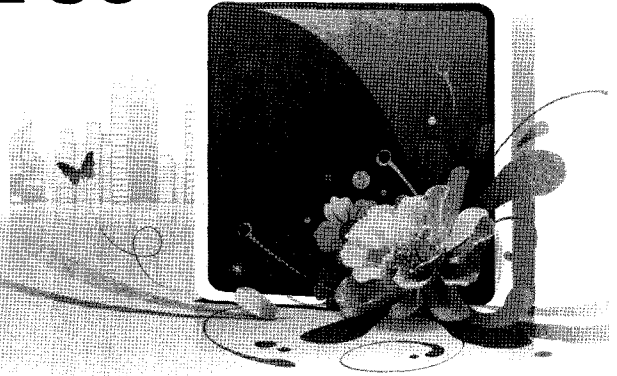


전자책 포맷 기술 및 표준 동향

이경호 | 연세대학교 컴퓨터과학과 부교수



1. 머리말

최근 들어 아마존 킨들, 애플 아이패드 등 전자책 관련 단말기의 보급이 증가하고 있다. 또한 전자책 콘텐츠의 다변화로 인해 그 어느 때보다 전자책 시장의 활성화에 대한 업계의 기대가 고조되어 있다. 전자책(electronic book)이란 책의 콘텐츠를 디지털 형태로 가공 및 저장한 출판물을 의미한다. 현재 전자책 시장이 활성화되고 있는 배경에는 전자책 단말기의 보급 확대, 전자잉크 등 가독성 높은 디스플레이 개발, 이동통신 및 무선 네트워크 서비스의 확대, 종이책 대비 전자책이 갖는 장점, 전자책 시장 선점을 위한 업체 간 경쟁, 전자출판물 시장의 급성장 등 여러 요인이 복합적으로 작용하고 있다.

PWC는 전 세계 전자책 시장이 2013년까지 89억 4,100만 달러 규모의 시장을 형성하면서 연평균 37.2% 성장할 것으로 전망하고 있다. 특히 미주권의 성장률은 23.3%인 반면 유럽권은 127.0%, 중국권은 135.2%의 높은 연평균 성장률을 보일 것으로 예측하고 있다.

이와 같은 전자책 시장 활성화에 대한 높은 기대와 관심에도 불구하고 전자책 포맷의 표준화, 저작권 보호를 위한 무단 복제 방지 기술, 전자책 상거래 및 검색을 위한 메타데이터 표준화, 그리고 높은 가독성을 지원하는 단말 및 소프트웨어 기술 등 기술적인 측면에서 선결되어야 할 과제가 남아 있다.

특히 전자책 포맷의 경우, 현재 전자책 업체마다 EPUB, PDF, AZW, HTML 등 다양한 포맷이 사용되고 있다. 이와 같이 다양한 전자책 포맷의 사용 결과, 전자책 업체에서는 각각의 포맷에 맞는 소프트웨어를 개발해야 하며, 전자책 소비자는 전자책을 읽기 위하여 여러 종류의 소프트웨어를 설치해야 한다. 또한 중요한 문제는 이러한 상황이 업체끼리의 중복 투자를 유발해 전자책 시장의 활성화를 가로막는다는 것이다.

따라서 미국과 유럽을 중심으로 전자책 포맷의 표준화에 대한 필요성을 인식하고, 도서유통업체, 전자책 솔루션 업체, 출판사 등을 중심으로 IDPF(International Digital Publishing Forum)¹⁾를 결성하고, XML 기반의 EPUB 표준을 개발했다. EPUB은 미국, 유럽 등 해외 시

1) <http://www.idpf.org/>

장에서 사실상의 전자책 표준으로 널리 사용되고 있다. 북미 시장의 경우, 전자책 사업자의 대부분은 EPUB 파일 형태의 전자책을 유통하고 있으며, 아마존 킨들을 제외한 대부분의 단말기가 EPUB 포맷을 지원한다.

