

디지털 콘텐츠를 위한 메타데이터에 관한 연구*

: INDECS를 중심으로

A Study on the Characteristic of Metadata for Digital Contents : Based on the INDECS

황상규 ; 한국국방연구원

윤세진, 오경묵 ; 숙명여자대학교 문헌정보학전공

Sang-kyu Hwang ; KIDA

Se-Jin Yun, Kyung-Mook Oh ; Sookmyung Womens University

웹 정보 자원은 서지 자료와는 달리 인터넷상에 널리 분산되어 있으며, 그 특성상 계속하여 동적으로 변화하기 때문에, 전통적인 메타데이터의 주제 접근방식으로 접근하기는 어렵다. 또한, 인터넷상의 웹 정보 자원에 폭발적 증가는 더 이상 메타데이터 전문가가 아닌, 저작물의 생산자가 직접 메타데이터를 만들어내는 생산자로서의 역할까지 수행하도록 요구한다. 디지털 콘텐츠를 위한 메타데이터는 인터넷 전자상거래에서처럼 보다 복잡하고 다양한 용도로 그 사용범위가 확대되어짐에 따라, 디지털 콘텐츠를 위한 새로운 메타데이터 표준안은 기존 메타데이터들의 역할인 대상에 대한 세부적인 유형과 특징을 기술한 형식적 정보(syntactic information) 전달기능을 제공함과 동시에, 어떠한 사건이 발생한 상황을 정확히 기술할 수 있는 의미적 정보(semantic information) 전달기능의 역할을 수행해야 한다.

1. 웹 정보 자원을 위한 메타데이터

웹 정보 자원의 가장 큰 특징은 정보원이 인터넷상에 분산되어 있다는 점과 동일한 정보자원이 상이한 형식의 포맷으로 표현될 수 있으며, 사용용도별 주제 분야에 따라, 해당내용을 기술하는 메타데이터의 종류와 형식이 각기 다양해질 수 있다는 점이다. 이러한 이유로 동일한 내용의 정보가 웹 상의 각기 다른 위치에

서 각기 상이한 형식의 문서포맷으로 존재할 때, 해당 포맷이 다르기 때문에 동일한 내용을 각기 서로 다른 정보자원으로 식별하게 되는 문제점이 발생하기도 한다. 또한, 인터넷상에 분산되어 있는 웹 정보 자원은 서지 자료와는 달리 그 성격이 동적이기 때문에, 전통적인 목록과 같은 주제 접근방식으로 접근하기는 불편하다. 그러므로 웹 정보 자원을 기술하기 위한 새로운 표준이나 기법이 필요하게 되었다.

* 이 연구는 과학기술부 지원 '01 숙명여자대학교 연구기반확충사업 (01-N6-02-01-A-01)에 의하여 지원되었음.

이러한 관점에서 인터넷상에서 흩어져 있는 다양한 정보원에 대해 국민들이 빠르고 쉽게 접근할 수 있는 방법을 생각하게 되었으며, 그 결과로 고안된 것이 정부에서 보유하고 있는 정보를 일반 국민들이 공유할 수 있는 시스템의 한 일례가 GILS(Government Information Locator Service)이다. GILS는 미국정부산하 각 기관들이 보유하고 있는 정보의 소재 여부와 위치 정보를 기술하며, 정보자원 획득 방법을 제공하게 된다. GILS의 경우, Core Element 요소들은 USMARC 포맷과 상호참조가 가능하도록 설계되어 있다. 그러나, 보다 나중에 설계, 개발된 AGLS(Australian Government Locator Service)에서는 USMARC 대신 더블링크어 메타데이터를 기초로 하고 있는데, 이는 이용자가 실제 접근하게 되는 대부분의 정보가 웹 정보자원의 형태이고, 이를 위한 보다 효과적인 서비스를 위하여 더블링크어 메타데이터의 가장 중요한 특성인 '단순성(simplicity)'을 중요시한 결과이다.

