

# 전자책 문서 변환을 위한 컨텐츠 대응 관계에 관한 연구

## A Study on Mapping Relations between eBook Contents for Conversion

고승규(Seung-Kyu Ko)\*, 임순범(Soon-Bum Lee)\*\*, 김성혁(Sung-Hyuk Kim)\*\*,  
최윤철(Yoon-Chul Choy)\*

### 초 록

전자책은 디지털의 여러 장점으로 인해 널리 사용되기 시작하고 있으며, 여러 시장조사 기관에서도 전자책 시장이 조만간에 급격히 성장할 것으로 예측하였다. 그러나 예상과는 달리 보안 문제와 다양한 전자책 컨텐츠 포맷 등의 이유로 시장의 활성화가 늦어지고 있는 실정이다. 보안과 관련된 문제는 DRM을 이용하여 해결하고 있으며, 다양한 포맷과 관련된 문제는 세계 각국에서 나름대로의 표준을 정의하여 해결하고자 노력하고 있다. 그러나 각국에서 제정한 표준은 각국 내부의 다양한 포맷 문제는 해결할 수 있으나 국가 간의 전자책 교환이나 가공을 어렵게 하였다. 따라서 이들 표준 컨텐츠 간의 공유나 교환을 위하여 표준 간의 변환기법이 필요하다. 이에 본 논문에서는 전자책 표준 간의 변환을 위하여 전자책 컨텐츠 간에 발생할 수 있는 대응 관계에 관하여 연구하였다. 먼저 전자책 컨텐츠 간의 대응 관계를 정의하기 위하여 전자책 컨텐츠 간의 대응과 일반 XML 문서에서 발생할 수 있는 대응을 분석하여 일곱 가지의 대응 관계를 정의하였다. 그리고 각각을 사상 수에 따라 분류하고 자동으로 생성할 수 있는 것과 그렇지 않은 것으로 구별하였다. 그리고 이를 이용하여 전자책 컨텐츠 간의 변환을 자동, 수동, 반자동으로 분류하였다. 또한 정의된 대응 관계를 사상 수에 따라 변환 클래스로 나누고, 각각에 대하여 변환 템플릿을 정의하여 변환 스크립트를 자동으로 생성하도록 하였다. 그리고 이를 실제 전자책 컨텐츠 변환에 적용하여 보았다. 본 논문에서 정의된 대응 관계와 변환 템플릿은 전자책 컨텐츠 간의 변환 뿐 아니라 일반 XML 문서 간의 변환에도 적용될 수 있으리라 예상된다.

### Abstract

By virtue of diverse advantages derived from digital media, eBook is getting started to use. And many market research agencies have predicted that its market will be greatly expanded soon. But against those expectations, copyright-related problems and the difficulties of its accessing inherited from various eBook content formats become an obstacle to its diffusion. The first problems can be solved by DRM technology. And to solve the second problems, each nation has published its own content standard format. But the domestic standards are useful only the domestic level, they still leave the problems in the national level. The variety of content formats has created a demand for mechanisms that allow the exchange of eBook contents. Therefore we study the mapping relations between eBook contents for conversion. To define the mapping relations, first we extract the mappings both between eBook contents and between normal XML documents. From those mappings, we define seven mapping relations and classify them by cardinality. And we analyze the classified relations, which can be generated by automatic, or not. Using these results, we also classify the eBook content conversion as automatic, semi-automatic, and manual. Besides, we provide the conversion templates for mapping relations for automatic generation of conversion scripts. To show the feasibility of conversion templates, we apply them to the eBook content conversion. Experiment shows that our conversion templates generate the conversion scripts properly. We expected that defined mapping relations and conversion templates can be used not only in eBook content conversion, but also in normal XML document conversion.

본 연구는 2002년도 정보통신부지원 대학기초연구지원사업에 의해 수행되었음. (CI-2002-076-0-3)

