

식별체계 기반의 과학기술분야 전자원문 연계시스템 설계 및 구현

이상환⁰ 신동구^{*} 김재수^{*} 최진영^{**} 정택영^{*}

*한국과학기술정보연구원 정보기술지원실

**고려대학교 컴퓨터학과

*(sanglee⁰, lovesin, jaesoo, tychung)⁰@kisti.re.kr

**choi@formal.korea.ac.kr

Design and Implementation of a Digital Contents Service System in Science and Technology Field Based on Identification System

Sanghwan Lee⁰ D.G. Shin J. S. Kim J.Y. Choi T.Y. Chung

*Dept. of S&T Information Sysgtem, Korea Institute of Science and Technology Information

**Dept. of Computer Science & Engineering, Korea University

요 약

정보통신 및 인터넷의 급속한 발전으로 디지털 콘텐츠의 유통이 활발해지고 있으나 디지털 콘텐츠에 대한 서비스 방식이 위치정보를 가지고 서비스하는 URL(Uniform Resource Locator)방식의 문제점과 기존의 물리적인 저작물이 디지털 콘텐츠로 전환되면서 기존의 식별자로는 디지털 콘텐츠의 특성을 충족시키는 식별이 미흡하고 한계가 있다. 따라서, 해외의 주요 디지털 콘텐츠 서비스기관의 식별체계 활용사례를 분석하고 그 결과를 토대로 한국과학기술정보연구원(KISTI) 고유 식별자인 KOI(Kisti Object Identifier)를 개발하고 개발된 KOI를 기반으로 과학기술분야 전자원문 연계시스템을 설계 및 구현하였다.

1. 서 론

정보통신 및 인터넷의 급속한 발전으로 디지털 콘텐츠의 전자 유통이 활발해지고 있으며, 기존의 물리적인 저작물들은 디지털 콘텐츠 형태로 제작되면서 콘텐츠 유통의 새로운 패러다임을 맞이하고 있다.

그러나, 디지털 콘텐츠에 대한 서비스 방식이 URL(Uniform Resource Location)방식으로 되어지기 때문에 사용자는 위치정보를 지속적으로 유지·관리해야하는 단점을 가지고 있다[6]. 이러한 URL의 문제점을 대처하기 위해 새로운 정보자원 식별 체계로 URN(Uniform Resource Name)이 개발되었다[4].

이러한 URN의 원리와 디지털 콘텐츠를 식별할 수 있는 식별 체계로서 1997년 DOI(Digital Object Identifier)가 개발되었다[2,3,6]. DOI는 URN의 명세[8]를 만족시키는 동시에 최대한 개방적인 구조를 채택하고 있으면서, 기존의 식별체계를 수용할 수 있기 때문에 지금까지 개발된 다양한 식별자들 가운데 가장 우수하다[5].

KISTI에서도 디지털 콘텐츠에 대한 서비스 방식을 URL방식으로 이루어지고 있기 때문에 디지털 콘텐츠 위치가 변할 경우 위치정보에 대한 수정이 바로 이루어지기 어렵고 사용자에게 잘못된 디지털 콘텐츠 위치정보를 제공하는 문제점을 가지고 있다.

본 논문에서 이러한 문제점을 개선하고 KISTI에서 보유하고 있는 과학기술분야 디지털 콘텐츠를 보다 효율적이고 체계적으로 서비스하기 위해 식별체계 기반의 서비스 시스템을 연구하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 디지털 콘텐츠 식별자인 DOI에 대하여 기술하고, 3장에서는 식별체계를 활용하고 있는 사례를 조사·분석한다. 4장에서는 KOI 구문구조와 KOI를 기반으로 한 시스템 구조에 대하여 기술하고 5장에서 결론과 향후 연구 방향에 대하여 기술한다.

