

한국전자책문서표준의 적합성 조항 및 전자책 표준 스키마의 정의

김재경*, 손원성*, 임순범**, 최윤철*

*연세대학교 컴퓨터과학과

**숙명여자대학교 멀티미디어학과

{ki187cm, sohnws, ycchoy}@rainbow.yonsei.ac.kr

sblim@sookmyung.ac.kr

Definition of Conformance Clause and Schema for Electronic Book Standard of Korea

Jae-kyung Kim*, Won-Sung Sohn*, Soon-Bum Lim**,

Yoon-Chul Choy*

*Dept. of Computer Science, Yonsei University

**Dept. of Multimedia, Sookmyung Women's University

요 약

지난 2001년 7월 제정된 한국전자책문서표준 EBKS(Electronic Book of Korea Standard) 1.0은 지난 2002년 7월 한국산업기술표준원의 산업표준심의를 거쳐 한국전자책 표준문서형 정의, 규격번호 KS X 6100으로 제정되었다. 따라서 KS X 6100을 사용하는 전자책 문서들은 표준의 목적과 명세에 대한 부합 여부를 판별하기 위한 적합성 검사 및 이에 기반한 인증 절차가 요구된다. 본 논문에서는 이러한 적합성 검사의 기준이 되는 적합성 조항(Conformance Clause)의 정의와 이에 따른 전자책 표준 스키마를 정의하였다. 따라서 본 논문에서 정의된 조항들을 통하여 전자책 저작자들은 전자책표준의 목적과 명세에 보다 적합한 전자책 문서들을 제작할 수 있으며, 서비스 제공자들은 적합성 검사 기법을 통한 자동화된 적합성 표준 스키마를 사용하여 효율적인 적합성 검사가 가능하다.

1. 서론

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 본 연구는 2002년도 정보통신부지원 대학교초연구지원사업의 지원에 의하여 수행되었음.

2002년 7월 한국산업기술 표준원의 산업표준심의를 거쳐 한국전자책 표준문서형 정의, 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.

전자책 업계에서 단일화된 KS X 6100 표준을 준수하는 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 확신할 수 있는 검사가 필요하다.

이는 적합성 검사에 의하여 수행될 수 있는데 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[3].

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.

적합성 조항의 정의가 이루어지면 이를 실제적으로 문서에 적용할 수 있는 기법이 필요한데 수많은 전자책 문서를 전문가가 일일이 검사할 수 없기 때문에 자동화된 기법을 이용하여 문서의 적합성 여부를 판단하거나 혹은 전문가가 빠른 판단을 내릴 수 있도록 보조 도구 역할을 하여야 한다.

이와 같은 적합성 조항 및 기법에 의한 적합성 검사는 다음 두가지 측면에서 요구된다. 먼저 한국 전자책문서 표준이 KS X 6100 표준으로 제정됨에 따라 표준 전자책 문서가 표준의 명세를 준수하는가의 여부에 대한 검사가 요구되며, 다음으로 전자책 업계 및 사용자간의 전자책 문서의 신뢰성 있는 유통 및 서비스를 제공하기 위한 상호 운용성의 보장 측면에서 적합성 검사는 필수적으로 수행되어야 한다.

본 논문에서는 이러한 적합성 검사를 수행하기 위해 적합성 검사의 기준 지침이 되는 적합성 조항을 정의하고 이에 기반하여 실제 문서에 적용할 수 있는 적합성 표준 스키마를 정의하였다.

2. 관련 연구

2.1 OEBPS

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 이루어져 있는데 이들의 내용은 문서가 XML 형식을 가지고 있는가의 여부, OEBPS의 DTD를 따르는가의 여부, 적합한 인코딩과 멀티미디어 파일의 MIME 지원여부, 그리고 필수 메타

데이터 태그의 사용여부 등이다[4].

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

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

2.2 E-CALS

E-CALS는 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[6].

문서 내용 검사는 카탈로그의 분류, 인자값, 수량 등이 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 에러없이 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[6] 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 이루어지고 있는데 이것은 표준 문서가 지켜야할 최소한의 필수 요소라고 할 수 있다.

KS X 6100의 경우 이러한 필수요소 외에 전자

책의 특성을 반영할 수 있는 요소가 필요하다. 즉 구문론적인 요소외에 전자책의 문헌정보학적 분류에 의한 의미적인 구조를 위한 검사가 필요하다. 다음 장에서는 본 논문에서 정의한 전자책 표준 스키마를 통한 검사 방법에 대하여 알아보도록 한다.

3. 전자책 표준 적합성 검사

KS X 6100은 국내 전자책 문서의 명확한 구조 정의로 효율적인 유통에 그 목적을 두고 있으며 다양한 문헌 유형을 반영하였으나 1개의 DTD로 구성되어 그 구조가 복잡하다.

따라서 전자책 저작자는 소설이나 시, 논문등과 같은 여러 유형의 전자책을 하나의 표준으로 제작할 수 있는 장점이 있으나 제작 경험 부족등으로 적절한 문서 유형 구조를 사용하지 못하여 전자책의 명확한 구조를 정의하기 힘든 단점이 있다.