현재 인터넷상의 웹 정보 자원에 폭발적 증가는, 더 이상 메타데이터를 한 단체나 개인의 일률적인 통제하에 생산, 유지 관리하는 것이 거의 불가능한 단계에 이르도록 하였다. 이에 대한 현실적인 대안으로, 메타데이터의 하나 혹은 다수의 표준 하에서 분산된 작업형태로 저작물의 생산자가 직접 그 저작물에 대한 메타데이터를 기술해야 한다. 따라서 메타데이터의 목록전문가가 아닐지라도 저작물의 생산자가 쉽게 작성할 수 있도록 하는 메타데이터의 '단순성'이, 새로운 형태의 정보원인 인터넷 환경에서는 메타데이터가 갖추어야 할 가장 중요한 특성이 된다. 그러나, 웹 환경 하에서 메타

데이터가 갖추어야 할 가장 중요한 특성으로 '단순성'만을 중시할 경우에는, 더 복잡하고 세부적인 메타데이터 정보요소들을 더블링크어 형식으로는 표현하지 못한다는 문제점이 발생하게 된다. 이는 디지털 콘텐츠를 판매하는 전자상거래에서 관련 메타데이터정보를 유지, 관리하는 경우 보다 심각해지는데, 예를 들어 저자가 세 명인 어떤 단행본이 미국판과 유럽판으로 동시에 출간되었고, 2년 후 일본어와 한국어로 번역판이 출간되었다. 그리고 제 3의 국가에서 불법 복사본으로 출판되고, 제 2판에서 내용이 부분적으로 수정되는 등, 하나의 자료가 여러 상이한 형태로 변화할 때, 이러한 모든 일련의 변화를 기존의 더블링크어의 15개 메타데이터요소들 만으로는 표현하기 힘들기 때문이다. 이를 위하여, 디지털 콘텐츠의 전자상거래를 위한 새로운 메타데이터 요소들이 정의되고 추가, 보완되게 되었으며, 그런 성과물의 하나가 전자상거래를 위한 메타데이터 INDECS (INteroperability of Data in E-Commerce Systems)이다[Godfrey2000]. INDECS는 특히, 저작물의 각각의 세부적인 유형과 특징을 기술하고, 저작권과 관련하여 계약, 권리 이전, 권리 행사 등, 누가 소유권을 가지는지를 설명하는 인증기능을 아울러 제공하고 있다.

2. 전자상거래를 위한 메타데이터 요소

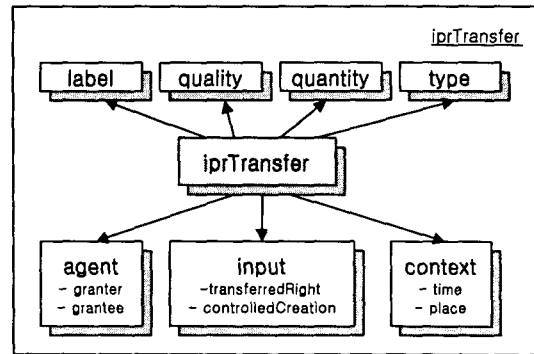
더블링크어 메타데이터는 주요 메타데이터 요소별로 세분화된 한정어를 확정하고, 각 요소별로 또다시 상세구분 한정어를 결정함으로써 요소의 의미를 보다 좁히거나 더욱 특정화시킨다. 이러한 더블링크어 메타데이터 요소는

일반적으로 그 특성상 3개의 그룹으로 나눌 수 있으며, 메타데이터 요소들이 세분화함에 따라 계층적 형태의 구조를 이루게 된다.