\* 연세대학교 컴퓨터 과학부

\*\* 숙명여자대학교 정보과학부

## 1. 서론

기존 종이책의 디지털 형태인 전자책은 디지털의 여러 장점으로 인해 점차 널리 사용되기 시작하고 있다. 실제로 앤더슨 컨설팅 사에서는 전자책 시장이 2005년에 23억 달러(2), IDC(International Data Corporation)에서는 미국 시장에서만 2004년에 4억1천4백만 달러에 달할 것(6)으로 예측하였다. 그러나 이러한 예상과는 달리, 저작권 문제와 다양한 콘텐츠 포맷으로 인하여 전자책 시장의 활성화가 늦어지고 있다. 현재, 저작권과 관련된 문제는 DRM을 이용하여 해결하는 방향으로 나아가고 있으나, 다양한 콘텐츠 포맷에 대한 문제는 아직까지 해결되지 않고 있다. 세계 각국에서는 이 문제를 해결하기 위하여 나름대로의 콘텐츠 표준 포맷을 제정하였으나 이는 각국 내부에서 사용되는 다양한 포맷만을 해결할 수 있으며, 각국의 표준 콘텐츠 간의 교환 및 가공을 어렵게 하였다. 이 문제를 해결하기 위해서는 각국의 표준을 망라하는 교환을 위한 표준을 제정하거나 표준 간의 변환 방법을 제공해야 한다. 변환 시중간 모델을 제정하여 변환하는 것이 일반적인 변환 방법이다. 그러나 각국의 전자책 표준은 각국의 실정에 맞게 제정되어 각 표준들에서 정의하는 전자책의 구조가 다르고, 따라서 관련된 전자책 표준을 전부 표현할 수 있는 중간모델을 정의하는 것은 쉽지 않다. 또한 이는 또 다른 표준을 제정하는 것과 유사하며, 각국의 표준이 전부 표현되어야 하므로 일관성도 부족하게 된다. 따라서 현재 상황에서 빠르고 정확한 표준 간의 변환을 지원하기 위해서는 표준 간의 변환 기법을 제공하

는 것이 바람직하다. 그리고 어느 방법을 선택하여도 콘텐츠 간의 대응 관계는 발생한다.

현재 전자책 콘텐츠 표준은 한국, 미국, 일본에서 제정되었으며, 중국 등 다른 나라에서도 진행 중에 있다. 현재 제정된 표준은 전부 XML(12)에 기반하여 제정되었다. XML 문서와 일반적인 문서와의 차이점은 XML에서는 명확한 문서 구조를 정의하고 있다는 점이다. 이러한 문서 구조는 문서의 일관성 있는 해석을 가능하게 하여 명확한 교환을 할 수 있다. 그러나 발표된 전자책 표준들은 각국의 전자책 시장의 현실만을 반영하여 자기 문서 구조가 상이하다. 이 문서 구조의 상이성은 전자책 문서 간의 변환을 어렵게 만드는 주된 요인이다.

이에 본 논문에서는 전자책 콘텐츠 간의 변환을 위하여 콘텐츠 간의 대응 관계를 정의하고, 이를 정보량의 측면에서 분류하였다. 그리고 분류된 각 대응 관계를 자동, 반자동, 수동으로 생성할 수 있는지 분석하여 전자책 콘텐츠 간의 변환을 분류하였다. 또한 본 논문에서는 분류된 각 대응 관계에 변환 템플릿을 제공하여 대응 관계로부터 변환 스크립트를 자동으로 생성할 수 있도록 하였다. 정의된 대응 관계 및 변환 템플릿은 전자책 콘텐츠 변환뿐 아니라 일반적인 XML 문서 변환에서도 이용될 수 있으리라 예상된다.

본 논문의 구성은 다음과 같다. 2장에서는 전자책 표준과 전자책 콘텐츠의 특성에 대해 살펴보고, 3장에서는 전자책 콘텐츠에서 발생할 수 있는 대응관계를 정의한다. 4장에서는 정의된 대응관계를 이용하여 변환 템플릿을 정의하고, 이를 전자책 콘텐츠 변환에 적용하였다. 마지막으로 5장에서는 결론에 대해 기술한다.

## 2. 관련연구

본 장에서는 전자책 콘텐츠 표준에 관해 간략히 설명하고, 전자책 콘텐츠 특성에 대하여 살펴본다.

### 2.1 전자책 콘텐츠 표준

현재 전자책 콘텐츠 표준은 한국의 EBK(eBook of Korea)에서 발표한 EBKS(eBook of Korea Standard)(4)(15), 미국의 OEBF(Open eBook Forum)에서 발표한 OEB PS(Open eBook Publication Structure)(7), 그리고 일본의 Jepa(Japanese Electronic Publishing Association)에서 발표한 JepaX(5)가 있다. 각 표준에 대해 간략히 살펴보면 다음과 같다.

- EBKS: 전자책 콘텐츠의 명확한 교환을 위해 2001년에 제정되었으며, 콘텐츠, 메타데이터, 스타일로 구성된다. 메타데이터는 더블린 코어(3)의 15가지 요소를 사용하고, 내용은 "cover", "front", "book", "back"으로 구성된다. 이 중 "body"는 "part", "chapter", "section"으로 구성되며, 최대 6단계까지의 깊이를 가질 수 있다. EBKS는 문서의 세부 구성까지 명확하게 DTD에서 정의하고 있다. 스타일은 XSL-FO(13)를 표준이 아닌 권고 사항으로 제시하고 있다.
- OEB PS: 전자책이 시장에 효과적으로 적용되는 것을 목표로 1999년 제정된 OEB PS는 패키지와 내용으로 구성된다. 패키지는 더블린 코어로 표현되는 메타데

이터와 문서구성 정보를 기술하고 있으며, 내용은 기본 문서 또는 확장 문서로 표현된다. 기본 문서는 XHTML(11)과 유사하며, 확장문서는 임의의 XML 문서가 된다. 스타일 정보는 CSS1(9)를 바탕으로 CSS2(10)의 일부 기능을 추가하여 정의되었다. OEB PS는 2003년에 버전 1.2로 갱신되었다.

