

지식정보자원 표준화 모델 연구

A Study of the Standardization Model for the National Knowledge and Information Resource

이창열(ChangYeol Lee)* 정의석(EuiSuk Jeong)**

초 록

한국정보문화진흥원이 관리하는 국가 지식정보자원은 여러 기관에 분산되어 있으며, 메타데이터 규격은 통합이 아니라 검색을 위한 개념적 수준의 권고 표준이었다. 그래서 데이터를 연계하거나 통합하는데 많은 문제가 발생하고 있다. 본 논문에서는 여러 기관에 분산된 지식정보자원에 대한 통합을 위하여 기존에 여러 기관에 분산된 메타데이터를 분석하여 문제점을 도출하고 이를 보완하며, 지속적으로 연계 및 통합할 수 있는 표준 모델을 제시하고자 한다.

ABSTRACT

National Knowledge and Information Resources of KADO(Korea Agency for Digital Opportunity and Promotion) were distributed to the several data centers. The metadata for the resources was the conceptual level recommended standard. It was not for the integration, but the retrieval. So it is not easy to integrate to the central metadata DB or connect metadata among the data centers. In this paper, we analysed the metadata of the several data centers and provided the integrated standard model for the central metadata DB.

키워드 : 메타데이터, KIX, ID, 표준, 통합 표준 모델
metadata, national Knowledge and Information eXchange(KIX), ID, standard,
the integrated standard model

* 동의대학교 컴퓨터공학과 교수 (lcy@deu.ac.kr)

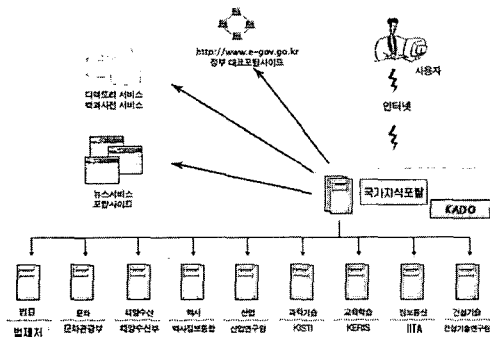
** 고려대학교 컴퓨터교육학과 박사과정/한국정보문화진흥원 (goodguy@kado.or.kr)

■ 논문접수일자 : 2006년 11월 17일

■ 게재확정일자 : 2006년 12월 15일

1. 서론

국가 지식정보자원은 <그림 1>처럼 국가 지식포털 시스템을 중심으로 5개 종합정보센터와 직접 연결된 4개의 전문정보센터를 통하여 760만개 이상의 정보가 서비스되고 있다(한국정보문화진흥원 2005).



(그림 1) 국가지식포털 연계 구조

이들 각 센터에서 서비스되는 데이터는 2004년까지 통합되지 않고 국지포(국가지식포털; <http://www/knowledge.go.kr>)의 검색을 통한 분산질의 형태로 정보에 대한 접근이 이루어지고 있었다.

이러한 지식정보자원에 대한 규격으로 한국정보사회진흥원(구 한국전산원)이 2001년 제시한 지식정보자원에 대한 규격(한국정보사회진흥원 2001)은 Dublin Core(DC) 기반의 가이드로 7개 요소(elements)를 제시하였으며, 2004년 3개 요소가 추가된 것(한국정보사회진흥원 2004)이다.

그러나 이들 규격은 각 센터에서 필수적으로 사용된다기보다 권장 표준이었고 교환용 데이

터로써 실제적으로 XML(eXtensible Markup Language) 형태의 반출이 지식정보자원으로 한국정보사회진흥원에서 요구된 적이 없었기 때문에, 구축된 데이터의 관리 상태나 품질에 대하여 실제적 검증이 이루어지지 않은 상태였다. 표준 또한 매우 느슨한 구조로 이루어져 있었다.

본 연구는 이러한 환경을 통하여 관리를 인수받은 한국정보문화진흥원(이하 KADO; Korea Agency for Digital Opportunity and Promotion)이 지식정보자원에 대한 메타데이터 통합 DB 구축을 위하여 종합정보센터의 데이터를 실제적으로 검사를 하고, 검사 결과를 바탕으로 이들 데이터에 대한 큰 변환 없이 데이터를 통합할 수 있는 모델 개발을 하는 것으로 진행되었다.