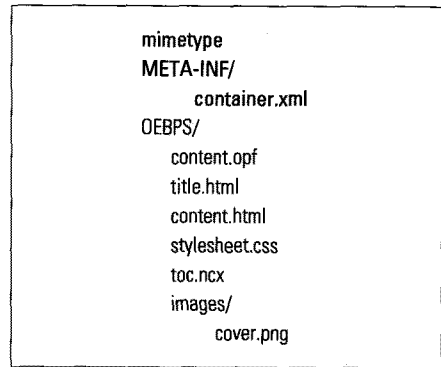
현재 우리나라도 국내 전자책 시장의 활성화, 전자출판물 관련 산업의 육성, 그리고 향후 관련 산업의 세계 시장 진출을 위해서 전자책의 표준화가 매우 필요한 실정이다. 본 고에서는 현재 전자책 업계의 사실상 표준인 EPUB의 특징을 소개하고, 국내 전자책 표준 제정의 필요성을 살펴보고자 한다.

■ 2. EPUB 개요

EPUB은 IDPF에서 제정한 XML에 기반한 개방형 전자출판물(electronic publication) 표준 포맷이다. EPUB은 자동공간조정(reflowable)이 가능한 전자출판물을 생성할 수 있도록 설계되었다. EPUB 콘텐츠는 화면 크기에 자동으로 최적화되어 보일 수 있기 때문에 전자책 단말기와 같은 모바일 기기에 적합하다. EPUB은 컨테이너 표준인 OCF(Open Container Format), 콘텐츠 마크업 표준인 OPS(Open Publication Structure), 패키징 표준인 OPF(Open Packaging Format)의 3개 명세로 구성된다.

2.1 OCF(Open Container Format)

EPUB 포맷을 구성하는 3개 명세 중에 하나인 OCF는 전자출판물의 교환, 전송, 유통, 저장을 위하여 전자출판물을 구성하는 파일들을 통합해 전송할 수 있는 방법을 제공한다. 이를 위해 전자출판물의 디렉토리 구조를 정의하고, 이를 단일의 Zip 파일로 압축하는 방법을 기술한다. Zip 파일로 압축되어 전송되는 EPUB 파



[그림 1] EPUB 파일의 디렉토리 구조의 예

일의 일반적인 디렉토리 구조는 [그림 1]과 같다.

[그림 1]과 같이 mimetype 파일은 EPUB의 mimetype인 'application/epub+zip'을 기술한다. 또한 META-INF 디렉토리는 패키징 파일([그림 1]에서 content.opf 파일에 해당)의 위치를 가리키는 container.xml 파일을 포함한다. META-INF 디렉토리는 추가로 컨테이너 수준의 메타데이터, 전자서명, 암호화, DRM(Digital Rights Management) 등에 대한 정보를 기술하는 파일을 포함할 수 있다.

2.2 OPF(Open Packaging Format)

전자출판물은 텍스트, 이미지 등의 다양한 종류의 미디어 파일로 구성된다. OPF는 전자출판물의 메타데이터 및 패키징 정보를 기술할 수 있는 명세이다. [그림 2]와 같이 OPF 패키지 문서는 출판물을 유일하게 식별할 수 있는 식별자(unique identifier), 출판물에 대한 메타데이터(metadata), 출판물을 구성하는 파일 목록인 manifest, 콘텐츠 파일 간의 출력 순서를 기술한 spine, 그리고 출판물의 구조 정보를 기술한 guide로 구성된다.