- JepaX: 전자책의 저장과 교환을 목적으로 1999년에 제정되었으며, 전체 구조는 메타데이터인 "bookinfo", 내용인 "cover", "front", "body", "back", 그리고 기타 정보를 표현하는 "jepainfo"로 구성된다. JepaX에서는 상위 구조만을 정의하고 그 하위 구조는 사용자가 정의하도록 하고 있다. 메타데이터는 자체적으로 이חס 가치를 정의하였고, 스타일 정보는 업체마다 적합한 스타일을 사용할 수 있도록 정의하지 않고 있다.

### 2.2 전자책 콘텐츠의 특성

콘텐츠 변환 측면에서의 전자책 표준 콘텐츠의 특성은 XML 문서로써 갖는 공통적인 특성과 각기 지니고 있는 개별적인 특성으로 나눌 수 있다.

XML 문서로써 갖는 공통적인 특성은 다음과 같다.

- XML 문서 종류마다 정의하는 문서 구조가 다르다. 예를 들어 책의 구조는 "메타정보"와 "본문"으로 정의되거나 "표지", "본문" 또는 "서론", "본론", "결론"으로도 정의될 수 있다. 이는 각 문서마다 표현하

는 정보의 영역이 일치하지 않기 때문에 발생한다.

- 같은 이름이라도 위치마다 의미가 다를 수 있다. 예를 들어 “name”이라는 엘리먼트는 저자 정보에 포함될 경우에는 저자 명을, 출판사 정보에 포함될 경우에는 출판사명을 의미한다. 이와 같이 위치에 따라 의미가 달라지는 엘리먼트가 XML 문서에서는 다수 존재한다.

이러한 점 때문에 일반적인 XML 문서에서 대응 관계를 자동으로 정의하기 어렵다. 문서 구조가 상이한 문서는 전문가가 분석하여 정의하면 해결될 수 있으며, 위치마다 의미가 달라지는 문서는 대응 관계 정의 시 경로를 이용하면 해결될 수 있다.

표준 전자책 콘텐츠로써 갖는 특징을 살펴보면 다음과 같다.

- 범위  
세 표준 모두 내용 부분과 메타데이터 부분을 정의하고 있다. 이 외에 OEB PS는 패키지라고 하여 내용을 어떻게 처리하는지에 대한 정보를 따로 갖고 있으며, JepaX는 스타일에 관한 정보를 포함하지 않고 있다. 이에 반해 EBKS는 스타일 정보를 정의하지는 않고 권고만 하고 있다.
- 문서 구조  
세 표준에서 정의하는 문서 구조의 수준이 전부 다르다. 먼저 EBKS에서는 구체적인 부분까지 정의하고 있으며, OEB PS에서는 문서구조를 거의 표현하지 않는다. 이에 비해 JepaX에서는 상위 부분에서만 문서 구조를 정의하고 하위 부분은 사용자가 정의하도록 한다. 이는

EBKS와 OEB PS의 중간 형태이다. 또한 EBKS와 JepaX에서는 사용자가 문서 구조를 확장할 수 있는 기능을 제공한다. 이 기능은 변환 시 새로운 대응 관계를 정의하게 한다.

전자책 콘텐츠 간에 발생하는 대응 관계는 위의 두 가지 전자책 콘텐츠 특징을 모두 반영해야 한다.

### 3. 대응 관계

전자책 표준 콘텐츠 간에 발생할 수 있는 대응 관계를 살펴보면 다음과 같다.

#### 3.1 대응 관계

전자책 콘텐츠에서 발생할 수 있는 대응들은 (1)[8]에서 정의되었다. (1)[8]에서 정의된 대응 관계를 좀 더 세분하고, 다른 XML 문서 간에 발생할 수 있는 대응들을 고려하면 XML 문서 간에 발생할 수 있는 대응 관계는 다음의 일곱 가지로 정의할 수 있다.

- 일치(Exactness)  
같은 정보를 표현하고 문서 구조도 같은 경우가 이에 해당한다. 엘리먼트의 이름은 같을 수도 있고 다를 수도 있다.
- 확장(Extension)  
원시 문서에서 하나의 엘리먼트로 표현되는 정보가 목적 문서에서는 여러 엘리먼트로 표현되는 경우가 이에 해당한다. 이 대응은 목적 문서가 원시 문서보다 더 많은 정보량을 갖고 있음을 의미한다.

