

한국전자책문서표준(EBKS 1.1) DTD의 정의

김재경^o, 손원성, 고승규, 임순범^{*}, 최윤철

연세대학교 컴퓨터과학과

^{*}숙명여자대학교 멀티미디어과학과

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

^{*}sblim@sookmyung.ac.kr

Definition of EBKS 1.1 DTD

Jae-Kyung Kim^o, Won-Sung Sohn, Seung-Kyu Ko, Soon-Bum Lim^{*}, Yoon-Chul Choy

Dept. of Computer Science, Yonsei University

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

요 약

한국전자책문서표준 EBKS(Electronic Book of Korea Standard) 1.0은 국내 전자책산업의 활성화를 목적으로 제정된 산업표준으로서 한국전자출판협회의 지원아래 지난 2001년 7월 제정된 바 있다. 표준 제정이후, 전자책 업계에서는 EBKS 사용시 나타나는 문제점 및 여러 가지 개선사항을 요구하였다. 또한 빠르게 발전하는 멀티미디어 산업의 특성상 변화된 정보를 표준에서 수용해야 할 필요성이 대두되었다. 따라서 본 논문에서는 EBKS 1.0 DTD의 문제점을 해결하고 새로운 요소 정의 및 구조를 재정의 하였다. 이를 통하여 표준 전자책 작성시 보다 명확한 구조와 의미정보, 그리고 추가된 멀티미디어 요소를 표현할 수 있도록 하였다.

1. 서론

국내 전자책 콘텐츠 표준인 EBKS 1.0[2]은 다양한 전자책 콘텐츠 구조를 명시적으로 표현할 수 있다. 따라서 교환시 의미 정보를 명확히 전달할 수 있는 장점을 지니고 있다. 그러나 다양한 문서 구조를 표현하기 위하여 여러 문서 요소를 지니고 있음으로 인해 사용자가 필요한 각 분야의 문서 구조가 명확하지 않은 점이 있다.

이러한 결론하에 본 연구진은 EBKS DTD 버전을 위하여 현존하는 EBKS 1.0 버전의 잘못된 오류 및 보다 개선된 구조를 정의하였다. 이러한 목표아래

EBKS 1.1 DTD는 기존 EBKS 1.0 버전의 의미요소 및 구조요소 중 의미적으로 명확하지 않거나 개선의 필요가 있는 부분들을 심도있게 검토하여 새로운 요소 정의 및 구조를 재정의 하였다. 또한 EBKS 1.0에서 제공하지 않았던 멀티미디어 요소들을 보강하기 위하여 새로운 요소들을 검토하고 이를 보강하였다.

2. DTD 요소별 변경 내용

본 절에서는 EBKS 1.1에서 추가된 요소 및 변경된 구조에 대하여 자세히 설명한다. DTD refinement는 크게 DTD 의미요소에 대한 추가 및 구조에 대한 단

순화 부분을 포함한다.

2.1 의미요소의 버전업

EBKS 1.1에서는 기존 EBKS 1.0 DTD에 포함되지 않았던 의미요소들을 추가 및 변경 그리고 삭제하였다. 의미요소에 대한 수정 기준은 다음과 같다.

- 전자책 교환과 동시에 저작을 목적
- EBKS 1.0을 통하여 정의할 수 없었던 요소 추가
- EBKS 1.0의 오류 수정
- 사용자들의 요구사항을 반영

본 논문에서는 전자책 표준 사용자들의 다양한 요구들을 분석하고 이를 충분히 반영하고자 하였다. 그러나 이러한 요구사항의 대부분은 문서의 구조적 요소가 아닌 스타일 요소로서 특히 HTML에서 보편적으로 사용되는 요소들을 주로 포함하였다. 따라서 이러한 요소들은 EBKS DTD의 설계목적과 위배되는 것으로 본 연구진은 대부분의 스타일 요소들을 분리하고 XML 문서가 가지는 의미적 구조를 보다 명확하게 하기 위해 EBKS 1.1 버전에서 변경된 의미적 요소들은 다음과 같다.

