

EBKS에 적용한 전자책 Annotation 모델링

고승규, 이현찬*, 임순범*, 최윤철

연세대학교 컴퓨터과학과, *숙명여자대학교 멀티미디어학과

eBook Annotation Modeling Applied on EBKS

SeungKyu Ko, HyunChan Lee, SoonBum Lee, YoonChul Choy

Dept. of Computer Science, Yonsei University

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

E-mail : {pitta, emerald, ycchoy}@rainbow.yonsei.ac.kr,

*sblim@sookmyung.ac.kr

요 약

기존의 전자책 환경에서 종이책과 구별되는 특징은 네트워크를 통한 저자와 독자, 출판사 간의 인터랙티브(interactive)한 정보 교환이 가능하다는 점이다. 이러한 교환은 기존의 종이책에서 사용하는 Annotation을 이용하면 가능하다. Annotation이란 원본 문서에 부가적으로 추가되는 정보를 의미한다. 그러므로 Annotation과 원본 문서는 밀접한 관계를 갖는데 기존의 Annotation 모델링은 원본 문서를 고려하지 않고 Annotation만을 별개로 모델링하였다. 이에 본 논문에서는 Annotation을 보다 효과적으로 활용하기 위하여 Annotation과 원본 문서를 동시에 표현하는 모델링에 대해 제안한다. 그리고 본 모델링은 전자책 표준인 EBKS에 기반하며, 모델링 결과를 웹 자원을 기술하는 표준인 RDF를 이용하여 표현한다.

1. 서론

정보화 시대를 맞이하여 주위의 많은 환경이 디지털환경으로 변화가고 있는 현재, 기존의 주요 정보 표현 수단인 종이책도 디지털 형태인 전자책으로 바뀌고 있다. 전자책은 디지털 형태이기 때문에 기존 종이책에 비해 여러 장점[7][8]들을 갖게 된다. 첫째, 인쇄나 유통 등의 과정들이 효율적으로 처리되고, 재고 부담의 문제가 없으며, 생산 즉시 독자에게 전달이 가능하다. 둘째 텍스트나 이미지 정보뿐 아니라 멀티미디어 정보 등이 가능하여 실감있는 정보 전달이 가능하다. 셋째 책의 수명이 기존의 종이책과 달리 영구적이고, 방대한 양의 책을 저장, 관리하는데 용이하다. 넷째 네트워크를 통한 저자와 독자, 출판사 간의 인터랙티브(interactive)한 정보 교환이 가능하다. 이러한 장점으로 인해 전자책 시장은 앞으로 더욱 커질 것으로 예상되며, 여러 기관에서도 유사한 예측을 하고 있다. 예를 들어 대표적인 시장조사 기관인 IDC(International Data Corporation)에 따르면 미국의 전자책 시장은 2001년 9백만달러에서 2004년 4억1천4백만달러로 급격히 증가할 것으로 예측하고 있으며, 앤더슨 컨설팅사에서도 2005년 전자책 시장은 23억달러에 달할 것으로 전망하고 있다.

앞에서 언급한 여러 장점 중에서 네번째 장점에 대한 지원은 현재 미비한 실정이다. 그 이유는

업계에서 자체적인 전자책 포맷을 사용하여 교환이 어려우며, 전자책에 사용자의 의견이나 생각을 표현하는 방법이 체계적으로 정의되어 있지 않기 때문이다. 또한 전자책 환경에서 추가된 정보를 공유하거나 교환하는 방법에 대해서는 연구가 진행되고 있지 않다. 그러나 전자책 환경에서 네트워크를 통한 의견 교환이야말로 전자책이 기존의 종이책과 구별되는 가장 큰 차이이다.

올해 초 전자책 콘텐츠에 대한 표준으로 EBKS가 발표되어 업체마다 독자적으로 사용하던 포맷에 대한 문제가 조만간에 해결되리라 예측된다. 또한 전자책에 대해 사용자의 생각이나 의견을 추가하는 방법은 종이책에서 사용하였던 방법인 Annotation을 이용하면 가능하다. 종이책과 구별되는 전자책 Annotation의 특징[1]은 Annotation에 대한 키워드 검색 및 공유가 가능하다는 점이다. 사용자는 Annotation에 대한 키워드 검색을 이용하여 원하는 정보를 빠르게 찾을 수 있으며, Annotation 공유를 통하여 다양한 의견 교환이 가능하다.