내용 (content)	지적재산 (intellectual property)	구현 (instantiation)
Title	Creator Publisher Contributor Right	Date Type Format Identifier
Subject		
Description		
Source		
Language		
Relation		
Coverage		

[표 2] 더블린코어 메타데이터 요소

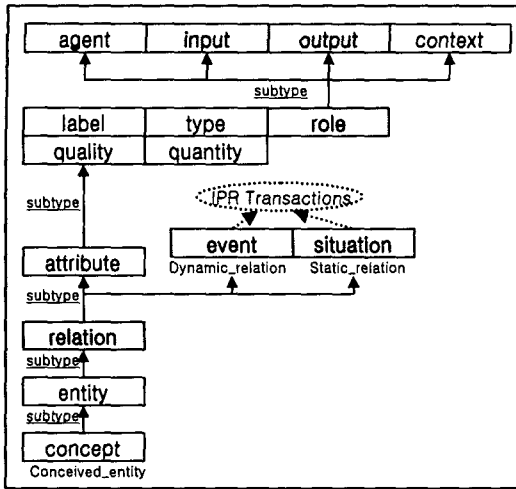
그러나, 더블린코어를 포함한 기존 메타데이터들의 가장 큰 단점은 대상의 세부적인 유형과 특징을 기술한 형식적 정보(syntactic information) 전달기능을 수행할 수 있으나, 어떠한 사건이 발생한 상황을 정확히 기술한 의미적 정보(issue of semantic identity) 전달의 기능을 제대로 수행하기 힘들다는 점이다. 예를 들어, 한글로 작성된 저작물 'A'가 2001년도에 PDF화일 형태로 전자 출판되었다면, 저작물 'A'의 언어가 한글이고, 출판년도가 2001년이며, 형식이 PDF라는 형식적 정보들은 기존 더블린코어의 메타데이터요소들 만으로도 충분히 표현 가능하다. 그러나 다양한 양상을 띄는 현실 상황에서 저작물 'A'가 영어로 된 저작물 'a'의 한글판인 경우, 발생한 상황의 이면에는 양수자(granter) '갑'이 양수인(grantee) '을'에게 저작물 'a'에 대한 '지적재산권(IPR)'을 '제공(iprTransfer)'하는 것에 '동의(agreement)'하는 '상황(situation)'이 발생했음을 전제한다[그림 1].



[그림 1] Indecs Diagram (iprTransfers)

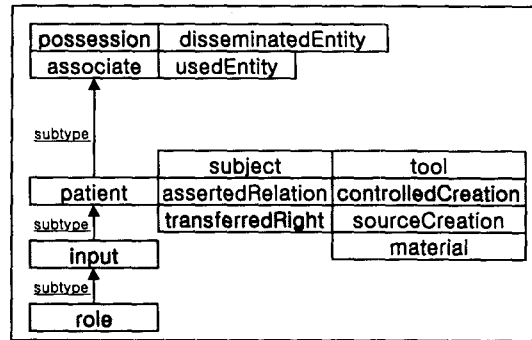
이는 전자상거래상에서 누가 소유권을 가지는지를 설명하는 저작권과 관련하여 발생하는 행위인 계약, 권리 이전, 권리 명시 등을 기존의 더블린코어의 메타데이터요소들 만으로 표현하기에는 그 표현상의 한계가 있음을 확인할 수 있다.

또한, 위와 같은 상황을 보다 깊고 면밀히 살펴보면, 메타데이터로 표현해야 할 어떠한 대상의 역할은 항상 고정적이기보다는 주어진 상황(situation)에 따라 그 대상의 메타데이터 엘리먼트 속성이 결정되는 것이 바람직하다는 점을 살펴볼 수 있다. 앞의 예에서 행위자 '을'은 '지적재산권(IPR)'을 넘겨받는 양수인(grantee) '을'인 동시에, 한글판 저작물 'A'를 출판하는 행위에서는 출판작업을 총괄하는 책임자(director) '을'로서의 역할을 하게 될 수 있다. 이러한 INDECS의 기본설계방침을 검토, 확인하기 위하여, 더블린코어 메타데이터 요소 중 지적재산권(Right)에 해당하는 INDECS의 지적재산권업무(IPR Transactions)와 관련된 주요 엘리먼트들을 INDECS 프레임워크 구조에 맞추어 계층화시켜 도식화하였다[그림 2,3,4].