- 축소(Reduction)
 

“확장”의 반대되는 경우로, 원시 문서에서 표현되는 정보를 표현하기 위하여 목적 문서에서는 원시 문서보다 더 적은 엘리먼트가 필요한 경우가 이에 해당한다.
- 생략(Omission)
 

원시 문서에서 표현하는 정보가 목적 문서에서는 표현되지 않는 경우이다. 즉, 원시 문서의 정보가 목적 문서에서는 필요하지 않다는 것을 의미한다.
- 대치(Replacement)
 

원시 문서의 엘리먼트에 해당하는 목적 문서의 엘리먼트가 존재하지 않을 경우에 사용되는 방법 중에 하나이다. 이 경우는 다음의 세가지 대응 관계 중의 하나로 표현된다. 첫번째는 새로운 엘리먼트를 정의하여 표현하는 “생성”이고, 두번째는 특정 엘리먼트의 내용으로 그 내용을 표현하는 “대치”이다. 세번째는 목적 문서에서 표현되지 않는 “생략”이다. HTML에 “author”라는 엘리먼트가 존재하지 않으므로 <h4>author</h4>와 같이 대응시키는 것이 “대치”의 예이다.
- 생성(Creation)
 

원시 문서의 정보를 표현하기 위하여 목적 문서에서 새로운 엘리먼트를 정의하는 경우가 이에 해당한다. 확장 방법이 표준마다 제각기 다르기 때문에 사용자 작업을 필요로 한다.
- 선택(Selection)
 

원시 문서에서 표현되는 정보에 해당하는 목적 문서의 엘리먼트가 여러 개일 경우가 이에 해당한다. 예를 들어 원시 문서의

“title”은 HTML 문서로 변환할 경우에 “h1”~“h6”중의 하나로 대응될 수 있다. 따라서 사용자의 선택이 필요하다.

### 3.2 전자책 콘텐츠 간의 대응관계

정의된 대응 관계를 이용하여 실제 전자책 콘텐츠 간의 대응을 EBKS를 중심으로 살펴보면 다음과 같다. 이 때 EBKS에서는 스타일을 권장하고, JapaX에서는 정의하지 않고 있으므로, 대응 관계를 메타데이터와 내용 측면에서 살펴본다.

#### EBKS 문서와 OEB PS 문서

EBKS는 문서 구조를 명확하게 정의하고, OEB PS는 구체적인 문서 구조를 지니고 있지 않다. 따라서 이들 간의 변환은 구조 문서와 비구조 문서 간의 변환이 된다. 먼저 EBKS 문서를 OEB PS 문서로 대응시킨 결과는 <표 1>과 같다. <표 1>을 살펴보면 정의한 대응 관계 중에서 여섯 개의 대응 관계가 발생하였다. OEB PS에서 EBKS의 자세한 문서 구조를 표현하기 위해서는 “생성”이 아닌 “대치”의 관계가 발생한다. 이는 OEB PS에서는 새로운 문서 요소를 정의할 수 없기 때문이다. <표 1>에서 “선택”의 경우를 살펴보면, “title”이 위치에 따라 의미가 달라지므로 경로를 이용하여서 대응 관계를 표현하고 있다. 메타데이터의 경우는 EBKS와 OEB PS가 전부 더블런 코어에 기반하고 있기 때문에 “일치”의 관계가 성립한다.

역으로 OEB PS 문서의 EBKS 문서로의 변환은 비구조 문서를 구조 문서로 변환하는 경

우이다. 비구조 문서인 OEB PS에서는 구체적인 구조 정보를 표현하는 엘리먼트가 없기 때문에 하나의 엘리먼트가 여러 의미로 사용될 수 있다. 예를 들어 같은 "h4"의 경우는 저자명이나 제목 등의 다중의 의미로 사용될 수 있다. 따라서 이 경우에는 대응 관계를 스키마 수준에서 정의하기 어렵고, 각 문서 객체마다 대응 관계를 정의해야 한다.

〈표 1〉 EBKS문서에서 OEB PS 문서로의 대응

구분	EBKS	OEB PS
일치	emph. p	em. p
확장	ebook	html, head, body
축소	artwork, fig	img
생략	ebks	-
대치	preface	<h4>PREFACE</H4>
선택	cover.title	h1
	section.title	h4

EBKS 문서와 JepaX 문서

JepaX는 상위 구조만 지니고 있고, 하위 구조는 사용자가 정의해야 한다. 따라서 JepaX의 구조 수준은 EBKS와 OEB PS의 중간 수준으로 볼 수 있다. 먼저 EBKS 문서를 JepaX 문서로 대응시킨 결과는 〈표 2〉와 같다.

