

Web상의 전자문서를 위한 메타데이터 모델의 제안 및 관리시스템의 개발

정 호 택[†] · 양 영 종[†] · 김 순 용[†] · 이 상 덕[†] · 최 윤 철^{††}

요 약

인터넷상에서 탐색 엔진을 사용하여 원하는 데이터에 정확하게 접근하기는 쉽지 않다. 왜냐하면 선택되는 데이터의 수가 많을 뿐만 아니라 해당 데이터와 관련된 충분한 정보를 제공하지 않기 때문이다. 메타데이터는 데이터에 대한 데이터로서, 데이터의 내용은 물론 데이터 자체에 대한 정보도 포함하고 있다. 사용자는 메타데이터를 검색함으로써 해당 데이터에 대한 충분한 정보를 얻을 수 있으므로 원하는 데이터에 보다 정확하게 접근할 수 있으며 그로 인한 데이터의 활용성도 증대될 것이다.

본 논문에서는 메타데이터와 관련된 기술 및 현재 표준화 작업이 진행중이거나 표준으로 채택된 메타데이터 모델들을 분석하였다. 이를 참조하여 인터넷상의 과학기술 분야의 논문, 보고서, 기술문서 등의 문서나 그래픽 이미지를 대상으로 하는 SeriCore 메타데이터 모델을 제안하였다. 제안된 SeriCore 모델에 따라 메타데이터를 효과적으로 생성, 저장, 검색할 수 있는 SeriCore 메타데이터 관리시스템을 설계하고 구현하였다.

Design of Metadata Model and Development of Management System for Electronic Documents on the Web

Hyo-Taeg Jung[†] · Young-Jong Yang[†] · Soon-Yong Kim[†] · Sang-Duk Lee[†]
· Yoon-Chul Choy^{††}

ABSTRACT

It is not easy to access to the required data from the Web by using search engines because there are too many data selected and they do not provide enough information related to the corresponding data.

Metadata is data about data. It includes information about data itself and contents of data as well. Users can acquire enough information about the corresponding data and access to the required data exactly using metadata, and therefore the data usability will be increased.

In this paper, several metadata technologies and metadata models that are already in process of standardization or adopted as standards are analyzed, and the SeriCore Metadata Model for documents such as papers, project reports, technical reports, abstracts, and manuals, and graphic images that are in the field of science technologies on the Web is proposed. The SeriCore Metadata Management System that can generate, store, and retrieve metadata effectively is designed and implemented based on the SeriCore Metadata Model.

[†] 정 회 원 : 시스템공학연구소 소프트웨어공학연구부

^{††} 손 회 원 : 연세대학교 컴퓨터학과 멀티미디어/그래픽스 연구실
논문접수 : 1997년 8월 5일, 심사완료 : 1998년 2월 2일

1. 서 론

사용자가 이미 구축되어 있는 데이터나 데이터 집합을 적절하게 사용하거나 네트워크를 통하여 필요한 데이터를 전송 받기 위해서는 해당 데이터에 대한 정확한 정보를 필요로 한다. 즉, 데이터 내용을 기술하는 정보뿐만 아니라 내용과는 상관없는 데이터 자체에 대한 정보도 필요하게 된다. 메타데이터는 바로 이러한 데이터를 일컫는 용어로서, 각 응용 분야의 특성에 따라 표준화되거나 모델링이 되어야 하며 이를 바탕으로 사용자의 요구에 맞게 구축되어야 한다.

메타데이터에 대한 인식이 제고되고 그 중요성이 강조됨에 따라 지리정보 시스템에서 사용되는 공간 지리 데이터에 대한 표준 모델 및 표준안 제정을 시작으로 메타데이터에 대한 연구가 여러 응용 분야에 걸쳐 활발히 진행되고 있다. 특히, 최근 들어 통신 기술의 발전과 인터넷의 등장으로 다양한 정보에 대한 요구가 증가하고 네트워크를 통한 정확한 정보의 접근에 대한 중요성이 강조됨에 따라 인터넷상의 자원을 대상으로 한 메타데이터의 표준화 작업 및 이를 지원하는 소프트웨어의 개발이 활발히 추진되고 있다. 또한 표준화 작업 과정에서 여러 가지 의견을 수렴하는 동시에 문제점을 해결하기 위한 방안으로 국제 학술대회나 워크샵 등이 빈번히 개최되고 있다.

본 논문의 2장에서는 메타데이터와 관련된 전반적인 내용을 살펴보고, 3장에서는 본 연구에서 참조한 표준 모델들의 특성과 내용을 분석한다. 특히 이들 모델과 관련되어 개최된 컨퍼런스나 워크샵 등을 통하여 메타데이터 표준 모델과 관련된 이슈 및 표준화 과정 등을 알아본다. 4장에서는 인터넷상의 과학기술 분야의 논문, 보고서, 기술문서 등의 문헌이나 그래픽 이미지에 적합한 메타데이터 모델을 제안하고 특성들을 기술하며, 5장에서는 4장에서 제안한 모델을 근거로 전용 편집기를 구현하여 메타데이터를 생성하고, 이를 UniSQL상의 데이터베이스로 구축한 후, 효율적으로 저장, 검색하기 위한 메타데이터 관리시스템을 구현한다. 6장에서는 결론으로 향후의 연구 방향이나 활용에 관하여 기술한다.

2. 메타데이터의 개요

2.1 메타데이터의 정의와 종류

메타데이터는 데이터에 관한 데이터(data about data)로 정의할 수 있다[1]. 방대한 양의 데이터 시스템을 사용하는 사용자들에게는 서로 다른 형태의 자료들로 인한 복잡성이 가장 큰 문제로 대두되었으며, 이를 해결하기 위해서는 해당 정보와 그 구성요소들에 대한 물리적이고 개념적인 세부사항을 저장하는 것이 필요하게 되었다. 그런 의미에서 메타데이터는 이미 존재하고 있는 데이터 집합 자체에 대한 정보는 물론 여러 가지 다양한 형태의 정보를 제공하기 위한 메카니즘으로 볼 수 있다[1]. 최근 들어서는 데이터나 데이터 집합을 효율적으로 접근하고 관리할 수 있도록 해 주는 데이터에 대한 정보를 총칭한 용어로 사용되고 있다[2]. 즉, 데이터가 생성됨에 따라 결정되는 데이터의 구조, 표현 기법, 참조 정보, 내용 서술 정보, 데이터의 접근, 획득, 배포, 활용에 관한 정보, 생성자 및 관리자 정보 등을 광범위하게 서술하는 데이터 집합 또는 정보를 의미한다.

메타데이터는 응용 분야에 따라 포함하고 있는 내용이 각각 다르기 때문에 일반화하여 분류하기는 매우 어려우나 관점에 따라 몇 가지 종류로 분류될 수 있다. R.M.Sumpter는 시스템 수준의 메타데이터와 응용 수준의 메타데이터로 구분하였다[2]. 시스템 수준의 메타데이터는 파일 시스템의 특성을 구현하거나 혹은 파일 시스템내의 데이터를 관리하기 위한 정보를 의미하는데, 계층적 저장 공간에서의 접근속도나 자료의 크기, 저장 공간의 위치, 데이터 블록의 저장 방법 등을 서술하는 정보가 될 수 있다. 응용 수준의 메타데이터는 사용자가 데이터를 운용하고 접근하는데 도움을 주는 사용자 데이터와 관련된 정보를 의미하며, 문서나 그래픽, 비디오 등의 내용에 관한 요약 정보, 혹은 각각의 데이터간의 관계를 나타내는 정보 등이 될 수 있다.

V.Kashyap, K.Shah, A.Sheth는 데이터 내용과의 관련성 여부에 따라 세 종류로 분류하였는데[3], 데이터의 내용에만 관계되며 인간의 인식이나 지식을 기반으로 생성되는 메타데이터(Content-dependent Metadata), 데이터의 내용을 서술하는 메타데이터(Content-descriptive Metadata), 데이터의 최종 갱신일이나 저장 공간 위치 등과 같이 내용과는 직접적으로 관련이 없는 메타데이터(Content-independent Metadata) 등이다.

K.Bohm과 T.C.Rakow는 멀티미디어 문서에 관한 메타데이터를 다음과 같이 분류하였다[4]. 즉, 데이터

의 포맷, 코딩방법, 압축기법 혹은 사용언어 등과 같이 멀티미디어 표현과 관계되는 메타데이터(Metadata for the Representation of Media Types), 멀티미디어 문서의 내용을 서술하는데 사용되며 미디어의 형태와 관계되는 메타데이터(Content-Descriptive Metadata), 문서의 내용이 포함하고 있는 전문지식 수준등과 같이 문서의 내용으로부터 부가적으로 유도되는 메타데이터(Metadata for Content Classification), 문서를 구성하고 있는 요소들간의 관계와 각 요소들의 역할을 서술하는 메타데이터(Metadata for Document Composition), 멀티미디어 요소의 상태 혹은 저작자의 이름이나 최종 갱신일 등과 같이 히스토리화 관계된 데이터(Metadata for Document History), 데이터의 접근과 위치에 관련된 메타데이터(Metadata for Document Location) 등으로 분류하였다.

