학위논문의 XML DTD 설계에 관한 연구

A Study of an XML DTD Design of Theses and Dissertations

안영선(Young-Sun Ahn)*, 이은철(Eun-Chul Lee)**, 오삼균(Sam-Gyun Oh)***

목 차

1 서 론
2 이론적 배경
2.1 XML DTD의 구성
2.2 선행연구 개관
3 연구의 방법
3.1 연구의 설계
3.2 XML DTD 설계
4 XML기반의 학위논문 DTD 설계
4 1 학위논문의 구조 분석
4 2 학위논문의 XML DTD 설계
5 결 론

초 록

본 연구는 이공계열과 인문사회계열의 석 · 박사 학위논문 24종의 논리구조를 분석하여 이를 바탕으로 XML 기반의 학위논문을 생성하여 활용할 수 있도록 XML DTD를 설계하였다. XML은 HTML의 단순성의 한계와 SGML의 복잡함을 해결하기 위해 등장한 메타언어로서, 웹상에서 구조화된 문서의 전송을 가능케 한다. 따라서 이 연구에서 설계한 DTD는 XML 기반의 학위논문을 구축하는데 좋은 출발점으로 활용될 수 있게 되기를 기대하고, 이렇게 구축된 학위논문은 효율적이고 체계적으로 논문의 지식을 축적 및 관리하고 이용자가 질적으로 향상된 구조검색을 수행하는데 도움을 주는데 목적이 있다.

ABSTRACT

XML, a flexible and concise markup language designed to solve both the limitation of HTML and the complexity of SGML, has capacity to exchange and search documents effectively using the DTD structures on the Web. The primary objective of this paper is to design an XML DTD based on the analysis of the logical structure of 24 theses and dissertations in the field of natural science, engineering, humanities, and social science. If theses and dissertations are constructed using the DTD designed in this paper, the effective management of theses and dissertations and more precise searching can be accomplished.

키워드 : XML, DTD, 학위논문 DTD, 전자문서

* 경희대학교 의학도서관
** 성균관대 문헌정보학과 교수
*** 성균관대 문헌정보학과 부교수
■ 논문 접수일 : 2000년 11월 17일

1 서 론

정보통신과 네트워크 기술의 발전은 고도화된 정보화 사회를 이끌고 있으며, 이러한 환경은 도서관의 문헌들을 전자문서로 변환시켜, 효율적으로 축적하고 이용할 수 있도록 촉진하고 있다.

특히, 디지털도서관은 많은 양의 전자문서 정보를 효율적이고 체계적으로 축적, 검색, 관리하여 이용자들에게 능동적으로 서비스하기 위한 노력을 기우리고 있다.

전자문서의 핵심은 사용하는 플랫폼에 독립적이며, 문서 정보의 전송과 교환이 용이하고, 문서가 나타내는 의미를 그대로 표현할 수 있어야 한다는 점이다. 전자문서의 대표적인 표준으로 SGML (Standard Generalized Markup Language), HTML(HyperText Markup Language), XML(Extensible Markup Language) 등을 들 수 있다. 이 전자문서 표준들 중 SGML은 그 구성이 매우 복잡하여 실제 구현이 매우 어렵기 때문에 현재까지는 독자적인 해결책에 의존하고 있는 상황이며, 웹에서 가장 많이 사용되고 있는 HTML은 문서의 구조를 표시할 수 있는 태그들이 너무 부족하고 문서의 구조와 출력이 밀착되어 있어서 전자문서의 핵심 사항들을 만족시키는 데에는 한계를 지니고 있다.

이의 대안으로 제시된 인터넷 전자문서의 표준이 XML이다. 이것은 웹 환경에 맞도록 변형한 언어로 문서의 교환과 전송의 용이성을 추가한 표준이다.

XML은 HTML이 정보의 표현만을 목적으로 한다는 점을 보완하고, SGML의 복잡함을 해결해 줄뿐만 아니라 XML 문서를 대상으로 구조검색이 가능하게 하여 이용자에게 보다 적합한 문헌을 제공할 수 있다.

따라서 도서관에서 전자문서를 만들 때 XML을 활용하게 되면 이용자들에게 더 나은 서비스를 제공할 수 있다. 특히, 매년 각 대학교에서 생산되고 있는 엄청난 양의 학위논문들을 도서관들이 개별적으로 수집 · 축적 · 이용한다는 것은 매우 어려운 일이다.

그러므로 각 도서관들이 학문적 가치가 높고 그 활용도가 높은 학위논문을 XML 기반으로 구축하고 이를 상호 공유한다면 공간활용, 자료보존 및 이용자 서비스에 있어서 매우 유용한 효과를 얻을 수 있다.

학위논문을 XML 문서로 만들기 위해서는 DTD(Document Type Definition)의 설계가 우선적으로 수행되어야 한다. DTD는 실제로 정보를 제공할 목적이 아니라 XML 문서의 생성에 목적을 두고 있으며, 구조검색 시 중요한 역할을 담당하기 때문에 정확한 문헌구조의 분석을 통해 좋은 DTD를 개발하는 것은 매우 중요하다.

따라서 하나의 문서 구조에 맞추어 많은 문서가 같은 형식으로 작성되어야 하는 상황에서는 XML 문서를 생성하기 전에 기본적으로 고유의 DTD 설계가 선행되어야 한다. 이렇게 구현된 DTD를 기반으로 학위논문을 XML 문서화한다면 학위논문의 원활한 유통과 정확한 구조검색을 통해 이용자에게 질적으로 향상된 서비스를 제공할 수 있을 것이다.

이에 본 연구는 XML 기반의 학위논문을 생성하여 활용할 수 있도록 학위논문의 논리적 구조를 분석하고, 이를 바탕으로 체계적으로 XML DTD를 설계하는 것을 목적으로 수행되었다.

2 이론적 배경

2.1 XML DTD의 구성

DTD는 어떤 특정한 형태를 갖는 문서에 적용하는 마크업 선언들의 집합으로 구성되며, XML 문서 자체를 구성할 수 있는 논리구조를 내포하고 있다. 이러한 DTD에는 작성하고자 하는 문서에서 사용할 태그의 종류와 태그의 순서, 태그의 발생 횟수 등 문서의 구조 정보가 정의되어 있다.

XML DTD는 태그들의 계층구조를 정의하고 표현하는 요소 선언(Element Declaration), 요소들의 속성명과 속성 값을 지정하는 속성 선언(Attribute Declaration), DTD나 XML 문서 내에서 문자열을 참조하거나 대체기능을 제공하는 개체 선언(Entity Declaration), 그리고 표기법(Notations), 주석(Comments), 처리 명령어(Processing Instructions), CDATA 섹션(Sections) 등으로 구성된다.

요소(Element)는 XML 문서의 기본이 되는 논리적 단위이며 XML 문서의 태그를 정의하는 것이다. XML 문서의 모든 내용들은 반드시 요소 안에 포함되어야 하며 XML 문서는 하나의 루트요소로 구성된다. 루트요소는 수많은 하위 요소가 포함될 수 있지만 그 자신은 둘 이상이 될 수는 없다.

속성(Attribute) 선언은 문서나 요소의 속성을 정의하는 것으로 각각의 속성들은 이름, 유형 그리고 기본 값의 3부분으로 선언된다. 속성 정의에서 처음 나오는 것은 속성이 적용될 요소의 이름이며, 한 요소에 대해 여러 개의 속성 정의가 가능하다. 속성 유형에는 'CDATA', '나열 값', 'ID', 'IDREF' 등이 있고, 속성 선언에 필요한 기본값은 '#REQUIRED', '#IMPLIED', '#FIXED', '열거형 기본 값' 이 있다.

개체(Entity)는 문서 내에서 참조될 수 있는 문서 집합의 단위이다. 문서에는 키보드로 입력할 수 없는 그리스 문자나 수학 기호, 그림 같은 것이 포함되거나, 같은 문자들을 여러 번 반복하여 입력해야 하는 경우에 개체를 사용한다. 개체는 문서 자체에서 사용하기 위한 텍스트 개체인 일반 개체(General Entity)와 DTD의 선언 부분에서 사용하는 파라미터 개체(Parameter Entity)로 크게 나눌 수 있으며, 참조 객체가 위치한 장소에 따라 내부 개체(Internal Entity)와 외부 개체(External Entity)로 나눌 수 있다.