〈표 2〉에서 위의 네 항목은 콘텐츠 간의 대응 관계이고, 아래 두 항목은 메타데이터 간의 대응관계이다. EBKS는 세밀한 문서 구조를 지니고 있으므로, JepaX로 대응시킬 경우에 "생

〈표 2〉 EBKS 문서에서 JepaX 문서로의 대응

구분	EBKS	JepaX
일치	emph, vita	em, <div type="vita">
생략	titlegrp	-
확장	title	head,head.title
생성	subject1	<div xtype="ebks:subject1">
일치	dc:creator	book-author
생성	dc:subject	<div xtype="dc:subject">

성"이 다른 관계보다 많이 발생한다. "생성"과 같은 경우에는 일관성 있는 해석이 가능하도록 하기 위하여 네임스페이스를 이용하여 생성되는 엘리먼트의 출처를 표현한다. 〈표 2〉의 경우에는 첫번째 "생성"의 경우에는 "subject1"이 EBKS에서 유래되었음을 나타내고, 두번째 경우에는 더블린 코어에서 유래되었음을 나타낸다.

역으로 JepaX 문서를 EBKS 문서로 변환하는 경우는 〈표 3〉과 같다.

〈표 3〉 Jepax문서에서 Ebks 문서로의 대응

구분	JepaX	EBKS
일치	front, p, em	front, p, emph
확장	title	titlegrp.title
축소	float, res	uri
선택	key	keyword:keyphrase
생략	head	-
생성	ruby	dc:identifier DTD 확장
일치	isbn	dc:identifier
생성	edition	<x-metadata name="jepax:edition">

〈표 3〉에서 위의 여섯 항목은 콘텐츠 간의 대응 관계이고, 아래의 두 항목은 메타데이터 간의 대응 관계이다. 콘텐츠의 경우는 여섯개의 대응 관계가 발생하였고, 메타데이터의 경우는 “대치”를 제외한 EBKS 문서를 JepaX로 변환하는 경우와 같이 “일치”와 “생성”이 발생하였다.

이와 같이 EBKS 문서를 중심으로 다른 두 표준 콘텐츠 간의 발생하는 대응관계는 정의된 일곱 가지 관계에 전부 속한다. 따라서 정의된 일곱 대응 관계는 전자책 콘텐츠 간의 대응 관계를 전부 표현할 수 있다.

#### 4. 변환 템플릿 및 실험

본 장에서는 정의된 대응관계를 이용하여 자동으로 변환 스크립트를 생성하기 위하여 변환 템플릿을 정의하였고, 이를 이용하여 전자책 콘텐츠를 변환하여 보았다.

##### 4.1 변환 템플릿

변환 템플릿을 정의하기 위하여 대응 관계를 사상 수에 따라 분류하면 〈표 4〉와 같은 열 가지의 변환 클래스로 분류할 수 있다.

〈표 4〉에서 “일대일” 클래스는 세가지 부클래스로 나뉜다. 첫번째는 일대일 대응인 “일대일”, 두번째는 원시 문서의 엘리먼트가 목적 문서의 특정 엘리먼트의 값으로 대응되는 “일대값”, 그리고 세번째는 원시 문서의 엘리먼트에 대해 목적 문서의 여러 엘리먼트가 대응되는

“일대선택”이다. “일대선택”과 “일대다”의 차이점은 “일대선택”은 여러 목적 엘리먼트 중에서 하나만 대응되고, “일대다”는 여러 엘리먼트가 함께 대응된다. 그리고 “다대다” 클래스는 실제 콘텐츠 대응에서는 발생하지 않지만 사상 수에 따라 클래스를 분류하면서 자연스럽게 정의된 것이다.

〈표 4〉 변환 클래스

클래스	대응관계	변환
영대일, 일대다	생성	EBKS&JepaX
일대영	생략	EBKS&JepaX, EBKS2OEB
일대일	일대일	EBKS&JepaX, EBKS2OEB
	일대값	EBKS2OEB
	일대선택	JepaX2EBKS, EBKS2OEB
일대다	확장	EBKS&JepaX, EBKS2OEB
다대영	생략	EBKS&JepaX, EBKS2OEB
다대일	축소	JepaX&2EBKS, EBKS2OEB
다대다	선택	

변환 클래스로부터 변환 스크립트를 자동으로 생성하기 위하여 변환 템플릿을 정의하였다. 변환 클래스는 대응 관계를 원시 엘리먼트와 목적 엘리먼트에 따라 정렬함으로써 얻을 수 있다. 변환 템플릿을 XSLT[14]의 예를 이용하여 설명하면 다음과 같다.

① 영대일, 영대다

Ex) from <subject1> of EBKS to JepaX

```
<xsl:template match="subject1">
  <xtype-desc element="div" value="
    ebks:subject1">
    first sub-structure level of
    section (EBKS)
  </xtype-desc>
  <div xtype="ebks:subject1">
    <xsl:apply-templates />
  </div>
</xsl:template>
```

② 일대영

이 경우는 변환할 필요가 없으므로 변환 스크립트를 정의할 필요가 없다.

