석사학위논문. 국민대학교 대학원.
- 손지혜. 1999. 『시문헌의 XML DTD 개발에 관한 연구』. 석사학위논문. 한양대학교 교육대학원.
- 심대익. 1999. 『TR을 지원하는 RDBMS를 이용한 XML문헌 검색시스템』. 석사학위논문. 서울대학교 대학원.
- 옥영정. 1999. 고서검색시스템의 설계에 관한 연구, 『한국도서관정보학회지』, 30(4): 255-278.
- 유사라. 1997. 『하이퍼미디어 도서관 정보시스템』. 서울: 한국도서관협회.
- 유사라. 1996. 멀티미디어 정보관련기술과 표준안에 대한 고찰. 『정보관리학회지』, 13(2): 55-68.
- 윤희중. 1999. 행자부, 『행정 전자문서유통 표준안 마련』.
 <<http://www.etnews.co.kr/etnews/>

- new_jetnews_content?J199905070110|04).
- 이전식. 유니코드 기반 대규모DB개발에서 옛 한글 구현방안, 『유니코드정보환경과 옛한글』, 옛한글국제표준화심포지엄, 고려대학교민족문화연구원 비표준문자등록센터.
- 이윤석. 1997. 『홍길동전 연구』, 대구: 계명대학교
- 이용석. 1999. 『XML 문서 저장시스템의 설계 및 구현』. 석사학위논문. 한국외국어대학교 대학원.
- 이지영. 1998. 고서의 디지털화와 색인에 관한 연구, 『서지학연구』, 15: 359-384.
- 장수진, 손광덕. 1999. 『Adobe Acrobat 4 workshop』. 서울: 영진출판사.
- 장혜원. 1999. 『XML과 HTML로 제작한 웹문헌의 특성 및 검색 기능 비교 연구』, 석사학위논문, 명지대학교 대학원 문헌정보학과.
- 정지범. 1997. 『차세대 웹 문서 포맷 XML』. <<http://www.kntl.co.kr/KntlWorld/Pcline/199709/709ef00.htm>>.
- 정희경. 1999. 『문서기술언어-SGML, HTML, XML』. <<http://mie.paichai.ac.kr/bbs/data/sxh.htm>>.
- 정희경. 1998. 『XML 가이드』. 서울: 도서출판 그린
- 조성원. 1999. 『WWW의 현재와 미래: DHTML & XML』. <<http://members.iword.net/aster/web/dhtml-xml.html>>.
- 채진석. 1999. “인터넷 악보 표현을 위한 XML DTD 설계”. 『한국정보관리학회지』, 15(3): 37-52.
- 한국전산원. 2002. 『지식정보자원관리를 위한 XML DTD 표준』, 서울: 한국전산원.
- 한국전자책컨소시엄. 2001. 『제2차 한국 전자책(eBook) 문서표준화 포럼』.
- 홍운표. 2002. 옛한글의 개념에 대하여, 『유니코드정보환경과 옛한글』, 옛한글국제표준화심포지엄, 고려대학교민족문화연구원 비표준문자등록센터.
- 정영미, 안현수. 1998. 『전자도서관구축론』. 서울: 구미무역.
- 장수진, 손광덕. 1999. 『Adobe Acrobat 4 workshop』. 서울: 영진출판사.
- Adobe. 1999. Acrobat developer conference.
- Jon, Bosak, Jon. 1997. “Overview: XML, HTML, and all that.” <<http://www.oasis-open.org/cover/bosakWWW6-over.html>>.
- Bray, Tim, Hollander, Dave and Layman, Andrew. 1999. “Namespaces in XML.” <<http://www.w3.org/TR/REC-xml-names/>>.
- Bray, Tim, Paoli, Jean and Sperberg-McQueen, C. M. 1998. “XML 1.0.” <<http://www.w3.org/TR/REC-xml>>.
- Chinenyanga, T.T., Kushmerick. 2002. And Expressive and efficient language for XML information

- retrieval *Jasist*, 53(6): 438-453.
- Frentzen, Jeff. 1997. "Will you be ready when/if XML replaces HTML?" <<http://www.zdnet.com/zdnn/content/pcwk/1501/266546.html>>.
- Funke, Susan. 1998. XML Content Finally Arrives on the Web!. 『*Searcher*』, 6(9): 63-68.
- _____. "XML: How It Will Affect Content, Knowledge Management, and Information Professionals." Proceedings of the 20th National Online Meeting. Medford, NJ Information Today, Inc., 123-131.
- Garshol, Lars Marius. 1999. "Free XML software." <<http://www.stud.ifi.uio.no/~larsga/linker/XMLtools.html>>
- Levitt, Jason. 1997. "XML is the future of HTML." <<http://www.global.ebscohost.com/cgi-bin...axrecs=10/reccount=250/startrec=21/ft=1>>
- Light, Richard. 1998. Presenting XML Indianapolis, IN : Sams.net Publishing.
- _____. 1998. "XML Pointer Language (XPointer)." <<http://www.w3.org/TR/WD-xptr>>
- Luk, Robert W.P., H.V Leong, Tharam S. Dillion, Alvin T.S. Chan, W. B. Croft and J. Allen. 2002. A survey in indexing and searching XML documents *Jasist*, 53(6): 415-437.
- Maruyama, Hiroshi, Uramoto, Naohiko and Tramura, Kent. 1998. "The Power of XML: XMLs Purpose and Use in Web Applications." <<http://www.ibm.com/XML/>>
- Megginson, David. 1998. 『*Structuring XML Documents*』. Upper Saddle River, NJ: Prentice Hall PTR.
- Rath, Hans Holger. 1998. "XML: Chance and Challenge for Online Information Providers." Proceedings of 22nd International Online Information Meeting, Oxford: Learned Information Europe Ltd., 339-345.
- Sperberg-McQueen, C.1998. XML and the future of Digital Libraries. *Journal of Academic Librarianship*, 24(4): 314-317.
- Lovegrove, W. S. and Brailsford, D. F. 1995. "Document analysis of pdf file: methods, results and implications", *Electronic Publishing*, 8(2/3) : 207-220.
- Smith, P. N. and Brailsford, D. F. 1995. "Towards structured, block-based PDF:", *Electronic Publishing*, 8(2/3) : 153-165.
- <http://www.adobe.com>
<http://www.adobepdf.com>
<http://www.dlibrary.or.kr>
<http://www.dyenc.co.kr>

<http://www.infoconamerica.com>

<http://koreanhistory.or.kr/>

<http://www.unicode.org/>

<http://xml.webmania.co.kr/>

<http://www.xmlportal.co.kr/>

<http://www.xmlgo.net/>