본 논문에서는 이러한 통합 모델을 도출하는데 근간이 되었던 종합정보센터의 데이터 분석 결과와 이를 통하여 궁극적으로 메타데이터를 통합을 할 수 있는 국가지식정보자원 메타데이터 표준(KIX; Knowledge and Information resource eXchange)을 도출하는데 고려되었던 사항에 대한 연구를 진행한 것이다.

2. 지식정보자원 분석

각 종합정보센터에 대한 DB 스키마 문서와 샘플 DB 데이터를 육안으로 확인하는 형태로 분석을 하였다. 분석의 관점은 이들 데이터를 앞으로 통합 관리하는데 발생할 수 있는 문제점을 도출하고, 이를 바탕으로 상호운용 가능한 통합 메타데이터 표준(KIX)을 정의하는데 있다.

○ 데이터 분석 대상

대부분의 기관이 XML 형태로 메타데이터 표준 구조를 가지고 있지만, 실제적으로 XML 형태로 변환하여 데이터를 제공한 적이 없어서 (기존에는 분산 검색만 작동되면 되었기 때문에) 자료 분석은 DB 자체를 대상으로 하였다. 분석 결과 XML 형태의 메타데이터 표준과 DB 스키마 사이 정확한 매핑이 이루어지지 않고 있었다.

- 대표적 사례로 반복 구조 문제가 있다. 9개 방문 기관의 메타데이터 표준에는 전부 반복구조가 정의되어 있지만, 그 중 2개 기관만 DB 스키마에 반복 구조가 가능하게 설계되었다.

그러므로 메타데이터 표준 분석은 의미가 없으며 DB 분석을 하여야 했다.

○ URL 표기

- 유지 문제 발생
 - * 콘텐츠를 보유하지 않는 종합정보센터에서 URL을 보유하고 있는 것은 전문정보센터에서 URL이 바뀌었을 때 해당 사항을 제때 파악 못하는 일관성 문제 야기된다.
 - * 또한 많은 전문정보센터가 DB 스키마 상으로 URL과 관련된 많은 필드를 가지고 있다. 키워드 추출용 파일 URL(주로 HWP), 화면 제공용 파일 URL(주로 HTML), 내부 관리용 파일 URL, 또한 URL을 웹 주소 URL, 내부 파일 폴더까지 URL 등 너무 복잡한 형태로 URL을 간직하고 있다.

○ 데이터 분량 : DB 필드는 선언되어 있지만 실제 데이터가 들어 있는 필드가 매우 적다. 지식정보자원 사업 자체가 구축한 레코드수로 평가하지, 넣어진 필드로 판단하지 않기 때문이다.

- 방향 : 데이터 품질을 보장할 수 있는 방안이 강구되어야 하며, 선택적인 메타데이터 요소를 필수로 변경하는 것이 가장 간단한 방법일 것이다.

○ 의미 해석 차이

특정 요소에 대한 잘못 이해, 기관 사이 다른 사용 등이 나타나고 있다. 예를 들어 “Type”과 “Format”을 혼합하여 쓰는 기관, “Created Date”를 디지털 콘텐츠가 아닌 원 자료(Offline)의 생성일로 쓰는 기관, 그리고 어떤 기관은 “Subject”를 분류로, 어떤 기관은 키워드로 해석하여 통합에 문제가 발생할 수 있다.

- 방향 : 데이터 해석에 대한 애매성을 없애기 위하여 요소에 대한 정의를 보다 엄격하게 한다.

○ 검색엔진 : 3개 기관에서 검색엔진을 DBMS 대신으로 사용하고 있다. 검색 엔진이 가지는 데이터 저장 기능을 이용하여 데이터를 관리하는 것이다.

○ 유니코드 사용 : 일부 기관은 유니코드(UTF-8)를 사용하고 있어서 KS 표준 코드(EUC-KR)를 사용하는 데이터와 상호 변환이 필요하다.

- 방향 : 본 연구에서는 유니코드를 사용할지, 아니면 KS 코드를 사용할지는 논의

로 한다. 단순히 2가지 코드를 상호 구별하고 변환할 수 있게, 데이터 교환시 해당 코드를 기술하게 해야 한다.

○ 모음 자료 : 학술지, 전집 등 자료 관리에 문제가 발생하고 있다. 대부분의 메타데이터 구조가 단행본 체계로 되어 있어서 계층적 구조를 가지는 자료에 대한 정보 기술에 정확성이 부족하게 되어 있다.