2.2 메타데이터의 모델링

메타데이터를 생성하기 위해서는 생성 규칙, 사용 용어, 표기법과 같은 문법을 정의하고, 구조 및 구성요소, 각 구성요소의 의미, 데이터 형태 및 데이터 영역 등을 정의해야 한다. 이처럼 데이터에 대한 메타데이터의 정형적인 모델을 정의하는 것을 메타데이터의 모델링이라 하며, 메타데이터의 생성, 저장, 관리의 효율성과 디지털 정보에 대한 접근, 검색, 교환의 활용성 등에 관련된 중요한 과정이나, 멀티미디어 정보에 대한 메타데이터 모델링에 있어서는 미디어별 특성을 적절하게 표현할 수 있는 포괄적인 데이터 모델을 채택하여 비디오나 오디오와 같은 비정형 디지털 데이터의 구조화를 가능하게 하여야 한다. 또한 메타데이터를 효율적으로 저장, 관리함으로써 네트워크를 통한 접근, 검색이 용이하여야 하며 공공성이 높은 정보에 대한 재활용성을 향상시킬 수 있어야 한다.

메타데이터의 모델링시에 고려해야 할 중요한 사항은 다음과 같다.

- 더블린 코어, GILS코어, FGDC의 CSDGM 등과 같이 확장성 및 범용성이 높은 표준적인 메타데이터 모델을 참조하고, 표현 방식에 있어서도 국제 문서 표준인 SGML 등과 같은 표준을 이용 하는 것이 바람직하다.
- 대상 미디어별 특성 요소를 수용할 수 있어야 한다. 즉, 정적이거나 동적인 특성은 물론 구조화의

정도나 미디어의 복합 여부 등을 고려하여야 한다.

- 대상 응용 분야의 특성 및 메타데이터의 활용 목적 등을 수용하여야 한다.
- 메타데이터의 저장, 색인, 관리 등을 위한 하부 저장 구조를 고려하여야 한다. 즉, 화일 시스템을 사용할 것인지, 혹은 데이터베이스를 사용할 것인지 결정하고, 그에 따른 데이터의 내부 구조도 결정하여야 한다.
- 효율적인 메타데이터의 검색이 가능하여야 한다. 일반 인터넷 검색 엔진의 사용이 가능한지, 혹은 하이퍼텍스트 기반의 브라우징이 가능한지 등을 고려할 수 있다.
- 메타데이터 구성요소의 특성을 고려하여야 한다. 즉, 요소를 특성에 따라 필수요소, 선택요소, 반복이 가능한 요소 등으로 구분할 수 있으며, 메타데이터 내용의 형식을 지정하는 요소나 속성 등을 고려할 수 있다.

2.3 메타데이터의 관리와 활용

메타데이터의 생성 방법으로는 메타데이터를 자동으로 생성, 추출하는 방법과 원시 데이터의 외형 묘사, 내용 서술, 요약 정보 등을 GUI기반의 메타데이터 편집기를 이용하여 태깅, 편집하는 방법이 있으며, 대부분의 경우는 두 가지 방법을 병행하여 메타데이터를 생성한다. 최근에 텍스트 데이터에 대한 메타데이터 생성 도구의 개발은 물론 비디오, 오디오, 이미지와 같은 비정형 멀티미디어 데이터에 대한 메타데이터의 자동 추출, 접근 및 처리 기법이 개발되고 있으나 [5][6][7][8], 무한히 많은 내용을 포함하고 있는 비디오 데이터로부터 의미있는 정보를 추출하기에는 아직 어려움이 많은 실정이다. 최근에는 메타데이터의 생성 및 관리시에 메타데이터 모델의 적합성을 검증하거나, 혹은 텍스트, HTML, SGML등과 같은 표준화된 양식으로 메타데이터를 출력시켜 주는 기능을 갖는 메타데이터 해석기 등이 개발, 이용되고 있다.

생성된 메타데이터의 저장, 변경 및 유지 관리를 위해서는 데이터의 보안성이나 무결성 등이 지원되는 하부구조를 선정하여야 한다. 예를 들면 메타데이터 검색의 효율성과 활용성을 고려하여 메타데이터를 분산 저장하거나 또는 집중 저장할 수 있으며, 메타데이터 전용 서버나 혹은 클리어링 하우스(Clearing House)를

이용하여 메타데이터를 공동으로 활용, 관리할 수도 있다. 또한 모델의 변경에 따른 메타데이터의 변경 관리 방안도 고려할 필요가 있다.

메타데이터를 활용하기 위해서는 GUI기반의 질의, 검색, 브라우징 기능을 제공하는 단독적인 응용 프로그램을 개발하거나, 메타데이터 서버 또는 클리어링 하우스 등에 저장된 메타데이터의 활용을 위해서 클라이언트/서버 환경하에서 멀티미디어 또는 하이퍼미디어 정보 시스템과 통합하여 활용하는 방안 등이 있다. 어떤 경우이든지, 메타데이터를 일반 정보검색 도구와 연계하여 활용하거나, 텍스트 뿐만 아니라 하이퍼링크 기능이 제공되는 HTML, SGML 등 다양한 양식의 표현 방법들을 제공할 수 있어야 한다. 최근에는 인터넷을 통한 정보의 공유 및 교환이 실용화됨에 따라, 공간 지리 정보를 포함한 각종 전문 분야 데이터 집합의 생성시 기존의 메타데이터 참조 모델을 적용하거나 별도의 표준 양식을 설정하고, 이에 의해 메타베이스를 구축, 활용하는 등의 표준화 작업이 활발히 이루어지고 있다.

3. 메타데이터 표준 모델의 분석

최근 여러 연구기관이나 단체에서는 대상 응용 분야의 특성에 따른 메타데이터 모델들을 표준으로 제안하였다. 제안된 모델은 여러 차례의 의견 수렴을 거쳐 계속 확장 보완되어 표준으로 정해지는 것이 보통이며, 이러한 과정에서 다른 모델과의 데이터 교환 및 상호운용성도 고려하게 된다. 본 장에서는 4장에서 제안하는 모델의 참조 모델로 활용된 표준 모델인 더블린 코어, GILS코어, CSDGM 모델의 내용 및 특성을 분석한다.

3.1 더블린 코어 (Dublin Core)

95년 3월 미국 오하이오주의 더블린에서는 네트워크로 연결된 자원을 표현하기 위한 메타데이터의 핵심 요소들을 도출하고, 메타데이터에 대한 요구조건, 접근방법 등을 결정할 목적으로 OCLC/NCSA(Online Computer Library Center/National Center for Supercomputing Applications) 메타데이터 워크샵이 개최되었다(1회 더블린 코어 워크샵)[9]. 참가자들은 인터넷과 같은 네트워크 환경에서 전자문서를 처리하는데 필요한 최소한의 메타데이터를 13개의 요소로 제안하고 이를 더블린 코어(Dublin Metadata Core

Element, 줄여서 Dublin Core)라 명명하였으며, 구문(Syntax)은 명세하지 않고 의미(Semantics)는 다양한 사용자가 이해할 수 있는 범위내에서 명확하게 정의하였다.

더블린 코어는 다음과 같은 특징을 가지고 있다.

- 트랜잭션 정보나 접근에 대한 고려를 전혀 할 필요가 없다.
- 사용자의 부가적인 표현을 허용함으로써 확장성을 제공한다.
- 다양한 응용을 위해서 구문에 제한을 두지 않는다.
- 모든 요소는 선택적(optional)이며 반복적으로 나타날 수 있다.
- 서로 다른 분야에서도 응용할 수 있도록 선택적인 수식자(qualifier)를 두어 요소의 정의를 변경할 수 있다.

<표1>에서는 더블린 코어 요소에 대한 설명 및 이들 요소를 이용한 메타데이터의 예를 보여주고 있다.

'96년 4월에는 영국의 Warwick 대학에서 더블린 코어의 원칙과 언어의 의미에 대한 상호운용성을 향상 시키며, 풍부한 표현을 지원하기 위한 확장성과 서로 다른 모델과의 연결을 위한 매커니즘을 정의할 목적으로 OCLC/UKOLN(UK Office for Library and Information Networking)메타데이터 워크샵이 개최되었다(2회 더블린 코어 워크샵)[10]. 참가자들은 더블린 메타데이터 코어 요소를 기초로 하여 더욱 광범위한 타입이 필요하며 이를 위하여 더블린 코어를 견고한 구문을 이용한 SGML DTD로 표현하며, 인터넷상의 문서에서 저작자가 생성한 메타데이터에 대한 표현을 삽입할 수 있도록 HTML 태그와의 매핑을 제안하였다. 또한 서로 다른 타입의 메타데이터들과의 상호교환을 위하여 메타데이터 객체를 규합할 수 있는 컨테이너 구조(일명 프레임워크)가 필요하며, 이를 MIME(Multipurpose Internet Mail Extension), SGML(Standard Generalized Markup Language), CORBA(Common Object Request Broker Architecture)환경에서 구현하기 위한 표현이 필요하다는데 의견의 일치를 보았다.

