

전자책 표준 적합성 도구의 개발

Developing a Test Tool for Electronic Book Standard Conformance

김재경(Jae-Kyung Kim)*, 손원성(Won-Sung Sohn)*, 임순범(Soon-Bum Lim)**,
김성혁(Sung-Hyuk Kim)**, 최윤철(Yoon-Chul Choy)*

초 록

한국문서형표준(KS X 6100)은 지난 2001년 7월 제정된 한국전자책문서표준(EBKS1.0)이 2002년 7월 한국산업기술표준원의 산업표준심의를 거쳐 제정되었다. 그러므로 한국문서형표준에 기반하여 작성된 전자책 문서들은 표준의 목적과 명세를 준수하여야 하며 표준의 준수 여부를 판별하기위해서 적합성 검사가 요구된다. 본 논문에서는 이러한 적합성 검사의 기준이 되는 적합성 조항(Conformance Clause)의 정의와 이에 따른 적합성 검사 도구(Conformance Testing Tool)를 개발하였다. 따라서 본 논문에서 정의된 조항들을 통하여 전자책 제작자들은 전자책표준의 목적과 명세에 보다 적합한 전자책문서들을 제작할 수 있으며, 서비스 제공자들은 적합성 검사 도구를 통한 자동화된 적합성 검사 절차를 사용하여 효율적인 적합성 검사가 가능하다.

ABSTRACT

EBKS(Electronic Book of Korea Standard) 1.0 which was established in July 2001 passed the industry standard deliberation of agency for technology and standards of Korea industry in July 2002. EBKS is now established as KS X 6100 representing the Korea standard of electronic book of Korea. Therefore, Conformance test and certification procedure is required whether electronic books and documents which follow KS X 6100 are conform to specification of the standard. In the paper, we define conformance clause which is a basis for conformance test and standard electronic book schema. As a result, electronic book writers can produce more conformable documents to the specification and goal of electronic book standard by using the conformance clause, and service provider can effectively perform the conformance test by using standard electronic book schema

키워드 : 적합성, 적합성 검사 도구, 전자책 표준, EBKS, KS X 6100

Conformance, Conformance Testing Tool, Electronic Book Standard, EBKS, KS X 6100

본 연구보고서는 정보통신부 정보통신연구진흥원에서 지원하고 있는 정보통신기초연구지원사업의 연구결과임.

* 연세대학교 컴퓨터과학부

** 숙명여자대학교 정보과학부

1. 서 론

지난 2001년 7월 제정된 한국전자책문서표준 EBKS (Electronic Book of Korea Standard) 1.0[1][2]은 국내 전자책산업의 활성화를 목적으로 제정된 산업표준이다. EBKS 1.0 버전은 전자책 콘텐츠의 명확한 교환을 지원함으로써 업계간의 콘텐츠 공유를 가능케 하고 관련 소프트웨어 기술개발에 대한 중복투자를 최소화하는 것을 목적으로 하고 있다.

한편 전자책 업계 표준인 EBKS 1.0 버전은 지난 2002년 7월 한국산업기술 표준원의 산업표준심의를 거쳐 한국전자책 표준문서형정의, 규격번호 KS X 6100으로 제정된 바 있다. 그 결과 현재 국내 전자책 업계에서는 HTML PDF 및 기타 XML 형식을 이용한 전자책 형식을 비롯하여, KS X 6100을 따르는 전자책 콘텐츠 또한 빠르게 그 수가 증가하고 있다.

현재 전자책 표준인 KS X 6100은 지난 2002년 7월 제정이 되었으므로 앞으로 전자책 표준 문서에 대해 KS표준에 적합한지의 여부를 검사하는 인증 절차가 확립되어야 한다. 즉, 전자책 표준 문서가 표준을 준수하는 문서로써 가치를 갖기 위해서는 기본적으로 표준의 문법을 준수하여야 하며 나아가 전자책으로서 적합한 의미적인 구조와 내용으로 구성되었는지를 검사하는 인증 절차를 정의하고 적용하여야 한다.

이것은 전자책 콘텐츠가 KS X 6100 표준의 명세와 목적에 부합하는가를 판별하는 적합성 테스트에 의하여 수행될 수 있다. 전자책 시장에서 단일화된 KS X 6100 표준을 준

수하는 전자책 콘텐츠를 제작하고, KS X 6100 표준을 사용하는 업체나 사용자들 간에 표준을 준수하는 콘텐츠를 유통시키기 위해서 상호간에 자신들이 사용하고 있는 콘텐츠가 표준 규격에 적합한 것임을 확인할 수 있도록 적합성 테스트를 통하여 인증서를 부여할 필요가 있다.

즉, 이와 같은 적합성 테스트는 EBKS의 KS X 6100 표준 확정으로 전자책 콘텐츠에 대한 표준 규격(명세)의 준수 여부에 대한 인증과 신뢰성 있는 디지털라이브러리의 구축 및 활용과 콘텐츠의 효율적인 유통 및 서비스를 위해서 요구되어 진다.

따라서 KS X 6100 표준의 제정으로 표준을 준수하는 신뢰성 있는 콘텐츠의 제작 및 유통을 위해서는 적합성 테스트를 통한 인증 절차의 확립이 반드시 필요하다.

또한 적합성 테스트를 통하여 신뢰성 있는 콘텐츠를 유통시킬 수 있을 뿐만 아니라 전자책 레지스트리 혹은 라이브러리 구축 시에도 효과적인 납본 및 활용이 가능하다. 예를 들어 전자책은 그 특성상 소설이나 시와 같이 여러 분류로 나누어 질 수 있으며 전자책 문서의 구조는 이러한 분류에 알맞은 구조로 표현되어야 한다. 현재 KS X 6100 표준에는 여러 분류의 문서 구조를 모두 표현 할 수 있도록 DTD가 정의되어 있으므로 전자책 문서가 그 분류에 따른 올바른 구조를 사용했는지의 여부를 적합성 테스트를 통하여 판별하여야 한다. 각 분류에 따라 적합한 구조로 정의된 문서는 레지스트리에서 효율적인 관리 및 활용이 가능하여 진다.

