

XML기반 디지털 콘텐츠 메타데이터 설계

나철수*, 김재인, 한대영, 김대인, 박준, 황부현
*전남대학교 전자컴퓨터공학부
e-mail:choulsu@daum.net

MetaData Design for Digital Contents based on XML

Choulsu Na*, Jaein Kim, Daeyoung Han, Daein Kim, Jun Park,
Buhyun Hwang
*Dept of Computer Science, Chonnam University

요 약

정보화 사회의 발전과 인터넷의 급속한 확산으로 인하여 디지털 콘텐츠에 대한 수요 및 생산되는 콘텐츠의 수가 기하급수적으로 증가하고 있다. 그러나 디지털 콘텐츠는 생산 및 수집 시기, 그리고 저장 방법 등에 따라 그 형태가 매우 다양하므로 디지털 콘텐츠를 표준화할 수 있는 방법에 대한 연구가 필요하다. 본 논문에서는 디지털 콘텐츠의 메타데이터 타입을 표준화하기 위한 새로운 메타데이터 타입을 정의한다. 제안하는 메타데이터 타입은 XML 형태로 기술하며 디지털 콘텐츠의 국제 표준 메타데이터인 더블링크어에 기초하여 정의한다. 제안 방법은 더블링크어에서 정의한 주요소에 역할 속성을 추가하여 콘텐츠 생산에 직간접적으로 영향을 준 생산자 정보를 표현할 수 있다. 또한 제안하는 방법은 콘텐츠 생산에 관련된 연관 콘텐츠 정보를 유지함으로써 디지털 콘텐츠 간의 연관 관계를 표현할 수 있다.

1. 서론

정보화 사회의 발전, 인터넷 기술의 확산 및 보급, 그리고 다양한 정보를 디지털화하기 위한 수요 증가로 인하여 다양한 형태의 디지털 콘텐츠가 무수히 생산되고 있다. 그러나 디지털 콘텐츠는 자료 수집 방법, 시기, 저장 방법에 따라 매우 다양한 형태로 표현된다. 그리고 디지털 콘텐츠 간에는 다양한 종류의 연관 관계가 존재하고 다양한 관계 정보를 표현할 수 있으며 형태를 표준화 할 수 있는 메타데이터(metadata) 정의에 대한 연구가 필요하다.[1]

메타데이터란 일반적으로 데이터의 데이터, 또는 전자 자원을 기술하는데 사용되는 데이터 요소를 의미하며, 표현하고자 하는 대상이 되는 정보 자원의 속성 및 특성, 그리고 다른 자원과의 관계를 기술함으로써 다양한 검색 방법 및 데이터 관리 방법을 제공한다. 즉, 메타데이터는 저장하는 객체에 대한 실제적인 정보는 아니지만, 저장하는 객체와 다른 객체와의 직간접적인 연관성을 표현하기 위한 데이터이다.[2]

인터넷 보급 및 각종 정보 처리 기술의 발전으로 인하여 다양한 유형의 자원들이 디지털화 되고 있으며 디지털 콘텐츠의 양적 성장은 기하급수적으로 증가하고 있다. 그러나 디지털 콘텐츠는 수집 및 생산 시기, 생산자, 저장 방법 등에 따라서 매우 다양하게 표현되며 또한 표현 방법의 비일관성으로 인하여 같은 콘텐츠가 서로 다르게 간주되는 문제가 있으며 문화에 대한 디지털 콘텐츠의 경우 콘텐츠 생산 과정에 다른 문화 자원과의 복잡한 연관 관

계를 갖을 수 있다. 또한 디지털 콘텐츠는 적용하는 환경과 요구에 맞게 적절한 메타데이터 타입을 요구하며, 이에 따라 각 분야에서 사용되는 메타데이터도 특정 분야의 요구에 맞게 개별적으로 정의된다. 그러므로 디지털 메타데이터를 표준화 하면서도 유연하게 디지털 콘텐츠가 내포하고 있는 다양한 연관 정보 및 특성을 표현할 수 있는 방법에 대한 연구가 필요하다.