96년 9월에는 CNI/OCLC(Coalition for Networked Information) 메타데이터 워크샵이 미국 더블린에서 개최되었는데, 제안된 더블린 코어와

Warwick 프레임워크를 구성하는 요소들을 정리, 확장함으로써, 네트워크로 연결된 이미지나 이미지 데이터 베이스에 대한 메타데이터의 요구사항들을 결정하였다 (제3회 더블린 코어 워크샵) [11].

97년 3월에는 OCLC, DSTC(Distributed Systems Technology Center), NLA(National Library of Australia)의 후원으로 호주의 캔버라에서 제4회 더블린 코어 워크샵이 개최되었는데, 더블린 코어를 확장하여 최종적으로 확정시키는 문제와 더블린 코어 요소들의 스키마와 서브 요소들의 표기 형태 등에 관한 구조적인 문제 및 Coverage, Relation, Management등의 요소에 관한 명확한 정의와 의미를 확정하는 문제들을 토의하였다[12].

96년 12월 현재 더블린 코어는 Description과 Rights Management의 두 요소가 첨가되어 모두 15개 요소로 구성되어 있다. 이처럼 더블린 코어는 워크샵을 통하여 여러 분야의 전문가의 의견을 수렴하여 보완, 확장되고 있으며, 앞으로도 일련의 워크샵을 통하

여 계속 발전되면서 표준으로 확정될 것으로 예상된다. 새로 추가된 Description과 Rights Management의 내용은 다음과 같다[13].

- Description : 문서나 이미지의 요약 정보를 포함한 자원의 내용에 관한 서술정보
- Rights Management : 저작권이나 사용 권한에 관한 내용

3.2 GILS코어 (GILS Core)

GILS(Government Information Locator Service)는 미국의 국가정보기반(NII : National Information Infrastructure)의 중요한 일부로서, 일반 사용자들이 원하는 정보를 직접, 간접으로 접근할 수 있도록 지원해 주는 공공 정보 서비스라고 정의할 수 있다[14]. GILS는 원하는 정보를 찾는데 필수적인 카탈로그 카드에 비유되는데 도서관에서 이용자들은 원하는 책이 어디에 있는지를 카탈로그 카드를 통하여 알 수 있는 것처럼, GILS는 연방정부를 통하여

〈표 1〉 더블린 코어 요소 및 생성된 메타데이터의 예
 〈Table 1〉 Dublin Core Elements and metadata

Element	Description
Subject	The topic addressed by the work (예)Subject = Metadata, Subject (scheme = LCSH) = computer science
Title	The name of the object (예)Title = A Construction of Meta- Modeling for electronic documents
Author	The person(s) primarily responsible for the intellectual content of the object (예)Author = Hyo-Taeg Jung
Publisher	The agent or agency responsible for making the object available (예)Publisher = SERI
Other Agent	The person(s), such as editors, transcribers, and illustrators who have made other significant intellectual contributions to the work (예)OtherAgent(scheme=Editor) = Yoon-Chul Choy
Date	The date of publication (예)Date = June 15, 1997
Object type	The genre of the object, such as novel, poem or dictionary (예)ObjectType = Thesis
Form	The physical manifestation of the object, such as PostScript file or Windows executable file (예) Form = MS Word 7.0 file
Identifier	String or numbered used to uniquely identify the object (예)Identifier(scheme=URL) = http://openlab.seri.re.kr/htjung/thesis.html
Relation	Relationship to other objects (예)Relation (ContainedIn) = URL:http://www.seri.re.kr
Source	Objects, either printer or electronic, from which this object is derived, if applicable (예) Source(scheme=ISBN) =0-201-63337-X
Language	Language of the intellectual content (예) Language = Korean
Coverage	The spatial locations and/or temporal durations characteristics of the object (예)Coverage(type=temporal, scheme=yyymmdd) = 19970615

공공의 정보 자원들을 명세하고 이러한 자원들 중에서 이용이 가능한 정보들을 표현함으로써 일반인들이 손쉽게 원하는 정보를 찾을 수 있도록 도와주는 실질적인 서비스를 제공한다[15]. 따라서 정보를 찾거나 검색하기를 원하는 사용자는 인터넷을 통하여 수십개의 연방정부의 GILS 사이트를 탐색함으로써 원하는 정보를 직접 찾을 수 있다.

사용자들이 연방정부를 통하여 정보를 찾는데 도움을 주는 것은 Locator이다. Locator는 실제적인 정보 자원이 아니라 자원에 대한 표현 정보로서 이용 가능한 정보의 종류, 정보가 위치하고 있는 장소, 정보에 접근하는 방법 등에 관한 내용으로 구성되어 있으며, 해당 정보에 대한 메타데이터를 의미한다. 이러한 Locator는 참조 문헌 등의 형태로 오래 전부터 존재하여 왔지만 모든 연방정부의 정보에 대한 서비스를 제공하지는 못하였다. 그러나 GILS는 정보의 탐색과 검색을 위하여 국제 표준의 네트워크 기술과 명세된 메타데이터 요소를 사용함으로써 서로 다른 Locator간의 상호 운용성이 가능해졌고, 이로 인하여 통합적인 연방정부 기반의 서비스를 제공하게 되었다. 이것은 연방정부가 관리하는 Locator 레코드의 모임을 의미하는 GILS 코어 요소들이 수준 높은 내용으로 구성되어 있기 때문이다. 즉, GILS가 유용하게 사용되기 위한 가장 중요한 요소로는 GILS 코어 레코드들이 정확하고 일관성이 있어야 하며 완전성 및 현재성을 지녀야 한다[16]. 또한 GILS코어 레코드들은 GILS코어 요소의 표준을 따르고 있으며 네트워크를 통한 상호 접근이 가능하다[17]. GILS 코어의 요소와 구조는 GILS 프로파일에 정의되어 있으며, 각 요소는 계층적 구조를 이루고 있다.

3.3 FGDC의 CSDGM

CSDGM(Content Standards for Digital Geospatial Metadata)은 미국의 연방지리정보 위원회(FGDC : Federal Geographic Data Committee)에서 디지털 공간 지리 데이터의 메타데이터 내용에 관해 제안한 표준안으로서, 지리정보 메타데이터와 관련된 용어 및 개념에 대한 정의를 제공하고 있다[18].

연방지리정보 위원회는 농무성, 상무성, 국방성 등 14개 정부 기관의 대표들로 구성되어 공간 데이터의 개발, 사용, 보급을 증진할 목적으로 설립된 위원회로서, 산하에는 부위원회와 실무 그룹들이 있다. 여기에서는 데이터의 내용, 질, 전송 등에 관한 표준안을 개발, 분

류하고 있으며, 또한 공간 데이터 구조의 분석, 분류 및 정보 교환, 데이터 전송 등에 관한 연구도 수행하고 있다. 이 위원회가 92년 6월에 개최한 포럼에서 참가자들은 공간 지리 데이터의 표준에 대한 필요성에 대한 공감대를 형성하였으며, 그 후 위원회에서 표준안을 작성하여 여러 차례의 공개적인 검토 및 테스트를 거쳐 현재 상태에 이르고 있다.

CSDGM의 전체 내용은 11개의 섹션과 용어해설, 참조문헌으로 이루어져 있으며, 각 섹션은 제목, 구문(Syntax), 의미(Semantics)로 나누어 정의되어 있다. 섹션0은 섹션들의 모임으로서, 서로 다른 개념들을 가지는 복합 요소들로 이루어져 있다. 각 섹션은 복합 요소 혹은 데이터 요소들로 이루어져 있으며, 각 요소들의 이름 뒤에는 계층 단계가 표시되어 있다.

CSDGM 표준에서 정의된 메타데이터들은 다음과 같은 특성을 가진다.

- 지리적인 위치를 나타내는 데이터 집합을 결정하는데 필요한 데이터이다. (유용성 : Availability)
- 데이터 집합이 특수한 필요성에 부합되는지를 결정하는데 필요한 데이터이다. (사용의 적합성 : Fitness for use)
- 명시된 데이터 집합을 얻는데 필요한 데이터이다. (접근성 : Access)
- 데이터 집합을 사용하고 처리하는데 필요한 데이터이다. (전이성 : Transfer)

CSDGM 표준안은 데이터 요소와 복합 요소들이 생성규칙(Production Rule)과 함께 여러 가지 조합으로 이루어져 있다

- 데이터 요소(Data Element) : 논리적인 최소 단위의 정보로서 데이터 이름, 정의, 할당된 값으로 구성되어 있다.