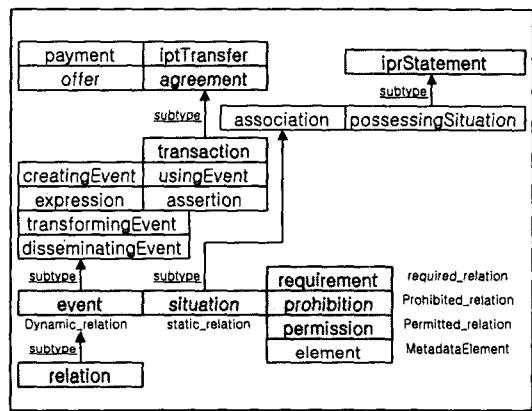


[그림 2] Indecs framework diagram (entity)

[그림 2]는 INDECS 프레임워크의 전체적인 구조를 보여주고 있는데, INDECS는 일반적인 관점(general view)에서 식별되어지는 대상인 entity와 entity간의 관계인 relation으로 대상을 모델링(data modeling)하게 된다. entity간의 관계인 relation은 크게 동적관계인 event와 정적인 관계인 situation으로 구성되며, 하나의 entity는 다른 entity간의 관계(relation)에서 어떠한 역할(role)을 수행하게 된다. 역할(role)은 크게 4부분으로 구성되어지는데, 어떠한 사건에서 벌어지는 행위의 역할(agent)과 행위의 소스가 되는 input, 행위의 결과로 만들어지는 output, 마지막으로 행위가 벌어진 시점과 장소를 설명하는 context로 구성된다.



[그림 3] Indecs framework diagram (role)



[그림 4] Indecs framework diagram (relation)

[그림3,4]는 크게 relation, event, situation, role이라는 4가지 범주별로 지적재산권업무(IPR Transactions)와 관련된 주요 엘리먼트들을 보여주고 있는데, 지적재산권 관련 엘리먼트들은 식별이 용이하도록 검게 표시하였다.

더블린코어 메타데이터와 비교하여 INDECS의 지적재산권과 관련된 메타데이터 엘리먼트의 특징을 살펴보면, 메타데이터요소들이 한 범주 내에 묶여있기 보다는 여러 범주에 흩어져 있음을 살펴볼 수 있다. 기존 더블린코어 메타데이터 요소들이 형식적 정보(syntactic information)을 제공하기 위해 설계되었음에 반하여, INDECS의 메타데이터요소들은 주어진

상황(situation)에 따라 그 대상의 메타데이터 엘리먼트 속성이 결정되기 때문이다. [그림 3]의 transferredRight는 지적재산권은 양도하는 상황(situation)인 iprTransfer의 소스로서의 역할을 하게 됨으로서, input이라는 범주에 속하게 되며, 마찬가지로 지적재산권은 양도하는 상황을 가리키는 iprTransfer는 situation이라는 범주에 속하게 된다.

