

# DDC를 이용한 DOI구성체계에 관한 연구\*

## A Study on the DOI Structure Using DDC

오경묵(Kyung-Mook Oh)\*\*, 황상규(Sang-Gyu Hwang)\*\*\*

### 초 록

DOI는 디지털 저작물에 식별자를 부여하는 URN의 명세를 따르면서 개방적인 구조를 갖고 있기 때문에 지금까지 개발된 다양한 식별자들 중 가장 우수하다고 할 수 있다. DOI가 대상 콘텐츠에 대한 어떠한 설명적 정보를 제공하지 못하기 때문에 DOI구성체계에서 접미부가 단순한 일련번호인지 혹은 의미를 담고 있는 식별자이어야 하는지 여부는 오랜기간 논란의 대상이 되었다. 비록 DOI의 접미부 번호를 부여함에 있어 아무런 제약이 없음에도 불구하고, 대부분의 DOI 등록기관들은 기존 식별체제인 SICI, ISBN, ISMN 등과 같은 국제표준을 DOI 접미부 번호부여에 사용해왔다. 이는 의미 전달을 위한 접미부의 구성방식에 있어 표준화된 구문형식이 존재하는지 여부와 상관없이 전통적으로 표준화된 식별자 채택의 중요성을 보여주고 있다. DOI 식별자는 식별체계의 지능화에 있어 중요한 수단이 될 수 있다는 가능성을 고려해야 한다. 이를 위해 본 논문에서는 DOI 식별자를 활용하여 대상 콘텐츠에 대한 주제정보를 전달할 수 있는 새로운 방식의 지능형 DOI 식별체계를 전통적인 주제표현 방식 중 가장 보편적인 DDC를 이용하여 개발하였다. 국내의 기존 사례로는 KERIS와 KDPC에서 사용하는 DOI 구현 방식을 본 연구의 목적에 맞게 사용하였다.

### ABSTRACT

The DOI is an opaque string or dumb number - nothing at all can be inferred from the number. However, some identifiers can include some intelligence in the number itself, which will tell the user subject and meaningful information about the entity that is being identified. Therefore whether the DOI suffix should be 'dumb' or 'meaningful' is a major issue in DOI syntax structuring. Although the choice of DOI Suffix is entirely optional, many DOI registrants have been using existing ISO standards such as SICI, ISBN, ISMN, etc. for identification of information resources. It is also important to consider that intelligence in a system of identifiers may be carried out by means of the DOI structure. It was suggested that there should be an explicit part of the DOI syntax which declares the meaningful information of the digital objects.

키워드 : 디지털 콘텐츠 식별자, 의미 식별자, 듀이 십진 분류법, 주제 분류

Digital Object Identifier(DOI), meaningful identifier, DDC, Subject Classification

\* 본 연구는 과학기술부 지원 2000 숙명여자대학교 연구기반 확충사업에 의하여 지원되었음.

\*\* 숙명여자대학교 정보과학부 문헌정보학 전공 조교수(kmoh@sookmyung.ac.kr)

\*\*\* 국방연구원 연구원(kid4@naver.com)

■ 논문 접수일 : 2000년 11월 30일

■ 게재 확정일 : 2001년 3월 6일

## 1 서 론

DOI(Digital Object Identifier)는 디지털 콘텐츠 객체에 대한 영구식별자로, 인터넷 상 다양한 종류의 디지털 문서(텍스트, 이미지, 사운드, 동영상 등)에 대한 고유성을 부여하고, DOI와 URL과의 변환연계를 통하여 인터넷 상의 해당 객체에 대한 위치 정보를 제공할 수 있다. 이를 위하여 CNRI(Corporation for National Research Initiatives)는 영구식별자인 DOI를 등록, 관리하는 기관으로서 DOI를 해당 URL로 변환해주는 핸들시스템을 운영하고 있다(Sun and Lannom 1999). DOI는 1997년 프랑크프르트 도서 박람회를 통해 처음 소개된 이후로, NISO표준(NISO 1999)으로 제정되었다. DOI의 관리는 IDF(International DOI Foundation)가 담당하고 있는데, IDF는 1998년 결성되어 DOI 체계의 개발과 정책을 수립하고, DOI 등록관리를 담당하고 있다. 그리고 1999년 11월 해외 주요 학술지 출판사들이 IDF와 협력하여 각종 학술지의 논문을 DOI를 기반으로 연계하는 프로젝트에 참여함에 따라, DOI는 인터넷 상에서 연계서비스를 제공하기 위한 중요한 연계 도구로 자리매김을 해가고 있다.

DOI는 디지털 저작물에 식별자를 부여하는 URN의 명세를 만족시키는 동시에 최대한 개방적인 구조를 채택하고 있다. 따라서 DOI는 URN의 특성을 간직하고 있으면서, 기존의 식별자 체계를 수용할 수 있기 때문에 지금까지 개발된 다양한 식별자들 가운데 가장 우수하다(Norman

