

전자책 문서표준 간의 변환 기법 연구

고승규*, 손원성*, 이경호**, 김성혁***, 임순범****, 최윤철*

*연세대학교 컴퓨터과학과, **미국 국립표준기술원(NIST) 객원 연구원

숙명여자대학교 문헌정보학과, * 숙명여자대학교 멀티미디어과학과

The Study on Conversions of eBook Document Standards

Seung-Kyu Ko*, Won-Sung Sohn*, Kyong-Ho Lee**, Sung-Hyuk Kim***, Soon-Bum Lim****, Yoon-Chul Choy*

*Dept. of Computer Science, Yonsei University, **IT Lab. National Institute of Standards and Technology,

***Dept. of Library and Information Science, Sookmyung Women's University,

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

요약

기존 종이책의 디지털 형태인 전자책은 종이책에 비해 인쇄와 유통, 저장 관리가 효율적이고, 멀티미디어 정보 표현 등이 가능한 장점을 지니고 있기 때문에 향후 시장이 급성장할 것으로 예측되고 있다. 이와 같이 급성장할 전자책 시장을 위해 미국, 일본 등 세계 각국에서는 전자책 문서 표준을 제정하여 시장을 선점하고자 노력하고 있으며, 국내에서도 한국전자책컨소시엄(EBK)에서 지난 5월 전자책 문서표준으로 EBKS를 발표하였다. 이와 같이 전자책에 관한 여러 표준들이 존재함에 따라 각 표준을 따르는 전자책 들에 대한 상호호환의 문제가 발생하게 된다. 이러한 문제를 해결하기 위해서 본 논문에서는 EBKS를 중심으로 다른 전자책 문서표준 간의 변환기법에 대해 연구하였다. 이 변환기법은 향후 전자책 교환을 위한 필수적인 기술이 될 것이다.

1. 서론

기존의 아날로그 형태의 종이책은 정보화 시대를 맞이하여 디지털 형태인 전자책으로 변해가고 있다. 디지털 형태의 전자책은 기존 종이책에 비해 인쇄와 유통, 저장 관리가 효율적이고, 멀티미디어 정보 표현 등이 가능한 장점[1][2]을 지니고 있다. 이러한 장점으로 인해 전자책 시장은 앞으로 급격히 팽창할 것으로 예상되고 있다.

이와 같이 급속히 팽창하는 전자책 시장을 위해 세계 각국에서는 전자책 문서 표준을 제정하여 시장을 선점하고자 노력하고 있다. 미국에서는 1999년 국립표준기술연구소 (NIST: National Institute of Standards and Technology)와 업계를 중심으로 전자책 컨소시엄 (OEBF: Open eBook Forum)을 구성하여 OEB 출판 구조 명세(OEB PS: Open eBook Publication Structure Specification)[4]를 발표하였다. 그리고 일본에서도 1999년 일본전자출판협회 (JEPFA: Japanese Electronic Publishing Association)을 중심으로 JepaX 0.9[5] 버전을 발표하였다. 국내에서는 지난 5월 한국전자책컨소시엄(EBK: eBook of Korea)에서 전자책 문서 표준(EBKS: Electronic Book of Korea Standard)[3]을 발표하였다. 특히 EBKS에서는 국내 전자책 시장에서 시급히 필요한 전자책 교환을 목표로 제정되었다.

향후 국내외로 전자책 시장이 활성화 되면 각 표준을 따르는 전자책이 다수 존재하게 될 것이고, 이러한 전자책 간의 교환이 이루어질 것이다. 이럴 때 전자책 문서표준 간의 변환은 필수적이다. 따라서 본 연구에서는 EBKS를 중심으로 미국의 전자책 표준과 일본의 전자책 표준간의 변환 기법에 대해 연구하였다. 본 논문의 구성은 다음과 같다. 2절에서는 전자책 관련 표준에 대해 소개하고, 3절에서는 EBKS를 중심으로 해외 표준 간의 교환 기법에 대해 설명한다. 그리고 4절에서는 결론 및 향후 연구방향에 대해 기술한다.

