

과학기술분야 전자원문 식별자 연구

이상환*, 신동구*, 김재수*, 최진영**, 정택영*

*한국과학기술정보연구원 정보기술지원실

**고려대학교 컴퓨터학과

e-mail:sanglee@kisti.re.kr

A Study on Digital Contents Identifier in Science and Technology Field

Sang-Hwan Lee*, Dong-Gu Shin*, Jae-Soo Kim*, Jin-Young Choi**, Taik-Yeong Chung*

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

**Dept of Computer Science, Korea University

요 약

정보통신의 발전으로 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. 또한, 학술지 위주의 식별자 연구가 활발하나 비학술지에 대한 식별자 연구는 부족한 실정이다. 따라서, 해외 주요 디지털 콘텐츠 서비스기관의 식별체계 활용사례를 분석하고 그 결과를 토대로 한국과학기술정보연구원(KISTI) 고유 식별자인 KOI(Kisti Object Identifier)를 개발하였다.

1. 서론

정보통신의 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 되어가기 때문에 사용자는 위치정보를 지속적으로 유지·관리해야하는 단점을 가지고 있다[6].

이러한 URL의 문제점을 대처하기 위해 새로운 정보자원 식별체계로 URN(Uniform Resource Name)이 개발되었다[4].

URN의 원리와 디지털 콘텐츠를 식별할 수 있는 식별체계로서 1997년 DOI(Digital Object Identifier)가 개발되었으며[2,3,6], 국내에서도 식별체계에 대한 연구가 활발하나 대부분의 연구내용이 DOI를 근간으로 하는 국내 학술지에 대한 연구이다[10,11].

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 학술지뿐 아니라 비학술지에도 적용 가능한 식별자를 개발하고, KISTI에서 보유하고 있는 과학기술분야 디지털 콘텐츠중 5종에 대한 메타데이터를 분석하여 본 논문에서 제시하는 식별자를 적용하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 디지털 콘텐츠 식별자인 DOI에 대하여 기술하고, 3장에서는 식별체계를 활용하고 있는 사례를 조사 및 분석한다. 4장에서는 KOI 구문구조와 메타데이터를 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 절

차에 따라 운영 관리하는 시스템을 디지털 콘텐츠 식별체계라고 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 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].

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.

DOI는 IDF에서 지정한 등록관리기관(DOI-RA : Registration Agency)에서 등록 관리한다. RA는 등록기관(Registrant)을 지정할 수 있으며, 이를 통하여 디지털콘텐츠 정보를 등록 받고 DOI를 부여하게 된다[7,13].

RA는 원활한 DOI 서비스를 위하여 IDC (Internet Data Center)에서 DOI 시스템을 운영해야 한다. 2004년 현재, 지정된 DOI-RA로는 전 세계적으로 CrossRef, Content Direction, Enpia Systems, Learning Objects Network, Inc, Copyright Agency Ltd., TSO(The Stationery Office), mEDRA, Nielsen BookData 등 8개 업체 및 단체이다[13].

3. DOI 식별자 조사 및 분석

3.1 DOI 식별자 구문구조 조사 및 분석

DOI 식별자의 구문구조는 접두부(Prefix)인 관리기관(Naming Authority)과 접미부(Suffix)인 로컬이름(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)

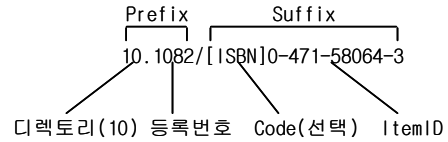


그림 1 DOI 구문구조 예

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].

CrossRef에 등록된 275개 기관 중 KISTI와 연관 있거나 DOI 등의 식별자로 디지털 콘텐츠를 서비스 중인 16개 기관을 선정하여 다음과 같이 조사하였다[12].

표 1 CrossRef 이용기관의 디지털 콘텐츠 식별자

조 사 기 관	식 별 자 예	구분
Wilson Select Plus	Accession No : BRD02039210	자체
ACM DL	10.1145/505282.505283	DOI
American Institute of Physics & American Physical Society	10.1063/1.1527703	DOI
American Mathematical Society	PII : S 0273-0979(02)00943-6	PII
ASCE	10.1061/(ASCE)0893-1321(2003)16:2(46)	DOI
Blackwell Publishing - STM	10.1034/j.1399-6576.2003.470101.x	DOI
Emerald MCB	10.1108/07363760310456919	DOI
Institute of Physics Publishing	10.1088/0305-4470/36/11/303	DOI
Kluwer Academic Publishers	Article ID : 5115103	자체
Lippincott	00043843-200329030-00006	자체
Nature Publishing Group	10.1038/422099a/10.1038/nature01474	DOI
Science	cgi/content/full/299/5605/392	자체
ScienceDirect	10.1016/S1093-0191(01)00128-9	DOI
Springer Verlag LINK	10.1007/s004930100016	DOI
ACM	10.1145/505282.505283	DOI
AIP Conference Proceedings	10.1063/1.1434614	DOI

3.2 DOI 식별자 메타데이터

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

Journal에 대한 DOI 식별자 메타데이터는 총 10개의 항목으로 이루어져 있다[10]. 이 중 lookup시 참조되는 필수 항목은 ISSN, Volume ID, Issue ID, Page ID 이며 나머지 항목은 Journal Title, First

Author, First Author 항목은 선택항목이다. 각 항목별 내용은 다음과 같다.

표 2 저널 DOI 메타데이터

구분	내용
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, Q1) 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 Edition항목을 사용하며 나머지 11개 항목은 같이 사용한다. 각 항목별 자세한 내용은 다음과 같다.