본 논문에서는 디지털 콘텐츠를 표준화하고 디지털 콘텐츠간의 다양한 연관 정보를 표현할 수 있는 XML 기반의 메타데이터를 제안한다. 제안하는 메타데이터는 디지털 콘텐츠의 국제 표준 메타데이터 중의 하나인 더블링크어(Dublin Core)에 기반하여 정의한다. 즉, 제안하는 방법은 더블링크어에서 정의한 주요소에 역할 속성을 추가하여 콘텐츠 생성에 직간접적으로 영향을 준 생산자에 대한 정보를 유연하게 표현할 수 있다. 또한 하나의 디지털 콘텐츠 생산 과정에 관련된 다른 디지털 콘텐츠와의 다양한 연관 정보를 유지함으로써 디지털 콘텐츠 생산에 관련된 다양한 연관 정보를 표현할 수 있다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 디지털 콘텐츠 메타데이터에 대한 관련 연구를 기술하고 3절에서는 제안하는 디지털 콘텐츠의 메타데이터를 정의한다. 끝으로 4절에서는 결론 및 향후 연구 방향을 기술한다.

2. 관련연구

인터넷 확산 및 정보 처리 기술의 발전으로 인하여 디

디지털 콘텐츠의 양적 성장은 매우 빠르게 증가하고 있으며 국내외 기관에서도 이러한 디지털 콘텐츠에 표준화하기 위하여 다양한 메타데이터를 정의하고 있다.

한국 문화 콘텐츠진흥원에서는 문화 자원에 대한 디지털 콘텐츠를 표준화하기 위하여 저작자명, 공동저작자명, 제작년도, 언어, 제작처명, 파일 위치, 키워드, 자료 유형 등의 15개의 주요소(main element)로 구성되는 메타데이터를 정의하였다. 또한 문화 관광부에서는 문화 예술 정보 표준으로 제작자, 날짜, 초록, 표현 형식, 언어, 주제 분류 등의 14개의 주요소를 정의하였다. 그러나 정의된 메타데이터는 모두 문화 자원 콘텐츠에 대한 것으로 자원의 기본 정보는 표현할 수 있지만 하나의 문화 콘텐츠 생성에 관련된 다양한 생성자(저작자) 정보를 표현할 수 없는 단점이 있다.[3][4][5]

국외의 사례로 웹을 통하여 다양한 디지털 콘텐츠를 서비스하고 있는 대영 박물관 및 스미소니언 박물관의 경우 소장하고 있는 문화 자원을 각각 12개 및 9개의 주요소로 표현하고 있다.[3] 그러나 이러한 기관들 역시 문화 자원에 대한 기본 정보만을 제공하므로 각 박물관에서 소장하고 있는 특정 형태의 콘텐츠만 표현 가능하며 다양한 문화 콘텐츠 자원의 생산자 정보를 표현할 수 없는 단점이 있다.

더블린코어에서는 50개국 20개 이상의 언어로 정의되어 다양한 분야의 디지털 콘텐츠를 표준화할 수 있는 국제 표준 메타데이터를 정의하였다. 더블린코어의 메타데이터는 그림 1과 같다.[6][7]