이와 같은 절차 및 활용은 전자책 업체나

사용자, 그리고 신뢰성 있는 레지스트리 구축을 위해서 적합성 인증 평가 항목의 정의가 이루어져야 한다. 이를 위해 본 연구에서는 표준의 명세와 목적에 부합하도록 하는 인증 평가 항목에 대한 연구를 수행하였다.

2. 적합성 개요 및 적합성 조항의 정의 방향

2.1 적합성 개요

적합성이라는 것은 테스트 대상이 표준이나 명세를 얼마나 충실하게 따르고 있는가의 여부이다. 적합성은 적합성 조항에 기반한 적합성 테스트로 판별할 수 있는데 적합성 조항이란 적합성을 만족하기 위한 표준의 요구사항이며 적합성 테스트는 표준 및 적합성 조항에 정의되어 있는 요구사항을 만족하는가에 대한 테스트를 말한다.

현재 KS X 6100에는 문서 구조를 정의하기 위한 문법적인 규칙만 정의되어 있으며 적합성 테스트를 위해서는 표준 DTD와 더불어 적합성 조항의 정의가 요구된다. 왜냐하면 적합성 테스트는 표준에서 명시된 적합성 조항에 대해서만 수행될 수 있기 때문에 테스트를 수행하기 위해서 적합성 조항의 정의는 필수적이라 하겠다.

2.2 적합성 검사를 위한 평가 항목

적합성 인증 평가 항목, 즉 적합성 조항은 전자책 콘텐츠 제작자들에게 요구되는 문서

의 구조와 콘텐츠 내용에 대한 최소한의 요구사항으로 이루어질 수 있다. 일반적으로 요구사항은 자연어로 되어있으며 이를 테스트 대상이 준수하는가의 여부는 명령어 형태인 테스트 슈트를 작성하여 판별한다.

현재 민간 최대 XML 표준화 단체인 OASIS에서는 적합성 위원회를 운영하며 XML, ebXML[11], SVG등의 적합성 조항 정의와 테스트 슈트 개발 등을 하고 있고, OEB PS[8]에서도 적합성 조항을 정의하여 이를 준수하도록 하고 있다. 또한 국내에서도 ECIF등과 같은 위원회에서 ebXML 적합성 테스트를 위하여 기초 연구 및 적합성 조항 정의 작업이 이루어지고 있는 상황이다. 이러한 관련연구들을 살펴보면 일반적으로 미국의 표준화 기구인 NIST의 적합성 테스트 지침을 따르고 있는데 이에 의하여 적합성 요구사항은 다음과 같은 프로파일 및 레벨로 표현하였다[7].

① 프로파일

프로파일은 표준 전체에서 특정 목적을 위해 갖추어야할 표준 명세의 일부분이다. 일반적으로 어떤 테스트 대상이 표준의 모든 명세를 만족하기 보다는 필요한 일부분을 만족하는 경우가 대부분인데 이러한 경우 필요한 특정 명세부분이 프로파일로 설정된다.

KS X 6100의 경우, 전체 명세는 모든 종류의 전자책 구조를 정의하도록 되어 있으므로 문헌정보학적인 분류별로 프로파일을 설정할 수 있다. 예를 들어 소설 콘텐츠의 경우 문서 구조를 위한 적합성 조항 프로파일을 설정할 수 있다. 따라서 KS X 6100표준에 적합성 조

항을 추가하기 위해서는 각 분류(프로파일)별로 표준적인 문서의 구조를 선정하여야 한다.

② 레벨

레벨은 반드시 준수되어야 하는 필수 요구 사항부터 연관되는 사항들로의 확장으로 정의된다. 이것은 테스트 대상이 만족해야 하는 필수적인 사항이 기본이 되며 레벨이 높아지면 필수 사항 외에 만족해야 할 사항들이 추가된다.

레벨1은 모든 콘텐츠에서 필수적으로 만족되어야 하는 사항이며 따라서 모든 프로파일은 레벨1을 포함하게 된다. 예를 들어 논문 프로파일을 선정하였다 할 때 레벨 1은 표준 DTD의 문법을 준수하는 것이 될 수 있으며 레벨 2는 레벨1을 만족하고 제목, 저자, 요약문, 서론, 본론, 결론, 참고문헌을 포함하여야 하는 것이 될 수 있을 것이다.

적합성 테스트는 적합성 조항에 기반하여 수행되므로 이 조항들의 정의는 신중히 선택되어야 한다. 규격이 정형화된 디지털 카탈로

그등과는 달리 전자책은 그 특성상 구조가 매우 유연하고 다양하기 때문에 충분한 전자책 및 종이책 구조와 내용을 분석하고 전자책 업체의 의견을 수용하여 실용성 있는 요구사항들을 정의해야 한다.

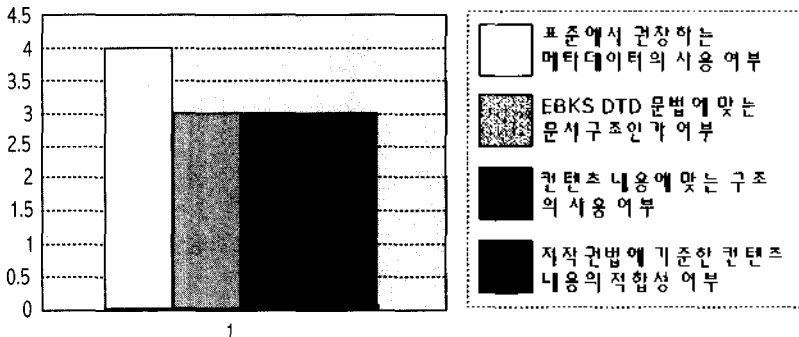
다음 <그림 1>는 설문조사를 통하여 전자책 업체들이 요구하는 적합성 조항의 방향을 조사한 것이다.