표기법(Notation)은 XML이 아닌 외부 소스로부터 오는 데이터가 필요할 경우 사용하는 것으로, 파서 이외의 다른 응용프로그램에게 처리를 넘겨주는데 필요하다.

주석문(Comment)은 특정 요소들과 마크업의 의미를 해석하는데 도움을 주는 것으로 '--'를 사용하여 정의한다. 요소 선언 형식의 내용 부분과 같은 CDATA 부분과 문서 맨 위 부분을 제외하고는 어디든지 사용할 수 있다.

처리 명령(Processing Instruction)은 XML 파서와 프로그램으로 전달되는 처리 방향과 정보를 유지하는 것으로서 '<?'로 시작되고 '?>'으로 끝을 표시해야 하고, XML 버전을 지정하는데 주로 사용된다.

CDATA 섹션은 XML 마크업 태그가 있는 문서를 파싱하지 않고 축적하는 경우에 주로 사용한다. 따라서 CDATA 섹션 안에 있는 모든 것이 파싱의 결과로 오류를 발생시키지 않는 방법을 제공하며, '<![CDATA['로 시작되고 ']]>'으로 끝나게 된다. 이 섹션 안에 있는 내용은,

DTD규칙에 의거하여 파서가 전혀 해석하지 말라는 명령을 내리는 것과 같다.

2.2 선행연구 개관

XML에 관한 연구는 1990년대 후반부터 본격적으로 시작되었다. 국내의 현재까지의 XML에 관한 연구를 XML의 특징을 SGML, HTML 등과 같은 다른 마크업 언어와 비교, 분석한 연구(김현희, 장혜원 1999), 전자 상거래, 디지털 도서관 등 특정 분야에서의 XML의 활용방안에 관한 연구(강지훈, 맹성현, 이만호 1999, 49-60), XML문서를 데이터 베이스에 축적하는 방법에 관한 연구(신은숙 1999; 류진영 1999) 및 XML문서를 검색할 수 있는 엔진과 XML 브라우저를 구현하는 연구(이은주 1999; 송종철 1998) 등으로 구분해 볼 수 있다.

그외에 본 연구와 관련이 있는 특정 유형의 문서구조를 분석하여 DTD를 설계한 연구를 들 수 있다. 이러한 유형의 연구로는 시집의 문헌분석을 통한 시 문서의 XML DTD를 개발한 연구(손지혜 1999), 연구보고서용 XML DTD를 개발한 연구(우항준 1999) 및 인터넷 악보 표현을 위한 XML DTD를 설계한 연구(채진석 1999, 37-52) 등이 있다.

외국에서의 연구도 국내의 연구동향과 크게 다르지는 않다. 외국의 선행연구들을 유형별로 보면, XML의 특성을 다른 마크업 언어와의 연관성 및 차이점 등을 비교한 연구(Sharpe 1999; Randall 1999), 출판사나 특정 기관에서 발행하고 있는 문헌의 유형에 맞게 DTD를 설계한 연구(Rath 1998; Lie and Sarela 1999), 웹 출판시 XML의 적용성에 관한 연구(Abiteboul 1999, 1-9) 등이 있다.

상기와 같이 다양한 유형의 문서에 대한 XML DTD 설계를 위한 연구가 진행되고 있으나 연구자들에게 매우 중요한 자료인 학위논문에 대한 XML DTD를 설계한 연구는 국내는 물론 외국에서도 찾아 볼 수 없었다. 또한 필자가 조사한 바로는 학위논문의 XML DTD를 설계함에 있어서 중요한 자료가 될 수 있는 학위논문 작성과 관련한 지침서가 국내외에서 개발되어 있지 않아, 국내외의 학위논문에 공히 적용할 수 있는 DTD를 설계하기에는 한계가 있었다.

3 연구의 방법

3.1 연구의 설계

본 연구에서는 XML DTD를 설계하기 위해서 석 박사 학위논문의 논리적인 구조분석을 수행한다. 문서 내의 어떠한 정보가 마크업으로 표현되어야 하는지를 정의하여 문서의 구성요소를 파악하고, 문서의 전체적인 계층구조상에서 요소들과 속성의 구축, 하위 계층의 요소와 속성 구축, 그리고 요소들간의 관계 정의 등을 수행한다. 또한 구조검색을 강화시키기 위해 논문의 주요 부분인 연구의 목적을 비롯하여 연구가설, 연구결과, 연구방법, 그리고 통계방법 등을 메타데이터 요소로서 별도로 이용할 수 있도록 요소를 추출한다. 이상과 같은 과정을 통해 학위논문을 위한 XML DTD를 설계한다.

3.2 XML DTD 설계

XML DTD를 기반으로 문서를 생성하기 위해서는 특정 문헌의 유형을 분석하는 것이 선행

되어야 한다. 따라서 본 연구에서는 문헌의 논리적인 구조를 분석하기 위해 석·박사 학위논문을 선정한다. 선정기준은 1999년까지의 박사학위 수여자의 수를 기준으로 상위 10위까지의 학교 중 서울 소재 6개 대학교를 대상으로 한다. 그 결과 표본으로 선정된 학교는 고려대학교, 서울대학교, 성균관대학교, 연세대학교, 중앙대학교, 한양대학교이다(교육부 1999). 선정된 각 대학교의 이공계열과 인문사회계열에서 1999년에 발행된 석사학위논문 1편, 박사학위논문 1편을 추출한 전체 24종과 각 대학교의 학위논문 규정을 참조하여 문헌의 논리구조를 분석하여 구성요소를 추출한다. 그리고 각 구성요소들의 요소를 정의하고 요소들간의 순서, 계층구조, 발생 빈도수 (부록1 참조) 등을 결정하여 이들 상호간의 관계를 표현한다. 또한 학위논문의 핵심적인 내용을 담고있는 연구의 목적, 가설, 방법, 결과, 통계방법을 요소로 추출한다 그리고 이를 기반으로 속성과 개체 등을 정의하여 DTD를 설계한다.

국내 각 대학의 학위논문은 매우 다양한 형태와 논리구조를 취하고 있으며, 특히 학문분야에 따라 논리구조가 매우 상이하다. 이러한 문제점을 극복하기 위해 상기와 같이 여러 대학과 학문분야의 학위논문을 대상으로 하여 분석·설계하였으나, 국내에서 발표되는 모든 학위논문에 일반화시키기에는 한계가 있을 것이다.

4 XML기반의 학위논문 DTD 설계

4.1 학위논문의 구조 분석

학위논문의 구조를 분석하는 것은 학위논문의 유형을 파악하여 논리적 구조를 추출해 내는 단계라고 할 수 있다. 학위논문은 크게 논문의 표제 면을 시작으로 서론에 들어가기 전까지의 논문 앞부분, 논문의 본 내용이 있는 서론을 시작으로 결론까지의 본문, 그리고 본문 다음부터 논문의 마지막까지를 포함하는 논문 뒷부분의 3부분으로 구성되어 있다. 일반적으로 논문 앞부분에는 표제면, 제출면, 인준면, 목차, 약기호면 등이 나타나고, 본문은 장, 절, 항, 목 및 단락 등으로 구성되며, 논문 뒷부분은 참고문헌, 부록 등이 있다. 초록의 유형 중 국문초록은 대개 논문의 앞부분에 나타나고 영문초록은 논문의 뒷부분에 나타났기 때문에 논문 앞부분과 뒷부분 모두의 구성요소로 보았다. 그리고 감사문은 논문의 앞부분보다는 뒷부분에서 더 많이 포함하고 있기 때문에 논문 뒷부분의 구성요소로 정의하였다. 또한 표, 그림, 수식, 각주 등은 논문 전체에서 유동적으로 나타났다. 〈표 1〉은 표본으로 선정한 각 대학교 학위논문의 논리적 구조와 그 출현순서를 나타낸 것이다.

학위논문의 본문은 주제에 따라 차이는 있지만 대개 연구의 목적, 연구 또는 실험 방법과 절차, 자료 또는 실험의 분석, 연구의 결과 등의 내용이 있다. 본 연구에서는 사회과학분야에 제한을 두어 시범적으로 이 부분을 메타데이터 요소로 추출하였다.