번호	요소명	정의
1	Title	제작자(CREATOR)나 발행자(PUBLISHER)가 자필에 부여한 제목. HTML 문서의 경우 <TITLE>, </TITLE> 태그 사이의 값이 해당된다.
2	Author or Creator	자원의 지적 내용에 책임을 진 인물이나 기관. 예) 저자(문헌), 화가(사진가/삽화가/시각자료)
3	Subject and Keywords	자원의 주제나 그 내용을 표현한 명사나 구. 제어어휘나 키워드의 사용을 위해 정의된 요소로서 외부 표준체계에 기반한 분류어휘(scheme-qualified classification data)나 통제어휘(scheme-qualified controlled vocabulary)가 될 수 있다.
4	Description	자원의 내용에 대한 텍스트 기술. 기술 이외에 연결정보(link)를 포함할 수도 있다. 예) 문헌의 초록, 시각자료의 내용기술
5	Publisher	현재의 형태로 자원을 이용 가능케 한 실체. 예) 출판사, 대학, 기업체 등
6	Other Contributor	CREATOR 요소에 명시된 인물이나 기관 이외에, 자원의 내용에 기여한 인물이나 기관. 예) 편집자(editor), 이기자(transcriber), 삽화가(illustrator), covernor 등
7	Resource Type	자원의 범주나 장르. 범주 유형을 나타내는 제어어휘 리스트(preliminary set of types)를 미리 정해 놓고 이 중에서 선택하는 것이 바람직하다. 예) 음레미지, 소울, 시, 토론자료, 기술보고서 등.
8	Date	자원이 현재의 형태로 이용 가능하게 된 날짜. 자원의 데이터 표현형식.
9	Format	사람 혹은 기계가 일정한 환경에서 해당 코드화된 데이터의 사용가능성을 결정하는 데 필요한 정보를 제공한다. 원칙적으로 단행본, 연속간행물 등의 비전자적 형식의 물리적 매체도 포함할 수 있다. TYPE 요소와 마찬가지로 형식을 나타내는 제어어휘집(preliminary set of formats)에서 선택하도록 하는 것이 바람직하다. 예) text/html, ASCII, Postscript, 실행파일(executable application), JPEG 이미지
10	Resource Identifier	자원을 고유하게 식별하기 위한 문자열 혹은 숫자. 예) 네트워크자원 - URL(Uniform Resource Locator), URN(Uniform Resource Name), 인쇄자료 - ISBN
11	Source	해당 자원의 출처가 된 원정보자료. 예) 세인스피어 소켓을 Xml 형식으로 코딩해 놓은 정보자료는 세인스피어의 원 소켓을 정보원으로 한다.
12	Language	자원의 내용을 기술한 언어
13	Relation	공식적인 관계를 가지면서 독립적으로 존재하는 자원들간의 관계를 표현하기 위한 요소. 예) image in a document, chapters in a book, items in a collection
14	Coverage	자원의 지리적, 시간적 특성을 나타내는 요소
15	Rights Management	저작권이나 정보이용권에 관한 공지, 또는 이러한 정보를 통적인 방법으로 제공해 주는 서브포의 링크(URL, URL, ...)를 포함한다. 정보제공자들이 자원에 대해 대한 이용 및 접근조건 혹은 저작권관련 정보를 연결할 수 있는 수단을 제공하는데 목적이 있다.

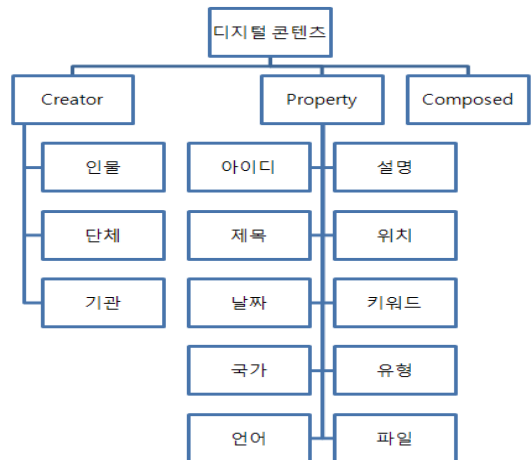
(그림 1) 더블린코어 메타데이터 요소

더블린코어에서 정의한 메타데이터는 다양한 분야의 디지털 콘텐츠를 15개의 주요소를 통하여 표준화하였으며 특정 분야에 쉽게 확장 적용될 수 있는 특성을 갖고 있다. 그러나 더블린코어에서 정의한 메타데이터는 디지털 콘텐츠 생성에 관련된 다양한 생성자 정보 및 생성자의 역할을 표현할 수 없다. 또한 문화 자원에 대한 디지털 콘텐츠의 경우 여러 문화 자원들 간의 포함 관계, 선행 관계, 참조 관계 등 다양한 연관 관계 표현에 한계가 있다는 단점이 있다.