이와 같은 결과에 기반하여 향후 KS X 6100의 명세에 다음과 같은 사항들이 적합성 조항으로 추가되고 적합성 테스트가 수행되었다.

- 표준 DTD 문법의 준수 여부
- 콘텐츠를 적절히 표현할 수 있는 메타 데이터 집합들
- 콘텐츠 분류에 맞는 적합한 구조의 사용 여부

다음 사항들은 KS X 6100의 목적에 포함되지 않거나 테스트 절차를 적용하기 어려운

EBKS 전자책 콘텐츠에 대한 인증 절차가 도입된다면 다음 중 어느 사항에 대하여 인증이 중점적으로 이루어져야 한다고 생각하십니까?(복수 선택)



<그림 1> 인증항목에 대한 설문 조사 결과

부분이므로 본 연구외의 다른 연구 분야에서 평가하는 것이 바람직하다.

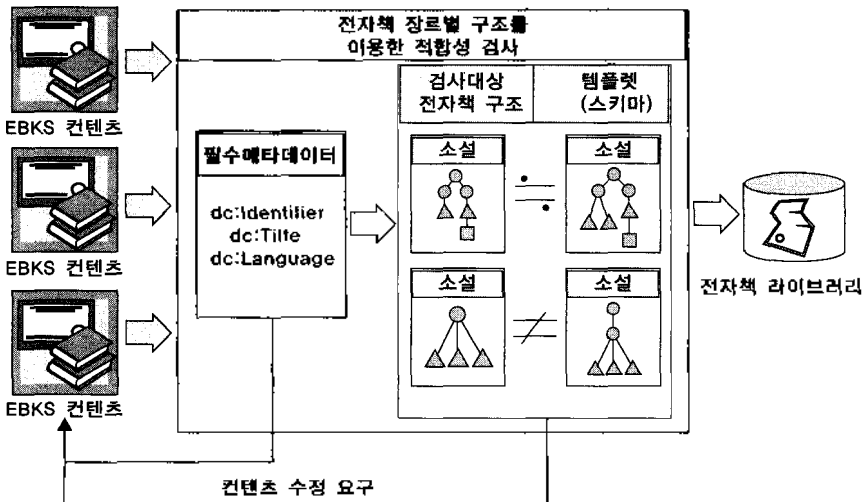
- 저작권법에 의한 콘텐츠 내용의 적합성
- 보안 인증 여부
- KS X 6100 문서 자체가 아닌 연관된 외부 콘텐츠의 적합성

표준 DTD 문법의 준수 여부에 대한 적합성 테스트는 기본적으로 전자책 표준을 따르는 모든 전자책 콘텐츠가 갖추어야 할 조건으로, 인증 대상 전자책 콘텐츠가 전자책 표준의 XML DTD를 준수하고 있는지를 판단하는 것이다. 전자책 표준은 다양한 콘텐츠 표현을 위해서 확장 가능한 구조를 제공하고 있는데, 콘텐츠 제작자가 전자책 표준에서 허용하는 확장 가능한 구조를 벗어나, 유효하지 않은 확장을 시도할 수 있다. 그러므로 전자책 표준 DTD에 대한 확장 여부에 대한 분석과 허용 범위에 대한 연구를 통하여 유효한

DTD만을 인증하여야 한다.

다음으로 현재 KS X 6100의 메타데이터는 웹 환경 및 디지털라이브러리에서 주로 쓰이고 있는 더블린 코어에 기반하고 있으며 콘텐츠 제작자가 더블린 코어 요소 중 필요한 것을 선택하여 사용하도록 하고 있다. 그러나 메타데이터를 이용하여 전자책 문서를 보다 정확하게 기술하기 위해서는 최소한의 메타데이터의 요소 집합을 정의해 주는 것이 바람직하다. 다만 현재 국내전자책을 위한 필수 메타데이터에 대한 연구가 미흡하므로 본 논문에서는 미국 전자책표준인 OEBPS 1.2의 필수 메타데이터 요소인 dc:Identifier, dc:Title, dc:Language를 채용하도록 한다. 본 논문에서는 전자책 콘텐츠의 구조에 대한 검사 적합성을 주로 다루고 있으며 국내 전자책을 위한 메타데이터의 필수요소 선정은 향후연구로 남겨두기로 한다.

마지막으로 문헌 장르별 의미정보를 이용하여 전자책 문서의 구조에 대해 적합성 검사



〈그림 2〉 문헌 장르별 의미정보를 이용한 구조 인증

절차를 적용하는데 이는<그림 2>와 같은 형식으로 이루어진다. 문헌 장르별 특성에 따른 검사 기법은 전자책 표준의 논리적 구조에 따른 인증을 통과한 콘텐츠에 대하여 수행하는 것으로, 콘텐츠가 문헌정보학적인 분류에 따른 특성을 의미적인 구조로 적절히 표현 및 구성하고 있는지에 대한 인증이다. 이를 위해 문헌 정보학적인 분류에 맞는 문서 구조 스키마를 정의하고 이에 기준하여 적합성 테스트를 수행한다.

이와 같은 적합성 테스트를 적용하기 위해 필요한 것은 먼저 해당 장르의 표준적인 구조를 표현 할 수 있는 템플릿, 즉 표준 문서 구조이다. 시, 소설 및 논문 등은 콘텐츠 특성상 문서 구조가 상이하기 때문에 각 장르의 특성을 반영하여 표준 구조를 선정하고 이에 기반한 유형별 전자책 스키마를 정의하였다.

따라서 표준 구조를 선정하기 위해서 현 전자책 업체에서 서비스되고 있는 전자책 콘텐츠, 기존 디지털라이브러리의 콘텐츠 및 종이책의 각 장르별 구조를 조사하고 공통되는 표준 구조를 선정하였다. 특히 전자책 업계의 의견을 수용하여 현실적으로 적용이 가능하고 사용이 용이한 구조를 다음과 같이 선정하였다.