- 방향 : 계층 구조를 가지는 콘텐츠를 기술할 수 있는 요소를 제시하여야 한다. 학술지인 경우 3단계(종, 권호, 기사) 메타데이터 구조를 가져야 한다. 이 경우 일반적으로 실제 콘텐츠를 가지는 단계는 기사(article) 메타데이터이다.

예를 들어 다음과 같은 단계별 데이터 요소가 필요하다.

* 종 정보

Creator가 일반적으로 없으며, Publisher에 관한 정보를 세부적으로 넣을 필요가 있다.

* 권호 정보

권호 설명, 목차, 권호 번호, 전체 페이지, 필요시 부제목(권호별로 제목이 있을 경우) 등이 필요하다.

* 기사 정보

필요 없는 것 : 출판사 등을 넣으면 안 된다.

필요한 것 : 요약, 기사 번호, 저자 등을 넣어야 한다.

3. 지식정보자원 표준 모델 방향

각 기관의 지식정보자원을 분석한 결과 품질 향상을 위하여 다음과 같은 단순하면서 현실적 방안을 제시한다.

○ 메타데이터 편집기 사용

많은 기관이 데이터 구축 자체가 지식사업으로 년 1회씩 구축하기 때문에 편집기를 사용하지 않고, 사업자가 배치(batch) 처리를 하고 있어서 데이터 값이 정확치 않아도 입력되고 있다.

표준 메타데이터를 반영한 편집기를 사용하여 메타데이터 DB를 구축하면, 입력에 대한 오류와 필수 입력 체크 등 데이터에 대한 어느 정도 수준의 품질이 보장이 되며 또한 관리자가 필요시 데이터를 확인할 수 있는 수단을 제공하게 된다.

○ 스킴(Scheme) 사용

정보를 직접 요소에 넣지 않고, 정보를 유형과 값(type, value)으로 2원화 시키고, 유형은 다시 스킴을 사용하여 일반화시킴으로써 정보의 확장성과 호환성을 유지한다(이창열 2006).

○ 표준안과 스킴 사항의 분리

ONIX(ON-line Information eXchange)(EDItEUR, 2005)나 LOM(Learning Object Metadata) (한국교육학술정보원, 2001)처럼 XML 표준과 스킴 데이터를 분리하여 관리하여야 한다. 그것은 정보가 가지는 유형이 늘어나도 XML 표준에 영향을 미치지 않게 만드는 것이다. 스킴 데이

터와 메타데이터 표준 규격을 분리 운영하는 것은 최소한 데이터 유형이 확장되어도 표준을 업그레이드 하지 않게 하는 장점이 있다.

○ URL 분리

디지털 콘텐츠를 소유하지 않은 기관은 해당 콘텐츠의 URL도 간직하지 않게한다. URL은 자주 변할 수 있는데, 콘텐츠를 소유하지 않는 기관이 메타데이터에 URL 정보를 간직하고 있으면, 추후 콘텐츠를 가진 기관의 URL이 변화시, 해당 정보를 알 수 없는 상황이 발생할 수 있기 때문이다. 대신에 해당 콘텐츠에 표준화되고 유일하게 접근할 수 있는 URN(Uniform Resource Names) 기반의 ID(K. Sollins, 1994)를 사용할 수 있게 한다.

품질향상을 위하여 제안되는 위의 방안을 반영하고, 추가적으로 기존 데이터에서 분석한 문제점은 다음과 같은 것이 있다 :

- 동일 요소의 다른 해석에서 발생하는 문제
- URL과 같은 데이터 관리 문제
- 단행본 구조에 저널이나 전집 등 자료를 표현하면서 발생하는 문제
- 온/오프라인 자료의 명확한 정의 부족에서 발생하는 문제

그러므로 위와 같은 문제가 발생하지 않게 하기 위하여 엄격한 의미적, 구문적 규제를 가진 KIX 설계 방안을 4장에서 제시하고자 한다.