본 논문에서는 디지털 콘텐츠를 표준화하고 콘텐츠간의 연관 정보를 표현할 수 있는 XML 기반의 디지털 콘텐츠 메타데이터를 제안한다. 제안하는 메타데이터는 더블린코어에 기반하며 콘텐츠 생산에 관련된 다양한 생성자를 표현하기 위하여 생성자 요소에 역할을 부여하여 정의한다. 그리고 디지털 콘텐츠 간의 연관 정보를 표현할 수 있는 연관 요소를 추가하여 메타데이터를 정의한다

3 XML 기반의 디지털 콘텐츠 메타데이터

본 논문에서 제안하는 디지털 콘텐츠를 표준화하기 위한 메타데이터는 Creator, Property, Composed의 세 개의 주요소로 구성된다. Creator 요소는 디지털 콘텐츠 생산에 직간접적으로 관련된 인물, 단체, 기관 정보를 나타낸다. 이 때 디지털 콘텐츠 생산에 기여한 생성자의 역할을 부여하여 정의함으로써 다양한 역할의 생성자 정보를 표현할 수 있다. Property 요소는 아이디, 제목, 제작날짜, 국가, 언어, 세부설명, 위치, 키워드, 유형, 파일 정보의 세부 요소로 나누어지며 디지털 콘텐츠에 대한 필수 항목 정보를 표현한다. 그리고 Composed 요소는 디지털 콘텐츠에 직간접적으로 영향을 준 다른 디지털 콘텐츠와의 연관 정보를 표현한다. 이 때 포함하는 정보는 해당 디지털 콘텐츠를 참조할 수 있는 키 정보를 저장한다. Creator, Property, Composed로 구성되는 메타데이터는 그림 2와 같다.



(그림 2) 디지털 콘텐츠의 메타데이터

제안하는 메타데이터를 적용하여 디지털 콘텐츠를 표현하는 예로 오페라의 유령(콘텐츠)을 표현하여 보자. 오페라의 유령의 Creator 요소로 감독(조엘 슈마허), 배우(제라드 버틀러, 에미 로섬), 각본(앤드류 로이드 웨버), 배급사(위너 브로스) 등의 정보를 역할과 함께 기술한다. 그리고 Property 요소는 오페라의 유령에 대한 필수 항목 정보를 입력한다. 이 때 Property 요소에 포함되지 않는 기타 내용들은 세부설명 요소에 기술한다. Composed 요소에는 오페라의 유령 제작에 영향을 준 다른 디지털 콘텐츠 정보로 원작, OST 정보 등을 기술한다. 영화 오페라의 유령을 주제로 한 디지털 콘텐츠의 메타데이터는 그림 3과 같이 기술한다.

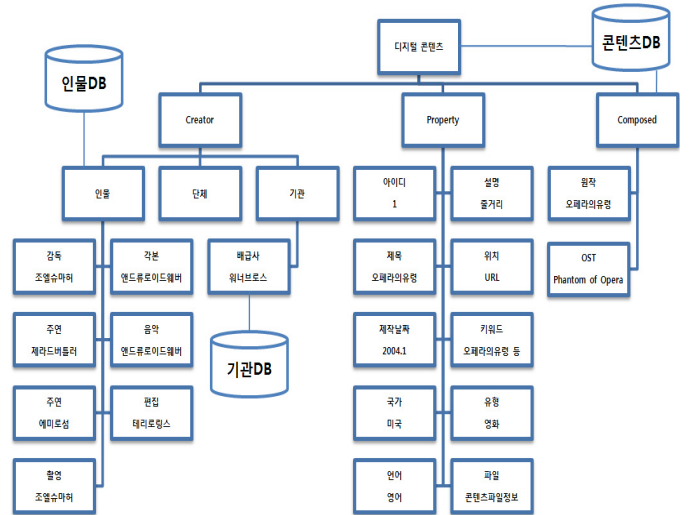
```

<Digital content>
<Creator>
  <인물>
    <role=감독> 조엘 슈마허 </role>
    <role=주연> 제라드 버틀러 </role>
    <role=주연> 에미 로섬 </role>
    <role=각본> 앤드류 로이드 웨버 </role>
    <role=촬영> 조엘 슈마허 </role>
    <role=음악> 앤드류 로이드 웨버 </role>
    <role=편집> 테리 로링스 </role>
  </인물>
  <단체/> </단체/>
  <기관>
    <role=배급사> 위너 브로스 </role>
  </기관>
</Creator>
<Property>
  <아이디> 1 </아이디>
  <제목> 오페라의 유령 </제목>
  <제작날짜> 2004.1 </제작날짜>
  <국가> 미국 </국가>
  <언어> 영어 </언어>
  <설명> 줄거리 및 세부 내용 </설명>
  <위치> 디지털 콘텐츠 URL </위치>
  <키워드> 오페라의 유령, 조엘 슈마허, 팬텀 </키워드>
  <유형> 영화 </유형>
  <파일> 콘텐츠 파일 정보 </파일>
</Property>
<Composed>
  <role=원작> 오페라의 유령 </role>
  <role=OST> The Phantom of the Opera </role>
</Composed>
</Digital content>
    
```