1) **Artwork** 엘리먼트의 속성요소에서 **caption**과 **copyright** 요소 추가정의

2) **Fig** 엘리먼트의 속성에 **copyright** 요소 추가정의

3) **Affiliation** 엘리먼트의 정의

4) **Back** 엘리먼트에 **pubfront** 요소 추가

5) 사용자 정의가 가능한 **math** 요소의 정의

6) **Section** 모델의 삭제

7) **Pub** 모델에 **note** 엘리먼트 추가

8) **Mediaobject**의 정의

9) **Inlinemediaobject** 엘리먼트의 정의

2.2 구조요소의 버전업

EBKS 1.1 버전에서는 엔터티 모델들의 구조를 단순화하고 명확화하기 위한 작업을 수행하였다. 이를 위하여 새로운 엔터티 모델을 추가하고 그 구조를 간략화하였으며 그 결과 전체 DTD 요소들의 구조가 다수 변경되었다. 본 절에서는 이에 대한 내용을 상세히 설명한다.

1) **Image** 모델의 변경

2) **Author** 모델의 그룹화

3) **Title** 모델의 그룹화

4) **Name** 모델에서 **address** 요소의 분류

5) **Pubfront** 모델의 변경

3. DTD Enhancement

본절에서는 EBKS와 관련된 문서 관련 표준과 최신 표준을 분석하여, 전자책 콘텐츠 환경에서 필요한 기능을 멀티미디어, 수식, 링크, 테이블 등의 차원에서 살펴본다.

3.1 멀티미디어

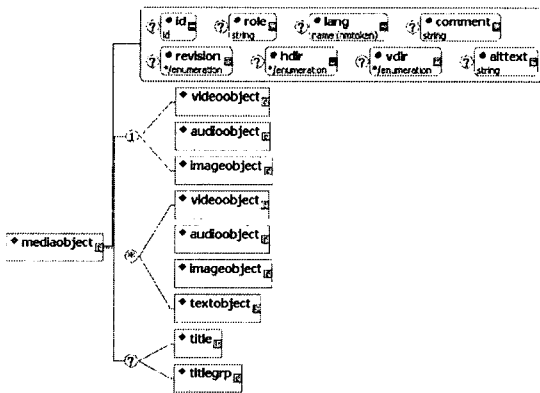
정보를 표현할 때에는 주로 텍스트를 이용하지만 효과적인 정보 전달을 위해서 이미지, 오디오, 동영상 등을 이용할 수 있다. 그러나 전자책 표준과 대부분의 문서 표준에서는 이러한 정보를 별도로 구별하지 않고 이미지 요소를 이용하여 표현한다.

EBKS에서는 멀티미디어 정보를 좀 더 명확하게 표현 및 교환할수 있도록 하기 위하여 RFC2046[5]를 참고로 미디어 정보를 텍스트, 이미지, 오디오, 비디오로 분류하고, 각각의 요소에 대해 정의하였다. 멀티미디어 정보를 표현하기 위해서 먼저 전체 미디어를 표현할 수 있는 **mediaObject**를 정의하고, 하위 각각 미디어 정보를 표현하는 **imageobject**, **audioobject**, **videoobject** 요소를 정의하였다. **Mediaobject**의 콘텐츠 모델은 [그림 1]과 같다.

“**mediaobject**”에서 실제로 내용을 표현하는 “**imageobject**,” “**audioobject**,” “**videoobject**”에 대해 살펴보면 다음과 같다.