OPF 패키지 문서의 메타데이터 엘리먼트는 출판물에 관한 정보를 기술한다. OPF는 더블린코어(Dublin Core)²⁾

2) <http://dublincore.org/documents/2004/12/20/dces/>

```

<?xml version="1.0" encoding="utf-8"?>
<package xmlns="http://www.idpf.org/2007/opf"
  unique-identifier="BookId" version="2.0">
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
    <dc:title>Alice in Wonderland</dc:title>
    <dc:language>en</dc:language>
    <dc:identifier id="BookId" opf:scheme="ISBN">123456789X</dc:identifier>
    <dc:creator role="aut">Lewis Carroll</dc:creator>
  </metadata>
  <manifest>
    <item id="ncx" href="toc.ncx" media-type="application/x-dtbnx+xml"/>
    <item id="cover" href="title.html" media-type="application/xhtml+xml"/>
    <item id="content" href="content.html"
      media-type="application/xhtml+xml"/>
    <item id="cover-image1" href="images/cover.bmp"
      media-type="image/bmp" fallback="cover-image2"/>
    <item id="cover-image2" href="images/cover.png" media-type="image/png"/>
    <item id="css" href="stylesheet.css" media-type="text/css"/>
  </manifest>
  <spine toc="ncx">
    <itemref idref="cover" linear="no"/>
    <itemref idref="content"/>
  </spine>
  <guide>
    <reference href="title.html" type="cover" title="Cover"/>
  </guide>
</package>

```

[그림 2] OPF 패키지 파일의 예

메타데이터 표준을 지원하며, 이중 title, identifier, language를 필수 요소로 지정한다. 특히 creator와 contributor에 상세정보를 기술하기 위하여 MARC relator code³⁾를 지원한다.

Manifest는 콘텐츠 문서, 이미지, 스타일시트, 폰트 등 출판물을 구성하는 파일의 목록을 기술한다. 또한 출판물은 OPS가 정의한 필수 지원 포맷이 아닌 파일을 포함할 수 있다. 이러한 경우, 해당 파일을 처리하지 못하는 리딩 시스템(reading system)⁴⁾을 위해 manifest에 해당 파일에 대한 대체정보(fallback)를 표현할 수 있

다. 예를 들어, [그림 2]에서 cover.bmp를 처리하지 못하는 리딩 시스템을 위해 필수 지원 포맷의 파일에 대한 대체 정보를 기술한다.

Spine은 manifest에 명시된 파일 중에서 콘텐츠 문서들의 집합 및 이들 간의 출력 순서를 기술한다. 특히 spine 엘리먼트는 manifest에 선언되어 있는 NCX(Navigation Center eXtended) 문서를 참조한다. OPF는 내비게이션의 용이성과 보다 나은 접근성을 위해 DAISY(Digital Accessible Information System)⁵⁾ 컨소시엄의 NCX를 지원한다. DAISY 컨소시엄은 시각 장애 등으로 기존의 종이

3) <http://www.loc.gov/marc/relators/>

4) EPUB 출판물을 입력으로 받아들이고, 이를 콘텐츠 소비자에 제공하는 시스템(하드웨어 및 소프트웨어)을 의미함.

5) <http://www.daisy.org/>

책을 읽을 수 없는 독자를 위해 적절한 데이터 포맷을 개발하고자 조직되었다. NCX는 장, 절 등과 같은 출판물의 계층구조를 기술할 수 있으며, 페이지 리스트를 제공해 특정 페이지로 이동할 수 있도록 하며, 그림 목록 등 논리적 구성 요소에 대한 정보를 제공한다.

NCX의 대안으로 제공 가능한 guide는 출판물의 논리적 구성 요소를 표현한다. 예를 들어, 독자는 guide 정보를 이용해 커버, 타이틀 페이지, 목차, 색인, 용어집(glossary), 사사(acknowledgements), 헌정사(dedication), 그림 목록, 표 목록, 서문, 참고문헌, 주석 등 출판물의 논리적 구성요소로 이동할 수 있다.

2.3 OPS(Open Publication Structure)

전자책 기술이 시장에서 보다 널리 성공하기 위해서는 보다 많은 전자책이 보급되어야 한다. OPS는 전자출판물의 콘텐츠를 표현하기 위한 표준으로 콘텐츠 생성자 및 공급자와 솔루션 개발자에게 콘텐츠 표현을 위한 공통 기준을 제공한다.

OPS는 소프트웨어 도구, 기존 콘텐츠, 출판 환경 및 시장 상황 등을 고려해 XHTML 1.1⁶⁾을 필수적인 콘텐츠 기술 언어로 지정했다. 따라서 저작자는 이미 익숙한 HTML 형태의 태그를 사용해 콘텐츠를 생성할 수 있다. 한편 OPS는 모바일 기기의 제한적인 컴퓨팅 자원 등을 고려해 XHTML 1.1 모듈을 선택적으로 필수요소로 지정하였다. HTML 1.1에서 OPS가 필수로 지정한 모듈은 <표 1>과 같다. 현재의 OPS 버전은 기본적으로 사용자 상호작용을 고려하지 않으므로 폼(form) 및 스크립트(script)와 관련한 모듈을 필수로 지정하지 않는다.