예) **Supplemental Information** 1.2.3

Other descriptive information

Type : text

Domain : free text

- 복합 요소 (Compound Element) : 데이터 요소와 다른 복합 요소로 이루어져 있으며, 모든 복합 요소는 데이터 요소 혹은 복합 요소가 될 수 있다.

예) **Citation** 1.1.

〈표 2〉 참조 모델들의 특징
 〈Table 2〉 Features of reference models

	더블린 코어	GILS 코어	FGDC의 CSDGM
적용대상	주로 텍스트 위주의 전자 문서	연방정부가 제공하는 정보의 자원	주로 디지털 공간 지리 데이터
구조	단일구조	계층적 구조, 요소는 세부 요소 포함 가능	계층적 구조, 요소는 세부요소 포함 가능
요소의 종류	목시적으로 선택요소, 반복가능 요소로 간주	선택요소, 반복가능, 제어가능 요소	단일/복합요소들이 생성규칙에 따름 (., =, +, , {}, (), >)
특징 (3.1~3.3 참조)	요소의 수가 작음 메타데이터 생성 용이	요소의 수가 많고 복잡함	요소의 수가 많고 복잡함

information to be used to reference the data set. *Compound*.

- 생성규칙 (Production Rule) : 복합 요소간 혹은 데이터 요소와 하위 단계의 복합 요소간의 관계를 표시하며, 기호 왼쪽은 서술자 (identifier), 오른쪽은 표현(expression)을 나타내며, 표현은 복합 요소 혹은 데이터 요소가 될 수 있다.

기호) = : is replaced by, produces, consist of
 + : and [|] : selection () : optional
 m{n} : iteration(repeated from "m" to "n" times)

3.4 참조 모델의 비교

참조 모델들의 특징을 정리하면 〈표2〉와 같다.

4. 전자문서를 위한 메타데이터 모델의 제안

최근 인터넷의 활발한 보급으로 사용자들은 네트워크를 통하여 문서뿐만 아니라 이미지, 비디오, 오디오 등과 같은 멀티미디어 데이터까지도 접근이 가능하게 되었다. 그러나 현재 인터넷 정보의 대부분은 문서로 되어 있으며, 멀티미디어 데이터들도 문서 정보에 하이퍼링크로 연결되어 간접적으로 접근되는 것이 일반적이다.

최종 사용자가 이러한 인터넷을 통하여 원하는 데이터를 정확하게 접근하기가 용이하지는 않다. 왜냐하면 인터넷에서 접근할 수 있는 데이터의 수가 방대할 뿐만 아니라, 선행 탐색 엔진을 사용하여 선택된 데이터들도

데이터 자체의 내용을 확인하지 않고서는 원하는 데이터인지를 정확하게 판별할 수가 없기 때문이다. 만약 데이터 자체의 내용이나 관련된 정보를 정확하게 서술해주는 메타데이터가 구축되어 있다면 원하는 데이터에 대한 접근이 훨씬 용이할 것이다.

본 장에서는 인터넷상의 과학기술 분야의 문서나 그래픽 이미지에 관한 메타데이터 모델인 SeriCore(註) 모델을 제안한다. 먼저 SeriCore모델의 특징과 기존의 표준 모델과의 차이점을 분석하고 구조 및 요소(element)와 속성(attribute)의 내용을 상세히 서술한 후 이를 SGML을 이용하여 표현한다. (註:본 논문을 작성한 주관기관인 SERI(Systems Engineering Research Institute)에서 제안한 전자문서를 위한 핵심적인 메타데이터 모델을 뜻함).

4.1 SeriCore모델의 특징

제안한 SeriCore모델은 인터넷상의 다양한 분야의 정보중에서 과학기술 분야와 관련된 각종 논문, 보고서, 기술문서와 같은 문서 및 그래픽 이미지를 위한 메타데이터 모델로서, 요소와 속성의 명칭, SGML태그명, 문법등이 정의되어 있다.

SeriCore 모델을 근거로 구현된 응용 프로그램을 이용하여 메타데이터를 생성, 관리하는 주체는 전문적인 정보 제공자나 정보 저작자는 물론 일반 사용자도 될 수 있다. 이를 고려하여 SeriCore 모델에서는 요소들이 간단한 계층적 구조를 이루고 있으며 요소의 수 및 깊이를 가능한한 제한하였다. 또한 더블린 코어의 13개 요소의 내용을 모두 포함하고 GILS코어 요소 중 관련된 요소의 일부를 포함함으로써 표준 모델의 요소

들을 적극 수용하였다.

제안한 SeriCore모델은 다음과 같은 특성들을 가지고 있다.

- 메타데이터 생성 대상은 과학기술 분야의 각종 논문, 보고서, 기술 문서, 자료집 및 그래픽 이미지 등이다. 이들은 HTML등과 같이 하이퍼링크가 가능한 문서는 물론 Postscript문서, 워드 프로세스 문서, 일반 텍스트 등과 같이 하이퍼링크가 없는 문서 화일들로서 전자화되어 인터넷상에 존재하는 정보들이다. 또한 하이퍼텍스트 문서에 연결된 그래픽 이미지도 포함된다.
- 계층적으로 구조화가 되어 있다. 즉 메타데이터의 내용과 특성을 고려하여 메타정보, 원시정보, 디지털정보, 검색정보로 구분되어 있으며, 각각은 하부에 적당한 깊이를 가지면서 관련된 내용들을 포함하고 있다. 이에 반해 더블린 코어나 GILS코어의 요소는 내용상으로 분류가 되거나 구조화가 되어 있지 않으며, 특히 더블린 코어의 요소는 계층적 구조를 가지고 있지 않다.
- 각 요소들은 필수요소, 선택요소로 구분되어 있으며, 또한 반복이 가능한 요소와 그렇지 않은 요소로 구분되어 있다. 즉, 반드시 생성되어야 할 요소와 생성자의 의도에 따라 선택적으로 생성될 수 있는 데이터로 구분되어 있다. 더블린 코어는 이러한 구분이 없으므로 모든 요소가 선택요소와 반복 가능한 요소로 간주되며, GILS코어는 선택요소, 필수요소, 반복 가능한 요소로 구분되어 있

다.

- 각 요소를 부가적으로 서술할 수 있는 속성을 정의하고, 속성값은 사용자가 입력하거나 혹은 미리 제시된 속성값 중에서 선택할 수 있다.
- 요소들의 수 및 깊이를 한정하면서도 정보 자체를 최대한 정확하게 서술할 수 있도록 구성되어 있다. 더블린 코어의 경우 일반적인 문서 정보를 대상으로 하며 요소의 수가 적고 간단하기 때문에 특정 응용 분야의 문서나 이미지를 정확하게 서술하기에는 부족하다.
- 문서의 표준인 SGML을 이용하여 표현하였으며, 메타데이터를 생성, 저장, 검색할 수 있는 편집기 및 관리시스템을 제공하고 있다. 그러므로 메타데이터 생성자는 요소가 의미하는 내용을 알고 있으면 응용 프로그램을 이용하여 용이하게 메타데이터를 관리할 수 있다.

<표 3>에서는 SeriCore모델과 더블린 코어의 특징을 비교, 정리하였다.

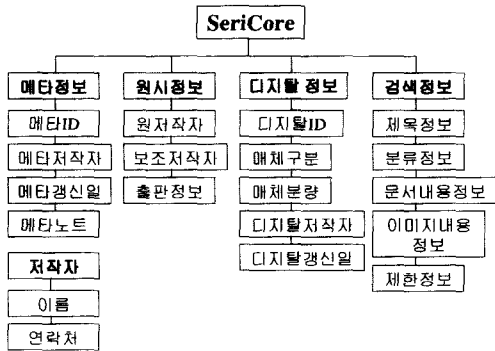
4.2 SeriCore모델의 구조 및 내용

메타데이터 모델은 네트워크를 통하여 최종 사용자가 원하는 정보에 대한 검색을 정확히 할 수 있으며, 원시 정보에 대한 메타데이터의 내용이 충분히 반영될 수 있도록 모델링 되어야 한다. 특히, 메타데이터의 작성이 정보 제공자뿐만 아니라 일반 사용자에게 의해서도 이루어질 수 있으므로, 메타데이터 모델은 전체 구조 및 세부 내용을 쉽게 파악할 수 있도록 구성되어야 하

<표 3> SeriCore와 더블린 코어의 특징
<Table 3> Features of SeriCore and Dublin Core

구분	SeriCore	더블린 코어
적용분야	과학기술분야의 문서 및 그래픽 이미지	텍스트를 위주로 한 전자문서
요소의 구조	계층적으로 구조화되어 있으며 요소는 세부요소를 가질 수 있음	계층적 구조가 아닌 단일구조임
요소의 종류	필수요소, 선택요소, 반복 가능한 요소로 구분됨	구분없이 묵시적으로 선택요소와 반복 가능한 요소로 간주됨
요소의 속성	속성이 미리 정의 되어있으며 속성값의 부여가 가능함	속성이 정의되어 있지 않음
요소의 수	적용분야를 고려하여 다양한 요소를 가짐	전자문서만을 고려하여 단순함
구현환경	SGML로 구현	MIME, SGML, CORBA 환경하에서 구현을 고려중임
기타특징	더블린 코어, GILS코어, FGDC 표준안의 장점을 이용	요소가 간단하므로 메타데이터의 생성이 용이함