3. 디지털 콘텐츠 메타데이터의 발전방향

지금까지 디지털 콘텐츠의 전자상거래를 위한 새로운 형태의 메타데이터의 하나로 INDECS의 특성에 대해 살펴보았다. INDECS의 가장 큰 특징으로는 어떤 상황에 대한 정확을 정확히 기술하기 위하여, 의미적 정보 전달의 기능을 강조하였음을 확인할 수 있었다. 또한, 메타데이터로 표현해야 할 어떠한 대상의 역할은 항상 고정적이기보다는 주어진 상황에 따라 그 대상의 메타데이터 엘리먼트 속성이 결정되는 것을 살펴볼 수 있는데, 이는 디지털 콘텐츠의 전자상거래를 위한 메타데이터 요소들은 사용 목적과 용도에 따라, 현재의 더블린 코어보다는 보다 세분화되고 다양해짐을 확인할 수 있다. 그러나, 새로 만들어진 메타데이터는 웹 환경 하에서 메타데이터가 갖추어야 할 가장 중요한 특성으로, 메타데이터의 전문가가 아닐지라도 저작물의 생산자가 쉽게 작성할 수 있도록 하는 기본적 속성인 '단순성(simplicity)'을 위배하게 된다. 또한 새로운 미디어 매체가 계속 개발되어지고, 전자상거래가 이루어지는 상황 역시 점점 더 복잡하고 다양해짐에 따라 메타데이터의 대상 범주와 항목수

는 점점 더 늘어나게 될 것이다. 이러한 문제 상황을 위한 메타데이터의 현실적인 기본개발 방침으로, 디지털 콘텐츠를 위한 메타데이터는 하나의 표준안보다는 대상주제와 사용목적에 따라 각기 세분화될 필요가 있다. 실례로, 전자상거래의 여러분야 중 저작권관리(Digital Rights Management)만을 대상으로 한 메타데이터인 XrML(eXtensible rights Markup Language)이 개발되었으며, INDECS 역시 실제 사용(Usage)을 위한 메타데이터라기보다는 저작물의 생산(creation)에 관계된 등록(register)을 위한 메타데이터표준안으로의 성격이 강하다.

마지막으로, 디지털 콘텐츠를 위한 메타데이터인 INDECS프레임워크의 구조적 특징과 개선방향에 대해 살펴보기로 한다. 더블린코어를 포함하여 기존의 메타데이터가 카테고리 범주 별로 나뉜 계층적 구조를 가진데 반하여, INDECS는 메타데이터 요소들을 주어진 상황(situation)의 역할에 따라 의미소들(semantic elements)을 그룹화, 계층화함에 따라, 상대적으로 구조적 명료성이 떨어진다는 단점을 가지고 있다. 앞선 예제에서도, INDECS의 지적재산권과 관련된 메타데이터요소들이 한 카테고리 범주 내에 묶여있기 보다는 여러 범주에 흩어져 있음을 살펴볼 수 있었다. 이는 결과적으로, 메타데이터의 전문가가 아닌 저작물의 생산자가 새로운 표준안에 익숙해지고, 일정 수준 이상의 메타데이터 생산자로서의 역할에 도달하는데 걸리는 학습시간이 길어지는 요인으로 작용하게 된다. 따라서, INDECS가 보다 효과적인 메타데이터 표준안으로 널리 이용되기 위해서는 꾸준한 INDECS 프레임워크의 구조

적 개선이 필연적이라 할 수 있겠다.

참고문헌

- 이석기. 1999. 디지털 도서관 구축을 위한 메타 데이터 코딩 방안. 『디지털도서관』, 1999년 겨울호, 74-90.
- GILS. "Global Information Locator Service (GILS) - making it easier to find all the information." [cited 2001. 5. 3].
<<http://www.gils.net/index.html>>
- Cunningham, A. 1998. "Enabling Seamless Online Access to Government", Metadata Unravelled Seminar, 26 August 1998.
<http://www.naa.gov.au/recordkeeping/gov_online/agls/Metadata_paper22sept98.html>
- Weibel, Stuart and Iannella, Renato. 1997. "The 4th Dublin Metadata Workshop Report."
D-Lib Magazine, June 1997. [online].
[cited 2001. 6. 17].
- Rust, Godfrey and Bide, Mark. 2000 "The <indec> Metadata Framework: Principles, model and data dictionary."
[cited 2001. 2.22].
<<http://www.indec.org/pdf/framework.pdf>>
- DCMI. 1999. "Dublin Core Metadata Element Set, Version 1.1: Reference Description."
[cited 2000. 12.5].
<<http://www.dublincore.org/documents/dces/>>