Paskin 1997). 그러나 DOI는 해당 콘텐츠의 유일성을 부여하는 영구적이고 고유한 식별자이기는 하지만, 대상 콘텐츠에 대한 어떠한 관련정보도 제공하지 않는다. 즉 DOI 자체는 단순한 '의미없는 식별자(dumb number)'로서, DOI가 부여된 해당 콘텐츠에 제목이 무엇인지, 콘텐츠의 타입은 무엇인지, 저자는 누구인지 등에 대한 관련 정보를 지니고 있지 않다. 따라서 DOI의 활용가치를 높이고 유용한 서비스를 제공하기 위해서는 DOI가 부여된 콘텐츠에 대한 관련 정보를 제공할 수 있는 일정한 수단이 반드시 필요하다(조소연, 조현주 2000). IDF에서는 이를 해결하기 위한 방법으로 메타데이터를 이용하는 방법을 고안하였으며, INDECS(Interoperability of Data in E-Commerce Systems)메타데이터 모델을 DOI를 위한 메타데이터 표준으로 채택하였다(IDF 2000). 그러나 현재 CNRI의 핸들시스템은 단순히 DOI를 URL로 변환되는 기능만을 제공하고 있으며, DOI와 메타데이터 간의 연계서비스는 제공하지 않고 있다. 또한 각 기관이나 단체마다 OCLC의 Dublin Core를 비롯하여 TEI의 TEI Independent Header 등 다양한 메타데이터가 사용되어진 경우에 비추어 볼 때, 실제 운영에 있어 메타데이터 간의 상호운영성 문제도 발생하게 된다. 이는 DOI핸들서버가 처리해야할 작업의 양이 증가함을 의미하며, 상대적으로 이용자에 대한 서비스의 처리 속도가 늦어지게 되는 요인으로 작용하게 된다. 따라서 DOI운영 시스템이 신속하고 효과적인 서비스를 제

공하기 위해서는 메타데이터와의 연계체계와는 별도로 DOI자체만으로 이용자에게 대상 콘텐츠에 대한 최소한의 의미정보를 제공할 수 있는 의미정보 식별자(meaningful identifier)구성체계가 마련되어야 한다.

의미정보 식별자 구성체계에 있어 기관마다 의미의 표현방식이 각각 서로 다를 경우에는 타 기관과의 연계가 불가능하게 되며, 따라서 전세계 공통적으로 통용될 수 있는 보편적인 의미전달방식이 필요하다. 또한, 식별자의 길이가 무한정 길어지는 것을 방지하고 DOI핸들서버에 처리속도를 향상시키기 위해서는 최대한 의미를 함축적으로 전달할 수 있는 표기법을 채택하여야 한다. 이를 위하여 기존 도서관 주제분류방식인 DDC번호를 통해 의미정보를 전달하는 새로운 DDC기반 DOI식별자(DDCDOI)를 개발하였다. 새로운 의미정보 식별자는 NISO의 표준을 기초로 구성되어 기존 DOI핸들시스템의 URL변환연계 기능을 그대로 활용하게 되며, 보다 자세한 정보를 제공하기 위하여 DOI메타데이터와의 연계서비스를 구축할 수 있다.

## 2 웹 정보자원의 식별

DOI식별자가 부여되는 대상인 디지털 저작물의 3가지 유형에 대해 살펴보고, 디지털 저작물을 위한 식별자가 갖추어야 할 기본적인 요건에 대해 살펴보기로 한다. DOI식별자를 이용하여 해당 콘텐츠에 접

근하는 연계방안과 DOI식별자의 활용사례로서 DOI등록, 변환, 검색, 유통관리, 참조링크 시스템의 실제 구축사례를 살펴보기로 한다.

### 2.1 저작물

과거의 경우, 저작물이라고 지칭할 때에는 보통 서지의 형태로 출판되는 인쇄물을 지칭하게 된다. 그러나 대상 저작물이 데이터베이스에 디지털화된 형태로 저장되고, 그 저장형태에 있어서도 PDF나 HTML과 같은 다양한 형태로 저장되면서 저작물의 형태에 있어 다양성이 생겨나기 시작하였다. 또한 인터넷이 보편화된 이후로 저작물은 다수의 인터넷 사이트에 복수로 존재할 수 있게 됨에 따라, 대상 저작물을 공간적, 시간적 표현 단계에 따라 구분지를 필요성이 발생하게 된다. IFLA모형을 기초로 Caplan과 Arms는 학술지 논문 형태의 창작물의 유형을 다음의 3가지 유형으로 구분 짓고 있다(Caplan and Arms 1999). 첫 번째 유형은 저작물(work)로서 저자의 창조적 활동의 산출물이라 할 수 있다. 두 번째 유형은 실현물(manifestation)로서 저작물(work)이 서지이나 디지털의 형태로 출판된 결과물을 지칭한다. 마지막 유형인 아이템(item)은 인쇄본 형태로 출판되어진 실현물 중 하나를 의미한다. 지금까지의 식별자에 관한 연구는 주로 저작물의 식별에 관한 연구가 주로 진행되어져 왔으나, 웹 상에 존재하는 디지털 형태에 저작물에 대해서 이러한 방식을 그대로 적

용하기에는 부적절하다. 웹 상에서는 동일한 저작물에 대하여 PDF나 HTML과 같은 다양한 형태와 날짜별로 다양한 버전이 존재할 수 있기 때문이다. 따라서 디지털 콘텐츠를 위한 식별자는 저작물 뿐만 아니라 실현물 혹은 아이টে에 대한 식별자로서의 역할을 동시에 수행할 수 있어야 한다.

웹 상에서 디지털 콘텐츠를 위한 식별자로 지금까지는 하이퍼링크 개념에 URL이 연계서비스를 제공하기 위한 식별자로서 사용되어져 왔으나, URL은 디지털 콘텐츠를 위한 식별자로서는 부적합하다. URL은 대상물에 대한 위치정보를 제공함으로써 실현물 혹은 아이টে에 대한 식별자로서의 역할을 수행할 수 있으나, 저작물 자체에 대한 영구적인 식별자로서의 기능을 수행할 수 없기 때문이다. 인터넷 환경 하에서 URL의 유효생명주기(valid life cycle)는 평균 75일로 조사(Kahle 1997)되었으며, 이는 URL이 디지털 대상물에 대한 영구적인 식별자로서의 기능을 제대로 수행할 수 없음을 보여 주는 것이다.