③ 일대일

a. 일대일

ex) from <cover> of EBKS to JepaX

```
<xsl:template match="cover">
  <cover>
    <xsl:apply-templates />
  </cover>
</xsl:template>
```

b. 일대값

Ex) from <preface> of EBKS to OEB PS

```
<xsl:template match="preface">
  <h4> Preface</h4>
  <p>
    <xsl:value-of select="preface" />
  </p>
  <xsl:apply-templates />
</xsl:template>
```

c. 일대선택

"일대일" 과 유사하지만 사용자의 선택이 필요하다.

④ 일대다

이 경우는 목적 문서로 변환할 때 연속된 엘리먼트의 집합으로 변환되는 것과 템플릿 형태로 변환되는 것으로 나눌 수 있다.

Ex1) from <ebks> of EBKS to OEB PS

```
<xsl:template match="ebks">
  <html>
    <head>
      <xsl:apply-templates
        select="title" />
    </head>
    <body>
      <xsl:apply-templates />
    </body>
  </html>
</xsl:template>
```

Ex2) from <title> of JepaX to EBKS

```
<xsl:template match="title">
  <titlegrp>
    <title>
      <xsl:apply-templates />
    </title>
  </titlegrp>
</xsl:template>
```

⑤ 다대영

이 경우는 "일대영" 의 경우와 같다.

⑥ 다대일

이 경우는 "일대일" 이 여러 개 있는 경우와 같다.

⑦ 다대다

이 경우는 "일대선택" 의 집합으로 생성될 수

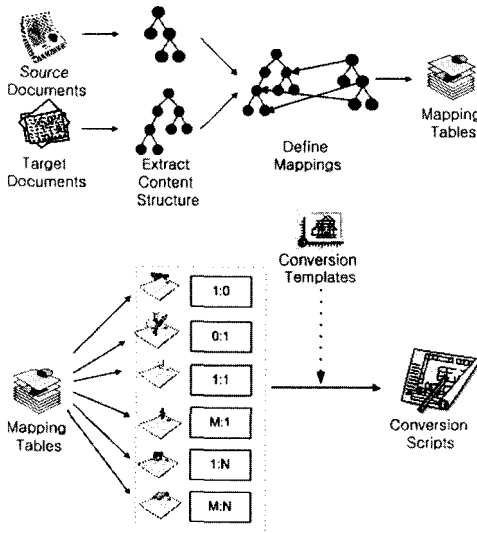


있다.

## 4.2 실험 및 결과

변환 템플릿을 이용하여 전자책 변환을 수행하는 과정은 <그림 1>과 같으며, 구체적인 변환 흐름은 다음과 같다.

- i) 원시 문서와 목적 문서를 분석하여 문서 구조를 추출한다.
- ii) 두 문서 구조 간의 대응을 정의한다.
- iii) 대응으로부터 대응 관계를 추출하고, 원시 엘리먼트와 목적 엘리먼트에 따라 정렬한다.
- iv) 정렬된 대응 관계를 변환 클래스로 바꾼다.
- v) 변환 템플릿을 이용하여 각 변환 클래스에 대해 변환 스크립트를 생성한다.



<그림 1> 변환 흐름도

이러한 변환 방법을 EBKS 문서를 OEB PS 문서로의 변환에 적용하여 보았다. 이 때, 생성

된 변환 스크립트는 <그림 2>와 같고, 원시 문서와 목적 문서는 각각 <그림 3>과 <그림 4>와 같다.

```

- <xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">
- <xsl:template match="/" />
- <xsl:apply-templates select="ebook" />
</xsl:template>
<xsl:template match="ebook">
- <HTML>
- <head>
- <TITLE>
- <xsl:value-of select="books/book/cover/titlegrp/title" />
</TITLE>
</head>
- <body>
- <xsl:apply-templates />
</body>
</HTML>
</xsl:template>
+ <xsl:template match="metainfo">
+ <xsl:template match="books">
- <xsl:template match="book">
- <xsl:apply-templates select="cover" />
- <xsl:apply-templates select="front" />
- <xsl:apply-templates select="body" />
- <xsl:apply-templates select="back" />
</xsl:template>
+ <xsl:template match="cover">
+ <xsl:template match="cover/artwork">
- <xsl:template match="cover/titlegrp">
- <xsl:apply-templates select="/title" />
- <xsl:apply-templates select="/subtitle" />
</xsl:template>
    
```