4.1.1 논문 앞부분

논문 앞부분은 논문의 표제면, 제출면, 인준면, 초록, 약기호면, 목차 등으로 구성된다. 표제면, 제출면, 인준면은 모든 논문에서 동일한 순서로 나타나고, 이 3부분은 학위구분, 논문의 제목, 저자명, 학교명, 학위수여년도 등 논문에 관한 주요 정보를 가지고 있다. 그러나 논문의 특성이나 각 학교의 논문 제출 형식의 차이에

〈표 1〉각 대학교의 학위논문 구조 및 출현순서

구조 \ 학교	고려대	서울대	성균관대	연세대	중앙대	한양대
논문앞부분	표제면 제출면 인준면 감사문 목차 영문초록 국문초록	표제면 제출면 인준면 국문초록 목차	표제면 제출면 인준면 목차	표제면 제출면 인준면 감사문 목차 국문초록	표제면 제출면 인준면 국문초록 목차 약기호면	표제면 제출면 인준면 국문초록 목차 약기호면
본 문	장 절 항 단락	장 절 항 목 단락	장 절 항 목 단락	장 절 항 단락	장 절 항 목 단락	장 절 항 단락
논문뒷부분	참고문헌 감사문	참고문헌 영문초록 감사문	참고문헌 부록 국문초록 영문초록	참고문헌 영문초록	참고문헌 부록 영문초록 감사문	참고문헌 부록 영문초록 감사문

따라 초록, 목차. 약기호면 등은 임의의 순서로 나타나며, 논문에 나타나지 않는 경우도 있다. 표제면, 제출면, 인준면은 외형적으로는 분리되어 있지만 공통적으로 포함하고 있는 내용들이 많다. 따라서 요소들의 중복을 피하기 위해 학위구분, 논문의 제목, 저자명, 학교명, 학과명, 전공명, 제출문, 인준문, 날짜 등을 독립적인 문헌요소로 추출하였다.

초록은 국문초록과 영문초록이 있으며, 논문의 앞부분과 뒷부분에 모두 출현하는 경우도 있지만 둘 중 하나만 나타나는 경우도 있다. 초록은 크게 초록의 제목과 초록의 내용이 서술되어 있다. 또한 초록의 본문에 들어가기 전에 논문의 제목, 부제목, 저자명, 학교명, 전공명, 학과명. 지도교수명 등이 유동적으로 보이지만, 표제면, 제출면, 인준면에서 공통적으로 포함되는 요소임으로 문헌요소로는 추출하지 않았다. 초록은 본문의 내용을 간략하게 나타내고 있으며 논문의 주요어를 나타내는 키워드는 각 학교마다 초록에 포함하는 경우도 있고 그렇지 않은 경우도 있다.

목차는 내용목차를 비롯하여 표목차, 그림목차, 부록목차 등이 있다. 이는 학위논문의 내용적 구성에 따라 여러 목차가 함께 나오는 경우도 있다. 목차의 구성은 목차를 나타내는 목차제목과 학위논문의 구성을 알 수 있는 내용들로 이루어져 있다.

약기호면은 약기호면을 나타내는 제목, 약어 및 기호와 그에 대응하는 설명이나 뜻으로 구성되어 있고, 감사문은 이를 나타내는 제목과 내용으로 이루어져 있다.

4.1.2 본문

학위논문의 내용을 이루는 본문은 일반적으로 서론을 시작으로 구체적인 연구의 내용, 종합적인 결론으로 이루어져 있다. 위의 내용을 담고있는 본문 전체는 장, 절, 항, 목 등이 반복적으로 나오며 다른 하위요소가 포함될 수 있는 구조로 되어있다. 각 장은 학위논문의 가장 상위의 구성요소로서 한 번 이상 보이며, 각 장의 제목, 장의 내용, 절로 구성된다. 절도 절의 제목, 내용, 항 등으로 구성되어 있다. 장은 한 번 이상 발생되고 특정 장은 여러 절을 포함하기도 하지만 절을 포함하지 않는 경우도 있다. 절도 마찬가지로 여러 항을 포함하는 경우도 있지만 그렇지 않은 경우도 있다.

장, 절, 항, 목 등에는 문자열, 각주, 수식, 표, 그림 등이 있다. 각주는 각주번호와 각주의 내용으로 이루어져 있고, 수식은 여러 수식이 하나의 제목 아래에 보이는 경우 또는 수식만 나오는 경우가 있었다. 표는 표의 내용과 표 제목, 표의 내용을 부가적으로 설명해주는 주석 등이 포함된다. 그림도 표와 유사한 구조로, 그림의 제목과 그림을 부가적으로 설명하는 주석 등으로 이루어져 있다.

4.1.3 논문 뒷부분

논문 뒷부분에는 참고문헌이 나오고 부록, 초록, 감사문 등은 논문에 따라 나오지 않는 경우도 있다. 참고문헌은 자료의 주요 정보가 되는 서명과 저자명을 비롯하여 출판사, 출판년도, 페이지수 등이 나온다. 연속간행물인 경우에는 저널명, 권, 호 등의 서지정보를 포함하며, 온라인 자료인 경우에는 URL를 구성요소로 한다. 부록은 크게 부록의 제목과 내용 등으로 구성되어 있고 내용은 문자열뿐만 아니라 표, 그림 등이 같이 구성되어 있다.

4.2 학위논문의 XML DTD 설계

학위논문의 XML DTD는 6개 학교의 이공계열과 인문사회계열에서 각각 석사학위논문 1편, 박사학위논문 1편을 뽑아 전체 24종의 문헌구조 분석을 토대로 요소와 속성, 개체 등을 추출하여 설계되었다. 각각의 요소들 중 서로 관련된 요소들을 모아 그룹화 하여 그들의 상위요소의 하위요소로 정의하였다. 그리고 논문의 주축이 되는 연구의 목적, 가설, 연구방법, 통계방법, 연구의 종합적인 결과와 사회과학분야의 자료분석에 많이 활용되는 t-Test와 상관관계 분석을 메타 데이터의 요소와 속성으로 정의함으로서 논문의 특정 내용부분이나 변인 등의 구조검색시 도움이 될 수 있도록 설계하였다.

〈표 2〉는 XML 문서 선언을 정의한 내용이다. XML 1.0 표준 규약에 따르면 모든 XML 문서는 '〈?xml'로 시작한다. 따라서 DTD의 맨 처음에 '〈?xml'을 선언하였고 속성으로 버전정보와 문자 인코딩 방식을 정의하였다.

〈표 3〉은 'Thesis'의 DTD를 나타낸다. 전체 학위논문을 나타내는 요소는 'Thesis'로 정의하고 이것의 하위요소는 논문 앞부분을 나타내는

〈표 2〉 XML 문서 선언

```
〈?xml version = "1.0" encoding = "euc-kr" ?〉
```

〈표 3〉 'Thesis'의 DTD

```
〈!ELEMENT Thesis (Front,Body,Back) 〉
〈!ATTLIST Thesis
        type (thesis|dissertation) "thesis"
        lang CDATA #REQUIRED
        year CDATA #REQUIRED 〉
```

〈표 4〉 연구결과의 DTD

```
〈!ELEMENT purpose (#PCDATA)* 〉
〈!ELEMENT hypothesis (#PCDATA)* 〉
〈!ELEMENT results (#PCDATA)* 〉
〈!ELEMENT methodology (#PCDATA|t|r)* 〉
〈!ELEMENT t (v1,v2,dv+,siglevel,tvalue)+ 〉
〈!ELEMENT v1 (#PCDATA) 〉
〈!ELEMENT v2 (#PCDATA) 〉
〈!ELEMENT dv (#PCDATA) 〉
〈!ELEMENT siglevel (#PCDATA) 〉
〈!ELEMENT tvalue (#PCDATA) 〉
〈!ELEMENT r (v1,v2,siglevel?,rvalue)+ 〉
〈!ELEMENT rvalue (#PCDATA) 〉
〈!ELEMENT statistics (#PCDATA)* 〉
```

'Front' 요소, 본문을 나타내는 'Body' 요소, 논문 뒷부분을 나타내는 'Back' 요소로 정의하였다. 'Thesis'의 하위요소들은 순서대로 한 번씩만 발생하기 때문에 발생지시자는 정의하지 않았다. 'Front', 'Body', 'Back' 요소는 '( )' 내에 정의하여 그룹화하고 순서대로 출현하기 때문에 연결자 ','를 부여하였다.