적합성 인증을 위한 장르별 전자책 구분 및 구조를 분류하기 위하여 각 분야에서 구조가 복잡한 책과 임의의 책을 무작위로 선택하여 이를 이용하여 문서의 구조를 분류하였다.

일반적으로 책은 논리적인 구조를 지니고 있는데 이러한 논리적인 구조는 분야에 따라 다르므로 먼저 전자책을 논리적인 구조에 따라 분류할 필요가 있다. 책을 분류하는 방법

에 대해 살펴보면 먼저 듀이십진분류법(DDC: Dewey Decimal Classification)이 있고, 이 방법을 한국 실정에 맞게 편집한 한국십진분류법(KDC: Korean Decimal Classification)이 있다. 이 방법은 주로 도서관에서 사용하는 방법으로 책을 “총류, 철학, 종교, 사회, 언어, 자연과학, 응용과학, 예술, 문학, 지리” 등의 주제별로 분류한다. 그리고 실제 온라인/오프라인 서점에서도 주로 주제별 분류를 사용한다.

전자책 업체에서는 온라인/오프라인 서점과는 다르고 회사마다 분류하는 방법이 확연히 차이가 나는데 이는 각 회사마다 강점을 띠고 있는 분야가 다르기 때문이다. 예를 들어 바로북닷컴의 경우에는 무협지나 환타지, 로맨스 분야의 전자책의 매출이 많아서 그 분야를 다른 분야와 별개로 세부적으로 분류하였다. 위의 분류에서도 바로북닷컴의 경우는 시를 따로 분류하였으나 다른 업체에서는 일반적인 서점의 경우와 같이 문학이라고 하여 문서의 구조를 반영하지 않고 분류하였다.

현재 도서관이나 온라인/오프라인 서점, 전자책 업체에서 수행하는 이러한 분류는 대부분 책의 주제나 내용을 기준으로 분류를 하고, 추가로 책의 판매량 정보를 바탕으로 분류하고 있다.

마지막으로 책을 분류하는 문서의 구조에 따라 분류하는 방법으로 TEI에서 사용하는 방법이 있다. TEI에서는 산문, 운문, 드라마, 스피치, 사전, 용어사전, 텍스트분석, 텍스트비평 등으로 책을 분류하였다. 본 워킹그룹에서는 이러한 TEI의 기준을 이용하고, 추가로 전자책 업체에서 많이 판매되는 분야 중 특별

히 스타일이나 문서의 구조가 복잡하다고 생각되는 분야를 포함하여 전자책을 산문, 운문, 실용서, 전집, 사전, 잡지, 스피치, 기타로 분류하였다.

이 중에서 실용서를 따로 분류하는 이유는 실용서가 문서의 구조나 스타일이 다른 분야와 구별이 많이 되기 때문이고, 일반 서점이나 전자책 업체에서도 이 분야에 대한 매출이 많기 때문이다. 그리고 잡지는 특별한 문서의 구조는 없으나 스타일이 복잡하여 따로 분류하였다.

전자책을 위와 같이 분류하여 각각에 대하여 종이책과 전자책의 문서 구조를 분석하였다. 전체적으로 cover, front, body, back의 네 가지 요소로 구별될 수 있었으며, body를 제외한 나머지 부분은 여러 분야에서 공통적으로 나올 수 있는 것들이었다. 또한 이와 같은 공통 요소외에 각 장르에 따라 구조가 상이하였는데 이에 따라 장르별 전자책 스키마를 정의하였으며 이는 다음장에서 논의한다.

3. 적합성 조항 및 스키마의 정의

3.1 적합성 조항의 정의

적합성 조항은 서론에서 논하였듯이 적합성 검사의 기준이 되는 항목들이다. 표준 전자책 문서들은 기본적으로 KS X 6100의 방법을 준수해야 하며 추가적으로 유형에 적합한 문서 구조를 사용해야 한다.

먼저 책을 분류하는 문서의 구조에 따라 분류하는 방법으로 앞서 살펴본 TEI에서 사용

하는 방법이 있다. TEI에서는 산문, 운문, 드라마, 스피치, 사전, 용어사전, 텍스트분석, 텍스트 비평 등으로 책을 분류하였다. 본 논문에서는 이러한 TEI의 기준을 이용하고, 추가로 전자책 업체에서 많이 판매되는 분야 중 특별히 문서의 구조가 복잡하다고 생각되는 분야를 포함하여 전자책을 산문, 운문, 시, 실용서, 전집, 사전, 잡지, 스피치, 기타와 같은 9가지의 문서 유형으로 구분하였다.

각각의 유형들은 적합성 검사를 위한 프로파일로 표현될 수 있으며 임의의 프로파일은 검사의 조건에 따라 다수의 레벨을 가질 수 있다. 다음 <표 1>의 예에서 레벨1 및 2는 모든 프로파일들이 공통적으로 만족해야 하는 문서의 적합성 조항이며 운문 프로파일의 전자책이 가지는 추가적인 적합성 조항들은 레벨3 이상에서 정의되고 있다.

본 논문에서는 적합성 조항을 구문론적인 방법과 의미론적인 방법으로 정의하고 있다. <표 1>의 예에서 볼 수 있듯이 운문 프로파일의 레벨 1 및 2에서는 구문론적으로 표준 문서가 갖추어야 할 유효성에 대한 적합성 조항을 정의하고 있으며, 레벨 3에서는 이러한 구문론적인 유효성이 보장된 상태에서 보다 명확한 문서의 구조를 위해 의미적인 조항들을 포함하고 있다.

상위 레벨의 조항을 만족하기 위해서는 하위 레벨의 조항을 기본적으로 만족해야 하며, 따라서 레벨이 높아질수록 보다 엄격한 제약이 따르며 상위 레벨의 조항을 만족할수록 표준의 목적과 명세에 더 적합한 문서이다.