<그림 2> 변환 스크립트

<그림 2>를 살펴보면 “확장”의 관계인 “ebook”이 “일대다” 변환 스크립트로 표현되어 있음을 알 수 있다. 그리고, 그 하위의 엘리먼트들은 나름대로의 변환 스크립트가 적용됨을 알 수 있다. <그림 4>를 살펴보면, <그림 3>의 “title”, “titlegrp”, “preface”처럼 “생략”, “일치”, “대치”에 해당하는 엘리먼트들이 적절히 변환되었음을 알 수 있다. 따라서 생성된 변환 스크립트는 EBKS 문서를 OEB PS 문서로 적절히 변환시키고 있음을 알 수 있다.

```

- <ebook>
  <metainfo />
  <books>
  - <book>
    - <cover>
      <artwork unit="96" un="cover.jpg" />
    - <titlegrp>
      <title>공통 영어</title>
      <subtitle>대학교 입학자격 검정고시 대비</subtitle>
      <subtitle>대입 검정고시 100% 합격 Series</subtitle>
    </titlegrp>
    - <authorgrp>
      - <author>
        <name>박남화</name>
        <role>편자</role>
      </author>
    </authorgrp>
    <note id="11">
    </cover>
  - <front>
  - <preface>
  - <subject1>
    <p>획정된 21세기를 맞이하여 광명일외국어학원이 탄생합니다. 21C, 보다 나은 미래를 준비하고 싶은 학생들을 위하여, 다양한 기술과 능력을 요구하는 이시대에 외국어 구술능력은 이제 필수입니다.</p>
    <p>광명일외국어학원은 체계적이고 과학적인 어학프로그램의 편성과 학생들을 가르치던 생각하는바듯하고 철저한 준비과정의 조성, 집중하고 체계한 실내환경 등 모든 조건에서 선발 앞서도록 노력한 결과 많은 인재를 배출하였습니다.</p>
  
```

〈그림 3〉 원시 EBKS 문서

```

<HTML>
- <head>
  <TITLE>공통 영어</TITLE>
</head>
- <body>
  chr />
  <center>
    
  </center>
  <h1 align="center" color="blue">공통 영어</h1>
  <br />
  <h5 align="center">대학교 입학자격 검정고시 대비</h5>
  <h5 align="center">대입 검정고시 100% 합격 Series</h5>
  <h5 align="right">박남화 : 편자</h5>
  <blockquote>획정된 교육 내용을 쉽고 간결하게 정리 각 단원마다 예상 문제 명선 수록 문제 해 따른 상세한 해설 제공으로 편자 노력</blockquote>
  <hr />
  <br />
  <h1 align="center">머리말</h1>
  <br />
  <p style="text-indent:12pt">획정된 21세기를 맞이하여 광명일외국어학원이 탄생합니다. 21C, 보다 나은 미래를 준비하고 싶은 학생들을 위하여, 다양한 기술과 능력을 요구하는 이시대에 외국어 구술능력은 이제 필수입니다.</p>
  <p style="text-indent:12pt">광명일외국어학원은 체계적이고 과학적인 어학프로그램의 편성과 학생들을 가르치던 생각하는바듯하고 철저한 준비과정의 조성, 집중하고 체계한 실내환경 등 모든 조건에서 선발 앞서도록 노력한 결과 많은 인재를 배출하였습니다.</p>
  <p style="text-indent:12pt">안으로도 저희 광명일외국어학원은 최상의 어학프로그램을 위한 다양한 연구와 개발을 통해 어학전문기관으로서의 역
  
```

〈그림 4〉 목적 OEB PS 문서

## 5. 결 론

변환이나 통합 작업에서 정의된 대응 관계를 이용하기 위하여 대응 관계를 자동으로 생성할

수 있는 것과 그렇지 않은 것으로 분류하면 (표 5)와 같다.

〈표 5〉 변환 클래스 생성 방법

변환 클래스		연산
영대일, 영대다		수동
일대영		자동
일대일	일대일	자동
	일대값	반자동
	일대선택	반자동
일대다		자동
다대영		자동
다대일		자동
다대다		반자동

“영대일”과 “영대다”는 “생성”의 관계로 수작업이 필요하고, “일대값”과 “일대선택”은 “대치”와 “선택”의 관계로 사용자의 선택이 필요하다. 그리로 나머지 클래스는 사용자의 작업이 필요 없다. 〈표4〉와 〈표5〉를 구조 정보의 측면에서 살펴보면 전자책 변환을 〈표 6〉과 같이 분류할 수 있다.

〈표 6〉 전자책 콘텐츠 변환

변환 방향	변환 방법	생성 방법
하향 변환	EBKS to OEB PS	(반)자동
동등 변환	EBKS 와 JepaX	반자동
상향 변환	OEB PS to EBKS	수동

구조 정보의 측면에서 보면 구조 정보를 지닌 EBKS와 JepaX 간의 변환은 동등 변환으

로 볼 수 있으며, 구조 정보가 없는 OEB PS와의 변환은 상향 또는 하향 변환으로 볼 수 있다. 그리고 각 변환의 생성 방법은 <표 4>와 <표 5>를 참조하면 자동과 반자동, 수동으로 나눌 수 있다. 초기값으로 사용자의 선택을 제공하면 "일대값"과 "일대선택"은 자동으로 간주할 수 있으며, 이는 하향 변환은 자동으로 생성될 수 있음을 의미한다.