이에 본 논문에서는 전자책 표준인 EBKS[5]에 기반하여 Annotation을 효과적으로 공유 및 교환할 수 있는 전자책 Annotation에 대해 모델링한다. 특히 생성된 Annotation 정보를 다른 Annotation 시스템과 효과적으로 공유 및 교환하기 위하여 웹 표준인 RDF(S)[6]를 이용한다.

본 논문의 구성은 다음과 같다. 2장에서 관련연구로 Annotation이 무엇이고, Annotation 모델이 어떠한 것이 있는지에 대해 살펴본다. 그리고 3장에서는 전자책 표준인 EBKS 환경에서의 Annotation 모델링에 대해 기술한다. 그리고 4장에서는 결론 및 향후 연구 방향에 대해 설명한다.

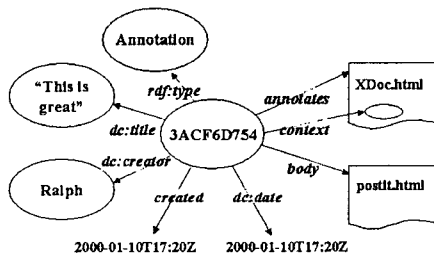
2. 관련연구

웹스터 사전에서는 Annotation을 “의견이나 설명을 위해 추가되는 노트”라고 정의하고 있으며, Annotea[4]에서는 메타데이터 유형으로 보고 있다. 본 논문에서는 Annotation을 “원본 문서에 추가되는 모든 정보”로 정의한다. 기존의 Annotation에 관한 연구는 다음과 같다.

Annotator[1]에서는 Annotation을 콘텐츠, 폼(form), 기능(functionality)으로 정의하였다. 콘텐츠란 사용자가 생성하는 Annotation 내용이 되고, 폼이란 밑줄, 하이라이트, 원 등의 Annotation의 형태가 된다. 그리고 기능은 이 Annotation이 어떠한 목적을 위해서 사용하는지를 의미한다. 일반적으로 사용자가 Annotation을 사용하는 목적은 기억하기 위한 것(remember), 자신의 생각에 도움을 주는 것(think), 그리고 자신의 생각을 명확히 하는 것(clarify)이고, 의견(comment)의 공유를 위해서도 많이 사용한다고 한다.

[2]에서는 Annotation을 여러가지 기능에 따라 분류하고 있다. 먼저 Annotation을 작성자 이외의 다른 사람들이 이해하는 정도에 따른 명시/암묵 Annotation, 생성한 Annotation을 계속 유지시킬 것인지에 따른 영구/임시 Annotation, 공개 유무에 따른 공개/개인 Annotation, 그리고 적용 범위에 따른 전역/기관/그룹/개인 Annotation 등으로 구별하고 있다.

[3]에서는 Annotation을 폼(form)과 기능에 따라 분류하였다. 폼은 인라인(inline)인지, 여백인지 등에 따라 구별하고, 기능은 미래의 주의가 필요한 부분(signal), 기억해야 할 부분, 문제와 관련된 정보, 해석, 전에 읽었던 부분, 내용과 관계없는 부분 등으로 구별하였다.



[그림 1] Annotea에서의 RDF 표현 예[4]

W3C에서 개발한 Annotea[4]에서는 [그림 1]과 같이 Annotation을 Annotation이 적용되는 자원, 자원 중에서 Annotation과 관련된 영역인 컨텍스트, 그리고 Annotation의 내용인 본문, 컨텍스트와 Annotation간의 관계로 정의하고, 기타 Annotation에 관한 생성자, 생성날짜, 수정날짜 등의 메타 정보를 정의하였다.

Annotea에서는 특히 컨텍스트와 Annotation과의 관계를 좀 더 세분하여 충고, 수정, 의견, 예, 설명, 질문, 참고 등으로 분류하였다. 이러한 Annotation과 컨텍스트의 관계는 향후 annotation의 교환 및 활용과 밀접한 관계가 있다. 그리고 Annotea에서는 이러한 표현을 웹에서 자원을 기술하는 표준인 RDF(S)[6], XLink[9], XPointer[10] 등을 이용하고 있다.

기타 Annotation 시스템으로는 여러가지[8]가 있으나 대부분 Annotator나 Annotea와 유사하다.