앞서 언급하였듯이 KS X 6100은 국내 전자책의 유통 및 교환을 위한 표준이므로 기본

〈표 1〉 운문 프로파일의 적합성 조항의 예

| 운문 프로파일 적합성 조항 | |
|----------------|--|
| 레벨 1 | 문서는 올바른 XML 선언을 가진다. |
| | 인코딩은 UTF-8이나 UTF-16을 사용한다. |
| | 문서는 KS X 6100 표준 DTD(//EBK/DTD eBook Exchange v.10//EN)을 준수한다. |
| | - 중략 - |
| 레벨 2 | 레벨 1의 조건을 만족한다. |
| | 메타데이터에 dc:Identifier dc>Title dc:Language가 반드시 있어야 한다. |
| | 외부 참조 파일은 DTD의 NOTATION을 따른다. |
| | - 중략 - |
| 레벨 3 | 레벨 2의 조건을 만족한다. |
| | 운문은 한개 이상의 장(chapter)로 이루어진다. |
| | 장은 한편 이상의 시(poem)로 이루어진다. |
| | 시는 한개의 제목(title)을 가지며 한개 이상의 연(stanza)으로 이루어진다. |
| | 연은 한개 이상의 시행(poemline)으로 이루어진다. |
| | - 중략 - |

적인 유효성 검증외에 문헌정보학적인 분류에 따른 전자책 문서의 의미적인 구조가 명확히 정의되어 표준 문서로서의 가치를 가지며 본 논문에서는 이를 위한 적합성 조항들을 제안하였다.

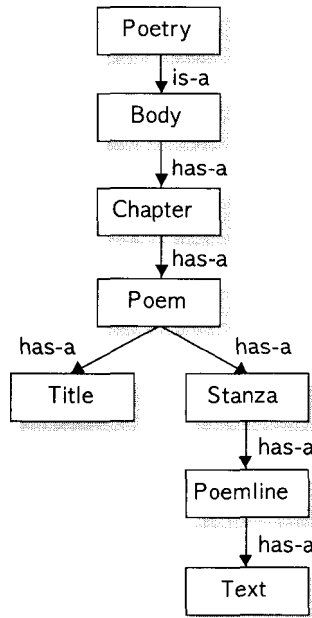
적합성 조항은 표준 명세의 일부분으로 포함되어야 하며 따라서 표준 문서가 이를 준수하는 가의 여부를 판별할 수 있는 기법이 필요하다. 즉 자연어로 이루어진 조항들을 실제 문서에 적용하여 검사할 수 있는 객관적인 기법을 정의하여야 하는데 다음 절에서 XML 스키마를 이용한 표준 스키마의 정의에 대해

여 설명하도록 한다.

3.2 전자책 표준 스키마의 정의

앞서 논한 바와 같이 적합성 조항은 구문론적인 조항과 의미론적인 조항이 있다. 여기서 구문론적인 조항에 대한 검사는 DOM[12]을 이용한 범용 파서 등으로 수행될 수 있는 것이므로 본 연구 보고서에서는 자세히 언급하지 않는다.

다음으로 의미론적인 적합성 조항에 대한 검사는 전자책 온톨로지 모델[6]에 의한 문서



〈그림 3〉 운문 적합성 조항의 온톨로지 예

```

<?xml version= 1.0* ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
...중략...
  <xs:element name = "poem">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="titlegrp" minOccurs="1" maxOccurs="1">
        <xs:element ref="stanza" minOccurs="1" maxOccurs="unbounded">
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="titlegrp">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="title" type="xs:string"/>
        <xs:any minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>

```

〈그림 4〉 시의 적합성 검사를 위한 전자책 표준 스키마

의 구조 분석 및 비교로 이루어진다. 온톨로지는 의미적인 구조의 연관관계를 나타낸 것

으로써 〈표 1〉의 레벨3의 조항은 〈그림 3〉과 같은 의미적인 모델로 표현할 수 있다.

본 연구에서는 이와 같은 모델에 따라 적합성 검사 대상 문서에서 적합한 문서 구조를 검사하기 위해 표준 전자책 XML 스키마를 <그림 4>와 같이 정의하였다. 표준 전자책 스키마는 각 문서의 분류에 따라 9가지로 정의되어 해당 분류에 해당하는 문서의 적합성 여부를 검사할 수 있다.

4. 전자책 적합성 검사 도구

4.1 관련 연구

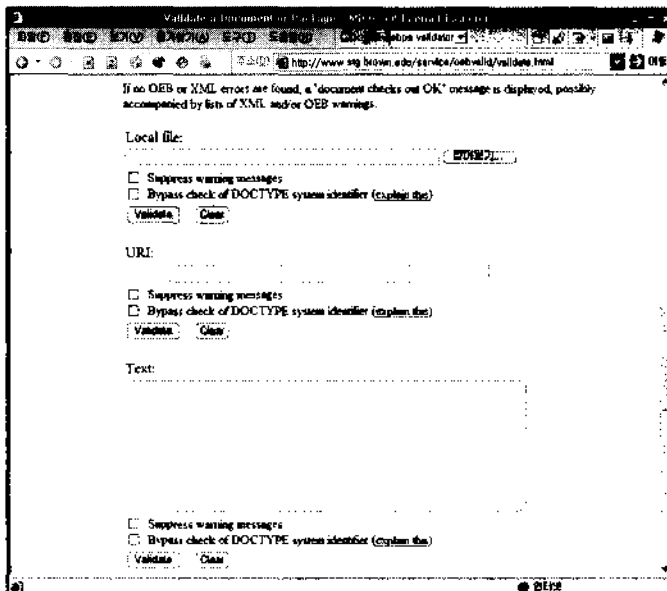
4.1.1 OEBPS의 적합성 검사 도구

미국 전자책 표준 포럼인 OEBF(Open E-Book Forum)에서 제정한 OEBPS의 표준 명세에 적합성 조항을 정의하고 있다. 조항은 문서, 출판, 그리고 출력 기기에 대해서 이루

어져 있는데 이들의 내용은 문서가 XML 형식을 가지고 있는가의 여부, OEBPS의 DTD를 따르는가의 여부, 적합한 인코딩과 멀티미디어 파일의 MIME 지원여부, 그리고 필수 메타데이터 태그의 사용여부 등이다.