본 논문에서는 이러한 점을 해결하기 위해 문헌 정보학적인 분류에 따라 전자책의 명확한 구조를 사용할 수 있도록 적합성 조항 및 검사 기법을 제안한다.

3.1 적합성 조항

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

먼저 책을 분류하는 문서의 구조에 따라 분류하는 방법으로 TEI에서 사용하는 방법이 있다. TEI에서는 산문, 운문, 드라마, 스피치, 사진, 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 갖추어야할 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 문서라고 할 수 있다.

[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.1.0//EN)을 준수한다. - 중략 - |
| 레벨 2 | 레벨 1의 조건을 만족한다. 메타데이터에 제작자(dc:creator)와 표제(dc:title)가 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 - |
| | - 중략 - |

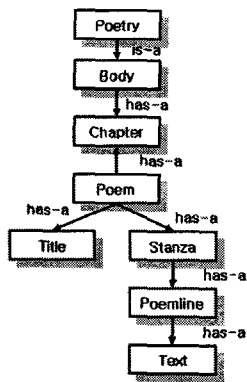
앞서 언급하였듯이 KS X 6100은 국내 전자책의 유통 및 교환을 위한 표준이므로 기본적인 유효성 검증외에 문헌정보학적인 분류에 따른 전자책 문서의 의미적인 구조가 명확히 정의되었을때 표준 문서로서의 가치를 가지며 본 논문에서는 이를 위한 적합성 조항들을 제안하였다.

적합성 조항은 표준 명세의 일부분으로 포함되어야 하며 따라서 표준 문서가 이를 준수하는가의 여부를 판별할 수 있는 기법이 필요하다. 즉 자연어로

이루어진 조항들을 실제 문서에 적용하여 검사할 수 있는 객관적인 기법을 정의하여야 하는데 다음절에서 XML 스키마를 이용한 표준 스키마의 정의에 대하여 설명하도록 한다.

3.2 전자책 표준 스키마

3.1절에서 논한 바와 같이 적합성 조항은 구문론적인 조항과 의미론적인 조항이 있다. 여기서 구문론적인 조항에 대한 검사는 범용 파서[6] 등으로 수행될 수 있는 것이므로 본 논문에서는 자세히 언급하지 않는다.



[그림 1] 운문 적합성 조항의 온톨로지 예

다음으로 의미론적인 적합성 조항에 대한 검사는 전자책 온톨로지 모델에 의한 문서의 구조 분석 및 비교로 이루어진다. 온톨로지는 의미적인 구조의 연관관계를 나타낸 것으로써 [표 1]의 레벨3의 조항은 [그림 1]과 같은 의미적인 모델[7]로 표현할 수 있다.

본 연구에서는 이와 같은 모델에 따라 적합성 검사 대상 문서에서 적합한 문서 구조를 검사하기 위해 표준 전자책 XML 스키마를 [그림2]와 같이 정의하였다. 표준 전자책 스키마는 각 문서의 분류에 따라 9가지로 정의되어 해당 분류에 해당하는 문서의 적합성 여부를 검사할 수 있다.

4. 결론 및 향후 연구방향

본 연구에서는 한국 전자책 표준인 KS X 6100의 명세를 준수하는 문서의 적합성 검사를 위한 적합성 조항과 기법을 제안하였다. 적합성 조항은 표준 문서가 기본적으로 준수해야할 구문론적인 문법과 문헌정보학적인 분류에 따라 준수해야할 문서의

의미적인 구조로 정의하였으며, 이러한 조항에 따라 온톨로지 모델을 구축과 적합성 기법을 제안하여 보다 표준의 목적에 적합한 문서를 검사할 수 있도록 하였다.

```

<?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: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 2] 시의 적합성 검사를 위한 전자책 표준 스키마

향후 연구 과제로는 적합성 조항과 표준 스키마 정의에 따른 도구의 개발로 다수의 전자책 표준 문서의 적합성 검사 절차를 자동화하여 신속하고 효율적인 검사가 이루어지도록 해야 할 것이다. 또한 적합성 조항과 검사 기법을 활용하여 전자책 문서의 분류 및 검색과 같은 활용 방안도 연구되어야 할 것이다.

참고문헌

- [1] Won-sung Sohn, et al, Standardization of eBook documents in the Korean industry, Computer Standards & Interface, vol.24, p45-60, 2002.
- [2] 손원성, 고승규, 이경호, 김성혁, 임순범, 최윤철, "한국 전자책 문서표준(EBKS)의 개발" 정보관리학회 학회지, 2001
- [3] GLOSSARY of Conformance Terminology, http://www.oasis-open.org/committees/ioc/glossary.htm
- [4] OEB Specification, http://www.openebook.org/doc_library/specifications.htm
- [5] Open eBook Validator, http://www.stg.brown.edu/service/oevalid/
- [6] James D. Kindrick, John A. Sauter, Robert S. Matthews, Improving Conformance and Interoperability Testing, IEEE, StandardView Vol.4, No.1, 1996.3
- [7] Allen Renear and David Dubin, Towards a Semantics for XML Markup, DocEng2002 ACM, November 2002