## 2.2 식별자

디지털 저작물을 위한 식별자는 저작물 자체에 대한 영구적인 식별자로서의 기능 뿐만 아니라 실현물 혹은 아이টে에 대한 식별자로서의 역할도 제공할 수 있어야 한다는 점을 언급한 바 있다. 디지털 저작물을 위한 식별자가 갖추어야 할 기본적인 요건에 대해 살펴보기로 한다. Caplan과 Arms은 참조연계서비스(reference linking

service)를 제공하기 위하여 식별자가 갖추어야 할 요건들을 다음과 같이 세 가지로 정리하였다(Caplan and Arms 1999). 이 중 세 번째 것은 참조연계 시 필수적인 속성이며, 처음 두 가지는 일반적인 식별자가 갖추어야 할 요건을 나열하고 있다.

① 영속성(Persistence) : 식별자는 영구적으로 사용될 수 있어야 하며, 신뢰성을 가지고 사용할 수 있도록 충분히 조직적이고 기술적인 체계를 갖추어야 한다. 영속성을 제공하기 위한 가장 중요한 요건은 무결성으로서 식별자의 부여체계에 있어 모호함이나 애매한 부분이 존재해서는 안 된다.

② 유일성(Uniqueness) : 하나의 식별자는 그 자신의 이름공간 내에서 유일해야 한다. 다시 말해 하나의 변환시스템 내에서는 서로 같은 동일형태의 식별자가 동시에 존재할 수 없다. 만약 주소공간 내에 여러 개의 변환시스템이 존재할 경우에는 그 중 하나가 새로 부여되는 식별자에 대해서 유일성을 제공할 수 있도록 관리기능을 제공할 필요가 있다.

③ 다중 변환(Multiple resolution) : 연계 서비스를 지원하기 위한 식별자 변환시스템은 하나의 식별자로부터 전체 혹은 다수 개 아이টে으로의 연계기능을 제공할 수 있어야 한다.

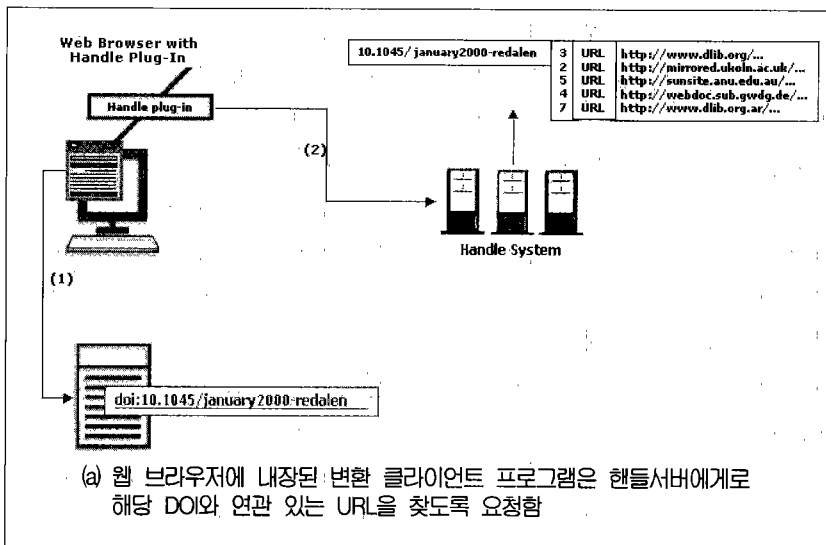
식별자로서 URL은 특정 아이টে에 대한 위치정보를 제공함으로써 전체 혹은 다수 개 아이টে으로의 연계기능은 제공할 수 없다. 이러한 이유 때문에 URL은 참조연계 서비스를 제공하기 위한 식별자로서는 부

적절하다. 이와는 별도로 인터넷상의 정보 자원 식별을 위한 식별기호 체계인 URN (Uniform Resource Name)은 정보자원의 대상 디지털 콘텐츠의 저장 위치, 서비스 시스템 등과는 상관없이 고유의 기호로 표현한 고유이름 부여체계로서 영속성과 유일성을 제공할 수 있는 식별기호이다. DOI(Digital Object Identifier)는 PURLs (Persistent URLs)와 함께 이러한 URN의 실험적 운영 체제로서 개발되었다. 그러나 PURL은 URL처럼 인터넷 자원의 위치로 직접 이동하는 것이 아니라 중간 변환 과정을 통하여 클라이언트에게 URL 정보를 알려주는 형태로, OCLC에서도 URN이 보편화되기 전 단계로 임시적 방법으로 간주하고 있다. DOI는 IETF에서 정의한 URN의 명세를 완벽히 만족시키면서 기존의 식별자 번호체계를 포함할 수 있으므로, 디지털 자원의 저작권 관리와 식별을 위한