OEBPS의 적합성 조항은 주로 XML 문서 로써의 유효성(Validity)을 지키도록 정의되어 있는데 이와 같은 조항들은 전자책 문서가 표준을 준수하기 위한 최소한의 요구사항이라고 할 수 있다.

OEBPS에 대한 적합성 검사는 <그림 5>과 같이 Brown 대학의 STG(Scholarly Technology Group)[9]에서 제공되는 도구로서 수행할 수 있다. 이 도구의 목적은 전자책 문서가 표준 문법을 만족하는지의 여부를 검증하고 이를 통해 신뢰성있는 문서의 출판을 목적으로 하고 있다.



<그림 5> Brown 대학의 OEBPS Validator

4.1.2 E-CALS의 적합성 검사

E-CALS[5]는 일본 국가 사업 중의 하나로 표준 전자 카탈로그이다. 표준 전자 카탈로그는 상품 생산자, 구매자, 그리고 판매자간에 유통되어야 하기 때문에 카탈로그가 표준 명세에 적합하고 올바른 데이터를 가지고 있는가의 여부를 검증하여야 신뢰성 있는 유통이 이루어질 수 있다.

E-CALS에서는 두가지 측면에서 적합성 검사가 이루어지는데 문서 구조 검사와 문서 내용 검사가 그것이다. 먼저 문서 구조는 다양한 저작자에 의해 작성된 SGML 형태의 문서 구조가 표준 문서 구조를 가지는지의 여부를 확인 하는 것이다. 이것은 범용 SGML 파서로 수행되며 문서가 문법적으로 유효(Valid)한 것인가만을 검사한다.

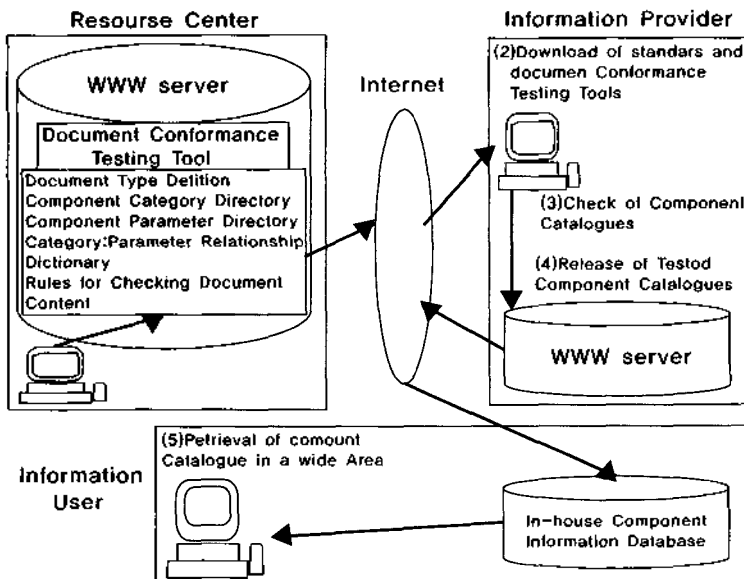
문서 내용 검사는 카탈로그의 분류, 인자값,

수량등이 표준 명세에 맞는가를 검사하는 것으로 10개의 규칙을 가지고 카탈로그 분류 코드와 이름, 인자의 이름과 값, 그리고 분류 코드와 인자의 관계가 올바른가에 대한 검사를 수행한다.

이를 통하여 표준 전자 카탈로그가 데이터베이스에 애러없이 저장되고 스키마에 일치함을 보장할 수 있고 사용자들은 신뢰성 있는 카탈로그를 사용할 수 있다.

이 외에 ebXML의 경우도 적합성 검사를 위한 프레임워크를 만들어 조항과 검사 기법등을 개발하는 중이며 STEP[4] 표준에서도 적합성 및 상호운용성 검사 기법을 제안하고 있다.

현재 이와 같은 관련 연구들의 경우 적합성 테스트는 표준 문서가 문법적인 규칙을 잘 따르고 있는가에 대해 주로 이루어 지고 있는데 이것은 표준 문서가 지켜야할 최소한의 필수 요소라고 할 수 있다.



〈그림 6〉 E-CALS의 적합성 검사 절차

4.2 전자책 적합성 검사 도구의 구현

적합성 검사 도구에서는 이와 같은 스키마를 이용하여 자동적으로 전자책 문서의 적합성을 검사하고 그 결과를 출력해주는 역할을 한다. 사용자 인터페이스는 <그림 7>과 같으며 인터넷 익스플로러에 ActiveX 형식으로 설치되어 전자책 저작자나 심의위원이 별도의 프로그램 설치 없이 해당 주소에 접속하여 도구를 쉽게 사용할 수 있도록 하였다.

사용자는 로컬 파일이나 문서의 웹 주소 및 전자책 문서인 XML 소스를 직접 입력할 수 있으며 원하는 프로파일 및 레벨을 설정하여 적합성 검사를 수행한다.