OPS는 래스터 이미지 포맷으로 JPEG, PNG, 그리고

<표 1> OPS의 XHTML 필수 모듈

모듈 명	엘리먼트
Structure	body, head, html, title
Text	abbr, acronym, address, blockquote, br, cite, code, dfn, div, em, h1-6, kbd, p, pre, q, samp, span, strong, var
Hypertext	a
List	dl, dt, dd, ol, ul, li
Object	object, param
Presentation	b, big, hr, i, small, sub, sup, tt
Edit	del, ins
Bidirectional Text	bdo
Table	caption, col, colgroup, table, tbody, td,tfoot, th, thead
Image	img
Client-Side Image Map	area, map
Meta-Information	meta
Style Sheet	style
Link	link
Base	base

GIF를 필수로 지정한다. 래스터 이미지와 비교하여 접근성, 검색기능, 작은 파일 크기 등의 장점을 갖는 벡터 이미지를 지원하기 위해 SVG⁷⁾를 필수 포맷으로 지정한다. 특히 OPS는 사용자 상호작용을 포함하는 콘텐츠를 대상으로 하지 않기에 SVG의 애니메이션 및 스크립트를 지원하지 않는다. 한편 필수 포맷이 아닌 객체의 삽입을 위해 object 엘리먼트의 사용을 권장한다. 예를 들어, 오디오, 비디오, flash 등은 object 엘리먼트를 이용하여 콘텐츠 문서에 추가할 수 있다.

OPS는 XHTML 1.1과 더불어 DAISY의 DTBook(Digital Talking Book)⁸⁾을 필수 콘텐츠 기술 언어로 지정한다. DTBook은 DAISY/NISO 표준⁹⁾으로 정의된 XML 스키

6) <http://www.w3.org/TR/xhtml-modularization/>

7) <http://www.w3.org/TR/SVG11/>

8) <http://www.niso.org/workrooms/daisy/Z39-86-2005.html>

9) ANSI/NISO Z39.86-2005

마이다. 특히 DTBook은 전자책을 위해 설계되었으며, XHTML과 달리 각주(footnote), 사이드바(sidebar), 주석(annotation), 페이지 번호 등의 구조를 표현할 수 있다. 전자책에 페이지의 시작 위치를 표기할 수 있기에 종이책에서의 해당 위치를 정확히 식별할 수 있다. 따라서 교육용 출판물 또는 구조화된 콘텐츠 표현에 적합하다. 미국의 경우, 장애인교육법¹⁰⁾에 의하여 많은 출판사들이 이미 DTBook 콘텐츠를 보유하고 있다. 한편 DAISY 컨소시엄은 DTBook 스키마의 올바른 사용을 위한 지침¹¹⁾을 제공한다.

XML은 현재 다양한 분야에서 표준 포맷으로 널리 사용되고 있다. 이에 OPS는 XML의 namespace 메커니즘을 기반으로 다양한 유형의 XML 문서를 콘텐츠로 활용할 수 있는 방법을 제공한다. 예를 들어, MathML¹²⁾ 형태의 수학기호를 OPS 콘텐츠 파일에 추가할 수 있다.

OPS는 콘텐츠의 포맷팅을 기술하기 위해 CSS2에 기반한 스타일 언어를 정의한다. 특히 XHTML 1.1과 같은 이유로 CSS2¹³⁾의 일부 기능을 삭제했으며, 페이지 레이아웃, 머리말, 꼬리말을 지원하기 위해 일부 기능을 추가했다. OPS는 임베디드 폰트를 지원하기 위해 CSS2의 font-face at-rule(@font-face)를 지원한다. 임베디드 폰트는 출판물의 구성요소로 manifest에 포함된다.