며 이를 위해서는 계층적인 구조가 가장 적합하다. 제안한 SeriCore 모델은 요소들과 속성들로 구성되어 있으며 각 요소들은 계층적 구조를 이루고 있다. 메타데이터의 내용을 크게 메타정보, 원시정보, 디지털정보, 검색정보로 구분하였으며, 각각은 관련된 내용들로 이루어져 있다. (그림1)은 SeriCore의 개략적인 구조를 나타내고 있다



(그림 1) SeriCore 모델의 개략적인 구조
(Fig. 1) Structure of SeriCore Model

4.2.1 구성 인자

- 요소 (Element)
 요소는 하위에 다른 요소들로 구성된 복합요소와 별도의 하위를 두지 않고 메타데이터의 내용을 저장하는 단일요소로 구성되어 있다. 저작자와 ID를 제외한 모든 요소는 독립적인 명칭을 가지고 있는데, 저작자와 ID는 동일한 형식을 가지며 명칭도 중복되거나 상위 요소에 따라 결정된다. 요소는 필수요소, 선택요소, 반복 가능한 요소로 구분되며, 요소의 내용은 메타데이터 생성자가 입력하거나 혹은 미리 제시된 값을 선택함으로써 결정된다.
- 속성 (Attribute)
 요소를 부가적으로 서술하는 속성은 요소당 한 개 이상 정의될 수 있으며 명칭은 중복이 가능하다. 속성의 내용은 메타데이터 생성자가 입력하거나 혹은 미리 제시된 값을 선택함으로써 결정된다.

4.2.2 메타정보

원시정보를 대상으로 생성된 메타데이터 자체에 대한 정보이며 필수요소이다.

- 메타ID : 메타데이터에 대한 유일한 식별자이다.
- 메타저작자 : 메타데이터를 생성한 저작자에 관한 정보이다.
- 메타갱신일 : 메타데이터를 생성하거나 수정한 최종 날짜이다.
- 메타노트 : 메타데이터 작성에 대한 참고적인 정보이다.

4.2.3 원시정보

메타데이터 생성의 대상이 되는 정보로서, 출판물의 형태나 네트워크로 접근할 수 없는 형태의 정보를 말하며 선택요소이다.

- 원저작자 : 원시정보를 생성한 저작자에 대한 정보로서 2인 이상이 될 수 있다.
- 보조저작자 : 원시정보를 생성하는데 부분적으로 보조한 저작자에 대한 정보로서 편집자, 번역자, 삽화가, 사진 제작자 등이 될 수 있다.
- 출판정보 : 원시정보가 출판된 문헌일 경우 수록된 원전명, 출판사, 출판일, 학회등의 행사명, 권호, 수록된 페이지에 관한 정보이다.

4.2.4 디지털정보

메타데이터 생성의 대상이 되는 정보로서, 네트워크로 접근할 수 있는 문서나 이미지 정보를 말하며 필수요소이다.

- 디지털 ID : 디지털정보에 대한 유일한 식별자이다.
- 매체구분 : 문서와 이미지인지를 구분하며 각각의 포맷에 대한 정보이다.
- 매체분량 : 정보의 양을 나타내며 단위는 Kbyte 나 A4면수로 나타낸다.
- 디지털저작자 : 디지털정보를 생성한 저작자에 대한 정보이다.
- 디지털갱신일 : 디지털정보를 생성하거나 수정한 최종 날짜이다.

4.2.5 검색정보

메타데이터 생성의 대상이 되는 원시정보나 디지털정보의 내용을 서술한 정보로서, 메타데이터 검색시 가장 유용하게 사용될 수 있으며 필수요소이다.

- 제목정보 : 두가지 이상의 언어가 고려된 반복 가

당한 요소로서 제목과 부제로 구성된다.

- 분류정보 : 다루고 있는 주제 분야에 대한 정보로서 문서나 이미지의 성격에 따른 타입 구분과 주제 분야를 입력하거나 선택할 수 있는 요소로 구성되어 있다.
- 문서내용정보 : 문서의 내용에 관한 정보로서 요약, 키워드, 본문에서 사용된 언어, 본문의 목차, 표나 그림에 대한 서술과 식별자, 참조문헌에 대한 서술과 식별자, 정보 생성을 지원한 후원기관 등에 관한 정보이다.
- 이미지내용정보 : 이미지의 내용에 관한 정보로서 내용 서술 정보, 계절이나 배경과 같은 속성의 종류 및 속성값으로 된 부가정보, 이미지내의

의미있는 텍스트인 캡션내용, 이미지의 배경이 되는 공간정보, 시대배경 등의 정보이다.

- 제한정보 : 정보에 대한 저작권, 접근제한, 사용제한에 관한 정보이다.

4.2.6 공통요소

메타저작자, 원저작자, 보조저작자, 디지털저작자의 하위 요소인 저작자에 공통으로 쓰이는 요소이다.

- 저작자 : 이름과 연락처로 구성되어 있다.
- 연락처 : 저작자에 관한 정보로서 직책, 부서, 직장, 한국식 주소, 외국식 주소, 우편번호, 출신국가, E-mail이나 홈페이지와 같은 네트워크 주소, 근무시간대, 전화번호나 팩스번호로 구성되어 있다.

<표 4>와 <표 5>에서는 SeriCore를 구성하고 있는

<표 4> SeriCore 요소의 설명
<Table 4> Description of SeriCore elements

요소명	태그명	설명 및 속성(@)	비고
메타정보	metaInfo	메타데이터에 관한 정보	(M)
• 메타ID	MetaId	@scheme(URN/URL/LCCN/ISBN/ISSN/SICI/MessageID /FPI/Others), @systemID	(M)
• 메타저작자	MetaAuthor	@type(Originator/DigiAuthor/WebBuilder/IP/Others)	(M)
- 저작자	AnAuthor	이름과 연락처로 구성	
• 메타갱신일	MetaDate	메타데이터를 생성하거나 수정한 최종 날짜	(O)
• 메타노트	MetaNote	참고적인 정보	(O)
원시정보	OrigInfo	네트워크로 접근할 수 없는 메타데이터 대상 정보	(O)
• 원저작자	Originator	@type(MainAuthor/CoAuthor/Others)	(M)(R)
- 저작자	AnAuthor	이름과 연락처로 구성	
• 보조저작자	OtherOri	@type(Editor/Translator/Illustrator/Photographer/others)	(O)(R)
- 저작자	AnAuthor	이름과 연락처로 구성	
• 출판정보	PubInfo	원시정보가 문헌형태일 경우	(O)
- 원전명	NamePub	원시정보가 게재된 도서명/가행물명, @ISBN Code	(M)
디지털정보	DigiInfo	네트워크로 접근할 수 있는 메타데이터 대상 정보	(M)
• 디지털ID	ResoId	@scheme(URN/URL/LCCN/ISBN/ISSN/SICI/MessageID /FPI/Others), @systemID	(M)
• 매체구분	Medium	@mediType(Document/Image/OtherType), @ediForm(html/ps/doc/ppt/txt/hwp/bmp/pcx/gif/jpg/Other)@	(M)
• 매체분량	Size	Kbytes, @PagesA4	(O)
• 디지털저작자	DigiAuthor	@type(Originator/MetaAuthor/WebBuilder/IP/Others)	(O)(R)
- 저작자	AnAuthor	이름과 연락처로 구성	
• 디지털갱신일	ModiDate	디지털정보를 생성하거나 수정한 최종 날짜	(O)