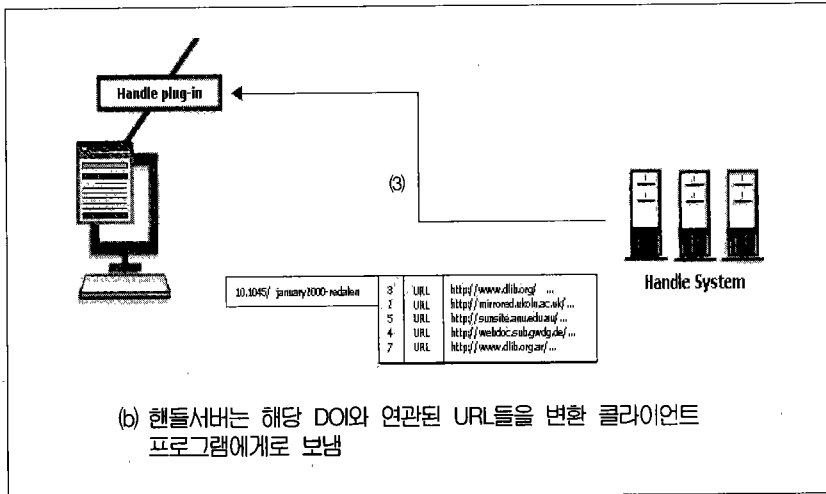
식별자로 사용할 수 있다.

### 3 DOI-URL연계변환

디지털 저작물을 위한 식별자는 해당 식별자로부터 전체 혹은 다수개의 아이템으로 연계기능을 제공할 수 있어야 한다. 즉, 식별자를 위한 연계서비스를 제공하기 위해서는 변환(Resolution)을 통해 사용자가 실제 콘텐츠에 접근할 수 있도록 해주는 시스템을 갖추어야 한다. <그림 1, 2, 3>은 CNRI(Corporation for National Research Initiatives)의 핸들시스템을 이용하여 DOI를 URL로 변환해주는 과정을 기술하고 있다(Larry 2000). 핸들시스템(Handle System)은 디지털 콘텐츠의 이름(Handle)을 저장하는 분산 컴퓨터 시스템으로서 DOI와 URL의 연계정보를 가지고 있으며,



<그림 1> DOI-URL연계변환 처리과정(a)



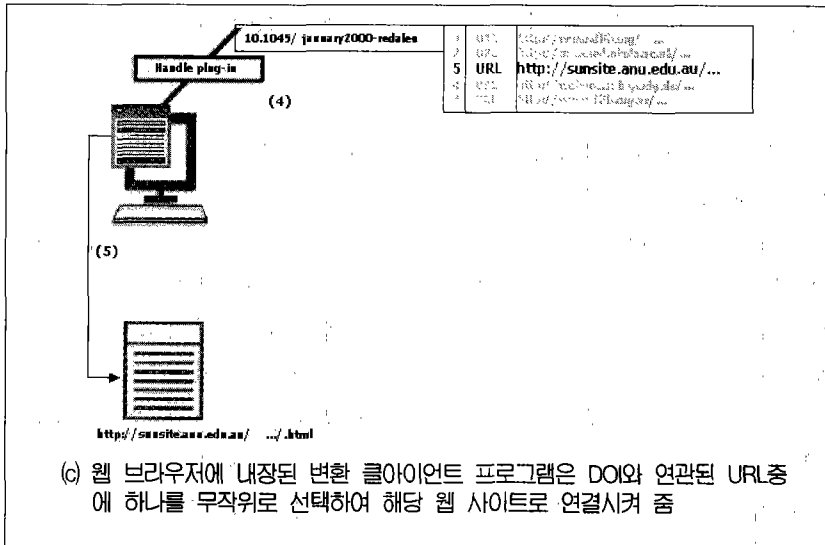
〈그림 2〉 DOI-URL연계변환 처리과정(b)

DOI 코드에 해당되는 디지털 콘텐츠의 위치(URL)로 변환해주는 역할을 한다. 사용자가 DOI를 클릭하면 웹 브라우저에 내장된 변환 클라이언트 프로그램은 이를 인식하여 사용자가 선택한 해당 DOI를 핸들시스템에 전송하게 된다(그림 1). 변환 요청을 받은 핸들시스템은 DOI와 관련된 URL을 이용자의 인터넷 브라우저에 전송(그림 2)하고, 변환 클라이언트 프로그램은 해당 웹사이트의 내용을 브라우저에 출력(그림 3)시켜 줌으로서 이용자는 콘텐츠 자체를 볼 수 있게 된다.

해당 DOI의 소유권자는 디지털 콘텐츠가 다른 곳으로 이동하거나 저작권 소유권자에 변경사항이 생겼을 경우에는, 변경내역 정보를 핸들시스템의 등록관리 서버에 변경, 기록함으로써 변경 후 접근하는 이용자는 새로운 사이트로 자동으로 연계된다.

초기의 CNRI의 핸들시스템에서는 하나의 DOI가 하나의 URL로 변환되는 기능만을 제공하고 있었다. 그러므로 하나의

저작물에 대한 여러 대상물, 다양한 버전, 혹은 복수 형식 중 선택하는 다중 변환기능의 필요성이 제기되었다(Paskin 1998). 그러나, 현재의 CNRI의 핸들시스템에서는 하나의 DOI가 여러 개의 URL로 변환되는 기능을 제공하며, 변환 클라이언트 프로그램은 여러 URL들 가운데 하나를 무작위로 선택하여 해당 웹사이트로 연결시켜 준다. 변환 클라이언트 프로그램이 여러 URL들 가운데 하나를 무작위로 선택(random)하여 해당 웹사이트로 연결시켜주는 방식으로 구현은 비교적 단순하나, 연계서비스에 있어서는 비효과적인 방식이며 추후 개선되어야 한다. 이를 해결하기 위한 보다 지능적이고 효과적인 연계방안의 하나로 근접거리 연계서비스를 생각해 볼 수 있다. 만약 이용자의 위치가 한국(.kr)이면서 DOI변환을 통해 얻은 첫 번째 URL주소위치가 영국(.uk)이고 두 번째 URL주소위치가 일본(.jp)이라면, 보다 가까이 위치한 일본으로 연결시켜주는 것이



〈그림 3〉 DOI-URL연계변환 처리과정(c)

보다 효과적이다. 이용자의 위치정보는 최초로 사용자의 웹브라우저가 DOI구문을 인식할 수 있도록 변환 클라이언트 프로그램(DOI plug-in)을 설치하는 과정에서 이용자로부터 획득할 수 있으며, 특정 아이템에 대한 위치정보는 변환 클라이언트 프로그램이 해당 URL의 구문을 분석함으로써 얻을 수 있다.