4. 지식정보자원 표준 개발

4.1 기존 체계와 비교

항목	기준안	KIX
방향	각 기관 자체 표준	기관 통합 표준
목표	자체적으로 DC 기반으로 정의하여 사용	기관 사이 의미적/구조적 통합에 중점
선택과 필수	기관의 자율에 맡김	필수와 선택의 엄밀한 정의 제시
수준	DC 기반 Top Level 교환 표준 제시	실제자료 분석을 통하여 상세 규격 제시
스킴 관리	언급 없음	KADO에서 통합 관리

〈표 1〉 지식정보자원 표준안 차이

2001년도에 제시한 지식정보자원 표준과 본 연구에서 개발한 지식정보자원 표준(KIX)에 대한 원칙적 차이가 〈표 1〉에 기술되었다.

4.2 스킴 사용

기존 표준안은 정보를 직접 기술하였지만, KIX는 〈표 2〉와 같이 정보가 가지는 유형과 값을 분리하였다.

KIX는 〈표 2〉와 같이 표현하기 위하여 스킴을 위한 정보를 별도로 관리하여야 한다. 즉 〈표 3〉에서와 같이 <dateType>을 위한 코드표를 분리하여 관리한다.

〈표 3〉과 같이 내용과 구조를 분리 함으로써

기존안 사례	KIX 사례
<Date> <Created> 2005-10-25 </Created> </Date>	<Date> <dateType>01</dateType> <dateValue>2005-10-25 </dateValue> </Date>

<표 2> 데이터 비교

dateType을 위한 표	
01 Created	디지털 자원 만든 날짜
02 Valid	디지털 자원의 유효일 (To)
03 Available	디지털 자원의 유효일 (From)
04 Issued	디지털 자원 발행일
05 Modified	디지털 자원 변경일
10 Published	원 정보자원 출판일
12 Source Modified	원 정보자원 변경일

<표 3> 스킴 코드 테이블 예

스킴 정보가 갱신되어도 KIX 구조에 영향을 미치지 않게 구성한다.

4.3 온라인/오프라인 정보 구별

지식정보자원의 대부분은 기존의 오프라인 자원에 대하여 온라인 콘텐츠를 구축하였다. 그러므로 메타데이터 정보를 기술할 때 오프라인 정보 자원과 온라인 정보 자원에 대한 기술을 명확히 구별하고, 오프라인 정보에 대한 관리 및 접근 정보를 제공할 수 있어야 한다.

이를 위하여 오프라인 자료를 위한 요소로 "Source" 요소를 별도로 정의하여 사용하고 있으며, 다른 요소는 전부 온라인 콘텐츠에 대한

사항으로 기술하였다.

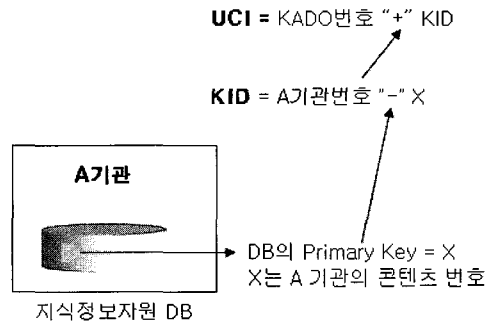
4.4 식별자 도입

4.4.1 KID 도입

KIX를 정의하면서 콘텐츠에 대한 URN(Uniform Resource Names) 기반 식별자로 KID(KIX Identifier)를 개발하였다.

KID는 기존 여러 센터에서 사용하는 ID를 그대로 보존할 수 있으면서 통합시 유일성을 보장하기 위하여 <그림 2>처럼 해당 기관 번호를 해당 기관 ID(해당 기관 DB의 Primary Key)에 추가하는 형태로 정의하였다.

KID의 구문 구조는 ABNF(Augmented Backus Naur Form) 형태(D, Crocker 1997)로 <그림 3>에 기술하였다.



<그림 2> KIX 식별체계 구조

KID	= InstituteCode Delimiter ContentID
InstituteCode	= 1 * alphaDigit
Delimiter	= "-"
alphaDigit	= ALPHA / DIGIT
ContentID	= alphaDigit * (alphaDigit / Others)
alphaDigit	= alphaDigit * (alphaDigit / Others)
Others	= "(" / ")" / "+" / "*" / "." / "=" / "@" / / ";" / "\$" / "_" / "!" / "%" / "'"

*InstituteCode : 기관코드
 *ContentID : 정보자원에 대한 해당 기관의 ID로 URL 표현을 사용하지 않고 해당 기관 DB의 Primary Key로 활용할 수 있는 ID를 의미한다.