2. 전자책 문서표준

전자책 문서와 관련된 표준으로는 현재 미국의 OEB PS와 일본의 JepaX 그리고 국내의 EBKS가 있으며, 표현 포맷은 모두 XML에 기반하고 있다.

2.1 OEB PS

OEB PS는 전자책의 효과적인 시장 적용을 목표로 하고 있으며. 구조는 전자책의 구성방법을 명시하는 OEB 패키지(OEB Package)와 실제 내용을 나타내는 OEB 문서(OEB Document)로 이루어진다.

OEB 패키지는 OEB 출판물의構성을 설명하고 있으며, 메타데이터, 구성 파일 목록(Manifest), 문서 배치구조(Spine), 탐방순서(Tour) 그리고 부가정보를 표현하는 안내정보(Guide)로 구성된다.

OEB 문서는 OEB 사양에서 정의한 구문만을 사용하여 표현된 기본 문서(basic document)와 임의의 XML 문서가 가능한 확장 문서(extended document)로 구별되며, 기본 문서는 HTML 4.0의 XML 형태인 XHTML 1.0을 기반으로 정의되었다.

메타데이터는 더블린코어[6]에 기반하고 있으며, 스타일 정보는 CSS1 중심으로, CSS2의 일부 기능 및 '머리말/꼬리말'과 다단 속성을 추가하여 정의되었다.

OEB PS는 전자책 교환보다는 전자책의 출력에 비중을 두어 개발된 표준으로, fallback 함수를 이용하여 단말기 환경에 따른 스타일 적용이 가능하다. 그리고 HTML 문서는 약간의 수정을 거치면 OEB 문서로 변환된다.

2.2 JepaX

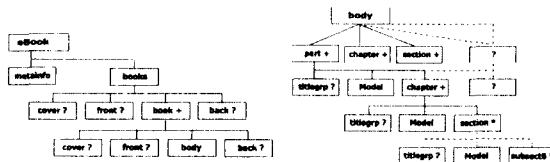
JepaX는 컨텐츠 생성이나 교환 포맷으로 사용하는 것이 목적으로 확장구조 및 레이아웃 정보를 표현하는 "jepaxinfo". 책에 대한 메타데이터를 표현하는 "bookinfo", 그리고 책의 구성 요소인 "cover", "front", "body", "back"으로 구성된다. JepaX에서는 문서의 구조를 정의하지 않고, 구조를 표현할 수 있는 방법으로 "div" 엘리먼트의 속성 "type"을 제공한다. 이 속성값으로 가정의된 항목은 "표지", "개요", "저자 소개", "문", "현사", "서문" 등이 된다. 사용자는 "div" 엘리먼트의 속성 "type"에 이러한 값을 이용하여 문서 구조를 표현할 수 있다.

메타데이터는 자체적으로 정의한 9가지의 정보(isbn, 제목, 부제목, 판정보, 등)로 구성되어 있으며, 스타일 정보는 언제마다 적합한 스타일을 사용하도록 하기위해 정의하지 않았다.

2.2 EBKS

EBKS는 전자책 컨텐츠에 대한 정확한 교환을 목적으로 재정되었으며 컨텐츠, 메타데이터, 스타일의 세 부분으로 구성된다.

명확한 컨텐츠 교환을 위하여 문서의 논리적인 구조를 [그림 1]과 같이 정의하였다. 이 구조는 기준의 문서 관련 표준인 ISO12083[7], TEI(Text Encoding Initiative)[8], DocBook[9] 및 시중에 판매되는 임의의 책들의 구조 정보를 분석하여 정의되었다. 그리고 EBKS에서는 정의된 문서 구조이외의 구조를 표현할 수 있는 확장 메커니즘을 제공한다.