#### 4 DOI 구성체계

핸들시스템을 통해 DOI를 URL로 변환해주는 연계서비스와 유사한 형태로, 한글 키워드 도메인 서비스는 기존의 인터넷 주소를 대신하여 특정단어를 입력하면 관련 홈페이지로 연결되는 서비스를 제공한다. 예를 들어 웹브라우저의 주소 입력창에 인터넷 주소를 모두 입력하지 않고 '한국일

보'라고 입력하면 한국일보 홈페이지(www.hankooki.com)에 연결된다. 그러나, 새로운 방식의 한글도메인주소체계는 기존의 URL주소체계에 비해 여러 가지 불편한 점이 존재한다. 첫째, 사용자가 입력한 한글주소를 기존의 영어식 주소(도메인 주소)로 전환하고, 이를 다시 IP주소로 전환하는데 따른 처리시간이 길어진다는 단점을 가지고 있다. 둘째, 한글도메인 주소를 이용하기 위해서는 한글키워드를 URL로 변환해주는 별도의 plug-in 프로그램을 설치하거나, 변환서버를 지원하는 새로운 네임서버로 주소를 변경해야하는 불편이 따른다. 이런 단점에도 불구하고, 새로운 방식의 한글도메인 주소체계의 이용자가 증가하는 가장 큰 이유는 일반 사용자에게 보다 이해하기 쉬운 서비스를 제공하기 때문이다. 그리고, "이해하기 쉬운 서비스"에서 가장 큰 핵심은 주소식별체계

가 얼마나 사용자에게 효과적으로 해당 사이트의 주제정보를 전달할 수 있는가 여부에 달려 있다.

URL주소체계 역시 DOI식별자와 동일하게 그 형식 자체로는 의미없는 식별자(dummy identifier)에 불과하다. 도메인 주소를 가지고 해당 웹사이트를 연결해주는 네임서버 역시 도메인 네임 자체의 의미를 식별하지 않으며, 단순히 일반적인 영어식 URL주소(도메인 주소)를 IP주소로 전환하는 역할을 수행한다. 그러나 새로운 URL주소를 부여하는 대부분의 등록자들은 URL주소자체가 일반사용자들에게 해당 사이트의 주제정보를 효과적으로 전달할 수 있도록 URL을 구성하기 위해 노력하고 있다. 예를 들어, A라는 펜팔클럽사이트의 인터넷주소가 “http://www.penpal.co.kr”일 때 URL주소는 해당사이트의 주제정보를 전달할 수 있는 의미 있는 식별자(meaningful identifier)역할을 수행하게 된다. 결국 A라는 펜팔클럽사이트의 인터넷주소가 한글도메인주소인 “http://펜팔친구”라면 일반사용자들에게 보다 효과적으로 의미 전달이 가능해지는 것이다. 이와 마찬가지로 DOI식별체계에서도 DOI가 디지털콘텐츠의 주제정보를 전달할 수 있는 의미 있는 식별자(meaningful identifier)

역할을 수행하게 될 때 DOI의 활용가치는 훨씬 더 커지게 될 것이다.

#### 4.1 DOI 형식

DOI의 식별기호(Identifier)는 접두부(prefix)와 접미부(suffix)로 구성되어 있으며, 접두부는 디렉토리 관리기관 코드인 <DIR>과 등록기관 코드인 <REG>만으로 구성되어 있다. DOI의 접미부 코드를 부여하는데 있어서는 임의부여를 원칙으로 하며, 기존 국제표준을 이용하거나 등록기관에서 임의적으로 부여한 일련번호가 사용되어지고 있다. 현재 디렉토리 관리기관 코드로 유효한 값은 핸들시스템을 운영하는 CNRI가 부여받은 ‘10’이 유일하며, 문자열의 최대 길이는 128자(UTF-8코드)로 제한된다. DOI의 구성체계는 정형화되어 있으며, NISO에서 정한 기본적 구분형태는 다음과 같다.

<DIR>.<REG>/<DSS>

국제 DOI재단(International DOI Foundation)과 핸들시스템을 운영하는 CNRI는 DOI자체를 의미없는 식별자(dummy identifier)로 간주하고 있다(Norman

<표 1> KERIS DOI번호부여방식

E3R/LS_Number	::=	<DIR>.<REG>"/"<ISDI>
<DIR>	::=	*(<<이름부여기관>>)
<REG>	::=	UTF-8로 표현된 모든 글자
<ISDI>	::=	(ISBN   ISSN   BICI   SICI)<ItemNo>+
<ItemNo>	::=	한글,숫자,영어를 포함한 UTF-8로 표현된 모든 글자