본 논문에서는 전자책 콘텐츠 변환에서 발생할 수 있는 대응 관계를 정의하고 이를 사상수에 따라 분류하였다. 그리고 각각에 대하여 변환 템플릿을 정의하여 자동적인 변환 스크립트 생성이 가능하도록 하였다. 그리고 이를 전자책 콘텐츠에 적용한 결과 콘텐츠를 적절히 변환시킴을 보여주었다. 또한 각 대응 관계를 생성하는 방법에 따라 자동, 수동, 반자동으로 분류하였고, 이를 이용하여 전자책 콘텐츠 간의 변환을 자동, 반자동, 수동으로 분류하였다. 본 논문에서 정의된 대응 관계와 변환 템플릿은 전자책 콘텐츠 간의 변환 뿐 아니라 일반 XML 문서 간의 변환에도 적용될 수 있으리라 예상된다.

---

## 참고 문헌

---

- [1] 고승규, 손원성, 임순범, 최윤철, "전자책 표준 콘텐츠 변환에 관한 연구", 정보처리학회 논문지, 정보처리학회 논문지 D-10, 2003.4.
- [2] AAP/Andersen Consulting Ebook Study, <http://www.publishers.org/dec2000anderson.ppt>, 2000
- [3] Dublin Core Metadata Element Set, Version 1.1: Reference Description, Internet RFC 2413, <http://purl.org/dc/elements/1.1>, 2001.
- [4] EBK (e-Book of Korea) Consortium, A Study of Korean Standardization of eBook documents, Technical Report, 2001.
- [5] Japanese Electronic Publishing Association (JEPA), <http://www.jepa.or.jp>, 2001.
- [6] Michael Bartlett, E-Book Market Set For Explosion - IDC Study, Newsbytes, 2000.12.19.
- [7] Open eBook Forum, Open eBook Publication Structure 1.0, <http://www.openebook.org>, 1999.
- [8] Seung-Kyu Ko, Won-Sung Sohn, Kyong-Ho Lee, Soon-Bum Lim, Yoon-Chul Choy, "A Study of Content Conversion between eBook Standards.", EC-Web, France, Sep. 2002, LNCS 2455, pp. 273-283.

Springer.

- [9] W3C Consortium, Cascading Style Sheets (CSS) level 1.0, <http://www.w3.org/TR/REC-CSS1-961217>, 1996.
- [10] W3C Consortium, Cascading Style Sheets level 2 CSS2, <http://www.w3.org/TR/1998/REC-CSS2-19980512>, 1998.
- [11] W3C Consortium, Extensible Hypertext Language (XHTML), <http://www.w3.org/TR/2000/REC-xhtml1-20000126>, 2000.
- [12] W3C Consortium, Extensible Markup Language (XML) 1.0, <http://www.w3.org/TR/1998/RECxml-19980210>, 1998.

- [13] W3C Consortium, Extensible Stylesheet Language (XSL), <http://www.w3.org/TR/2000/CR-xsl-20001121>, 2000.
- [14] W3C Consortium, XSL Transformations (XSLT), <http://www.w3.org/TR/1999/REC-xslt-19991116>, 1999.
- [15] Won-Sung Sohn, Seung-Kyu Ko, Kyung-Ho Lee, Sung-Hyuk Kim, Soon-Bum Lim, Yoon-Chul Choy, "Standardization of eBook documents in Korean Industry," Computer Standards & Interface, Vol. 24, pp. 45-60, 2002, Elsevier.

## 저자소개



고승규 (E-mail : pitta@rainbow.yonsei.ac.kr)  
1992. 연세대학교 컴퓨터과학과 졸업(학사)  
1994. 연세대학교 컴퓨터과학과(석사)  
2003. 연세대학교 컴퓨터과학과(박사)  
2003 ~ 현재 연세대학교 산업기술 연구소 객원 연구원  
관심 분야 : 웹문서 통합, 변환, 정보추출, 정보검색



임순범 (E-mail : sblim@sookmyung.ac.kr)  
1982. 서울대학교 계산통계학과 졸업(학사)  
1983. 한국과학기술원 전산학과 졸업(석사)  
1992. 한국과학기술원 전산학과 졸업(박사)  
1989 ~ 1992. (주)휴먼컴퓨터 이사/연구소장  
1992 ~ 1997. (주)삼보컴퓨터 부장  
2001 ~ 현재 숙명여자대학교 정보과학부 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 교수  
2000 ~ 현재 전자상거래 통합포럼 전자카타로그 기술위원회 위원  
2001 ~ 현재 EBK 전자책문서표준화위원회 위원  
관심 분야 : 디지털도서관, 텍스트인코딩, 전자카타로그, 문서구조화



최운철 (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 : 멀티미디어 문서처리(SGML/XML), 가상환경, eLearning