[그림 1] EBKS의 전체 구조

메타데이터는 더블린코어의 15개 기본 엘리먼트를 기반으로 하며 한정어(Qualifier)를 사용하지 않는다. 스타일은 XSL-FO를 표준이 아닌 권고 사항으로 제시하고 있다.

EBKS는 OEB PS와 JepaX와는 달리 고정된 문서 구조를 정의하고 있으며, 이를 이용하여 전자책의 명확한 교환 및 가공이 가능하다. 또한 한국 문현의 특징인 세로쓰기 및 다단편집을 위하여 XSL-FO를 권장하고 있으며, 여러 문서 구조에 대한 확장 방법을 제공한다.

3. 전자책 문서표준 간의 변환

EBKS는 문서의 논리적인 구조를 정의하고 있는 점이 OEB, PS나 JepaX와의 가장 큰 차이이다. 그러므로 변환 시 구조 정보를 반영할 수 있는지 여부가 가장 큰 이슈가 된다. 본 절에서는 EBKS와 다른 두 표준파의 변환을 컨텐츠, 메타데이터, 스타일 측면에서 각각 살펴본다.

3.1 EBKS와 OEB PS

- ① 컨텐츠 : 먼저 EBKS 문서를 OEB PS 문서로 변환하는 경우에 대해 살펴본다. OEB PS의 기본문서는 기본적으로 XHTML 1.0으로 구성되어 있어서 문서에 대한 명확한 논리적인 구조를 지니고 있지만 않다. 그러므로 논리적인 문서 구조를 지닌 EBKS 문서를 OEB 기본문서로 변환하는 경우와 같은 하향변환(down-conversion)의 경우는 EBKS 문서의 각 구성 요소를 OEB 기본 문서의 특정 엘리먼트로 매핑시키면 된다. 예를 들어 [표 1]과 같이 EBKS 문서에서 장제목을 OEB 기본 문서의 “h1”~“h6”중에서 어떠한 엘리먼트로 변화해야 하는지 지정해야 한다. 이 때 이 매핑 규칙은 사용자마다 다를 수 있으므로 기본적으로 하나의 규칙을 정한 후 이를 사용자마다 적합하게 사용(customizing)할 수 있도록 해야 한다. [그림 2]는 이러한 매핑 정보를 XSLT로 표현한 예이다. 이러한 과정은 EBKS 문서에 스타일을 적용시키는 경우와 유사하다. OEB 확장 문서는 위의 XML문서가 가능하기 때문에 EBKS 문서가 그대로 OEB 확장 문서가 될 수 있다. 그러므로 이 경우에는 특별한 변환이 필요 없다.

변환된 OEB 기본 문서를 EBKS문서로 역변환하는 경우는 첫 변화 시 문서의 구조 정보를 상실하였기 때문에 원본 EBKS 문서로의 역변환은 불가능하다. 이때에는 EBKS 문서를 OEB 문서로 변환하는 과정에서 “div”나 “span” 엘리먼트를 이용하여 문서의 구조 정보를 표현하면 이 정보를 이용하여 EBKS 문서로 역변화이 가능하다.

OEB 문서를 EBKS로 변환하는 경우도 두 가지로 나누어 살펴볼 수 있다. 먼저 기본 문서의 경우와 같이 구조가 없는 문서를 구조가 있는

문서로 변환하는 상향변환(up-conversion)인 경우에는 자동으로 변환할 수 없으며 사용자가 각각의 엘리먼트를 어떤 엘리먼트로 변환할지 지정하여야 한다. 이 경우는 위의 EBKS 문서를 OEB 기본 문서로 변환하는 경우와 달리 각 엘리먼트에 대해 명확한 매핑 정보를 구축하기가 어렵다. 왜냐하면 OEB 기본 문서의 “h3” 엘리먼트가 EBKS의 장제목, 절제목, 부절제목, 저자, 내용 등의 여러 엘리먼트로 변화가 가능하기 때문이다. 그러므로 이러한 경우에는 매핑정보를 지정하는 것이 어렵기 때문에 문서 전체에 대해 수동으로 변환을 해야 한다. OEB 확장 문서를 EBKS 문서로 변환하는 과정은 문서의 구조를 지닌 문서를 다른 구조 문서로 변환하는 경우와 같다. 이러한 경우는 매핑 정보를 이용하여 변화가 가능하다.