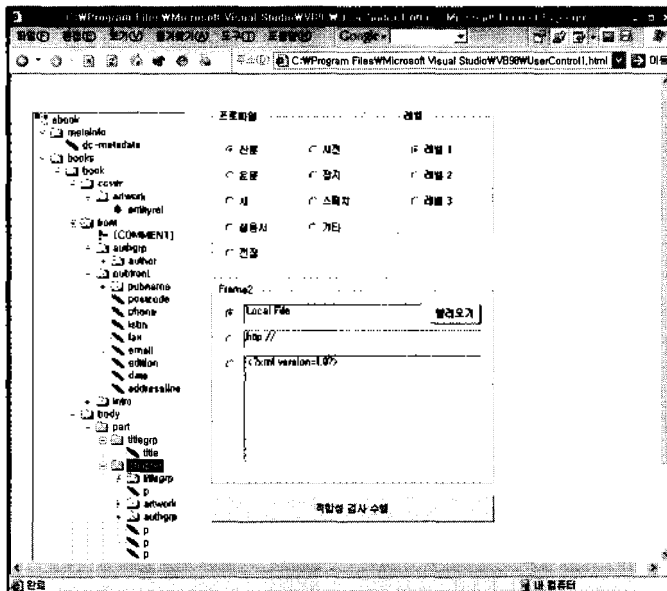
검사의 결과는 대상 문서가 선정된 프로파일과 레벨의 조항을 위배했을 경우 오류가 발생된 부분을 출력하고 조항을 만족할 수 있는 수정 지침을 출력하여 사용자가 적합성 조항

에 맞는 문서를 제작성할 수 있도록 한다.

예를 들어 사용자가 검사 대상 파일을 불러들여 레벨1의 메뉴를 선택하고 적합성 검사를 수행하면 본 도구는 내장 파서에 의하여 DTD의 유효성을 검사한다. 검사 결과는 다음과 <그림 8>과 같이 메시지로 출력되어 대상 문서가 레벨1의 적합성 검사를 통과하였는지의 여부를 사용자에게 알려준다.

사용자는 위 검사결과에 따라 수정사항이 있을 경우 문서를 수정하고 다음 단계의 검사를 수행할 수 있다. 레벨2를 선택하고 검사를 수행했을 경우는 레벨1과 레벨2에 해당하는 항목에 대한 검사를 수행하고 이에 대한 결과를 출력한다.

레벨3에 대한 검사는 문서의 유형에 따라 수행되므로 사용자는 대상 문서의 프로파일, 즉 장르를 선택한다. 대상 문서는 사용자가 선택한 프로파일에 해당하는 스키마를 기준



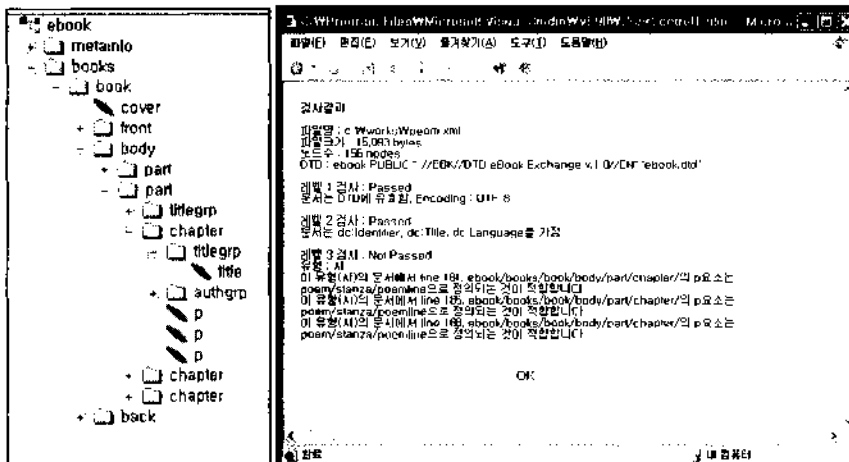
<그림 7> 인증 도구의 인터페이스

| | |
|---|---|
| <p>검사결과</p> <p>파일명 : c:\work\Wpeom.xml 파일크기 : 15,093 bytes 노드수 : 156 nodes DTD : ebook PUBLIC "-//EBK//DTD eBook Exchange v.1.0//EN" "ebook.dtd" 레벨 1 검사 : Passed</p> | <p>검사결과</p> <p>파일명 : c:\work\Wpeom2.xml 파일크기 : 15,001 bytes 노드수 : 152 nodes DTD : ebook PUBLIC "-//EBK//DTD eBook Exchange v.1.0//EN" "ebook.dtd" 레벨 1 검사 No: Passed</p> <p>에러내용</p> <p>Error no linr 70 at position 26</p> <p>DTD에 의하면 어레이소 <my Tag> 내용이 잘못되었습니다. </author></authgrp><my Tag/> 이 위치에는 p, delist, orgaddr, indaddr, bq, prestr, date, biblist, biblist, biblio, authgrp, author, keyword, keyphras, mdrllag, citation, poem, name, head, table, list, fomula, disformula, disformulagr, ma, emph, noteref, [noteref, figref, tableref, artref, appref, ceterel, biblioref, secret, chapter, fomularel, anwork, figgrp, note, xref, uri, otherinfo, section(이/가) 외여 합니다.</p> |
|---|---|

(a) 적합한 문서

(b) 부적합한 문서

〈그림 8〉 레벨1의 검사 결과



(a) 검사대상 문서구조(시)

(b) 검사 결과

〈그림 9〉 레벨3의 검사 결과

으로 적합성 검사가 수행된다. 예를 들어 시 장르이면서 〈그림 9〉의 (a)와 같은 구조를 가지는 전자책 문서는 한국전자책문서표준에서 시 문서를 표현할 때 사용하도록 정의되어 있는 <poem>, <poemline>과 같은 요소를 사용하지 않고 일반 문단을 표현하는 <p>를 사용하

였다. 이경우 레벨1 및 레벨2의 적합성 검사는 통과할 수 있으나 레벨3의 스키마 구조에는 위배되기 때문에 〈그림 9〉의 (b)와 같은 결과를 얻는다.

사용자는 이와 같은 결과를 참조하여 검사 대상 문서가 레벨별로 어느정도까지 적합한 문서인지 자동적으로 판별할 수 있다. 본 논문에서 개발한 도구는 전자책 제작자의 경우 표준에 적합한 문서를 작성하는데 이용될 수 있으며, 전자책 인증기관에서는 적합성 검사의 보조도구로서 활용될 수 있다.