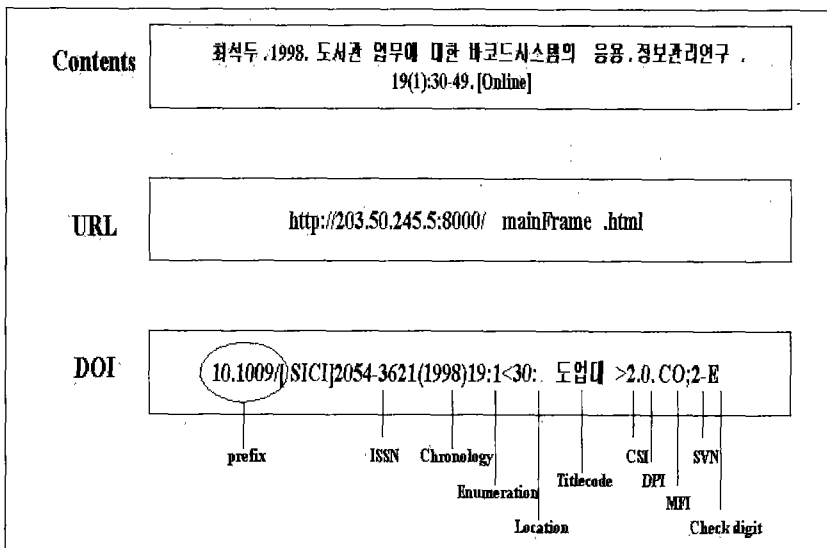
Paskin 2000). 이는 핸들시스템이 해당 DOI코드의 의미분석을 시도할 경우, DOI를 URL로 변환하는 과정이 복잡해져 응답시간이 길어질 수 있기 때문이며, 접미부 코드 부여방식을 특정 의미정보를 기반으로 일정한 형태로 정형화할 경우 DOI의 활용범위가 크게 제한되기 때문이다. 그러나 DOI등록서버가 별도로 운영되는 것처럼 해당 DOI코드의 의미분석만을 담당하는 서버를 별도로 운영할 수 있으며, 접미부 코드 부여방식의 일부만을 정형화할 경우에는 DOI의 활용성을 크게 제한하지 않는다. 실제로 여러 DOI 등록기관들은 접미부를 구성하는데 있어 ISBN, ISSN, ISAN 등과 같은 국제 ISO 표준을 사용하고 있으며, 미국국립표준협회(NISO)에서는 ISDI(International Standard Digital Identifier)를 사용할 것을 권장하고 있다.

## 4.2 기존사례

의미가 부여된 DOI식별체계의 최근 국내 사례로 국제 ISO표준을 토대로 DOI를 부여하는 방식과 등록기관 임의로 일련번호를 부여한 방식을 살펴보기로 한다. 국제 ISO표준을 토대로 DOI를 부여하는 방식의 예로 SICI를 기반으로 한 한국교육학술정보원(KERIS)의 DOI번호부여방식을 살펴보고, 등록기관 임의로 부여하는 방식으로는 한국 데이터베이스진흥센터(KDPC)의 DOI 구분 식별 코드를 살펴보기로 한다.

### (1) KERIS DOI번호부여방식

본 논문에서 인용할 KERIS DOI번호부여방식은 학술지 논문 연계시스템(E3R/LS)에서 실제 설계, 구현된 방식으로, 접미부 코드 부여방식에 있어 SICI기



<그림 4> KERIS의 DOI예제

반 번호부여체계를 토대로 한글처리를 위한 확장형 SICI를 적용하고 있다(한혜영 2000). <표 1>은 KERIS의 DOI번호부여방식을 BNF (Backus-Naur Form)로 표시한 것으로서, 의미전달을 위해 SICI의 번호체계를 접미부에 해당하는 <ISDI>에 그대로 적용하게 된다.

SICI는 학술지 논문의 식별자로서 ANSI NISO 표준 Z39.56으로 개발되었다. 지금까지 DOI가 연속간행물의 식별자로 먼저 활용됨에 따라, 인터넷상에서 SICI기반 DOI 번호 부여체계의 다양한 외국사례를 찾아볼 수 있다. <그림 4> KERIS의 DOI 실제예제를 토대로 DOI코드의 접미부를 구성하는 SICI의 주요 구성 항목들을 살펴보기로 한다.

ISSN	9자리 문자(하이픈 포함)
Chronology	연대. YYYYMMDD (YYYY : 연도, MM : 월, DD : 일)
Enumeration	통권번호 체제. (ex. 제 19권 제 1호)
Location	게재논문이 시작되는 학술지 권호 내의 위치를 나타내는 것으로 쪽 번호 혹은 프레임 번호를 말한다.
Title code	서명코드 길이는 SICI가 영문자 6바이트이므로 한글은 3자로 하며 UTF-8 코드를 사용한다. 한자인 경우는 한글로 변환한 후 기술한다.
CSI	code structure identifier 1 = serial item(예:권호) 2 = serial contribution

(예:게재논문)