검색정보	RetrInfo	메타데이터 대상 정보의 내용으로 검색정보	
· 제목정보	TitleInfo	제목정보	(m)
- 제목들	Title	@language(Korean/English/Japanese/Others)	(m,r)
* 제목	MainTitle	제목	(m)
* 부제	SubTitle	부제	(o)
· 분류정보	ClassInfo	주제 분류에 관한 정보	
- 타입구분	Category	@docCategory(TechnicalReport/ResearchReport/Paper/Manual/Abstract/OtherDoc), @imgCategory(Graphic/Photo/Drawing/OtherImg)	(m) (m)
- 주제분야	SubjectTerm	대상정보가 속한 분야	
* 입력분야	SubUnconTerm	분야를 메타데이터 생성자가 입력	(m)
* 선택분야	m	분야를 제시된 리스트로부터 선택	(o,r)
· 문서내용정보	subConTerm	대상정보가 문서일 경우	(o,r)
- 요약	docuInfo	내용 요약, @language(Korean/English/Japanese/Others)	(o)
- 키워드	abstract	중심되는 단어	(m,r)
- 본문언어	keyword	본문에서 사용되는 언어	(m,r)
- 본문목차	langMainText	문서의 차례	(o)
- 표그림	content	그림이나 표가 있을 경우	(o,r)
* 서술	figure	내용 설명	(o,r)
* ID	figureDesc	@scheme(URN/URL/LCCN/ISBN/ISSN/SICI/MessageID/FPI/Others), @systemID	(m) (o)
- 참조	figureId	참조 문헌이 있을 경우	(o,r)
* 참조자료	biblio	참조 문헌	(m)
* ID	biblioDesc	@scheme(URN/URL/LCCN/ISBN/ISSN/SICI/MessageID/FPI/Others), @systemID	(o)
- 후원기관	biblioid	문서 작성을 지원한 기관	(o)
· 미지내용정보	fundOrg	대상정보가 이미지일 경우	(o)
- 이미지서술	imageInfo	이미지의 내용	(m)
- 부가정보	resoDesc	부가정보	(o,r)
* 속성	addAttrVal	계절 혹은 배경등의 속성	(m)
* 속성값	addAttr	속성에 대한 값	(m)
- 캡션내용	addVal	이미지내의 의미 있는 텍스트	(o,r)
- 공간정보	caption	이미지 내용의 배경 정보	(o)
* 장소	spatial	지역/지역명	(o,r)
* 좌표	place	@WBC, @EBC, @NBC, @SBC	(o)
- 시대배경	coordinate	시대/시기	(o)
· 제한정보	timePeri	대상정보의 제한정보	(o)
- 저작권	constInfo	저작권에 관한 내용	(o)
- 접근제한	copyright	접근에 관한 내용	(o)
- 사용제한	accessConst	사용에 관한 내용	(o)
	useConst		

〈표 5〉 SeriCore 공통 요소의 설명
 〈Table 5〉 Description of SeriCore common elements

요소명	태그명	설명 및 속성(@)	비고
저작자	AnAuthor	저작자에 관한 정보	
· 이름	Name	저작자의 이름	(m)
· 연락처	ContactPnt	연락처에 관한 정보	(o)
연락처	ContactPnt	연락처에 관한 정보	

· 직책	Position	직위/직책/직무	①
· 부서	Department	소속 부서명	②
· 직장	OrgName	직장명	③
· 한국식주소	AddKorean	한국식 주소	④
· 외국식주소	addForeign	미국식 주소	⑤
· 우편번호	postCode	우편번호	⑥
· 국가	country	저작자의 국명	⑦
· N/W주소	networkAdd	N/W주소	⑧
- E-Mail	email	E-Mail주소	⑨
- 홈페이지	homePage	홈페이지의 URL	⑩
· 근무시간	hourServ	근무시간대	⑪
· 전화	telephone	전화번호	⑫
· 팩스	fax	팩스번호	⑬

각 요소 및 속성들을 나타내고 있다. 특히 <표5>에서는 SeriCore에서 공통으로 사용되는 저작자와 연락처에 관한 내용이다. 속성은 @로 표시하였으며, 비교란의 ③은 필수요소를, ④는 선택요소를 ⑦은 반복 가능한 요소를 나타낸다.

4.3 SGML을 이용한 SeriCore DTD

서로 다른 이기종 시스템간에 정보의 손실없이 문서의 전송, 저장, 처리를 위해 제정된 ISO 8879인 SGML(Standard Generalized Mark-Up Language)은 복합 문서에 대해 개괄적 마크업 언어의 구분만을 정의한 메타언어이다. SGML문서는 SGML 선언부(Declaration), 문서 타입 정의부(Document Type Definition), SGML 실제 문서부 등 세가지 요소로 구성된다.

SGML 선언부는 SGML문서를 구성하는데 필요한

문자 집합과 코딩 규약 등을 정의함으로써, SGML문서를 다른 시스템에 전송할 때 해당 문서가 어떤 기준으로 작성 되었는가에 대한 정보를 포함하고 있다. 문서 타입 정의부는 문서의 논리구조와 이와 관련된 속성들을 선언하는 부분으로, 같은 형태를 갖는 모든 문서에 적용되는 마크업 선언들의 집합으로 구성된다. 문서 타입 선언, 요소 선언, 엔티티 선언, 속성 선언, 표기법 선언, 주석 선언, 단축 참조 선언 등으로 구분된다. SGML문서부는 문서의 실제 내용을 정의하는 부분으로서, 문서 내의 상호 연결된 요소들로 구성된다. DTD에서 선언된 각 요소들의 구성을 나타내는 태그를 이용하여 문서를 작성하게 된다.

제한한 메타데이터 모델은 SGML 구문을 이용하여 SeriCore DTD로 구현하였으며, 그 내용은 <표6>과 같다.

<표 6> SGML을 이용한 SeriCore DTD
<Table 6> SeriCore DTD using SGML

```

<!-- SeriCore DTD for Metadata Description of Technical Documents and Images -->
<!ENTITY %id.scheme "URN | URL | LCCN | ISBN | ISSN | SIC | MessageID | FPI | Others" >
<!ENTITY %id.d "metaId | resold | figureId | biblioid" >
<!ENTITY %a.language "language (Korean | English | Japanese | Others) #IMPLIED" >
<!ENTITY % a.docCategory "TechnicalReport | ResearchReport | Paper | Manual | Abstract | OtherDoc" >
<!ENTITY % a.imgCategory "Graphic | Photo | Drawing | OtherImg" >
<!ENTITY % a.mediType "Document | Image | OtherType" >
<!ENTITY % a.mediForm "html | ps | doc | ppt | txt | hwp | bmp | pcx | gif | jpg | OtherForm" >
<!ENTITY % a.metaAuthor "type (Originator | DigiAuthor | WebBuilder | IP | Others) #IMPLIED" >

<!ENTITY % a.originator "type (MainAuthor | CoAuthor | Others) #IMPLIED" >
<!ENTITY % a.otherOri "type (Editor | Translator | Illustrator | Photographer | Others) #IMPLIED" >
<!ENTITY % a.digiAuthor "type (Originator | MetaAuthor | WebBuilder | IP | Others) #IMPLIED" >
    
```

```

<!ELEMENT (%id.d:) -- (#PCDATA) >
<!ATTLIST (%id.d:) scheme (%Id.Scheme:) #IMPLIED
                systemId CDATA #IMPLIED >

<!-- ***** Main Part ***** -->
<!ELEMENT SeriCore -- (metaInfo, origInfo?, digiInfo, retrInfo) >

<!-- ***** Part 1 : metaInfo (Metadata Information) ***** -->
<!ELEMENT metaInfo -- (metaId, metaAuthor, metaDate?, metaNote?) >
<!ELEMENT metaAuthor -- (anAuthor) >
<!ATTLIST metaAuthor %a.metaAuthor: >
<!ELEMENT metaDate -- (#PCDATA) >
<!ELEMENT metaNote -- (#PCDATA) >
<!-- ***** Part 2 : origInfo (Original Resource Information) ***** -->
<!ELEMENT origInfo -- (originator+, otherOri*, pubInfo?) >
<!ELEMENT originator -- (anAuthor) >
<!ATTLIST originator %a.originator: >
<!ELEMENT otherOri -- (anAuthor) >
<!ATTLIST otherOri %a.otherOri: >
<!-- Publication Information -->
<!ELEMENT pubInfo -- (namePub, publisher?, pubDate?, nameEvent?, volNumEd?, pages?) >
<!ATTLIST namePub isbnCode CDATA #IMPLIED >
<!ELEMENT (namePub | publisher | pubDate | nameEvent | volNumEd | pages) -- (#PCDATA) >

<!-- ***** Part 3 : digiInfo (Digital Resource Information) ***** -->
<!ELEMENT digiInfo -- (resoId, medium, size?, digiAuthor*, modiDate?) >
<!ELEMENT medium -- EMPTY >
<!ATTLIST medium mediType (%a.mediType:) #IMPLIED
                mediForm (%a.mediForm:) #IMPLIED >
<!ELEMENT size -- EMPTY >
<!ATTLIST size Kbytes NMTOKEN #IMPLIED
                pagesA4 NMTOKEN #IMPLIED >
<!ELEMENT digiAuthor -- (anAuthor) >
<!ATTLIST digiAuthor %a.digiAuthor: >
<!ELEMENT modiDate -- (#PCDATA) >
<!-- ***** Part 4 : retrInfo (Retrieval Information) ***** -->
<!ELEMENT retrInfo -- (titleInfo, classInfo, (docuInfo | imageInfo), constInfo?) >
<!-- 4.1 : Title Information -->
<!ELEMENT titleInfo -- (title)+ >
<!ELEMENT title -- (mainTitle, subTitle?) >
<!ATTLIST title %a.language: >
<!ELEMENT (mainTitle | subTitle) -- (#PCDATA) >
<!-- 4.2 : Classification Information -->
<!ELEMENT classInfo -- (category, subjectTerm) >
<!ELEMENT category -- EMPTY >
<!ATTLIST category docCategory (%a.docCategory:) #IMPLIED
                imgCategory (%a.imgCategory:) #IMPLIED >
<!ELEMENT subjectTerm -- (subUnconTerm+ | subConTerm+) >
<!ELEMENT (subUnconTerm | subConTerm) -- (#PCDATA) >
<!-- 4.3 : Document Content Information -->
<!ELEMENT docuInfo -- (abstract+, keyWord+, langMainText?, content*, figure*, biblio*, fundOrg?) >
<!ELEMENT abstract -- (#PCDATA) >
<!ATTLIST abstract %a.language: >
<!ELEMENT figure -- (figureDesc, figureId?) >
<!ELEMENT biblio -- (biblioDesc, biblioid?) >
<!ELEMENT (keyWord | langMainText | content | figureDesc | biblioDesc | fundOrg) -- (#PCDATA) >
<!-- 4.4 : Image Content Information -->
<!ELEMENT imageInfo -- (resoDesc, addAttrVal*, caption*, spatial?, timePeri?) >
<!ELEMENT addAttrVal -- (addAttr, addVal) >
<!ELEMENT spatial -- (place*, coordinate?) >
<!ELEMENT coordinate -- EMPTY >