2. 관련연구

사물을 식별하는 방법에는 여러 가지 있다. 예를 들면 사람에게는 태어날 때부터 부여받은 고유한 번호인 주민등록번호가 있다. 이처럼 디지털 콘텐츠에도 각각을 대표할 수 있는 영속적인 식별자를 부여하고, 이와 관련된 각종 메타데이터를 조직하고 절차에 따라 운영 관리하는 시스템을 디지털 콘텐츠 식별

체계라고 한다.

기존의 물리적인 저작물이 디지털화 되면서 기존의 식별체계로는 디지털 콘텐츠의 특성을 충족시키는 미흡하게 되었다. 디지털 콘텐츠의 특성상 각각의 오브젝트 즉, 각 장별이나 페이지별, 그림이나 표 등을 따로 식별할 필요성이 있다. 이러한 디지털 환경을 충족하기 위해 개발된 것이 DOI이다[7,9].

1997년 URN의 원리를 응용한 식별체계로서 DOI가 개발되었으며[2,3,6], ISBN과 같이 기존의 물리적인 저작물 형태에 부여되었던 식별기호를 디지털 저작물에 적용하는데 있어서의 한계를 극복하기 위한 것이다[6].

DOI는 구문구조(Syntax)가 가장 먼저 개발되었으며 2000년 5월에는 미국 국가표준인 ANSI/NISO Z39.84-2000으로 제정되었다[1]. DOI는 단순한 식별자로서 뿐만 아니라 모든 디지털 저작물의 이용, 관리 체계로서 응용되고 있다.

DOI 식별자의 구문구조는 접두부(Prefix)인 관리기관(Naming Authority)과 접미부(Suffix)인 로컬 이름(Local Name)으로 구성되어 있으며, 접두부와 접미부에 대한 구분은 "/"를 사용한다[7]. DOI의 구문구조는 다음 그림과 같다[1,6,7].

DOI = Naming Authority (Prefix) / Local Name (Suffix)

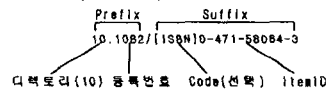


그림 1 DOI 구문구조 예

접두부는 [DOI 등록관리기관 번호],[DOI 등록자 번호]로 구성되며, DOI 등록관리기관이 DOI 등록자(기관)를 부여하게 된다. 접미부는 슬래쉬(/)뒤에 표시되며 디지털 콘텐츠의 등록자가 자체적으로 부여하는 지정코드이다. 통상 ISMN, ISWC, ISBN 등과 같은 각 콘텐츠별 국제표준코드를 활용하여 콘텐츠 등록자가 자유롭게 지정할 수 있다[7,9].

3. DOI/JOI 식별자 현황조사 및 분석

3.1 DOI 식별자 조사 및 분석

3.1.1 CrossRef 이용기관의 식별자 구문구조 조사 및 분석
CrossRef에 등록된 275개 기관 중 KISTI와 연관 있거나

DOI 등의 식별자로 디지털 콘텐츠를 서비스 중인 16개 기관을 선정하여 다음과 같이 조사하였다[10].

표 1 CrossRef 이용기관의 디지털 콘텐츠 식별자

조사기관	식별자	구분
Wilson Select Plus	Accession No : 88202089210	지체
ACM DL	10.1145/805282.505283	DOI
American Institute of Physics	10.1063/1.1527703	DOI
American Physical Society	PII : S 0273-0979(02)00943-6	PII
ASCE	10.1091/(ASCE)1098-1321(2003)18:2(46)	DOI
Blackwell Publishing - STM	10.1084/j.1369-6578.2003.470107.x	DOI
Emerald WCB	10.1108/07382780310450918	DOI
Journals of Physics Publishing	10.1088/0305-4470/28/11/303	DOI
Kluwer Academic Publishers	Article 10 : 515103	지체
Lippincott	00549843-200320030-0006	지체
Nature Publishing Group	10.1028/422098a/10.1028/PagInfo0147a	DOI
Science	cpj/content/full/2008/5808/592	지체
ScienceDirect	10.1016/S1003-0181(01)00128-9	DOI
Springer Verlag Link	10.1007/s004990100616	DOI
ACM	10.1145/805282.505283	DOI
AIP Conference Proceedings	10.1063/1.1434914	DOI

CrossRef 이용기관에 대한 디지털 콘텐츠 식별자 중 가장 많이 사용되는 DOI 식별자 중 접미부는 연간형 태인 Journal과 Proceeding은 ISSN, Year, Vol, Issue, StartPage등을 조합하여 DOI 식별자를 구성하였으나, DOI 식별자의 접미부에 ISBN과 ISRN은 거의 사용이 되고 있지 않은 것으로 조사되었다.