5. 결 론

본 연구에서는 한국 전자책 표준인 KS X 6100의 명세를 준수하는 문서의 적합성 검사를 위한 적합성 조항과 기법을 제안하였다. 적합성 조항은 표준 문서가 기본적으로 준수해야 할 구문론적인 문법과 문헌정보학적인 분류에 따라 준수해야 할 문서의 의미적인 구조로 정의하였으며, 이러한 조항에 따라 적합성 검사 도구를 개발하여 보다 표준의 목적에 보다 적합한 문서를 자동적으로 검사할 수 있도록 하였다.

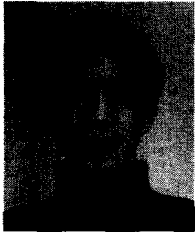
이와 같은 적합성 검사 기법을 이용하여 표준에 기반한 콘텐츠 제작 및 검사시 소요되는 인력과 시간을 줄이고 신뢰성있는 표준 전자책 콘텐츠를 유통시켜 표준의 활성화를 통한 제작 및 서비스의 필요한 중복 투자를 방지할 수 있다.

향후 연구 과제로는 적합성 조항과 검사 기법을 활용하여 전자책 적합성 검사도구에서 자동적으로 검사대상 문서의 구조를 표준에 적합한 구조로 변환하여 적용시키는 방안이나 전자책 문서의 분류 및 검색과 같은 활용 방안도 연구되어야 할 것이다.

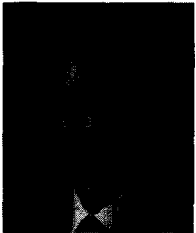
참 고 문 헌

- [1] Won-sung Sohn, et al, Standardization of eBook documents in the Korean industry, Computer Standards & Interface, vol.24, p45-60, 2002.
- [2] Won-sung Sohn, et al, Development Electronic Book of Korea Standard(EBKS), Korea Society of Information Management, 2001.
- [3] Beverly L. Harrison, E-Books and the Future of Reading, Proc. IEEE Computer Graphics and Applications, vol. 20, no. 3, pp. 32-39, 2000.
- [4] James D. Kindrick, John A. Sauter, Robert S. Matthews, Improving Conformance and Interoperability Testing, IEEE, StandardView Vol.4, No.1, 1996.3
- [5] Makoto Imamura, Sharing SGML/XML Document Information Conformable to Business Standards, Parallel Processing, International Workshops, pp. 381-386, 1999
- [6] Allen Renear and David Dubin, Towards a Semantics for XML Markup, DocEng2002 ACM, November 2002
- [7] William Gilbert, Conformance Requirements Version 1.0, OASIS Technical Committee.
- [8] OEBPS Specification, http://www.openebook.org/doc_library/specifications.htm
- [9] Open eBook Validator, <http://www.stg.brown.edu/service/oebvalid/>
- [10] Text Encoding Initiative, <http://www.tei-c.org>
- [11] ebXML - Enabling A Global Electronic Market, <http://www.ebxml.org>
- [12] W3C Recommendation REC-DOM-Level-1-19981001, Document Object Model (DOM). World WideWeb Consortium, 1998, <http://www.w3.org/TR/1998/REC-DOM-Level-1-19981001>.

저 자 소 개



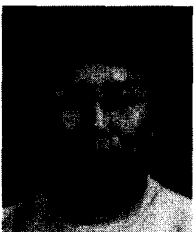
김재경 (E-mail : ki187cm@rainbow.yonsei.ac.kr)
 2000. 단국대학교 화학과(학사)
 2002. 연세대학교 컴퓨터과학과(석사)
 2003~현재 연세대학교 컴퓨터과학과 박사과정
 관심 분야 멀티미디어 문서처리(XML), 웹 Annotation 생성 및 응용



손원성 (E-mail : sohnws@rainbow.yonsei.ac.kr)
 1998. 동국대학교 컴퓨터공학과(학사)
 2000. 동국대학교 컴퓨터공학과(석사)
 2003~현재 연세대학교 컴퓨터과학과 박사과정
 관심 분야 멀티미디어 문서처리(XML), 웹 Annotation 생성 및 응용



임순범 (E-mail : sblim@sookmyung.ac.kr)
 1982. 서울대학교 계산통계학과(학사)
 1983. 한국과학기술원 전산학과(석사)
 1992. 한국과학기술원 전산학과(박사)
 1989~1992. (주)휴먼컴퓨터 이사/연구소장
 1992~1997. (주)삼보컴퓨터 부장
 2001~현재 숙명여자대학교 멀티미디어학과 조교수
 관심 분야 컴퓨터 그래픽스, 멀티미디어 응용, 전자출판(폰트, 전자책, 사이버교재)



김성혁 (E-mail : ksh@sookmyung.ac.kr)
 1975. 연세대학교 문헌정보학과(학사)
 1983. 미국 Case Western Reserve University(정보학석사)
 1992. 연세대학교 대학원 문헌정보학과(박사)
 1986~1991. 대신증권 전산실장
 1998~1999. 첨단학술정보센터 소장 직무대행
 1991~현재 숙명여자대학교 정보과학부 부교수
 2000~현재 전자상거래 통합포럼 전자카타로그 기술위원회 위원
 2001~현재 EBK 전자책문서표준화위원회 위원
 관심 분야 디지털도서관, 텍스트인코딩, 전자카타로그, 문서구조화



최윤철 (E-mail : ycchoy@rainbow.yonsei.ac.kr)
 1973. 서울대학교(학사)
 1975. Univ. of Pittsburgh(석사)
 1976. Univ. of California, Berkeley(석사)
 1979. Univ. of California, Berkeley(박사)
 1979~1982. Lockheed 사 및 Rockwell International 사 연구원
 1990~1991. University of Massachusetts 교환교수
 2002~2003. 일본 게이오대 방문교수
 1984~현재 연세대학교 컴퓨터과학과 교수
 관심 분야 아바타 인터페이스, 멀티미디어 문서처리, 컴퓨터 그래픽스, eLearning