기존의 Annotation 시스템에서의 Annotation 모델은 대부분 컨텍스트와 분리된 Annotation을 모델링하였다. 그리고 출력 시에 Annotation을 원문의 컨텍스트에 위치시키는 기법에 대한 연구가 많이 진행되었다. 이는 Annotation의 교환이나 활용보다는 원문에서 Annotation을 사용하기 위한 방법에 초점을 맞추어 왔기 때문이다. 그러나 Annotation은 원문의 부가적인 정보로 원문의 컨텍스트와는 함께 있어야 보다 효과적인 정보를 표현할 수 있다. 그러므로 본 논문에서는 전자책 원문과 Annotation을 함께 모델링하여 Annotation 필터링이나 검색 시 보다 효과적인 Annotation 처리가 가능하도록 하였다.

3. Annotation 관리 및 활용을 위한 모델링

본 논문에서는 전자책 환경에서 효과적으로 Annotation을 공유 및 교환할 수 있도록 하기 위하여 RDF에 기반한 모델링을 제안한다. 특히 Annotation은 컨텍스트와 떨어질 수 없는 밀접한 관계가 있기 때문에 기존의 방법과는 달리 컨텍스트를 포함하여 Annotation을 모델링하였다. 그리고 컨텍스트는 전자책 표준인 EBKS에 기반한다. 본 절에서는 먼저 모델링에 대해 설명하고, 제안된 모델링을 효과적으로 이용하기 위하여 웹 자원을 기술하는 표준인 RDF로 표현한 예에 대해 설명한다.

3.1 Annotation 모델

본 논문에서는 Annotation을 포함한 전자책을 다음과 같이 정의한다.

Annotated eBook = <*eBook, Annotation, relation*>

“eBook”은 EBKS에 기반한 전자책을 의미하며, “Annotation”은 사용자가 생성하는 Annotation을 의미하며, “relation”은 전자책과 “Annotation”간의 관계를 의미한다.

먼저 컨텍스트를 포함하는 전자책은 메타데이터와 실제 내용으로 구성된다.

eBook = <*metadata, content*>

EBKS에서의 메타데이터는 “<metadata>”와 본문, 두 군데에서 나온다. 본 모델에서는 먼저 “metadata”에 있는 정보를 우선하며, 필요 시 본문에 있는 메타 정보를 이용한다. 그리고 본 메타데이터는 더블린코어에 기반한 정보와 확장 메타데이터로 구성된다.

metadata = <*Dublin Core, extension*>

그리고 EBKS는 XML 기반이므로 “content”는 구조 정보와 텍스트 내용으로 구성된다. 그리고 Annotation과 연관되는 컨텍스트는 이러한 “content”의 한 인스턴스가 된다.

content = <structure, text>
context is an instance of a content

구조 정보는 EBKS에 기반한 정보로 “author”, “reference” 등으로 구성되며, 텍스트는 문장들로 구성되고, 문장들은 단어들로 구성된다.

Structure=<author, picture, formulas, reference, ...>
Text = <sentences>
sentences = <words>

전자책의 실제 내용인 “body”부분에 대해서는 의미가 있다고 판단되는 “Author”, “picture”, “inCommentary”, “formulas”, “reference”, “index”, “keyword” 등에 관한 구조 정보만을 반영하고, 장이나 절 등의 관한 구조 정보는 실제 전자책을 접근하도록 하여 본 모델링에서는 제외시켰다.

본 모델에서의 Annotation은 형태와 생성날짜, 내용으로 구성된다.

Annotation = <style, date, note, creator >

형태는 밑줄, 하이라이트, 원 등의 Annotation의 형태를 의미하며, “creator”는 Annotation을 생성한 사람을 의미하고, 노트는 실제 Annotation의 내용을 나타낸다.

“Relation”은 실제 전자책과 Annotation 간의 관계를 나타낸다

Relation = <anchor, functionality, access, annotation, category >

Relation : eBook(context) → annotation

“anchor”는 컨텍스트를 의미하며, 전자책이나 저자도 가능하며, 이 경우에는 서평이나 저자평이 된다. 그리고 기능은 본 모델에서 정의하는 디폴트 기능과 사용자가 확장할 수 있는 기능으로 나뉜다.

Functionality = <default, user_extended>

디폴트는 기존의 종이책에서 많이 분류되는 방법[1][2]을 참조하였으며, 다음과 같이 정의될 수 있다.