3.1.2 CrossRef에서 서비스 중인 DOI 관련 메타데이터

CrossRef에서 서비스 중인 Journal에 대한 DOI 식별자 메타데이터는 총 10개의 항목으로 이루어져 있다[10]. 이중 lookup 시 참조되는 필수 항목은 ISSN, Volume ID, Issue ID, Page ID이며 나머지 항목은 Journal Title, First Author, First Author 항목은 선택항목이다. 각 항목별 내용은 다음과 같다.

표 2 Journal DOI 메타데이터

구분	내	용
ISSN	Print and/or electronic ISSN delimited by a "."	
Title	The full journal title or an abbreviation	
Volume ID	Numerical or text (e.g. fall, 01) volume specifier	
Issue ID	Numerical or text issue identifier	
Page ID	Number or text page identifier (leading 0s are removed)	
Year	A number.	
Author	First author's sur name	
Type	full_text, abstract_only or bibliographic_record	
Unique key	A unique key that will be echoed back with the query results.	
DOI	The DOI (always left blank in the query)	

CrossRef에서 서비스 중인 Books/Conference Proceedings에 대한 DOI 식별자 메타데이터는 총 12개의 항목으로 이루어져 있다. Books와 Conference Proceedings를 동일 메타데이터에서 관리하고 있으나, Books에서만 사용하는 항목으로 Edition 항목을 사용하며 나머지 11개 항목은 같이 사용한다. 각 항목별 자세한 내용은 다음과 같다.

표 3 Books/Conference Proceeding DOI 메타데이터

구분	내	용
ISSN	Print and/or electronic ISSN delimited by a "."	
Serial Title	The full serial title or an abbreviation	
Volume Title	The full book title	
First Author	First author's sur name	
Volume ID	Numerical or text (e.g. fall, 01) volume specifier	
Edition	Numerical or text edition identifier (books only)	
Page	Number or text page identifier (leading 0s are removed)	
Year	A number.	
Component #	Chapter, section or part inside the book/conf. proceeding	
Type	full_text, abstract_only or bibliographic_record	
Unique key	A unique key that will be echoed back with the query results.	
DOI	The DOI (always left blank in the query)	

3.2 JOI 식별체계 조사 · 분석

3.2.1 JOI 구분구조

JST(J-Stage)에서 사용 중인 JOI 식별자의 구분구조는 DOI 식별자와의 호환이 가능한 형태로 구성되어 있으며, 접두부와 접미부는 첫 번째 "/"에서 구분된다[11].

접두부는 DOI에서 [DOI 등록관리기관 번호],[DOI 등록자 번

호]로 구성되어 있던 것을 JST.JSTAGE로 사용하고 있다.

접미부는 저널정보/VolumeID.StartPage로 구성되어 있다. 저널정보는 약기명으로 표기된 저널명을 사용하고 있으며, "/" 뒷부분은 VolumeID.StartPage로 구성되어 있다. 또한 CrossRef에 등록된 JST의 DOI 번호는 JOI 식별자 구분구조에서 두 번째 "/"를 "."로 변환하면 된다.

3.2.2 JOI 식별자 메타데이터

JOI 식별자는 JST 링크센터에서 관리되고 있다. JOI는 Journal만을 대상으로 식별하고 있는데 Journal에 대한 JOI 식별자 메타데이터는 총 15개 이상의 항목으로 이루어져 있다. 각 항목별 상세내용은 다음과 같다.