```

```

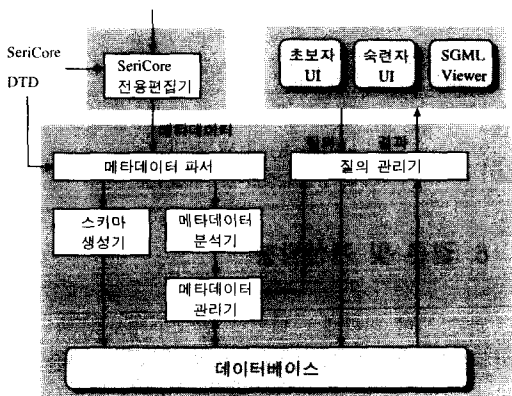
<!ATTLIST coordinate WBC NMTOKEN #IMPLIED
                  EBC NMTOKEN #IMPLIED
                  NBC NMTOKEN #IMPLIED
                  SBC NMTOKEN #IMPLIED >
<!ELEMENT (resoDesc | addAttr | addVal | caption | place) -- (#PCDATA) >
<!ELEMENT timePeri -- (#PCDATA) >
<!-- 4.5 : Constraints Information -->
<!ELEMENT constInfo -- (copyright?, accessConst?, useConst?) >
<!ELEMENT (copyright | accessConst | useConst) -- (#PCDATA) >

<!-- ***** Common Part ***** -->
<!-- Common Compound Element : An Author -->
<!ELEMENT anAuthor -- (name, contactPnt?) >
<!ELEMENT name -- (#PCDATA) >
<!-- Common Compound Element : Contact Point -->
<!ELEMENT contactPnt -- (position?, department?, orgName, addrKorean?, addrForeign?, postCode?,
                        country?, networkAdd?, hourServ?, telephone?, fax?) >
<!ELEMENT networkAdd -- (email?, homePage?) >
<!ELEMENT (position | department | orgName) -- (#PCDATA) >
<!ELEMENT (addrKorean | addrForeign | postCode | country) -- (#PCDATA) >
<!ELEMENT (email | homePage | hourServ | telephone | fax) -- (#PCDATA) >
    
```

5. SeriCore 메타데이터 관리시스템의 개발

5.1 시스템의 구성

SeriCore 메타데이터 관리시스템은 크게 메타데이터를 생성하는 부분, 생성된 메타데이터를 해석하고 스키마를 생성하여 UniSQL 데이터베이스로 저장하는 부분, 사용자 인터페이스를 통해 검색하는 부분으로 구성된다.



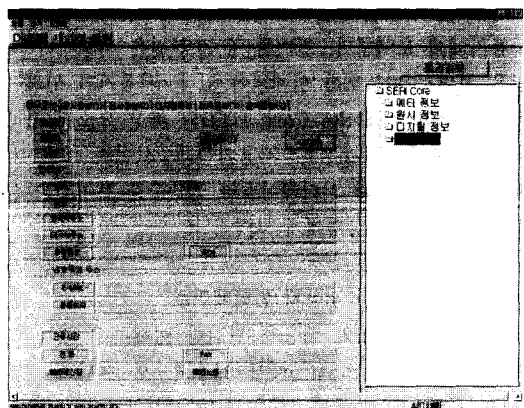
(그림 2) SeriCore 메타데이터 관리시스템의 구성
(Fig. 2) Structure of SeriCore metadata management system

5.2 SeriCore 메타데이터 전용 편집기의 구현

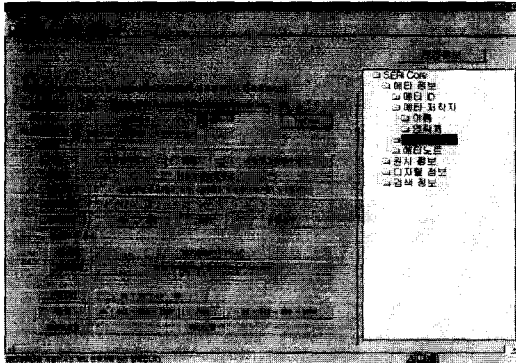
SeriCore 메타데이터 편집기는 Visual C++ 4.2

및 Visual Basic 4.0의 OCX를 이용하여 구현되었다. 편집기의 화면은 SeriCore의 요소를 계층적으로 볼 수 있는 트리정보 부분과 상위계층의 메타정보, 원시정보, 디지털 정보, 검색정보의 입력값을 볼 수 있는 부분으로 구성되어 있다. (그림3)은 트리정보의 SERI Core를 클릭한 경우의 화면이다.

트리정보에서 임의의 요소를 클릭하면 그 요소의 상위계층 요소를 볼 수 있으며 해당되는 메타데이터 입력을 위한 대화상자가 출력한다. (그림4)는 실제로 메타데이터를 입력하고 난 후의 화면을 보여주고 있다.



(그림 3) SeriCore 편집기 초기화면
(Fig. 3) SeriCore metadata editor



(그림 4) 입력된 메타데이터의 예
(Fig. 4) Input metadata

입력된 메타데이터는 저장시에 자동적으로 SGML 태그가 생성되어 SGML DTD형식으로 저장된다. <표 7>은 저장된 SGML 인스턴스의 일부를 보여주고 있다.

<표 7> SGML 인스턴스의 예
<Table 7> SGML instance

5.3 데이터베이스의 생성 및 질의

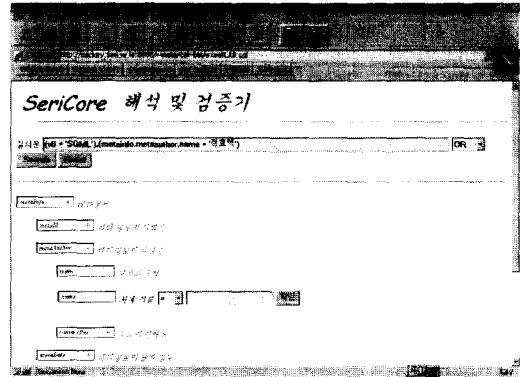
생성된 SGML인스턴스를 이용하여 객체 지향 데이터베이스 관리시스템인 UniSQL을 이용하여 데이터베이스의 스키마를 생성하고 메타데이터를 파싱한다. 파싱된 결과로부터 메타데이터를 분석하고 DTD에서의 각각의 요소를 객체로 생성하여 데이터베이스에 적재함으로써 메타데이터에 대한 데이터베이스가 생성된다.

질의는 사용자가 입력한 값을 이용하여 UniSQL에 적합한 질의문을 만들고, 이를 실행하여 질의 결과를 사용자에게 보여준다. 질의문 중 조건절은 사용자가 확인을 하면서 만들도록 설계되어 있다. 사용자가 원하는 정보를 찾기 위해 SeriCore DTD의 특정 요소에 검색어를 넣으면 from절은 사용자에게 보이지 않으며, where절은 사용자 인터페이스의 질의문을 보여주는 상태줄을 통하여 사용자에게 보여진다.

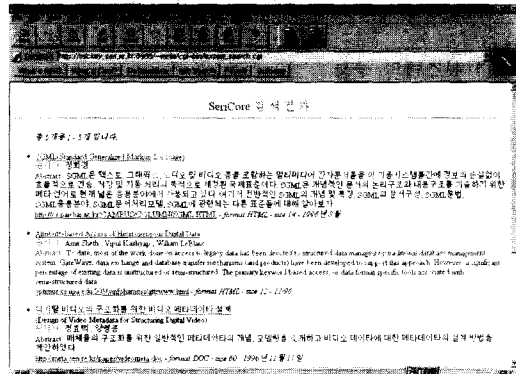
5.4 검색을 위한 사용자 인터페이스

SeriCore 요소의 내용을 완전히 이해하고 있는 숙련자들을 위해서 SeriCore를 계층적으로 제시하고 있으며, 하위 요소가 있는 경우 +를 표시하였다. 사용자가 원하는 요소에 질의값을 입력하면 그에 따른 질의문이 생성되어 화면에 표시되며 입력한 질의값들에 대한

AND나 OR검색을 사용자가 직접 선택할 수 있다. (그림5)에서는 사용자 인터페이스를 사용하여 질의를 하는 예를 보여주며, 검색 결과는 (그림6)과 같다.