Default = { forRemind, forThinking, forPrivate, forReadingAssistant }

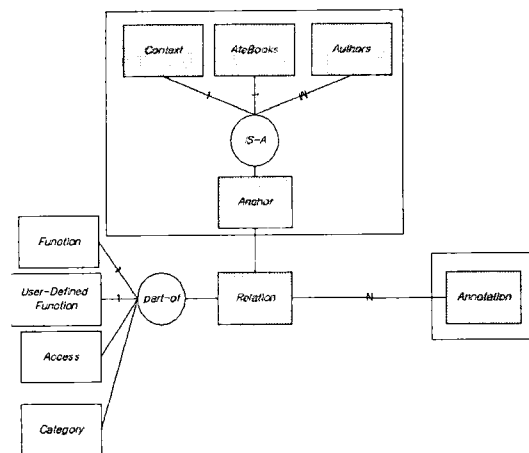
즉 기억하기 위한 것, 생각하기 위한 것, 개인적인 것, 도움이 되는 정보를 표현하기 위한 것으로 구별된다. 본 모델링에서는 다중 사용자간의 Annotation 공유나 교환을 고려하기 때문에 기본적으로 Annotation은 공유 가능한 것으로 간주한다. 그래서 개인적인 것을 따로 구별하였다. 그리고 사용자는 이러한 디폴트 기능을 나름대로 확장할 수 있다. 즉 “forRemind”의 하위로 시험에

나올 부분, 삭제, 중요한 부분 등의 기능 정보로 좀 더 확장할 수 있다. 이렇게 확장함으로써 다른 Annotation 시스템과 Annotation 정보를 교환할 경우에 최소한 디폴트 수준에서의 교환이 가능하도록 하였다. 또한 본 논문에서는 Annotation의 활용도를 높이기 위하여 의미정보를 표현하기 위하여 “category”를 정의하였다. 이 “category”는 본 모델링에서 정의한 것을 바탕으로 하며 크게 두 레이어로 구성된다.

- 상위 레이어 : EBKS의 모델
 Author, publisher, reference 모델 등
 - 하위 레이어 : 실제 내용을 표현하는 엘리먼트
 Author 모델: 이름, 이메일, 주소 등
 Publisher 모델: 이름, 대표자, 회사 약력 등
- 현재 “category”은 본 모델링에 국한된 정보만을 표현하고 있다. 이 정보는 향후 ontology를 이용하여 일반적인 정보도 표현할 수 있도록 할 예정이다.

3.2 RDF 표현

본 논문에서는 제안한 모델링을 시스템에 독립적으로 표현하기 위하여 웹자원 기술 표준인 RDF를 이용한다. 먼저 RDF로 표현하기 위해서는 정의한 모델링 요소 간의 관계를 추출하여야 한다. 여러 요소 중에서 “relation”에서 가능한 관계에 대해 살펴보면 [그림 2]와 같다.



[그림 2] “relation”에서 가능한 관계

[그림 2]를 살펴보면, “Annotation”과 “anchor”가 “relation”으로 연결되는 것을 알 수 있다. 즉 Relation은 기능, 접근 권한, 분류로 구성되며, “anchor”와 “Annotation”간을 연결시켜주는 것을 알 수 있다. 그리고 “anchor”는 컨텍스트와 전자책, 저자 등이 가능함을 알 수 있다. 이와 같이 모델링 요소 간의 관계를 추출하여 RDF 형식으로 표현하면 다음과 같다.

```
<? xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/TR/REC-rdf-syntax#"
  xmlns:rdfs="http://www.w3.org/TR/WD-rdf-schema#"
>
<!-- Address Model -->
<rdf:Description rdf:ID = "AnchorModel">
  <rdf:type rdf:resource="http://www.w3.org/TR/REC-rdf-
    syntax#Class"/>
  <rdf:subClassOf rdf:resource="http://www.w3.org/TR/REC-rdf-
    syntax#Resource"/>
  <rdf:label>Anchor Model</rdf:label>
  <rdf:isDefinedBy = ""/>
</rdf:Description>

<rdf:Description rdf:ID = "Context">
  <rdf:type rdf:resource="http://www.w3.org/TR/REC-rdf-
    syntax#Class"/>
  <rdf:subClassOf rdf:resource="#AnchorModel"/>
  <rdf:label>Context</rdf:label>
  <rdf:isDefinedBy = ""/>
</rdf:Description>

<rdf:Description rdf:ID = "AtedBookModel">
  <rdf:type rdf:resource="http://www.w3.org/TR/REC-rdf-
    syntax#Class"/>
  <rdf:subClassOf rdf:resource="#AnchorModel"/>
  <rdf:label>Annotated eBook Model</rdf:label>
  <rdf:isDefinedBy = ""/>
</rdf:Description>
...
```