표 4 JOI 식별자 메타데이터

구분	내	용
블로그컨텐츠번호	JST 블로그컨텐츠 부여한 unique한 번호	
저널컨텐츠번호	JST가 Journal에 부여한 번호	
Article no	JST가 Article에 부여한 번호	
Journal name	JST가 Journal에 부여한 국가명, journal name	
ISSN	Journal의 ISSN no	
Vol	Journal의 Volume no	
Year	Journal의 출판연도	
Issue	Journal의 Issue no	
StartPage	Article의 StartPage	
EndPage	Article의 EndPage	
JOI	JST가 link center가 부여한 JOI 식별자	
URL	JOI의 상세 URL 정보	
DOI	CrossRef에서 사용하는 식별자	
DOI	CrossRef에서 사용하는 식별자	
PNID	Pubmed에서 사용하는 식별자	

4. 전자원문 연계시스템 설계 및 구현

4.1 KISTI 고유 식별자(KOI) 개발

KISTI에서 서비스하고 있는 디지털 콘텐츠의 종류와 수가 점차 늘어나고 있으며 대내외적으로 상호 디지털 콘텐츠의 활용에 대한 요구가 증가하고 있어 이에 대한 체계적인 관리와 서비스 요구되어지고 있는 상황이다.

KISTI가 보유하고 있는 디지털 콘텐츠에 유일한 식별자를 부여함으로써 사용자에 식별자만으로도 원하는 디지털콘텐츠를 서비스 받을 수 있도록 효율적이고 체계적인 디지털 콘텐츠 유통체계를 구축하기 위하여 KISTI 고유 식별자인 KOI를 개발하였다.

KOI 식별자의 구분구조는 접두부와 접미부로 구성되어 있으며, 이는 DOI와 JOI와의 호환성을 고려하였으며 접두부와 접미부는 "/"으로 구분된다.

접두부는 [KOI 등록기관]으로 구성되는데 등록기관이 KISTI 인 경우 kisti.std로 구성된다. 그러나 향후 DOI를 부여할 경우에는 접두부는 [KOI 등록기관]부분을 해당 등록기관의 고유번호로 사용하게 된다. 접미부는 학술지 형태인 경우와 비학술지인 경우로 구분되며 KOI 구분구조는 표 5, 6과 같다.

표 5 KOI 식별자 구분구조

구분	내	용
Prefix	- 등록기관 고유번호 (kisti.std)	
Suffix	- 등록기관 자체부여 기사고유번호 - 학술공지 : 자료유형,출판년,연호정보,시작페이지, [매체유형] - 비학술공지 : 자료유형,권리정보 [매체유형]	
KOI 예	- 학술공지 : kisti.std/j.1234-5678.v50i8.293.pdf - 비학술공지 : kisti.std/r.12345678.111.kisti.std/h.02100004.doc kisti.std/i.99999.223.pdf, kisti.std/i.99999.351	

표 6 KOI 접미부 구조

구분	내	용
저널정보	• 형식 : s(문자 3자리) : [(Journal)] - 회의지정보 : o (Conference) • 대상 : - 학술공지 : r (RMO Report) - 전문인력DB : h (Human Resource) - 연구보고서 : r (RMO Report) - 개별연구보고서 : i (Individual Publication)	
출판년	• ISSN, OCLC 등 출판정보 식별자가 없는 학술지 • 출판정보에 대한 식별자가 있는 경우 KISTI에서 부여한 식별자 사용 • 기본 0자리 숫자이지만 자료유형에 따라 1~3자릿수는 가변적임	
권호정보	• 형식 : v(권호ID) : i(호번지정용 번호)	
시작페이지	• 형식 : 시작페이지(SICI 호번지정용 번호)	
권리정보	• 연구보고서 : 서비스 중인 학리번호와 발행연도, 호번지 정보 • 전문인력DB : 호번지정보와 해당 호번지 호번지 정보, 호번지 정보 • 개인연구보고서 : 전문인력DB에서 활용되는 호번지번호와, 호번지 호번지 호번지 정보	
매체유형	• 형식 : 매체(문자 3자리ID) • 매체유형 default 값은 PDF 형 • Type1 : 전자책, 매체유형 - PDF: pdf, - TIF: tif, - DJV: djv, - HTML: htm, - HTML: hmo, - OOC: doc, - PPT: ppt, - XLS: xls, - DVI: dtd • Type2 : 가상본서 유형(호번지ID) - 가상본서: bse, - 가상본서: doi	