학위논문은 박사 및 석사학위논문이 있기 때문에 논문의 유형을 구분해 주었고, 논문을 기술한 언어를 나타낼 수 있도록 하였다. 학위논문에 기술되는 날짜는 동일한 학기라도 각 대학교의 논문규정과 심사기간 등에 따라 다르게 나타나므로 통일된 연도가 필요하다. 따라서 'Thesis' 요소의 속성으로 정의한 것은 논문의 유형과 작성된 언어, 학위수여년도이다.

학위논문의 유형을 구분해 주는 속성 이름은 'type'으로 정의하였다. 학위논문은 두 종류가 있으므로 속성 'type'을 'thesis'와 'dissertation' 나열값으로 정의하여 이 두 값 중 하나를 선택할 수 있도록 하였다. 그리고 석사학위논문이 박사학위논문보다 수적으로 더 많기 때문에 기본값은 'thesis'로 정의하였다. 논문을 기술한 언어를 나타내는 속성명은 'lang'이고 이 속성의 유형은 'CDATA', 기본값은 '#REQUIRED'로 하여 사용된 언어의 종류를 필수적으로 기술할 수 있도록 하였다. 이용자의 검색에 있어서 주요 선택사항인 날짜는 각 학교의 기준에 따라 다르다. 따라서 이를 해결하기 위해 학위수여년

도를 'year' 속성으로 하여 일관성 있는 연도를 나타낼 수 있도록 정의하였다.

연구의 목적, 가설, 연구결과, 연구방법, 그리고 통계방법을 서술한 부분을 요소로 정의하여 구조검색을 할 때 도움이 되고자 〈표 4〉와 같이 'purpose', 'hypothesis', 'results', 'methodology', 'statistics'를 요소로 정의하였다. 'methodology'는 자료의 분석방법 중 t-Test를 의미하는 't'와 상관관계분석을 나타내는 'r'을 하위요소로 하여 DTD를 설계하였다.

4.2.1 파라미터 개체 선언

파라미터 개체는 DTD 내에서만 사용되는 개체로 반복적으로 발생하는 DTD를 참조 구분자 '%'와 ';'을 사용하여 정의한다. 〈표 5〉는 DTD에서 파라미터 개체로 선언한 것을 나타내며 모두 4가지이다. 'data'는 표그룹을 나타내는 'tblgrp', 그림그룹을 나타내는 'figgrp', 수식그룹을 의미하는 'formulagrp' 요소들이 연결자 '|'로 그룹화 되어있는 내용모델을 정의하는 부분에 사용되는 개체이다. 'pr'는 '#PCDATA', 파라미터 개체인 '%data;'와 각주 요소인 'footnote'를 연결자 '|'로 그룹화 하여 파라미터 개체로 선언한 것이다. 'language'은 특정 요소의 언어 속성을 정의할 때 사용할 수 있는 개체이며, 'insert' 개체는 그림파일 등의 이진 데이터를 문서안에 포함시켜 주기 위해 선언해야 할 속성을 대신하여 사용할 수 있는 개체이다.

4.2.2 학위논문의 앞부분

논문 앞부분의 요소는 표제면, 제출면, 인준면의 구성요소를 모두 포함한 'info', 초록을 나타내는 'abstract', 목차인 'toc', 약기호면을 의미하는 'itemize' 요소로 구성된다. 문헌분석 결과, 'toc'와 'itemize'는 논문 앞부분에 보여짐으로 'front'의 하위요소로만 정의하였다. 국문초록은 대개 논문 앞부분, 영문초록은 논문 뒷부분에 포함되어 있기 때문에 'abstract' 요소는 논문 앞부분과 뒷부분에 모두 나올 수 있도록 정의하였다.

각 대학교 또는 각 논문마다 'abstract', 'toc', 'itemize' 요소 등은 임의의 순서로 나타나며, 나타나지 않는 요소도 있다. 따라서 〈표 6〉과 같이 'toc', 'abstract', 'itemize' 요소를 연결자 '|'로 그룹화하여 발생지시자 '*'로 정

〈표 5〉 파라미터 개체 선언

```
〈!ENTITY % data 'tblgrp|figgrp|formulagrp' 〉
〈!ENTITY % pr '#PCDATA|%data;|footnote' 〉
〈!ENTITY % language 'lang CDATA #REQUIRED' 〉
〈!ENTITY % insert 'name ENTITY IMPLIED' 〉
```

〈표 6〉 논문 앞부분의 DTD

```
〈!ELEMENT Front (info,(toc|abstract|itemize)*) 〉
```

의하여 임의의 순서로 발생하거나 발생하지 않아도 되도록 DTD를 설계하였다.

4.2.2.1 논문의 표제면, 제출면, 인준면

논문의 표제면, 제출면, 인준면은 논문의 주요정보를 담고 있는 요소로서, 학위구분, 논문의 제목, 부제목, 지도교수명, 논문의 저자명, 학교명, 학과명, 전공명, 날짜, 제출문, 인준문, 논문 심사위원 등으로 구성되어 있다. 따라서 〈표 7〉과 같이 이 부분을 'info' 요소로 정의하고 그 하위요소는 학위구분을 나타내는 'degree', 논문이 기술된 언어와 동일한 언어의 제목과 부제목, 다른 언어로 표현된 제목과 부제목을 그룹화한 제목그룹인 'thesistitlegrp' 요소, 논문의 저자명과 지도교수명, 저자의 소속 등을 알 수 있는 저자그룹인 'authgrp' 요소, 날짜를 표현하는 'date' 요소, 제출문을 표현하는 'submitstate' 요소, 인준문을 나타내는 'admitstate' 요소, 논문 심사위원을 의미하는 'approver' 요소로 정의하였다. 학위구분 요소는 학위수여회수를 나타내는 'frequency'를 속성명으로 지정하고 필요에 따라서 정의할 수 있도록 기본값을 '#IMPLIED'로 설계하였다.

학위논문의 제목은 논문을 작성한 언어로 쓰여진 제목과 부제목, 다른 언어로 된 제목과 부제목으로 나타난다. 〈표 8〉은 이 제목들을 하나로 모은 논문제목그룹의 DTD이다. 제목그룹인 'thesistitlegrp'의 하위요소는 논문에 필수적으로 나오는 제목인 'title', 유동적으로 나타나는 부제목인 'subtitle', 논문에 기술된 언어와 다른 언어로 표현한 제목인 'altitle'과 부제목인 'alsubtitle' 요소이다. 'title' 다음에 'subtitle', 'altitle', 'alsubtitle'이 유동적으로 출현함으로 발생지시자 ','로 정의하였고, 'subtitle', 'altitle', 'alsubtitle'은 그룹화하여 연결자 '|'로 정의하고 임의의 순서로 나올 수 있도록 발생지시자 '*'으로 설계하였다. 또한 이 요소들이 문헌 인스턴스를 삽입할 수 있도록 '#PCDATA'로 정의하였다.