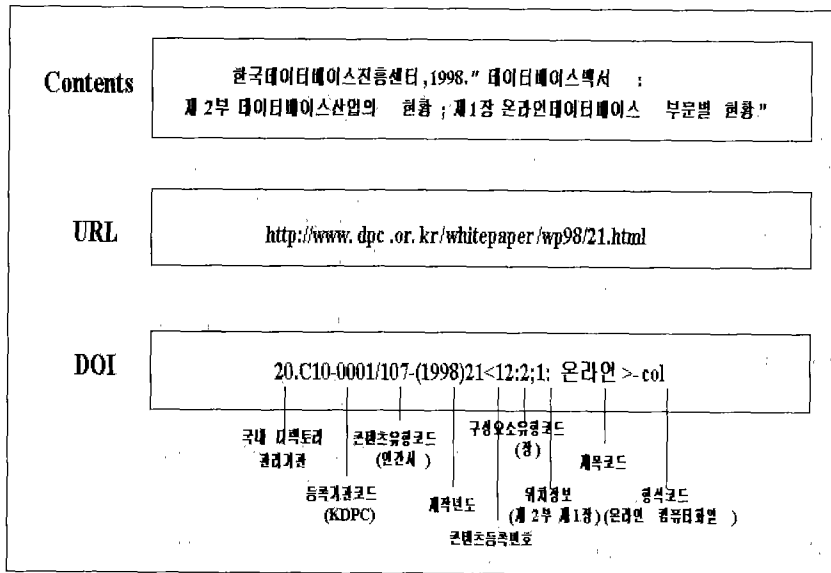
- 3 = serial contribution  
(로칼 번호에 의해 식별 가능함)
- DPI Derivative Part identifier  
0 = serial item 혹은 contribution  
1 = 연속간행물 item 혹은 contribution을 위한 목차  
2 = serial item 혹은 contribution의 색인  
3 = serial item 혹은 contribution의 초록

MFI	출판형태를 식별하는 코드
SVN	SICI번호를 작성시 사용한 SICI표준의 버전을 표시
Check Digit	SICI 일관성과 유효성을 검사하기 위해 사용하며, SICI 문자에 의해 계산됨

KERIS의 DOI번호부여방식은 SICI를 토대로 접미부 코드를 부여하게 되며, SICI는 자동적으로 생성될 수도 있으므로 DOI 역시 자동적으로 생성될 수 있다. KERIS DOI번호는 사용자에게 연속간행물에 대한 생성연도, 권호 혹은 게재논문을 식별할 수 있는 정보를 제공하게 되나, 6바이트 크기의 서명코드는 해당 콘텐츠에 대한 주제정보를 전달하기에는 미흡하다. 또한 SICI기반 DOI번호 부여방식은 다양한 종류의 디지털 콘텐츠를 위한 보편적인 DOI 식별체계로는 부적절한 실정이다.

## (2) KDPC DOI번호부여방식

KDPC의 DOI번호부여방식은 한국케이



〈그림 5〉 KDPC의 DOI예제

터베이스진흥센터(KDPC)를 디렉토리 관리기관으로 가정하고 '20'을 디렉토리 번호로 사용하는 새로운 방식의 Prefix와 Suffix 식별코드를 개발하였다(김세정, 안계성 2000). KDPC DOI번호부여방식에서는 등록기관 코드인 <REG>를 부여하는 등록기관 분류 원칙을 새로 제안하였으며, 접미부 코드를 부여방식 역시 KDPC만의 독자적인 구성체계를 개발하였다.

KDPC DOI번호부여방식에서는 디지털 콘텐츠 보유기관을 대분류, 중분류, 소분류로 구분하고, 코드길이는 대분류는 알파벳 1문자, 이하 분류 단위에서는 아라비아 숫자 2자리씩 부여하였다. 실제 예로서 '국가정부기관'은 A, 중분류인 '행정기관'은 A10, 소분류인 '중앙 정부기관의 등록기관코드'는 A1003이 된다. KDPC DOI의 접미부 코드는 다음과 같이 콘텐츠 유형코드, 콘텐츠

식별번호, 콘텐츠 구성요소 식별코드, 형식 코드의 4가지 데이터 항목으로 구성된다.

① 콘텐츠 유형코드

콘텐츠유형을 나타내며, 지적콘텐츠의 유형을 크게 '서적 및 일반 간행물', '음악 및 음향물', '영상물', '소프트웨어 프로그램물'로 대분류하고, 각각의 소분류 항목에 대한 콘텐츠의 유형코드를 개발하였다.

② 콘텐츠 식별번호

'제작년도', '콘텐츠 등록순 일련번호'의 2부분으로 구성되며, 제작년도가 제시되지 않을 경우라도 빈괄호로 표기한다.

③ 구성요소 식별코드

'구성요소 유형코드', '위치정보', '제목 코드'의 3부분으로 구성되며, 콘텐츠 보유기관이 임의적으로 결정할 수 있다.

④ 형식코드

지적콘텐츠의 형식을 식별하기 위한 것

〈표 2〉 DDC-DOI 번호표기 방법

ImageDOI_Number	::=	<DIR>.<REG>/<DSS>
<DSS>	::=	<SUBJECT>/<Serial_Number>
<SUBJECT>	::=	<SUBJECT>/[<SUBJECT_TYPE>]<SUBJECT_Number>   [<SUBJECT_TYPE>]<SUBJECT_Number>
<SUBJECT_TYPE>	::=	(DDC KDC LCC CC)
<DIR>	::=	{ 디렉토리 관리기관 번호 }
<REG>	::=	{ 등록기관번호 }
<SUBJECT_Number>	::=	{ 해당 분류방식을 따른 주제분류번호 }
<Serial_Number>	::=	{ 공급자가 부여한 일련번호 }
 (Example.1) 10.054/[DDC]020/1997-0320004.01		
(Example.2) 10.054/[DDC]005.7525/[DDC]005.754/ALCS99-03104.01		

으로 크게 패키지(Package) 형태의 콘텐츠와 객체(Objects) 형태의 콘텐츠로 구분하고 ANSI/NISO Z39.56(1996)에서 제공되고 있는 서브셋 코드를 사용한다.