4.2 KOI 메타데이터

KOI 식별자 메타데이터는 총 12개의 항목으로 구성하였다. 국내 학술잡지와 국내회의자료 등 학술지 형태의 자료는 ISSN, VolumeID(권), IssueNO(호), StartPage(시작페이지)로 구성된다. 국내연구보고서와 개인연구발표자료 등 비학술지 형태의 자료는 제목, 저자, 년도로 구성된다. 전문인력DB의 경우는 주민등록번호와 개인 ID로 구성된다. KOI 식별자 메타데이터 구조는 다음과 같다.

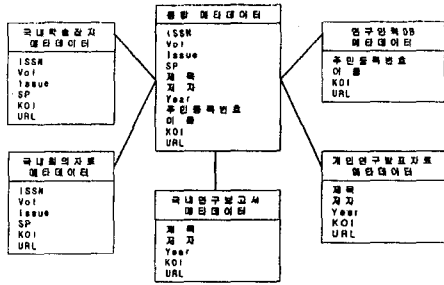


그림 2 KOI 메타데이터

4.3 KOI 기반 과학기술분야 전자원문 연계 서비스

KISTI 대표 홈페이지와 타 정보서비스 시스템에서 디지털 콘텐츠 서비스를 하는 경우 KOI를 활용한 lookup, 변환, KOI 등록에 대한 절차를 그림 3과 같이 구성하였다. 그림에서 보는 바와 같이 KISTI가 보유 또는 연계서비스를 하고 있는 전자원문을 획득하기 위해서 기존에는 KISTI 홈페이지를 통해서만 가능했으나 KOI 기반으로 서비스를 제공하면 KISTI 외부에서 개발 운영되고 있는 타 정보서비스 시스템에서도 KOI를 직접 lookup 및 변환이 가능하다는 것을 알 수 있다.

그림 3에서 절차①은 이용자가 정보를 검색한 결과에 대하여 원문 링크를 조회하기 위하여 필요한 항목들을 전달하는 단계이며, 절차②는 절차①에서 요청된 항목을 KOI 메타데이터에서 lookup을 수행한다. 절차③은 lookup 결과인 KOI 정보를 반환하는 단계이다.

절차(a)와 절차(b)는 이용자가 해당 전자원문을 획득 또는 열람하기 위해 KOI 링크정보가 있는 하이퍼링크 또는 버튼을 클릭하면 KOI에 해당하는 실제의 전자원문을 제공하기 위한 변환과정을 수행한다. 이 변환과정은 KOI에 해당하는 전자원문의 실제 위치(URL)로 변환되어 이용자에게 전자원문이 제공된다.

절차(1)은 신규 디지털 콘텐츠가 생성되었을 때에 KOI를 등록하는 것으로서 등록 요청시에 KOI 번호를 부여하고 식별 메타데이터 생성 및 해당 전자원문의 저장을 수행하게 된다.