논문 저자의 지도교수와 저자명, 논문 저자의 소속을 모아 저자그룹화 하였다. 〈표 9〉는 저자그룹의 DTD이다. 이 저자그룹을 의미하는 'authgrp' 요소의 하위요소는 지도교수를 의미하는 'supervisor' 요소, 논문의 저자를 나타내는 'author' 요소, 저자의 소속을 표현하는 'org' 요소이다. 이 요소들은 '|' 연결자로 그룹

〈표 7〉 'info'의 DTD

```
<!ELEMENT info
(degree|thesistitlegrp|authgrp|date|submitstate|admitstate|approver)*>
<!ELEMENT degree (#PCDATA) >
<!ATTLIST degree frequency CDATA #IMPLIED >
<!ELEMENT thesistitlegrp (title,(subtitle|altitle|alsubtitle)*) >
<!ELEMENT authgrp (supervisor|author|org)+ >
<!ELEMENT date (#PCDATA) >
<!ELEMENT submitstate (#PCDATA) >
<!ELEMENT admitstate (#PCDATA) >
<!ELEMENT approver (#PCDATA) >
```

〈표 8〉 논문제목그룹의 DTD

```
〈!ELEMENT thesistitlegrp  (title,(subtitle|altitle|alsubtitle)*) 〉
〈!ELEMENT title        (#PCDATA|footnote)* 〉
〈!ELEMENT subtitle      (#PCDATA) 〉
〈!ELEMENT altitle       (#PCDATA) 〉
〈!ELEMENT alsubtitle    (#PCDATA) 〉
```

〈표 9〉 저자그룹의 DTD

```
〈!ELEMENT authgrp    (supervisor|author|org)* 〉
〈!ELEMENT supervisor (#PCDATA) 〉
〈!ELEMENT author   (#PCDATA) 〉
〈!ELEMENT org      (school|dept|major)* 〉
〈!ELEMENT school   (#PCDATA) 〉
〈!ELEMENT dept     (#PCDATA) 〉
〈!ELEMENT major    (#PCDATA) 〉
```

화 되어 있고 순서에 상관없이 임의로 출현할 수 있도록 정의하였다.

'supervisor' 요소와 'author' 요소는 인스턴스를 삽입할 수 있도록 '#PCDATA'로 정의하였다. 저자의 소속을 나타내는 요소인 'org'는 학교명 'school' 요소, 학과명 'dept' 요소, 그리고 전공명을 나타내는 'major' 요소를 하위요소로 포함하도록 설계하였다. 이 요소들도 열거된 순서에 상관없이 임의의 순서대로 나올 수 있도록 발생지시자 '*'로 정의하였다. 'school', 'dept', 'major' 요소들은 문자 데이터를 삽입할 수 있도록 '#PCDATA'로 정의하였다.

4.2.2.2 목차

목차는 논문의 구성을 개괄적으로 볼 수 있는 부분으로 내용, 표, 그림, 부록 등의 제목이 기술된다. 〈표 10〉은 목차의 DTD를 나타낸다. 모든 목차를 나타내는 요소는 'toc'이고, 하위요소는 내용목차인 'contents', 표목차인 'tables', 그림목차인 'figures', 부록목차인 'appendices'이다. 내용목차는 논문에 항상 한 번 나오며, 표목차, 그림목차, 부록목차 등은 유동적으로 보인다. 따라서 'tables', 'figures', 'appendices'를 그룹화 하여 연결자 '|'와 발생지시자 '*'로 정의하였다. 'contents' 요소의 하위요소인 'title' 요소는 '차례' 등과 같은 내용목차의 제목을 표현하기 위해 사용되고, 'partcont' 요소는 'Ⅰ.서론', '3.1연구의 설계' 등과 같은 논문의 본문을 이루는 제목을 제외한 '국문초록', '참고문헌', 등의 제목을 위한 것이다. 'chapcont'는 본문의 장을 기술할 때 사용되는 요소이고, 'seccont' 요소는 장 다음의 절부터 항, 목 등을 위한 것이다. 'partcont', 'chapcont', 'seccont' 요소는 인스턴스를 삽입할 수 있도록 '#PCDATA'로 정의하였다. 또한 본문의 'title' 요소의 'ID'로 지정된 곳을 참조하여 본문으로 이동할 수 있고 목차에 기술되는 '페이지수'의 역할을 할 수 있도록 'partcont', 'chapcont', 'seccont' 요소

〈표 10〉 목차의 DTD

```
〈!ELEMENT toc (contents,(tables|figures|appendices)*) 〉
〈!ELEMENT contents (title,(partcont|chapcont|seccont)*) 〉
〈!ELEMENT partcont (#PCDATA) 〉
〈!ATTLIST partcont part_id IDREF #IMPLIED 〉
〈!ELEMENT chapcont (#PCDATA) 〉
〈!ATTLIST chapcont chap_id IDREF #IMPLIED 〉
〈!ELEMENT seccont (#PCDATA) 〉
〈!ATTLIST seccont sec_id IDREF #IMPLIED 〉
〈!ELEMENT tables (title,partcont+) 〉
〈!ELEMENT figures (title,partcont+) 〉
〈!ELEMENT appendices (title,partcont+) 〉
```

〈표 11〉 초록의 DTD

```
〈!ELEMENT abstract (p|title|keywords)* 〉
〈!ATTLIST abstract %language;
                   variable (inclusion|exclusion) #IMPLIED 〉
〈!ELEMENT p (%pr;)* 〉
〈!ELEMENT keywords (#PCDATA) 〉
```

에 'IDREF' 속성을 정의하였다. 'tables', 'figures', 'appendices' 요소의 하위요소인 'title'은 목차의 제목을 위해 사용되며 'partcont' 요소는 논문에 나오는 표제목, 그림 제목, 부록제목을 기술하기 위한 것이다.

4.2.2.3 초록

일반적으로 초록은 본문의 내용을 개략적으로 서술하며, 키워드가 포함되어 있는 경우도 있다. 〈표 11〉은 초록의 DTD를 보여준다. 초록을 나타내는 'abstract' 요소는 'p' 요소, 'title' 요소와 키워드를 나타내는 'keywords' 요소를 하위요소로 포함하였다. 또한 초록은 기술된 언어에 따라 유형이 구분되기 때문에 이를 구별할 수 있도록 파라미터 개체 'language'를 이용하여 정의하였고, 초록부분에서 변인에 관한 내용이 있는지의 여부를 검색하고자 하는 이용자를 위해 'variable' 속성을 정의하였다.

4.2.2.4 약기호면

약기호면은 약어나 기호가 나오고 그에 따른 설명으로 되어있다. 〈표 12〉는 약기호면의 DTD를 나타낸다. 약기호면 요소명인 'itemize'는 하위요소로 제목을 나타내는 'title', 약어나 기호를 의미하는 'term', 설명을 나타내는 'description' 요소로 구성하였다. 'title' 요소는 한 번 발생하고 나머지 'term', 'description' 요소는 반복해서 발생할 수 있도록 발생지시자 '*'로 정의하였다.

4.2.3 학위논문의 본문

학위논문 본문의 구조는 장, 절, 항, 목 등으

〈표 12〉 약기호면의 DTD

```
〈!ELEMENT itemize (title,(term|description)*) 〉
〈!ELEMENT term (#PCDATA) 〉
〈!ELEMENT description (#PCDATA) 〉
```

로 구성되어 있으며 각 장은 한 번 이상 발생되고, 다시 장의 제목, 장의 내용, 절로 구성된다. 절도 다시 절의 제목, 내용, 항으로 구성되는, 계속 반복되는 구조로 되어 있다. 표, 그림, 각주 등은 문자열과 같이 본문에 반복적으로 나타난다.

〈표 13〉은 본문의 DTD를 나타낸다. 이 본문을 나타내는 요소 'Body'는 'Chapter' 요소가 한 번 이상 계속 발생하는 내용모델로 정의하였다. 'Chapter' 요소의 하위요소는 'title', 'p', 그리고 장에 포함되는 절, 항, 목 등을 나타내는 'Section' 요소로 구성하였다. 'title' 요소에 '#PCDATA'와 'footnote'를 같이 선 언하여

〈표 13〉 본문의 DTD

```
〈!ELEMENT Body (Chapter)+ 〉
〈!ELEMENT Chapter (title|p|Section)* 〉
〈!ELEMENT title (#PCDATA|footnote)* 〉
〈!ATTLIST title
        body_id ID #IMPLIED 〉
〈!ELEMENT Section (title|p|Section)+ 〉
〈!ATTLIST Section
        sec_order CDATA #IMPLIED 〉
〈!ELEMENT p (%pr;)* 〉
〈!ELEMENT footnote (#PCDATA)* 〉
〈!ATTLIST footnote
        ref_id IDREF #IMPLIED 〉
〈!ELEMENT tblgrp (title|table|note)*〉
〈!ELEMENT table (#PCDATA) 〉
〈!ATTLIST table tbl_id ID #IMPLIED
            %insert; 〉
〈!ELEMENT note (#PCDATA)〉
〈!ELEMENT figgrp (title|fig|note)*〉
〈!ELEMENT fig (#PCDATA)〉
〈!ATTLIST fig fig_id ID #IMPLIED
            %insert; 〉
〈!ELEMENT formulagrp (title|formula|note)*〉
〈!ELEMENT formula (#PCDATA)〉
〈!ATTLIST formula formula_id ID #IMPLIED
            %insert; 〉
```