3. 맺음말

EPUB 표준은 북미 및 유럽을 중심으로 주로 전자책 표준 포맷으로 빠르게 보급되고 있다. 특히 개방형 표준인 EPUB은 XML의 장점인 재사용성, 확장성, 공개성, 간결성 등을 제공할 수 있기 때문에 전자책 표준

포맷으로 적합하다. 또한 XHTML을 필수 콘텐츠 언어로 지원하기에 현재 웹에서 보편적으로 사용되고 있는 HTML 기반의 콘텐츠를 비교적 용이하게 전자책으로 변환하여 서비스할 수 있다. 특히 OPF의 NCX 및 guide를 이용하여 문서의 논리적 구조 정보를 표현할 수 있으며 스타일 언어를 이용하여 콘텐츠와 스타일을 분리할 수 있다.

한편 EPUB 표준은 제정된 지 수년이 지나 현재의 웹 표준, 웹 브라우저 기술, 시장의 요구사항 등을 충분히 반영하지 못하고 있다. 이에 IDPF는 2010년도에 EPUB의 개정을 위한 워킹그룹을 결성해 개선작업을 진행 중이다. 현재 EPUB의 개선을 위해 논의 중인 사항의 일부를 요약하면 다음과 같다.

- 전자책은 물론이고 신문 및 잡지 등 다양한 유형의 전자출판물 지원
- 리치미디어(rich media) 및 상호작용(interactivity) 지원
- 한중일 등 아시아 국가의 언어적 특수성 지원
- 정교한 수준의 페이지 레이아웃 지원
- 웹 브라우저가 구현하는 표준 기술 지원
- NCX 등 내비게이션 요소의 개선
- ONIX, PRISM 등 전자출판물 관련 메타데이터 표준 지원
- 학술 서적의 상호운용성 증대를 위한 MathML 지원

특히 EPUB은 대체정보 기술을 통한 확장 메커니즘을 제공하지만, 비디오 등 리치미디어나 퀴즈, 십자말 풀이(crossword puzzle) 등의 사용자 상호작용을 표현할 수 있는 방법을 구체적으로 기술하지 않는다. 이러

10) Individuals with Disabilities Education Improvement Act (IDEIA)

11) <http://www.daisy.org/structure-guidelines>

12) <http://www.w3.org/TR/MathML2/>

13) <http://www.w3.org/TR/CSS2/>

한 기능은 전자교과서 및 전자잡지를 표현하거나 전자책이 단순히 종이책의 디지털 형태가 아닌 새로운 미디어로 진화하기 위해 필요하다. 현재 EPUB은 신문 및 잡지의 기본적인 구성단위인 기사(article)를 구체적으로 고려하지 않으며 잡지 등이 요구하는 정교한 수준의 페이지 레이아웃을 지원하는데 제한적이다.

현재 국내 전자책 업체의 대부분은 EPUB 형식을 지원한다. 그러나 국내 전자책 표준의 부재로 인하여 전자책 단말기마다 동일한 EPUB 파일을 다르게 처리 및

출력하는 등 호환성에 문제점을 갖는다. 이에 국내 전자책 시장의 활성화 및 관련 산업의 육성은 물론이고, 향후 관련 산업의 세계 시장 진출을 위해서 전자책의 표준화는 매우 필요한 실정이다. 따라서 정부의 전폭적인 지원을 기반으로 전자책 및 출판 관련 이해당사자의 연합체를 구성하여 글로벌 마켓 트렌드에 전략적으로 대응하기 위한 체제를 마련하고 전자출판물 관련 표준을 제정할 필요가 있다. TTA

정보통신용어해설

TP 시프트 족

Time Place Shift Tribe, -族 [관리용어]

시간(time)과 공간(place)을 자유롭게 넘나들며(shift) 본인이 원하는 시간과 원하는 장소에서 보유한 콘텐츠를 즐기거나 상호 공유하며 소비하는 새로운 문화계층.

스마트폰과 같은 이동형 스크린 등장과 3스크린 서비스가 실용화되면서 생겨나고 있다.