〈그림 3〉 KID의 ABNF 표기

4.4.2. UCI로 확장

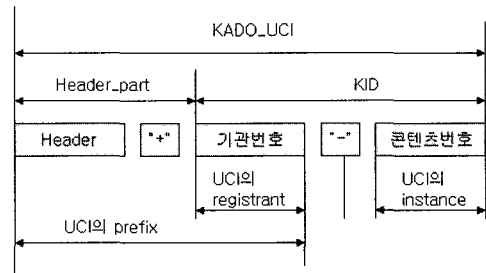
KID는 지식정보자원에 대한 유일한 식별체계이지만, 국제적 표준도 아니며 IANA (Internet Assigned Numbers Authority; www.iana.org)로부터 이름공간(Namespace)을 할당받은 것도 아니다. 반면에 UCI(Universal Content Identifier)(S. Kang 2005)는 IANA로부터 Namespace를 할당 받았으며, 국가 URN 식별체계 단일화에 의하여 현재 여러 기관이 보유하고 있는 콘텐츠 정보를 UCI 기반으로 연계하고 있다. KISTI(www.kisti.re.kr)인 경우 내부 식별자인 KOI (Knowledge Object Identification) (김재수 2006)는 내부적으로는 사용하고, 외부 연계를 위하여 UCI를 사용하는 형태로 ID 관리 체계를 가져가고 있다. KADO는 이러한 원칙에 따라 KID의 외부 연계 체계로 UCI를 도입하기로 하였다.

UCI 도입은 KID 개념을 보존하면서 연계의 편리성을 위하여 〈그림 4〉에 기술한 것처럼 개념적으로 KID 앞 부분에 KADO 번호를

붙이는 것으로 UCI를 정의하였다.

UCI 구분 구조, KID 구분 구조, 그리고 각 지식정보자원 관리 센터의 콘텐츠 번호를 명확히 구별하면서 연계하기 위하여 적용된 UCI 구조는 〈그림 2〉와 〈그림 4〉에서처럼 정의하였다.

〈그림 4〉처럼 UCI는 기존 KID에 “Header_part”가 추가된 형태이며, KID에서 사용하는 기관 코드는 UCI의 Prefix에 registrant와 일치하게 설계하였다. 예를 들어 살펴보자.



〈그림 4〉 KADO UCI 구조

- 가정
 콘텐츠 보유기관 : 한국교육학술정보원 (기관번호 KERIS)
 KADO 기관번호 : I123
 콘텐츠번호 : 1234
- ID 정보
 KERIS의 ID : 1234
 KID : KERIS-1234
 UCI : I123+KERIS-1234
- XML 표현
 <Identifier>
 <IDType>01</IDType> <!-- UCI -->

<IDValue>I123+KERIS-1234</IDValue>
</Identifier>

4.5 KIX 구조

4.5.1. 요소 선정

KIX는 DC와 비교하여 볼 때 다음과 같은 개념이 추가로 고려되었다 :

- 데이터 관리를 위한 정보
- 정기간행물, 전집 등 계층 구조를 가지는 자료 정보
- 온라인 자료와 오프라인 자료에 대한 명확한 기술과 접근 정보
- 데이터에 대한 엄격한 의미 정의

4.5.2. 표현 방법

KIX에 대한 상세 구조는 <표 4>와 같다. 일부 상위레벨 요소에 대하여 설명하면 다음과 같다 :

- Management : 데이터를 전송하는 기관과 보내는 자료 정보, 그리고 서비스 상태를 기록한다.
- Identifier : KADO가 UCI 사업을 하면서 필수적으로 정의하여야하는 "UCI" 값을 반드시 기술한다.
* 다른 곳에서 동일 구조를 재 사용하기 때문에 색칠하였음.
- Title : "mainTitle"이 한글 제목이 아닐 시, 반드시 "alternativeTitle" 내용을 한글로 기술하게 함.
- creator : 주저자
- Contributor : 도서의 부 저자, 삽화가, 편집자, 감독, 비평가 등에 대한 정보를

기술한다.