〈표 14〉 논문 뒷부분의 DTD

```
〈!ELEMENT Back (biblist|appendix|abstract|thanks)* 〉
〈!ELEMENT biblist (title,ref+)〉
〈!ELEMENT ref (title|name|otherinfo)*〉
〈!ELEMENT appendix (p|title)*〉
〈!ELEMENT thanks (title,p+) 〉
```

제목 인스턴스를 구축하고 각주를 달 수 있도록 하였고, 요소의 속성을 'ID'로 정의하여 위치를 지정해 주었다. 'footnote' 요소의 속성을 'IDREF'로 정의하여 참고문헌의 ID의 기본값을 참조함으로서 인용문헌의 서지정보를 볼 수 있는 참고문헌으로 이동할 수 있도록 설계하였다. 'title'의 하위요소인 '#PCDATA'와 'footnote'는 반복적으로 나타나지만 둘 중 한 요소만 나오는 경우도 있으므로 발생지시자 '*'로 선언하였다. 'Section' 요소도 'Chapter'와 같이 'title'과 'p'를 하위요소로 정의하고 다시 'Section' 요소로 선언하여 절, 항, 목 등이 반복될 수 있도록 하였다. 'Section'의 속성으로 절, 항, 목 등을 필요에 따라 구별하여 사용할 수 있도록 'sec_order'를 작성하였다.

본문에 반복적으로 등장하는 표, 그림, 수식 등은 이들을 나타내는 제목 요소인 'title', 그리고 'table', 'fig', 'formula' 요소와 이들을 부연 설명해주는 'note' 요소를 그룹화하여 반복적으로 나타나고 임의의 순서로 보여질 수 있도록 구성하였다. 실제 문서안에 표, 그림, 수식 등을 삽입하기 위해 'table', 'fig', 'formula' 요소들의 속성에 'ENTITY'를 정의하였다.

4.2.4 학위논문의 뒷부분

논문 뒷부분은 참고문헌, 부록, 초록, 감사문으로 구성되어 있다. 부록, 초록, 감사문은 유동적으로 보이며 초록은 대체로 영문초록이 나타난다. 〈표 14〉는 논문 뒷부분의 DTD이다. 논문의 뒷부분을 나타내는 'Back' 요소는 참고문헌인 'biblist' 요소, 부록인 'appendix' 요소, 초록인 'abstract' 요소, 그리고 감사문인 'thanks' 요소를 하위요소로 갖는다.

4.2.4.1 참고문헌

참고문헌은 자료의 주요정보인 서명과 저자명을 비롯하여 출판사, 출판년도, 저널명, 권, 호, 페이지수, 온라인 자료의 URL 등으로 이루어져 있다. 〈표 15〉는 참고문헌의 DTD를 나타낸다. 참고문헌을 의미하는 'biblist' 요소는 하위요소로 참고문헌의 제목을 표현하는 'title' 요소와 참고문헌의 서지정보를 생성하기 위해 정의된 'ref' 요소로 구성하였다. 이 요소의 하위요소는 제목인 'title' 요소가 자료를 의미하는 'ref' 요소보다 먼저 나오기 때문에 ',' 연결자로 정의하고, 참고문헌은 하나 이상의 자료들로 이루어졌기 때문에 'ref' 요소에 발생지시자를 '+'로 선언하였다. 이 'ref' 요소에 'bib_id'를 정의하여 각주의 'ref_id'가 참조할 수 있도록 정의하여 서지정보로 이동할 수 있도록 설계하였다.

서지정보의 제목은 'title', 저자명은 'name', 그 외 정보는 'otherinfo'의 하위요소로 그룹화 하였다. 'name'은 저자의 역할을 설명할 수 있도록 속성으로 정의하였다.

〈표 15〉 참고문헌의 DTD

```
〈!ELEMENT biblist (title,ref+)〉
〈!ELEMENT ref (title|name|otherinfo)*〉
〈!ATTLIST ref bib_id ID #IMPLIED〉
〈!ELEMENT name (#PCDATA)〉
〈!ATTLIST name role CDATA #IMPLIED〉
〈!ELEMENT otherinfo (pub|pubdate|edition|journal|vol|no|page|url)*〉
〈!ELEMENT pub (#PCDATA) 〉
〈!ELEMENT pubdate (#PCDATA) 〉
〈!ELEMENT edition (#PCDATA) 〉
〈!ELEMENT journal (#PCDATA) 〉
〈!ELEMENT vol (#PCDATA) 〉
〈!ELEMENT no (#PCDATA) 〉
〈!ELEMENT page (#PCDATA) 〉
〈!ELEMENT url (#PCDATA) 〉
```

'otherinfo'의 하위요소는 출판사를 나타내는 'pub', 출판년도인 'pubdate', 판사항인 'edition', 저널명인 'journal', 권과 호를 나타내는 'vol'과 'no' 요소, 문헌의 페이지수를 의미하는 'page', 온라인 자료의 위치인 'url'이다. 이 'otherinfo'의 하위요소들은 모두 나타나지 않을 수 있다. 따라서 연결자 '|'와 발생지시자 '*'으로 정의하였다.

4.2.4.2 부록

부록은 제목과 부록의 내용으로 이루어져있다. 〈표 16〉은 부록의 DTD를 보여준다. 부록 요소인 'appendix'는 'title'과 'p' 요소를 하위요소로 포함하고 있다. 'appendix' 요소는 속성으로 ID를 정의함으로서 'partcont' 요소의 IDREF가 참조하여 실제 부록의 내용으로 이동할 수 있도록 설계하였다.

4.2.4.3 감사문

감사문은 제목과 여러 단락으로 이루어져 있는 내용으로 구성된다. 〈표 17〉은 감사문의 DTD를 보여준다. 감사문 요소명은 'thanks'로

〈표 16〉 부록의 DTD

```
〈!ELEMENT appendix (title|p)*〉
〈!ATTLIST appendix app_id ID #IMPLIED〉
```

〈표 17〉 감사문의 DTD

```
〈!ELEMENT thanks (title, p+) 〉
```

정의하고 하위요소로 'title'과 'p'를 포함하였다. 'p' 요소는 한 번 이상 발생하기 때문에 발생지시자 중 '+'로 정의하였다.

이 연구에서 설계한 학위논문의 요소, 속성, 개체들에 대한 이해를 돕기 위해서 각 대학 학위논문을 보다 체계적으로 분석하여 〈부록 1〉을 작성하였다. 이 부록은 각 대학학위논문의 XML DTD 요소와 속성의 출현빈도 나타내며, 〈부록 2〉는 본 연구에서 설계한 학위논문의 XML DTD를 본문에서 분할해서 제시하고 설명한 것을 총괄적으로 한 곳에 모은 것이다.

5 결 론

XML은 HTML의 단순성의 한계와 SGML의 복잡함을 해결하기 위해 등장한 메타언어로서, 인터넷상에서 구조화된 문서를 전송 가능하도록 설계된 전자문서의 표준이다. 따라서 각 도서관들이 학문적 가치와 그 활용도가 높은 학위논문을 XML 기반으로 구축하고 이를 상호 이용한다면 공간활용, 자료보존 및 이용자 서비스에 있어서 질적으로 향상된 서비스를 제공할 수 있을 것이다.

이렇게 XML 기반의 학위논문을 생성하기 위해서는 DTD 설계가 우선적으로 수행되어야 한다. DTD는 문서들의 구조를 정의하는 것으로 XML 문서 작성시 핵심이 되는 부분이다. 그러나 아직까지는 학위논문을 위한 XML DTD에 관한 연구가 미흡한 실정이다.