▶ 이미지(imageobject)

정보를 텍스트 형태가 아닌 이미지 형태로 처리되는 경우를 의미한다. 즉 “bmp”, “gif”, “pdf”, “ps” 등으로 표현되는 정보가 이에 해당한다. RFC2046에서는 “ps”를 응용프로그램으로 분류하였으나 EBKS에서는 정보를 처리하는 방법이 아닌 표현 정보에 따라 미디어를 분류하였다. 따라서 “ps”, “pdf” 등은 정보를 이미지 형태로 바꾸어 전달하는 것으로 간주하여 “imageobject”에 속하게 되며 실제 콘텐츠는 “imagedata”를 이용하여 표현한다. 구체적인 파일 형식은 “imagedata”의 “format”을 이용하여 표현하고, 사용되는 파일 형식은 “video.notation.class”에 정의되어 있다. 사용자가 정의된 형식 이외에 사용하고자 하는 형식이 있으면, “local.vdeo.notation.class”에 추가하고 사용하면 된다. 갱신된 EBKS는 기존의 EBKS의 노테이션에서 “SVG”[7]를 추가로 지닌다.



[그림 1] Mediaobject의 콘텐츠 모델

▶ 오디오(audioobject)

정보를 오디오 형태로 표현하는 경우가 이에 해당한다. “audioobject”도 “imageobject”와 마찬가지로 실제 콘텐츠는 “audiodata”에서 표현되며, 출력되지 못하는 경우를 고려하여 “alttext”를 제공한다. 실제 콘텐츠는 “audiodata”의 “entityref”나 “uri”를 이용하여 표현한다. 구체적인 파일 형식은 “audiodata”의 “format”을 이용하여 표현하고, 사용되는 파일 형식은 “audio.notation.class”에 정의되어 있다. 사용자가 정

의된 형식 이외에 사용하고자 하는 형식이 있으면 “local.audio.notation.class”에 추가하고 사용하면 된다. 갱신된 EBKS에서는 “wav”, “mp3”, “modi” 등의 오디오 노테이션들이 추가되었다.

▶ 비디오(videoobject)

정보를 동영상으로 표현하는 경우가 이에 해당한다. 어린이를 위한 동화책은 플래쉬, 쿼타임 등과 같은 움직이는 영상으로 표현되는 경우가 많은데, 이러한 동영상 정보는 “videoobject”로 표현된다. “videoobject”도 다른 미디어 요소와 마찬가지로 구체적인 내용은 “videodata”에서 표현되며, “alttext” 속성을 제공한다. 구체적인 파일 형식은 “audiodata”의 “format”을 이용하여 표현하고, 사용되는 파일 형식은 “audio.notation.class”에 정의되어 있다. 사용자가 정의된 형식 이외에 사용하고자 하는 형식이 있으면, “local.audio.notation.class”에 추가하고 사용하면 된다. 정의된 노테이션은 “swf”, “mov”, “mpeg” 등이 있다.

3.2 수식

전자책 관련 표준에서는 수식은 별도로 정의되어 있지 않고 있고, 관련 문서 표준에서는 필수적인 요소만을 정의하고 있거나, 다른 표준을 사용할 것을 권고하고 있다. 예를 들어 DocBook에서는 최소한의 요소만을 정의하고, ISO12083에서는 별개의 Math DTD를 정의하고, TEI 지침에서는 ISO12083의 수식을 사용할 것을 권장하고 있다. ISO12083에서 정의한 DTD는 정교한 수식을 표현할 수 있는 것이 아니라 일반적으로 많이 사용되는 수식을 표현한다.

현재 EBKS에서는 수식을 “formula”로 정의하고 있으며, 수식을 이미지로 처리할 것을 권고하고 있다. 왜냐하면 EBKS는 수식보다는 전자책의 표현에 초점을 맞춘 전자책 표준이기 때문이다. 이는 다른 관련 표준의 경우도 마찬가지이다. 전자책 관련 표준은 수식을 지원하지 않고 있으며, DocBook이나 ISO12083 등의 문서 관련 표준에서도 간단한 형태의 수식만을 정의하고 있다.

W3C에서는 XML이 널리 사용됨에 따라 XML로

수식을 표현할 수 있는 수식 권고안인 MathML[3]을 1999년 7월에 발표하고, 버전 2.0을 2003년 10월에 발표하였다. MathML은 정교한 수식 표현이 가능할 뿐 아니라 XML을 제정한 제정한 XML 표준이다. 따라서 XML에 기반한 많은 표준들이 W3C에서 제정한 MathML을 지원할 것으로 예상된다.