(그림 5) 사용자의 질의 예
(Fig. 5) User query



(그림 6) 검색 결과
(Fig. 6) Result of retrieval

6. 결론 및 개선방향

최근 들어 인터넷은 다양한 분야의 정보를 멀티미디어 형태로 전달할 수 있다는 점에서 매우 각광을 받고 있다. 이에 따라 방대하고 다양한 인터넷 자원 중에서 사용자가 원하는 데이터를 정확하게 검색할 수 있고, 여러 사용자들이 데이터를 공유하면서 재활용할 수 있는 효율적인 방법이 차츰 필요하게 되었다.

기존의 탐색 엔진을 사용하여 데이터를 검색하는 경우, 원하는 데이터에 정확하게 접근하기가 쉽지 않다.

왜냐하면 탐색 엔진에 의해 선택되는 데이터의 양이 많을 뿐만 아니라 데이터의 내용이나 데이터 자체에 대한 정보를 정확하게 제공하지 않음으로써, 직접 데이터의 내용을 확인하지 않고는 원하는 데이터인지를 판별하기가 힘들기 때문이다.

그러나 탐색엔진과 연계하여 구축되어 있는 메타데이터를 검색할 경우에는 사용자가 원하는 데이터인지 아닌지를 쉽게 판단할 수 있는 정보들을 제시하여 줌으로써, 데이터의 내용을 직접 확인하지 않고도 정확하게 원하는 데이터에 접근할 수 있을 것이다. 또한, 여러 사람들이 공유할 가치가 있는 유익한 데이터에 대해서 메타데이터가 구축되어 운용된다면, 보다 많은 사람들이 손쉽게 데이터에 접근함으로써 데이터의 활용성이 증가하며 정보화 사회에도 기여할 것이다.

응용 분야에 따라 데이터가 제공하는 내용이 상이해듯이 메타데이터가 표현하는 내용도 모두 상이하기 때문에 응용 분야의 특성이나 사용 목적에 따라 특성을 충분히 반영해 줄 수 있는 메타데이터를 모델링하는 것이 무엇보다 중요하다. 현재 선진 외국에서는 메타데이터 모델에 대한 연구가 활발히 진행중이며, 이를 이용한 서비스도 일부 시행되고 있으나 국내에서는 이 분야의 연구가 거의 전무한 실정이다.

본 논문에서는 현재 표준화 작업이 진행중이거나 표준으로 채택되어 널리 사용되고 있는 몇 가지 모델을 분석하고, 이를 참조로 하여 인터넷상의 과학기술 문서나 이미지에 관련된 자료를 대상으로 한 메타데이터 모델인 SeriCore 모델을 제안하고, 이에 기반을 둔 메타데이터 관리시스템을 설계하고 구현하였다. SeriCore 모델은 주로 더블린 코어와 GILS 코어의 요소를 기반으로 응용 분야의 특성에 따른 요소를 추가하였으며, 문서에 대한 표준인 SGML의 신택스를 이용하여 표현함으로써 표준모델과의 호환성을 고려하여 모델링 되었다. 현재 메타데이터 편집기와 메타데이터 관리시스템이 분리되어 구현되어 있으며 상용 SGML 브라우저를 이용하여 메타데이터의 내용을 볼 수 있으나, 향후에는 모든 기능이 통합된 시스템으로 발전되어야 할 것이다.

SeriCore 모델은 인터넷상의 과학기술 분야에 적합한 모델이지만 다양한 응용 분야에 대한 메타데이터 모델링시에 참조 모델로 활용 가능할 것으로 생각한다. 향후에는 SeriCore 모델을 확장, 발전시켜야 함은 물론 이를 참조로 하여 다양한 응용 분야에 적합한 여러 메타데이터 모델들을 개발함으로써 인터넷 정보 자원에

대한 정확한 접근과 효율적인 재활용을 도모하여야 할 것이다.

참 고 문 헌

- [1] Shelley, E.P. and Johnson, B.D., Metadata : Concepts and Models, Proceedings of the 3rd National Conference on the Management of Geoscience Information and Data, pp.4.1-5, 1995.
<http://www.spirit.com.au/earthwear/Papers/AMF95/Shelley&Johnson.html>
- [2] R.M. Sumpter, Whitepaper on Data Management, Feb. 1994.
- [3] V. Kashyap, K. Shah, and A. Sheth, Multimedia for building the MultiMedia Patch Quilt, Multimedia Database System : Issue and V.S. Subrahmaniam Eds. Springer-Verlag, 1995.
- [4] K. Bohm and T.C. Rakow, Metadata for Multimedia Documents, ACM SIGMOD Record 23, 4 Dec. 1994.
- [5] Metadata and WWW Mapping Home Page
<http://www.blm.gov/gis/nsdi.html>
- [6] Creating and Contributing Metadata to WISCLINC.
<http://www.state.wi.us/agencies/wlib/sco/pages/creating.html>
- [7] J. Griffioen, R. Yavatkar and R. Adams, Automatic and Dynamic Identification of Metadata in Multimedia, First IEEE International Metadata Conference, April 16-18, 1996.
<http://www.nml.org/resources/misc/metadata/proceedings/adams/paper.html>
- [8] Intelligent Archive
http://www.llnl.gov/liv_comp/ia.html
- [9] OCLC/NCSA Metadata Workshop Report
http://www.oclc.org:5047/oclc/research/publications/weibel/metadata/dublin_core_report.html
- [10] L. Dempsey, S.L. Weibel, The Warwick

Metadata Workshop : A Framework for the Deployment of Resource Description, D-Lib Magazine, July/August, 1996.

<http://www.dlib.org/dlib/july96/07weibel.html>

- [11] CNI/OCLC Metadata Workshop, Workshop on Metadata for Networked Images

<http://www.oclc.org:5046/conferences/imagagedata/>

- [12] The 4th Dublin Core metadata Workshop <http://www.dstc.edu.au/DC4/>

- [13] Dublin Core metadata Element Set : Reference Description.

http://purl.org/metadata/deublin_core_elements#title

- [14] E.J.Christian, GILS - What is it? Wheres it going, D-Lib Magazine, Dec 1996.

<http://www.dlib.org/dlib/december96/12christian.html>

- [15] What is GILS?

<http://www.dtic.mil/defense/locator/morein.html>

- [16] Guidelines for the Preparation of GILS Core Entries - Introduction

<http://www.dtic.dla.mil:80/gils/documents/naradoc/intro.html>

- [17] The Government Information Locator Service (GILS)

<http://info.er.usgs.gov/public/gils/gilsdoc.html>

- [18] Content Standards for digital geospatial metadata FGDC

<http://geochange.er.usgs.gov/pub/tools/metadata/standard/overview.html>



정 호 택

1986년 경북대학교 전자공학과 졸업 (학사)

1997년 연세대학교 컴퓨터공학과 졸업 (석사)

1986년~1987년 삼성전관 컴퓨터 사업부

1987년~현재 시스템공학연구소 개방형S/W연구실 선임연구원

관심분야 : 인터넷/인트라넷, 멀티미디어, 분산시스템



양 영 종

1979년 서울대학교 과학교육과 졸업 (학사)

1995년 영국 Essex University 전산학과 졸업 (석사)

1983년~현재 시스템공학연구소 개방형S/W연구실 선임연구원

관심분야 : 소프트웨어공학, 인트라넷, 객체지향, 멀티미디어



김 순 용

1987년 서울산업대학교 전자계산학과 졸업 (학사)

1996년 대전대학교 컴퓨터공학과 졸업 (석사)

1981년~현재 시스템공학연구소 개방형S/W연구실 선임연구원

관심분야 : 데이터베이스, 소프트웨어공학, 멀티미디어



이 상 덕

1980년 부산대학교 대학원 경영학과 졸업 (석사)

1978년~현재 시스템공학연구소 개방형S/W연구실 실장

관심분야 : 인트라넷, 데이터베이스, 개방형 시스템



최 윤 철

1773년 서울대학교 전자공학과
졸업(공학사)

1975년 6월 Univ. of Pittsburgh
(공학석사)

1979년 6월 Univ. of Californ-
ia, Berkeley, Dept. of
IE&OR(공학박사)

1979년 8월~1982년 7월 Lockheed 및 Rockwell
International사 책임연
구원

1982년 9월~1984년 1월 Univ. of Washington 전
산학과 박사과정

1990년 9월~1992년 1월 Univ. of Massachusetts
연구교수

1984년 3월~현재 연세대학교 컴퓨터과학과 교수

관심분야 : 멀티미디어, 하이퍼미디어, 지리정보시스템
(GIS)