이에 본 연구는 XML 기반의 학위논문을 생성하여 활용할 수 있도록 학위논문의 논리구조를 분석하여 이를 바탕으로 XML DTD를 설계하였다. DTD를 설계하기 위한 문헌분석 대상이 된 자료는 이공계열과 인문사회계열의 석·박사 학위논문 24종이었으며 문헌들의 계층적인 구조뿐만 아니라 이용자가 학위논문을 대상으로 구조 검색에 도움을 줄 수 있도록 시범적으로 논문의 핵심부분인 연구의 목적, 가설, 연구방법, 연구의 결과 등을 메타데이터 요소로 이용할 수 있도록 요소와 속성으로 정의하였다.

따라서 본 연구에서 설계한 DTD는 XML 기반의 학위논문을 구축하는데 활용할 수 있으며, 이렇게 구축된 학위논문은 효율적이고 체계적으로 축적, 관리 및 유통되고 이용자가 질적으로 향상된 구조검색을 수행하는데 도움을 줄 수 있을 것으로 기대된다.

참고문헌

강지훈, 맹성현, 이만호. 1999. 분산 디지털 도서관 시스템에서 XML을 이용한 가상문서의 표현과 처리. 『제2회 디지털도서관 컨퍼런스 논문집』, 49-60.

교육부. 1999. 『교육통계연보』. 서울: 교육부.

김현희, 장혜원. 디지털도서관 문서 양식으로서의 XML과 HTML의 특성 및 검색 기능 비교 연구. 『정보관리학회지』, 16(2): 105-134.

류 광. 1999. 『Professional XML Applications』. 서울: 정보문화사.

류진영. 1999 『관계형 데이터 베이스에서 XML 문서 저장 방안 연구 및 구현』. 석사학위논문, 부산대학교 대학원.

손지혜. 1999 『시문서의 XML DTD개발에 관한 연구』 석사학위논문, 한양대학교 교육대학원.

송종철. 1998. 『Well-formed Document 처리를 위한 XML 브라우저 설계 및 구현에 관한 연구』. 석사학위논문, 광운대학교 대학원.

신은숙. 1999. 『XML 기반문서 디지털화 기술에 관한 연구』. 석사학위논문, 단국대학교 산업대학원.

우항준. 1999. 『전자도서관을 위한 연구보고서용 XML DTD 개발』. 석사학위논문, 동국대학교 산업기술환경대학원.

이은주. 1999. 『다중 검색방식을 지원하는 XML 기반 재사용 라이브러리 시스 템』. 석사학위논문, 서울대학교 대학원.

채진석. 1999. 인터넷 악보 표현을 위한 XML DTD 설계. 『정보관리학회지』. 16(4): 37-52.

한국전산원. 1998. 『공공부문의 DTD 개발 지침』. 서울: 한국전산원.

Abiteboul. Serge. 1999. "On Views and XML." *Proceedings of the Eighteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. 1-9.

"Extnsible Markup Language 1.0 Recommendation10-February-1998. 〈http://www.w3.org/TR/REC-xml〉

"Frequently Asked Questions about the Extensible Markup Language: The XML FAQ." 〈http://xml.css.co kr/xml/xmlfaq_ko.html〉

Lie, Hakon W. and Jane Sarela. 1999. "Multipurpose Web Publishing using HTML, XML, and CSS." *Communication of ACM*, 42(10) : 95-101

Randall. Neil. 1999. "A Second Chance for Web Markup." 〈http://www.zdnet.com/pcmag/pctech/content/16/19/it1619.001.html〉

Rath, Hans H.. 1998. "XML: Chance and Challenge for Online Information Providers." 〈http://www.oasis-open.org/html/xml_chance_challenge.html〉

Sharpe. Bruce. 1999. "XML: It's Not Your Farher's HTML." 〈http://www.webreference.com/authoring/languages/XML/intro〉

Suzuki, Junichi and Yoshikazu Yamamoto. 1998. "Managing the software design documents with XML." *Proceedings on the Sixteenth Annual International Conference on Computer Documentation*, 127-136.

〈부록 1〉 학위논문의 XML DTD 요소와 출현빈도

범례/ ◎: 출현하지 않음, ○: 1회 출현, ● : 2회 이상 출현
?. 0회 또는 1회 출현, *: 0회 이상 출현, + 1회 이상 출현, 빈칸: 1회 출현

1. 학위논문의 앞부분

요소						학교	고려대	서울대	성균관대	연세대	중앙대	한양대	출현빈도
			degree				●	●	●	◎	○	●	*
Thesis	Front	info	thesistitlegrp	title			○	○	○	○	○	○	
				subtitle			○	◎	○	◎	○	◎	?
				altitle			○	○	○	◎	○	○	?
				alsubtitle			◎	◎	◎	◎	◎	◎	?
			authgrp	supervisor			○	○	○	◎	○	○	?
				author			○	○	○	○	○	○	
				org	school		○	○	○	○	○	○	
					dept		◎	○	○	○	○	○	?
					major		○	◎	○	◎	○	◎	?
			date				○	●	●	○	○	○	*
			submitstate				○	○	○	○	○	○	
			admitstate				○	○	○	○	○	○	
			approver				●	●	●	●	●	●	+
		toc	contents	title			●	●	●	●	●	●	+
				partcont			●	●	●	●	●	●	+
				chpcont			●	●	●	●	●	●	+
				seccont			●	●	●	●	●	●	+
			tables	title			○	○	○	○	○	○	
				partcont			●	●	●	●	●	●	+
			figures	title			○	○	○	◎	◎	○	?
				partcont			○	●	●	◎	◎	●	*
			appendices	title			◎	◎	○	◎	○	◎	?
				partcont			◎	◎	●	◎	○	◎	*
		abstract	p	tblgrp	title		◎	◎	◎	◎	◎	◎	?
					tbl		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				figures	title		◎	◎	◎	◎	◎	◎	?
					fig		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				formulargrp	title		◎	◎	◎	◎	◎	◎	?
					formular		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				footnote			◎	◎	◎	◎	◎	◎	?
				#PCDATA			●	●	◎	●	●	●	*
			title				○	○	◎	○	○	○	?
			keywords				◎	●	◎	●	◎	◎	*
		itemize	title				◎	○	◎	◎	◎	◎	?
			term				◎	●	◎	◎	◎	●	*
			description				◎	●	◎	◎	◎	●	*

2. 학위 논문의 본문

요소						학교	고려대	서울대	성균관대	연세대	중앙대	한양대	출현빈도
Thesis	Body	Chapter	title	#PCDATA			●	●	●	●	●	●	+
				footnote			◎	◎	◎	◎	◎	○	?
			p	tblgrp	title		◎	◎	◎	◎	◎	○	?
					tbl		◎	◎	◎	◎	◎	○	?
					note		◎	◎	◎	◎	◎	◎	?
				figgrp	title		◎	◎	◎	◎	◎	◎	?
					fig		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				formulargrp	title		◎	◎	◎	◎	◎	◎	?
					formular		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				footnote			○	◎	◎	◎	◎	◎	?
				#PCDATA			●	●	●	●	●	●	+
			Section	title			○	○	○	○	○	○	
				p	tblgrp	title	●	●	●	●	●	●	+
						tbl	●	●	●	●	●	●	+
						note	◎	◎	●	◎	◎	◎	*
					figgrp	title	◎	●	●	◎	◎	●	*
						fig	◎	●	●	◎	◎	●	*
						note	◎	●	◎	◎	◎	●	*
					formulargrp	title	◎	◎	◎	●	◎	●	*
						formular	◎	◎	◎	●	◎	●	*
						note	◎	◎	◎	●	◎	◎	*
					footnote		●	●	●	●	●	●	+
					#PCDATA		●	●	●	●	●	●	+
				Section			●	●	●	●	●	●	+