- Classification : 한국정보문화진흥원에서 정의한 분류 체계인 KC(Knowledge Classification) (고영만 2005)을 기반으로 정의한다.
- keywords : 국지포에서 키워드는 매우 중요한 요소를 차지하고 있으며, KADO의 요구사항으로 정의되었다.
- publisher : 출판사
- Date : 콘텐츠와 관련된 각종 날짜. 다른 곳에서 동일 구조를 재 사용하기 때문에 표시(색칠)하였음.
- Type : 종합정보센터에서 정의한 모든 유형을 지원하는 형태로 정의함
- Collection : 본 부분은 저널, 전집 등과 같은 유형을 기술하기 위하여 정의한 부분이다.
- Description : 목차, 요약, 서문, 비평 등에 관한 요소를 정의한다.
- audience : 독자 정보
- File : Thumbnail 파일, 원 파일, 관련된 파일 등을 기술한다.
- language : 언어
- Source : 오프라인 자료에 대한 정보를 기술한다. 오프라인 자료 보유 기관, 접근 방법, 보관 정보 등
- Coverage : 본 자료와 관련된 역사적, 공간적 정보를 기술한다.
- Relation : 본 자료와 관련된 다른 자료를 기술한다. 예를 들어, 참고 자료, 유사 자료 등
- Rights : 본 자료와 관련된 권리 사항을 기술한다.

〈표 4〉 KIX 요소 구조

상위레벨요소	중간레벨요소	하위레벨요소	반복
Management			1..1
	Notification		1..1
		notificationType	1..1
		notificationAgency	1..1
		notificationID	0..1
		notificationDate	1..1
		notificationDescription	0..1
	serviceAgency		1..1
	serviceStatus		1..1
	serviceDescription		1..1
Identifier			1..n
	IDType		1..1
	IDValue		1..1
Title			1..1
	mainTitle		1..1
	alternativeTitle		0..1
creator			1..1
Contributor			0..1
	contributorType		1..1
	contributorName		1..1
	contributorDescription		0..1
Classification			1..1
	classificationType		1..1
		classificationText	or 구조
		classificationCode	
keywords			1..1
publisher			1..1
Date			1..n
	dateType		1..1
	dateFormat		0..1
	dateValue		1..1
Type			1..n
	typeClass		1..1
	typeValue		1..1
Collection			0..1
	collectionType		1..1
	collectionID	ID 구조 사용	0..1
	titleOfCollection		1..1
	numberOfCollection		0..1
	sequenceWithinCollection		0..1

〈표 4〉 KIX 요소 구조

상위레벨요소	중간레벨요소	하위레벨요소	반복
Description			1..n
	descriptionType		1..1
	descriptor		0..1
	descriptionContent		1..1
audience			0..1
File			0..n
	fileType		0..1
	fileName		0..1
	fileFormat		1..1
language			0..1
Source			0..1
	sourceID	ID 구조 사용	0..1
	sourceAgency		1..1
	sourceCountry		0..1
	sourceDate	Data 구조 사용	0..n
	keepingDescription		0..1
	accessCondition		0..1
	contactInformation		0..1
Coverage			0..1
	coverageType		1..1
	coverageValue		1..1
	coverageDescription		0..1
Relation			0..n
	relatedType		1..1
	relatedID	ID 구조 사용	0..n
	relatedTitle		1..1
	relatedDescription		0..1
Rights			1..n
	rightsType		1..1
	rightsHolder		1..1
	rightsDescription		0..1
local	any		0..1

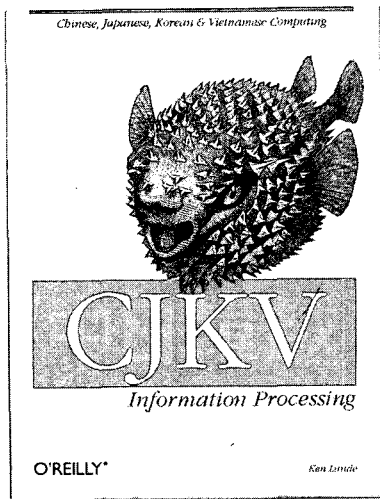
- local : KIX를 여러 기관에서 사용하는 데 확장 정의하여 사용할 수 있게 한다.