[그림 3] 제안된 모델링을 RDF로 표현한 예

본 논문에서는 [그림 3]의 경우와 같이 제안한 모델링에서 요소들간의 관계를 추출하여 RDF 형식으로 표현하였다. 이 때 RDF(S)에서 제공하는 기본적인 관계((sub)class, (sub)property, domain, range 등)만으로 표현하기 어려운 부분이 있다. 예를 들어 특정인이 저자인 동시에 독자일 경우에 이러한 사람을 표현하기 위해서는 동일(equal)이라는 관계가 필요하다. 이와 같이 RDF에 없는 관계로는 존재성(∃, 3) 및 인스턴스 관계 등이 있다. 그래서 본 논문에서는 동일과 존재성의 관계를 추가하여 전자책 Annotation을 표현하였다.

본 모델링은 Annotation을 전자책, Annotation, Relation으로 각각 분리하여 모델링함으로써 저장과 관리가 효과적이며, category 정보를 이용함으로써 Annotation을 보다 효과적으로 공유 및 교환할 수 있다.

5. 결론 및 향후 연구방향

본 논문에서는 전자책 환경에서 효과적으로 의견 교환 등을 하기 위하여 컨텍스트를 포함하여

Annotation을 모델링 하였다. 본 모델링은 기존의 다른 Annotation 모델링과는 달리 Annotation만이 아닌 컨텍스트를 포함하여 모델링하여 Annotation의 의미정보를 보다 효과적으로 표현할 수 있도록 하였으며, EBKS에 기반하여 Annotation의 category를 표현함으로써 Annotation 교환 시 최소한의 수준에서 일관성 있는 교환이 가능하다. 그리고 제안된 모델링을 웹 자원을 기술하는 표준인 RDF형식으로 표현함으로써 활용도를 높일 수 있도록 하였다. 그리고 표현 시 RDF에서 가능하지 않은 동일성과 존재성이라는 관계를 추가하여 표현하였다.

향후 본 모델링을 이용하여 실제로 전자책 환경에 적합한 시스템을 구현할 예정이다. 그리고 의미정보를 보다 효과적으로 이용하기 위하여 ontology에 기반한 표현 방법에 대해 연구 중에 있다.

참고문헌

- [1] Iliia A. Ovsianikov, "Annotation Technology," International Journal of Human-Computer Studies v.50 n.4, 1999.
- [2] Catherine C. Marshall, "Toward an Ecology of Hypertext Annotation", Proceedings of HyperText, ACM, Pittsburgh 1998
- [3] Catherine C. Marshall, "Annotation: From Paper Books to Digital Library", Proceedings of the 2nd ACM International Conference on Digital Libraries, Philadelphia, 1997.
- [4] José Kahan, Marja-Riitta Koivunen, Eric Prud'Hommeaux, and Ralph R. Swick, "Annotea: An Open RDF Infrastructure for Shared Web Annotations", in Proc. of the WWW10 International Conference, Hong Kong, May 2001.
- [5] EBK(e-Book of Korea) Consortium, "A Study of Korean Standardization of eBook documents," Technical Report, 2001.
- [6] D. Brickley and R.V. Guha (eds.). "Resource Description Framework (RDF) Schema Specification 1.0". CR, W3C, Mar. 2000. <http://www.w3.org/TR/2000/CR-rdf-schema-20000327>.
- [7] 문화관광부, "디지털시대의 전자책(e-Book) 발전방향," 2000.
- [8] R. M. Heck, S. M. Luebke, and C. H. Obermark. A Survey of Web Annotation Systems, 1999. http://www.math.grin.edu/~luebke/Research/Summer1999/survey_paper.html
- [9] S. DeRose, E. Maler, D. Orchard, and B. T. (eds.). XML Linking Language (XLink). WD, W3C, Feb. 2000. <http://www.w3.org/TR/2000/WD-xlink-20000221>.
- [10] S. DeRose, R. D. Jr., and E. M. (eds.). XML Pointer Language (XPointer). WD, W3C, Dec. 1999. <http://www.w3.org/TR/1999/WD-xptr-19991206>.