3. 학위 논문의 뒷부분

요소						학교	고려대	서울대	성균관대	인제대	중앙대	한양대	출현빈도
Thesis	Back	biblist	title				○	○	○	○	○	○	+
			ref	title			●	●	●	●	●	●	+
				name			●	●	●	●	●	●	+
				otherinfo	pub		●	●	●	●	●	●	+
					pubdate		●	●	●	●	●	●	+
					edition		●	●	●	○	○	●	+
					journal		●	●	●	●	●	●	+
					vol		●	●	●	●	●	●	+
					no		●	●	●	●	●	●	+
					page		●	●	●	●	●	●	+
					url		○	●	◎	●	◎	◎	*
		appendix	p	tblgrp	title		◎	○	●	◎	◎	◎	*
					tbl		◎	○	●	◎	◎	◎	*
					note		◎	○	◎	◎	◎	◎	?
				figgrp	title		◎	◎	◎	◎	◎	●	*
					fig		◎	◎	◎	◎	◎	●	*
					note		◎	◎	◎	◎	◎	◎	?
				formulargrp	title		◎	◎	◎	◎	◎	◎	?
					formular		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				footnote			◎	◎	◎	◎	◎	◎	?
				#PCDATA			◎	●	●	◎	◎	●	*
			title				◎	○	○	◎	◎	○	?
		abstract	p	tblgrp	title		◎	◎	◎	◎	◎	◎	?
					tbl		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				figgrp	title		◎	◎	◎	◎	◎	◎	?
					fig		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				formulargrp	title		◎	◎	◎	◎	◎	◎	?
					formular		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				footnote			◎	◎	◎	◎	◎	◎	?
				#PCDATA			◎	●	●	●	●	●	*
			title				◎	○	○	○	○	○	?
			keywords				◎	●	◎	●	◎	◎	*
		Thesis	p	tblgrp	title		◎	◎	◎	◎	◎	◎	?
					tbl		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				figgrp	title		◎	◎	◎	◎	◎	◎	?
					fig		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				formulargrp	title		◎	◎	◎	◎	◎	◎	?
					formular		◎	◎	◎	◎	◎	◎	?
					note		◎	◎	◎	◎	◎	◎	?
				footnote			◎	◎	◎	◎	◎	◎	?
				#PCDATA			●	●	●	◎	●	●	*
			title				○	○	○	◎	○	○	?

〈부록 2〉 학위논문의 XML DTD

```
〈?xml version="1.0" encoding="euc-kr"?〉

〈!-- XML Document Type Definition for Theses. --〉

〈!-- BASIC DOCUMENT ELEMENTS --〉

〈!ENTITY % data 'tblgrp|figgrp|formulagrp' 〉
〈!ENTITY % pr '#PCDATA|%data;|footnote' 〉

〈!ELEMENT tblgrp (title|table|note)*〉
〈!ELEMENT table    (#PCDATA) 〉
〈!ATTLIST table
          tbl_id  ID  #IMPLIED
          %insert; 〉
〈!ELEMENT note    (#PCDATA)〉
〈!ELEMENT figgrp (title|fig|note)*〉
〈!ELEMENT fig    (#PCDATA)〉
〈!ATTLIST fig
          fig_id  ID  #IMPLIED
          %insert; 〉
〈!ELEMENT formulagrp (title|formula|note)*〉
〈!ELEMENT formula    (#PCDATA)〉
〈!ATTLIST formula
          formula_id  ID  #IMPLIED
          %insert; 〉

〈!ELEMENT footnote  (#PCDATA)* 〉
〈!ATTLIST footnote
          ref_id IDREF #IMPLIED 〉

〈!-- METADATA ELEMENTS --〉
〈!ELEMENT purpose  (#PCDATA)* 〉
〈!ELEMENT hypothesis  (#PCDATA)* 〉
〈!ELEMENT results  (#PCDATA)* 〉
〈!ELEMENT methodology  (#PCDATA|t|r)* 〉
〈!ELEMENT t (v1,v2,dv+,siglevel,tvalue)+  〉
〈!ELEMENT v1 (#PCDATA) 〉
〈!ELEMENT v2 (#PCDATA) 〉
〈!ELEMENT dv (#PCDATA) 〉
〈!ELEMENT siglevel (#PCDATA) 〉
〈!ELEMENT tvalue (#PCDATA) 〉
〈!ELEMENT r (v1,v2,siglevel?,rvalue)+ 〉
〈!ELEMENT rvalue (#PCDATA) 〉
```

```
<!ELEMENT statistics (#PCDATA)* >

<!-- ATTRIBUTE DEFINITIONS -->

<!ENTITY % language 'lang CDATA #REQUIRED' >
<!ENTITY % insert 'name ENTITY #IMPLIED' >

<!-- THE DOCUMENT STRUCTURE -->

<!ELEMENT Thesis (Front,Body,Back) >
<!ATTLIST Thesis
        %lanaguage;
        year CDATA #REQUIRED
        type (thesis|dissertation) "thesis" >

<!-- FRONT ELEMENTS -->

<!ELEMENT Front (info,(toc|abstract|itemize)*) >
<!ELEMENT info (degree|thesistitlegrp|authgrp|date|submitstate|admitstate|approver)* >
<!ELEMENT degree (#PCDATA) >
<!ATTLIST degree
        frequency CDATA #IMPLIED >
<!ELEMENT thesistitlegrp (title,(subtitle|altitle|alsubtitle)*) >
<!ELEMENT subtitle (#PCDATA) >
<!ELEMENT altitle (#PCDATA) >
<!ELEMENT alsubtitle (#PCDATA) >
<!ELEMENT authgrp (supervisor|author|org)* >
<!ELEMENT supervisor (#PCDATA) >
<!ELEMENT author (#PCDATA) >
<!ELEMENT org (school|dept|major)* >
<!ELEMENT school (#PCDATA) >
<!ELEMENT dept (#PCDATA) >
<!ELEMENT major (#PCDATA) >
<!ELEMENT date (#PCDATA) >
<!ELEMENT submitstate (#PCDATA) >
<!ELEMENT admitstate (#PCDATA) >
<!ELEMENT approver (#PCDATA)* >
<!ELEMENT toc (contents,(table|figures|appendices)*) >
<!ELEMENT contents (title,(partcont|chapcont|seccont)*) >
<!ELEMENT partcont (#PCDATA) >
<!ATTLIST partcont part_id IDREF #IMPLIED >
<!ELEMENT chapcont (#PCDATA) >
<!ATTLIST chapcont chap_id IDREF #IMPLIED >
<!ELEMENT seccont (#PCDATA) >
<!ATTLIST seccont sec_id IDREF #IMPLIED >
<!ELEMENT table (title,partcont+) >
```

```
<!ELEMENT figures (title,partcont+) >
<!ELEMENT appendices (title,partcont+) >
<!ELEMENT abstract (title|p|keywords)* >
<!ATTLIST abstract
        %lanaguage;
        variable (inclusion|exclusion) #IMPLIED >
<!ELEMENT keywords (#PCDATA) >
<!ELEMENT itemize (title,(term|description)*) >
<!ELEMENT term (#PCDATA) >
<!ELEMENT description (#PCDATA) >

<!-- BODY ELEMENTS -->

<!ELEMENT Body (Chapter)+ >
<!ELEMENT Chapter (title|p|Section)* >
<!ELEMENT title (#PCDATA|footnote)* >
<!ATTLIST title
        body_id ID #IMPLIED >
<!ELEMENT Section (title|p|Section)* >
<!ATTLIST Section
        sec_order CDATA #IMPLIED >
<!ELEMENT p (%pr;)* >

<!--BACK ELEMENTS -->

<!ELEMENT Back (biblist|appendix|abstract|thanks)* >
<!ELEMENT biblist (title,ref+) >
<!ELEMENT ref (title|name|otherinfo)* >
<!ATTLIST ref
        bib_id ID #IMPLIED >
<!ELEMENT name (#PCDATA) >
<!ATTLIST name
        role CDATA #IMPLIED>
<!ELEMENT otherinfo (pub|pubdate|edition|journal|vol|no|page|url)* >
<!ELEMENT pub (#PCDATA) >
<!ELEMENT pubdate (#PCDATA) >
<!ELEMENT edition (#PCDATA) >
<!ELEMENT journal (#PCDATA) >
<!ELEMENT vol (#PCDATA) >
<!ELEMENT no (#PCDATA) >
<!ELEMENT page (#PCDATA) >
<!ELEMENT url (#PCDATA) >
<!ELEMENT appendix (p|title)*>
<!ATTLIST appendix
        app_id ID #IMPLIED >
<!ELEMENT thanks (title,p+) >
```