[표 1] EBKS와 OEB 문서 간의 매핑 테이블 예

EBKS	OEB
body/chapter/title	H2
cover/titlegrp/title	H1
chapter//Item	li
...	...

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>XHTML 예제</title>
  </head>
  <body>
    <h1>Hello World</h1>
    <h2>Hello World</h2>
    <h3>Hello World</h3>
    <ul>
      <li>Hello</li>
      <li>World</li>
    </ul>
    <ol>
      <li>Hello</li>
      <li>World</li>
    </ol>
    <table border="1">
      <caption>Hello World</caption>
      <thead>
        <tr>
          <th>Hello</th>
          <th>World</th>
        </tr>
      </thead>
      <tbody>
        <tr>
          <td>Hello</td>
          <td>World</td>
        </tr>
      </tbody>
    </table>
    <div style="text-align: center;">
      
    </div>
  </body>
</html>

```

[그림 2] 변환 스크립트(XSLT)

- ② 메타데이터 : OEB PS와 EBKS에서는 메타데이터를 표현할 때 더블린코어를 이용한다. 그러므로 더블린코어로 표현된 메타데이터에 대해서는 자동 변환이 가능하다. 그리고 더블린코어 이외의 와콤 메타데이터에 대해서도 EBKS와 OEB PS의 확장 방법이 유사하기 때문에 자동 변환이 가능하다.

메타데이터 이외의 OEB 패키지의 구성 파일 복록(Manifest), 문서 배치구조(Spine), 탐방순서(Tour), 안내정보(Guide)에 대해 살펴보면, OEB 패키지의 Manifest와 OEB의 Spine 정보는 EBKS의 일부 엔티티로 표현되며, Tour나 Guide 정보는 EBKS의 확장 메커니즘을 이용하여 표현할 수 있다.

- ③ 스타일 : EBKS에서는 스타일로 XSL-FO를 권고하고 있고, OEB에서는 CSS에 기반한 스타일 정보를 정의하고 있다. XSL-FO는 다만 세로쓰기, 페이지 지원 등 CSS보다 기능이 풍부하기 때문에 EBKS의 스타일 정보를 CSS로 변환하여 표현할 경우에 표현되지 못하는 것이 존재한다. 그러나 CSS에서 표현 가능한 것은 대부분 XSL-FO에서도 표현 가능하다. 현재 EBKS에서는 스타일에 대해서 참고자료인가 때문에 CSS 사용은 권장하지 않다.

[그림 3-4]는 원복 EBKS 문서와 변환된 OEBK 문서의 예이다.

[그림 3] EBKS 문서

[그림 4] 변환된 OEB 문서

3.2 EBKS와 JepaX

① 컨텐츠 : JepaX는 “div”태그의 “type”속성에 문서의 구조 정보를 표현하고 있으나 EBKS와 같이 구체적으로 정의한 문서 구조는 없다. 그러나 문서에 대해 사용자가 문서의 논리적인 구조를 정의하므로 EBKS와 JepaX간의 변환은 EBKS와 OEB PS 간의 변환과는 달리 구조 정보를 지닌 문서간의 변환으로 볼 수 있다. EBKS는 고정된 하나의 논리적인 구조를 정의하고 있으며, JepaX도 “div” 엘리먼트의 속성 “type” 값의 가능한 집합을 정의하고 있다. 이와 같이 각 표준이 표현할 수 있는 논리적인 구조는 공통적인 것도 존재하며, 그خلف자 않은 것도 존재한다. 이와 같이 공통적인 부분에 대해서는 고정된 매핑 정보 구축이 가능하고, 한쪽에만 있는 요소에 대해서는 각 표준의 확장 방법을 이용하여 매핑이 가능하다. 특히 JepaX에서의 확장은 경우에는 “div” 엘리먼트의 “xtype” 속성 값으로 표현하기 때문에 자동 변환이 가능하다. [표 2]는 EBKS와 JepaX간의 매핑테이블이다. 상단 부분은 공통요소에 대한 매핑 정보이고, 하단의 두 열은 각 표준의 확장 방법을 이용하는 경우이다.