(그림 3) 디지털 콘텐츠 표현 예

그림 3에서 표현하는 디지털 콘텐츠의 메타 데이터는 하나의 콘텐츠 생산에 일곱 명의 인물과 하나의 기관이 참여하였으며 콘텐츠 생산에 다른 두 개의 콘텐츠가 각각 원작 및 OST 형태로 관여하였음을 알 수 있다. 그리고 이러한 생산된 콘텐츠에 영향을 준 두 개의 디지털 콘텐츠는 다시 그림 3과 유사한 형태로 각각 정의된다.

설계한 디지털 콘텐츠의 메타데이터는 크게 생성자 테이블, 디지털 콘텐츠 테이블, 그리고 역할 및 관계 테이블로 나누어 저장된다. Creator 요소에 대한 데이터는 생성자의 종류에 따라 각각 인물, 단체, 기관 형태로 나누어 생성자 테이블에, Property 요소에 속하는 정보는 디지털 콘텐츠 테이블에 저장된다. 그리고 역할 및 관계 테이블은 생성자의 역할 및 디지털 콘텐츠간의 연관 정보를 저장한다. 그림 3의 예는 실제로 그림 4와 같은 형태로 저장된다.



(그림 4) 메타데이터를 지원하는 테이블 구성

4. 결론 및 향후 연구과제

본 연구에서는 디지털 콘텐츠 정보를 표준화하고 다양한 생성자 및 콘텐츠 간의 연관 관계를 표현할 수 있는 XML 기반의 메타데이터를 정의하였다. 설계한 메타데이터는 Creator 요소에 역할을 추가하여 정의함으로써 디지털 콘텐츠 생성에 관련된 인물, 단체, 기관 정보를 표현할 수 있다. 또한 디지털 콘텐츠간의 다양한 연관 관계 정보를 Composed 요소를 통하여 정의할 수 있다. 그리고 설계한 디지털 콘텐츠의 메타데이터에 맞추어 영화 콘텐츠를 표현함으로써 다양한 생성자 정보 및 디지털 콘텐츠간의 연관 정보를 효율적으로 표현할 수 있음을 알 수 있었다. 향후 연구 방향은 디지털 콘텐츠 종류에 따른 다양한 생성자 역할 및 속성, 연관 관계에 대한 시소러스를 정의하고 이에 기반하여 메타데이터 타입을 개선하고자 한다.

참고문헌

- [1] 황상규 외. “디지털 콘텐츠를 위한 메타데이터에 관한 연구”, 2001.
- [2] 조윤희, 황동열. “디지털콘텐츠 집중관리를 위한 표준화에 관한 연구”, 2003.
- [3] 황동열. “미술 정보 데이터베이스 구축을 위한 데이터 요소에 관한 연구”, 1992.
- [4] 인디애나 대학 디지털 음악 라이브러리 (<http://www.dml.indiana.edu/>).
- [5] 한국데이터베이스진흥원. “온라인디지털 콘텐츠 시장의 현황과 전망”, 2003.
- [6] Becktt, David. “IAFA Templates in Use as Internet Metadata”, 2003.
- [7] DCMI. “Dublin Core Metadata Element Set, Version 1.1: Reference Description”, 1999.