표 3 Book/Conference Proceeding DOI 메타데이터

구분	내용
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, Q1) 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)

4. 과학기술분야 전자원문 식별자 개발

4.1 KISTI 고유 식별자(KOI)

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

KISTI가 보유하고 있는 디지털 콘텐츠에 유일한 식별자를 부여함으로써 사용자에게 식별자만으로도 원하는 디지털콘텐츠를 서비스 받을 수 있도록 효율

적이고 체계적인 디지털 콘텐츠 유통체계를 구축하기 위하여 KISTI 고유 식별자인 KOI를 개발하였다.

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

접두부는 [KOI 등록기관]으로 구성되는데 등록기관이 KISTI인 경우 kisti.std로 구성된다. 그러나 향후 DOI를 부여할 경우에는 접두부는 [KOI 등록기관]부분을 해당 등록기관의 고유번호로 사용하게 된다. 접미부는 학술지 형태인 경우와 비학술지인 경우로 구분된다.

학술잡지의 구문구조는 "등록관리기관/자료유형.종정보.기사정보.최종Item.매체유형"이고, 비학술잡지의 구문구조는 "등록관리기관/자료유형.관리정보.매체유형"이다. KOI 구문구조는 표 5, 6과 같다.

표 4 KOI 식별자 구문구조

구분	내용
Prefix	- 등록기관 고유번호.(kisti.std)
Suffix	- 등록기관 자체부여 기사고유번호 - 학술잡지 : 자료유형.종정보.권호정보.시작페이지.[매체유형] - 비학술잡지 : 자료유형.관리정보.[매체유형]
KOI 예	- 학술잡지 : kisti.std/j.1234-5678.v50i6.293.pdf - 비학술잡지 : kisti.std/r.12345678.tif, kisti.std/h.92100004.det kisti.std/i.99999.223.pdf, kisti.std/i.99999.351

표 5 KOI 접미부 구문구조

구분	내용
자료유형	• 형식 : a(문자 1자리) • 대상 : - 학술잡지 : j(Journal), - 회의자료 : c(Conference) - 연구보고서 : r (R&D Report), - 전문인력DB : h (Human Resource) - 개별연구발표자료 : i (Individual Publication)
종정보	• ISSN, CODEN 등 종정보를 식별하기 위한 식별자 • 종정보에 대한 식별자가 없는 경우 KISTI에서 부여한 식별자 사용 • 기본 8자리 숫자이지만 자료유형에 따라 자릿수는 가변적임
권호정보	• 형식 : v2h호(SICI 표준지침을 따름)
시작페이지	• 형식 : 시작페이지(SICI 표준지침을 따름)
관리정보	• 연구보고서 : 서비스 중인 관리번호와 발행기관코드 포함 • 전문인력DB : 주민등록번호와 일대일 매칭되는 개인ID와 자료ID 포함 • 개인연구발표자료 : 전문인력DB에서 활용하는 주민등록번호와 일대일 매칭되는 개인ID 포함
매체유형	• 형식 : aaa(문자 3자리이내) • 매체유형 default 값은 PDF로 함 • Type1 : 전자원문 매체유형 - PDF: pdf, - TIF: tif, - XML: xml, - HTML: htm, - HWP: hwp, - DOC: doc, - PPT: ppt, - XLS: xls, - DVI: tpd • Type2 : 가상문서 유형(전문인력DB) - 기본정보: bas, - 상세정보: det

4.2 KOI 메타데이터

KOI 식별자 메타데이터는 총 12개의 항목으로 구성하였다. 국내학술잡지와 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.

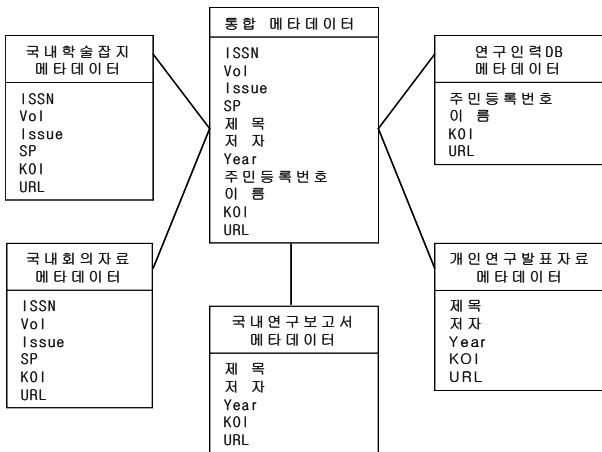


그림 2 KOI 메타데이터

5. 결 론

본 논문에서는 KISTI에서 서비스하고 있는 과학 기술분야 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 구축 할 수 있게 된다.

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

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 if metadata supply chains in DOI-based, value-added services," ICSTI Forum, No.30, 1999.
 [4] K. Sollins, L. Masinter, "Fuctional Requirments for Uniform Resourece 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, "URN Stntax," RFC2141, 1997.
 [9] 안계성, 이재진, 「디지털 콘텐츠 식별자 구문구조 연구」, 정보통신부, 2000.
 [10] 오경목, 황상규, "DDC를 이용한 DOI 구성체계에 관한 연구," 정보관리학회지 제18권 제1호, 2001.
 [11] 한혜영, 정동영, "국내 학술지 논문의 DOI기반 연계시스템 구축에 관한 연구," 정보관리학회지 제 17권 제4호, 2000.
 [12] <http://www.crossref.org>
 [13] <http://www.doi.org>