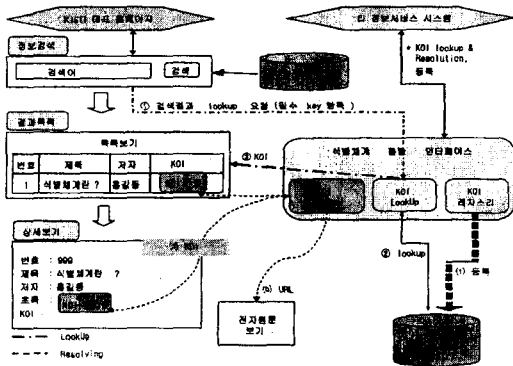


그림 3 서비스 구조

그림 4는 KOI lookup 프로세스와 전자원문 변환 프로세스

수행 절차를 나타낸다. KOI lookup 프로세스는 검색된 자료에 대한 전자원문의 KOI정보를 획득하기 위해 KISTI 정보 서비스 시스템과 통합 API를 통해 메타데이터 값을 전달받아 로컬 KOI Registry에 lookup한다. 전자원문 변환 프로세스는 로컬 KOI Registry에서 획득한 KOI정보를 이용하여 실제 전자원문의 위치로 변환하는 프로세스를 나타낸다.

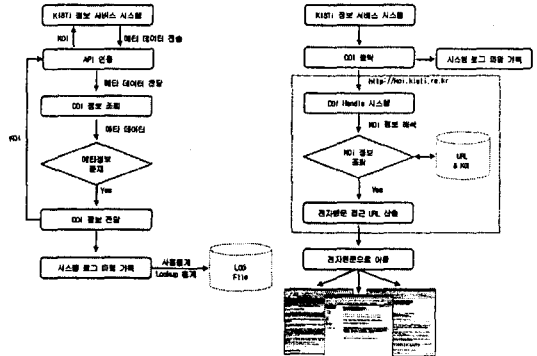


그림 4 KOI lookup 및 전자원문 변환 프로세스 절차

그림 3의 서비스 구성도와 그림 4의 전자원문 연계 서비스의 요소기능인 KOI lookup 프로세스 절차, 전자원문 변환 프로세스 절차 등을 통해 KOI를 기반으로 한 과학기술분야 전자원문 연계 서비스 시스템을 구현하였다.

5. 결론

본 논문에서는 KISTI에서 서비스하고 있는 과학기술분야 디지털 콘텐츠에 대한 식별체계에 대하여 논하였다. KISTI에서 서비스하고 있는 과학기술분야 디지털 콘텐츠에 KOI 부여하고, 활용함으로써 사용자에게 정확한 디지털 콘텐츠 위치정보를 제공하게 되어 디지털 콘텐츠를 효율적이고 체계적으로 서비스할 수 있는 체계를 구축할 수 있게 된다.

또한 KOI를 활용하여 과학기술분야 디지털 콘텐츠를 서비스함으로써 DOI 식별자를 활용하여 해외학술계 분야의 디지털 콘텐츠를 서비스 중인 CrossRef 등과의 디지털 콘텐츠에 대한 호환성이 높아졌다.

향후 연구과제로는 KOI 식별자를 활용한 디지털 콘텐츠 유통·관리체계 구축이 필요하며, 최근에 한국전산원에서 개발한 UCI(Universal Content Identifier)식별체계에 대한 연계·활용 방안에 대하여 지속적인 연구가 필요하다.

6. 참고문헌

- [1] ANSI/NISO, "Syntax for the Digital Object Identifier," Z39.84-2000, 2000.
- [2] Brain Green & Mark Bide(1997), "Unique Identifiers: a brief introduction," BIC and EDItEUR, 1997.
- [3] John Erickson, "The role of metadata supply chains in DOI-based, value-added services," ICSTI Forum, No.30, 1999.
- [4] K. Sollins, L. Masinter, "Functional Requirements for Uniform Resource Names," RFC1737, 1994.
- [5] Norman Paskin, "Information Identifiers," Learned Publishing, 10(2), 1997.
- [6] Norman Paskin. "DOI: Current Status and Outlook," D-Lib Magazine, Volume 5 Number 5, May 1999.
- [7] Norman Paskin, "DOI Handbook," DOI, 2003
- [8] R. Moats(1997), "URN Stntax," RFC2141, 1997.
- [9] 안계성, 이재진, 「디지털 콘텐츠 식별자 구분구조 연구」, 정보통신부, 2000.
- [10] <http://www.crossref.org>
- [11] <http://www.jstage.jst.go.jp>