따라서 EBKS에서는 MathML을 사용하는 경우가 증가할 것으로 예상하고, 수식을 이미지로 표현하는 경우뿐 아니라 MathML을 이용하여 표현하는 경우도 지원하기 위하여 MathML에 관한 노테이션을 다음과 같이 정의하였다.

```
<ENTITY % math PUBLIC "-//W3C//DTD MathML
2.0/EN"
```

```
"http://www.w3.org/TR/MathML2/dtd/mathml2.dtd"
```

```
>
```

3.3 기타

멀티미디어와 수식 이외에 링크와 테이블과 관련된 사항들을 살펴보면 다음과 같다.

ISO12083이나 DocBook[1]과 같은 문서 관련 표준에서는 링크가 내부 링크와 외부 링크로 나뉘고, 전자책 관련 표준에서는 구별 없이 하나의 링크만을 사용한다. 이에 비해 EBKS와 문서 관련 표준에서는 링크를 문서 내부를 참조하는 링크인 "xref"와 문서 밖의 외부를 참조하는 경우인 "uri"로 정의하였다. EBKS에서는 내부 링크인 "xref"에서는 "idref"를 이용하여 문서 내부의 요소를 가리키고, "uri"에서는 "ulink"를 이용한다.

테이블은 크게 두가지 종류로 구별할 수 있다. 하나는 CALS에 기반한 테이블로 행이나 열에 따라 의미정보 부여가 가능하다. 그리고 두번째는 HTML 테이블로 테이블에 의미정보를 부여할 수는 없으나 사용하기 쉽다. 관련 표준을 살펴보면 문서 관련 표준에서는 전부 CALS 테이블[4]을 사용하고, 전자책 표준에서는 HTML 테이블을 사용하고 있다. EBKS에서는 테이블을 논리적으로 정의하고, 이에 따라 의미정보 추출이 가능한 CALS에 기반한 테이블 모델을 사용하고 있다. 이는 XML의 논리적인 정보가 자연

스럽게 표현된 형태이다. 그리고 CALS 테이블은 HTML테이블로 변환이 용이한 반면에 그 역변환은 불가능하므로 XML에 기반한 표준에서는 CALS 테이블을 이용하는 것이 바람직하다.

4. 결론

본 논문에서는 EBKS 1.1 DTD 정의를 위하여 EBKS 1.0 버전의 잘못된 오류를 개선하고 다양한 기능을 제공하고자 하였다. 이를 위하여 기존 EBKS 1.0과는 달리 1.1 버전에서는 전자책 교환과 동시에 저작을 목적으로 의미요소들을 추가하였으며 동시에 EBKS 1.0을 통하여 정의할 수 없었던 요소들을 분별하여 새로 정의하였다. 특히 빠르게 발전하는 멀티미디어 요소를 표현하기 위하여 미디어 오브젝트 모델을 정의하였다. 또한 EBKS 1.0의 오류들을 수정하는 동시에 사용자들의 요구사항을 반영하여 전반적인 구조요소들을 재정리 및 간결화하였다.

[참고문헌]

- [1] DocBook, the official home page for DocBook: The Definitive Guide, <http://docbook.org/>, January 2004
- [2] EBKS version 1.0, EBK 표준화위원회 워킹그룹, <http://orange.yonsei.ac.kr/ebook/>, 2001
- [3] Mathematical Markup Language (MathML) Version 2.0 (Second Edition), W3C Recommendation, <http://www.w3.org/TR/2003/REC-MathML2-20031021/>, October 2003
- [4] CALS table model Document Type Definition, SGML Open Technical Memorandum TM 9502:1995, <http://www.oasis-open.org/specs/tm9502.html>, October 1995
- [5] RFC2046: Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", N. Freed and N., <http://www.ietf.org/rfc/rfc2046.txt>, November 1996.
- [6] TEI, Text Encoding Initiative, <http://www.tei-c.org>, September 2003
- [7] Scalable Vector Graphic, World Wide Web Consortium, <http://www.w3.org/SVG>