4.6 KIX 샘플

- 샘플 정보

* KERIS에서 작성한 신규 메타데이터

- 정보를 KADO에 통지하는 상황임.
- * O'Reilly 출판사에서 만든 CJKV Information Processing이라는 도서에 관한 것
 - * 일부 사항은 임의로 작성한 것임
 - * 샘플 콘텐츠 이미지



〈그림 5〉 샘플 콘텐츠 이미지

- 설명이 추가된 샘플 KIX 데이터

```

<?xml version = "1.0" encoding="euc-kr"?>
<kix
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xmlns="http://www.kado.or.kr/2006/KIX"
xsi:SchemaLocation="kix.xsd"

<Management>
<Notification> <!-- 초기 통지 -->
<notificationType>01</notificationType>
<notificationAgency>KERIS
</notificationAgency>
<notificationDate>2005-10-25

```

```

</notificationDate>
<notificationDescription>신규 파일로 통보합니다
</notificationDescription>
</Notification>
<serviceAgency>KERIS</serviceAgency>
<!-- 즉시 서비스 가능 -->
<serviceStatus>04</serviceStatus>
<serviceDescription>본 디지털 자료는 허가를 받은
사람이나, 유료로 서비스를 제공하고 있습니다
</serviceDescription>
</Management>

<Identifier>
<IDType>01</IDType> <!-- UCI -->
<IDValue>G001+KERIS.12345</IDValue>
</Identifier>
<Identifier>
<IDType>21</IDType> <!-- ISBN -->
<IDValue>1565922247</IDValue>
</Identifier>
<Identifier>
<IDType>32</IDType> <!-- URL -->
<IDValue>http://www.keris.or.kr/sample.hwp
</IDValue>
</Identifier>

<Title>
<mainTitle>CJKV Information
Processing</mainTitle>
<alternativeTitle>한.중.일 베트남 정보 처리
</alternativeTitle>
</Title>
<creator>Ken Lunde</creator>
<Contributor> <!-- 편집자 -->
<contributorType>B01</contributorType>
<contributorName>Peter Mui
</contributorName>
</Contributor>
<Classification> <!-- KADO 분류 -->
<classificationType>01</classificationType>

```

```

    <classificationCode>HZ</classificationCode>
  </Classification>

  <keywords>컴퓨터 정보처리 한글 유니코드</keywords>
  <publisher>O'Reilly & Associates, Inc.
</publisher>
  <Date> <!-- Created -->
    <dateType>01</dateType>
    <dateFormat>01</dateFormat>
    <dateValue>2005-10</dateValue>
  </Date>
  <Date> <!-- Issued -->
    <dateType>04</dateType>
    <dateValue>2005-10-25</dateValue>
  </Date> .
  <Type> <!-- UCI Type -->
    <typeClass>01</typeClass>
    <typeValue>Digital</typeValue>
  </Type>
  <Type> <!-- UCI Mode -->
    <typeClass>02</typeClass>
    <typeValue>Visual</typeValue>
  </Type>
  <Type> <!-- KERIS 유형 -->
    <typeClass>05</typeClass>
    <typeValue>M</typeValue>
  </Type>
  <Description> <!-- 간략 해설 -->
    <descriptionType>02</descriptionType>
    <descriptionContent>한국, 중국, 일본, 베트남에
서 사용하는 2바이트 기반의 Unicode를 처리하는데
필요한 기술을 서술하였다</descriptionContent>
  </Description>
  <Description> <!-- 비 구조화된 목차 -->
    <descriptionType>04</descriptionType>
    <descriptionContent>
      1. CJKV Information Processing Overview
      2. Writing System
      3. Character Set Standards
      4. Encoding Methods
    </descriptionContent>
  </Description>
  </descriptionContent>
</Description>
  <audience>08</audience> <!-- 전문가 -->
  <File> <!-- 메인 파일 -->
    <fileType>01</fileType>
    <fileName>sample.jpg</fileName>
    <fileFormat>application/pdf</fileFormat>
  </File>
  <language>eng</language>
  <Source>
    <sourceID>
      <IDType>21</IDType>
      <IDValue>1565922247</IDValue>
    </sourceID>
    <sourceAgency>KERIS</sourceAgency>
    <sourceCountry>KR</sourceCountry>
    <sourceDate>
      <dateType>04</dateType>
      <dateFormat>00</dateFormat>
      <dateValue>1999-01-01</dateValue>
    </sourceDate>
    <keepingDescription>습도 낮고, 음지에 보관
  </keepingDescription>
    <!-- 방문시 상시 열람 가능 -->
    <accessCondition>01</accessCondition>
    <contactInformation>연구개발부 홍길동 02-222-
3333</contactInformation>
  </Source>
  <Coverage> <!-- 공간적 사항 -->
    <coverageType>02</coverageType>
    <coverageValue>한국</coverageValue>
    <coverageDescription>한국내에서만 판매 가능
  </coverageDescription>
  </Coverage>
  <Relation> <!-- 버전 정보 -->
    <relatedType>05</relatedType>
    <relatedID>
      <IDType>21</IDType>
      <IDValue>8979142269</IDValue>
    </relatedID>
  </Relation>