[표 2] EBKS와 JepaX 간의 컨텐츠 매핑 테이블

JepaX	EBKS
jepax	ebook.
bookinfo	jepaxinfo

cover	cover
front	front
body	body
back	back
title	title
subtitle	subtitle
div type="chapter"	chapter
...	...
div type="program list"	논리구조 확장
div xtype="subsect1"	subsect1

② 메타데이터 : JepaX에서는 자체적으로 9가지의 메타데이터들을 정의한다. 이 중에서 5가지는 EBKS와 동일하지만 나머지 4가지는 상이하다. 그러므로 JepaX에서 EBKS로 변환할 경우에는 EBKS의 메타데이터 확장 메커니즘을 이용하면 되고, 반대의 경우에는 JepaX의 “div”엘리먼트를 이용하여 메타정보를 표현할 수 있다. [표 3]은 EBKS와 JepaX간의 메타데이터 맵핑 테이블이다.

[표 3] EBKS와 JepaX 메타데이터 매핑 테이블

EBKS	JepaX
Identifier	isbn
Title	book-title
Creator	book-author
Date	pub-date
Publisher	publisher

③ 스타일 : JepaX에서는 스타일 정보를 표현하고 있지 않기 때문에 스타일에 대한 변화는 존재하지 않는다.

4. 결론 및 향후 연구방향

본 논문에서는 국내의 전자책 표준인 EBKS를 기준으로 해외 전자책 표준인 OEB PS와 JepaX간의 변화에 대해 살펴보았다. 이러한 변화는 결국에는 구조정보를 지닌 문서와 그렇지 않은 문서 간의 변화, 구조문서 간의 변화으로 귀결된다. 앞으로 각국에서 전자책이 많이 제작되는 상황을 고려해 볼 때 본 논문에서 살펴본 변화기법에 바탕을 둔 매체 도구의 개발이 시급히 요구되고 있다. 또한 현재 전자책을 제작할 때 대부분 상용 워드프로세서를 이용하는 현실을 고려할 때 상용 워드프로세서로 제작되어진 형식과 EBKS간의 변화기법에 대한 연구가 필요하다.

5. 참고문헌

- [1] 문화관광부, "한국전자책 산업 발전방안 연구," 2000.
 - [2] 손원성, 고승규, 이경호, 김성혁, 임순범, 최윤철, "전자책 패권 기술과 발전 동향," *한국멀티미디어학회지* 제5권 제1호, 한국멀티미디어학회, 2001.3.
 - [3] EBKE (Book of Korea) Consortium, "A Study of Korean Standardization of eBook documents," Technical Report, 2001.
 - [4] Open eBook Publication Structure 1.0, Open eBook Forum (OEBF), <http://www.openebook.org>, 1999.
 - [5] Japanese Electronic Publishing Association (JEPFA), <http://www.jepfa.or.jp>, 2001.
 - [6] Dublin Core Metadata Element Set, Version 1.1: Reference Description, <http://purl.org/dc/elements/1.1/>.
 - [7] ISO 12083 DTD, Information and Documentation - Electronic Manuscript Preparation and Markup, Geneva, Switzerland, 1993.
 - [8] Text Encoding Initiative (TEI), <http://www.tei-c.org/>.
 - [9] DocBook, <http://www.oasis-open.org/docbook/>