```

```

<relatedTitle>정보처리</relatedTitle>
<relatedDescription>한글 정보 처리를 위한 가이
드라인</relatedDescription>
</Relation>
<Rights> <!-- 출판권 -->
<rightsType>02</rightsType>
<rightsHolder>KERIS</rightsHolder>
<rightsDescription>본 권리는 디지털 콘텐츠에 대
한 출판권으로 현재 KERIS는 자유롭게 제공하고 있
습니다 </rightsDescription>
<Rights>
<local>
  <test>본 파일은 샘플 파일입니다</test></local>
</kix>

```

5. 결 론

여러 센터에 분산된 지식정보자원을 통합(메타데이터만)하는 것은 쉬운 작업이 아니다. 센터들 사이에 제시된 기존의 표준은 검색을 위한 권고적 수준이기 때문에 통합을 위한 엄격한 스킴이나 의미에 대한 언급이 부족하였다.

본 연구에서는 기존에 분산된 센터들이 보유한 데이터를 확인 분석하여 다음과 같은 문제를 확인하였다 :

- 기관 사이에 동일 요소의 다른 해석에서 발생하는 문제,
- URL과 같은 데이터 관리 문제,
- 단행본 구조에 저널이나 전집 등 자료를 표현하면서 발생하는 문제,
- 온/오프 라인 자료의 명확한 정의 부족에서 발생하는 문제
- 그리고 데이터에 대한 품질에 문제

위와 같은 문제를 해결하기 KIX 설계에 다음과 같은 원칙을 적용하였다 :

- 스킴 사용
- 스킴 값과 표준 규격의 분리
- 엄격한 데이터 요소 의미 정의
- URL 사용 제한
- 저널, 전집 등 모음 콘텐츠 표현 지원
- 온/오프라인 자원에 대한 명확한 구분
- 데이터 전송 정보 기록

설계된 KIX는 한국정보문화진흥원의 통합 메타데이터의 표준으로 2006년부터 적용되고 있으며, 종합정보센터와 전문정보센터의 요구 사항을 지속적으로 수렴하여 업그레이드가 이루어져야 할 것이다.

참 고 문 헌

고영만. 2005. 메타데이터와 온톨로지. 『제3회 메타데이터 표준화 세미나』 : 3-14.

김재수, 권이남. 2006. NIST 식별메타데이터, 『제5회 메타데이터 표준화 세미나』 : 107-120.

- 손진곤. 2005. 고등교육정보및 저작권 관리를 위한 메타데이터(KEM v3.0) 연구. 『KERIS 연구보고서 KR 2005-27』: 71-82.
- 이창열. 2006. 『메타데이터 이해』. 국립중앙도서관 디지털정보자원관리과정 교육 훈련 교재 2006-전문 8 : 39-67.
- 정의석. 2006. 『지식정보표준 로드맵』. 한국정보문화진흥원.
- 한국정보문화진흥원. 2005. 『지식정보자원 통합 메타데이터 구축 지침서』. 버전 1.0, KADO 지침 05-01 : 전체.
- 한국정보문화진흥원. 2005. 『국가지식정보 메타데이터 구축.유통.관리 표준체계 연구』. KADO 연구 05-04 : 49-59.
- 한국정보사회진흥원(구 전산원). 2001. 『지식정보자원관리 메타데이터 표준안』.
- 한국정보사회진흥원(구 전산원). 2004. 『지식정보자원관리 메타데이터 표준안 (개정안)』.
- 한국정보사회진흥원. 2006. 『UCI 명세서 2.0』. 한국정보사회진흥원(구 한국전산원) : 1 - 20.
- D. Crocker. 1997. "Augmented BNF for Syntax Specifications." *ABNF. IETF RFC 2234*.
- EDITEUR. 2005. ONIX for Books : Product Information Message Overview and Data Elements Releases 2.1 revision 02.
- K. Sollins. 1994. "Functional Requirements for Uniform Resource Names." *IETF RFC 1737*.
- S. Kang. 2005. Using Universal Content Identifier (UCI) as Uniform Resource Names (URN) RFC4179.