KDPC의 접미부 DOI번호부여방식은 SICI와 유사한 형태이나, DOI가 다양한 디지털 콘텐츠에 적용될 수 있음에 착안하여 다양한 디지털 콘텐츠유형을 세분화하였다는 이점을 가지고 있다. 그러나 KDPC의 제목코드는 SICI의 서명코드에 서처럼 약어형태로 기술함에 따라 해당 콘텐츠에 대한 주제정보를 전달하기에 부족하다. 또한 Suffix 코드 부여방식에 있어 기존 국제표준을 활용하는 대신에 독자적인 Suffix 코드체계를 채택함으로써 콘텐츠 등록 및 관리에 있어서 타 등록기관과의 혼란을 야기할 수도 있을 것이다.

### 4.3 DDC를 이용한 DOI구성체계

기존의 디지털콘텐츠에 대한 의미정보

가 부여된 DOI식별체계는 각 기관마다 해당 콘텐츠에 대한 주제정보를 전달하기 위한 영역이 충분히 할당되어 있지 못하며, 의미체계의 구성방식 또한 각기 서로 다르기 때문에 보편적인 의미전달이 불가능하다는 문제점을 가지고 있다. 이를 위하여 디지털 콘텐츠에 대한 주제정보를 전달하기 위한 수단으로, 전세계적으로 의미가 통용될 수 있는 새로운 방식의 DOI구성체계를 제안하고자 한다.

새로운 DOI구성체계가 디지털 콘텐츠에 대한 주제정보를 표현함에 있어 가장 중요한 두 가지 사항은 표현된 의미정보가 이용자간에 원활한 정보교환을 위하여 동일양식으로 의미정보를 표기해야 한다는 점과 의미정보를 표현할 수 있는 공간이 128자(UTF-8코드)로 제한된다는 점이다. 다시 말해서, 전세계적으로 누구나 동일한 대상을 동일한 의미로 이해할 수 있는 표기법을 채택해야 하며, 의미정보를 최대한 짧고 효과적으로 표현할 수 있는 방법

을 제시해야 한다. 본 연구에서는 이를 위해 전세계에 이용자가 원본 디지털콘텐츠에 대한 의미정보를 서로 교환할 수 있는 표준 방식으로 도서관분류방식을 기초로 한 DDC-DOI번호부여방식을 개발하였다. <표 2>는 새로운 방식의 DOI구성체제인 DDC-DOI번호부여방식을 BNF (Backus-Naur Form)로 정의하였다. <표 2>의 예제에서는 의미정보를 표현하는 데 있어 여러 가지 도서관 분류방식 가운데 DDC를 기반으로 DDC-DOI를 구성해 보았다. DDC는 전세계적으로 가장 널리 사용되는 도서관 분류방식인 동시에 보다 구체적인 분류가 필요한 경우에 이용자 스스로 분류항목을 세분화함으로써 보다 자세한 의미정보의 표현이 가능하다는 이점을 가지고 있다. <표 2>에서 DDC-DOI의 식별자 구조는 접두부(Prefix)와 접미부(Suffix)로 구성되어 있다. (Emample.1)의 접두부에서 '10'은 디렉토리 관리기관 코드이며, '054'는 ALCS라는 등록기관번호이다. DDC-DOI의 식별자 구조에 가장 큰 특징은 접미부가 일정한 형식에 따라 의미정보를 표현하

고 있는 주제영역(SUBJECT)과 각각의 등록기관이 임의로 부여 가능한 일련번호(Serial\_Number)로 구성된다는 점이다. 이는 기존의 각각의 등록기관에서 독자적으로 채택해온 접미부 구성방식을 그대로 수용할 수 있다는 장점을 지닌다. 현재 주제영역의 타입(SUBJECT\_TYPE)은 UDC를 제외한 대부분의 도서관 분류방식을 허용하고 있다. 이는 UDC의 경우, 주제분류번호 구성 시 DDC-DOI의 필드 구분을 위한 특수문자 '/' 사용을 허용하기 때문에 자동화된 DOI처리 시 주제영역(SUBJECT)과 일련번호(Serial\_Number)항목간의 모호성을 없애기 위해 제외하였다. 새로운 DOI구성체제에서 디지털 콘텐츠에 대한 주제정보는 <표 2>의 (Emample. 2)에서처럼 전체 DDC-DOI의 길이가 128글자를 넘지 않은 범위 내에서 1개 이상의 의미정보를 표현할 수 있다. 공급자가 부여하게 되는 일련번호(Serial\_Number)부여 방식에 있어서는 전체 길이가 128글자를 넘지 않은 범위 내에서 자유로이 일련번호를 구성할 수 있다.

<표 5> DDC-DOI 확장모델

예제. 1. DDC-DOI + KERIS 혼합모델(<그림 4>의 예제 참조.)

DDC(020) is Library and information sciences.  
->10.1009/[DDC]020/[SICI]2054-3621(1998)19:1<30:도업대>2.0.CO:2-E

예제. 2. DDC-DOI + KDPC 혼합모델(<그림 5>의 예제 참조.)

DDC(005.74) is Data files and databases.  
->20.C10-0001/[DDC]005.74/107-(1998)21<12:2:1온라인>-col

도서관분류방식을 기초로 한 DOI번호 부여방식인 DDC-DOI번호표기방법의 가장 큰 장점은 사람 뿐만 아니라 기계인 컴퓨터가 스스로 대상 콘텐츠에 대한 의미정보를 해독할 수 있다는 점이다. 구체적인 활용 가능한 사례로 DDC-DOI가 부여된 원본 디지털 콘텐츠들을 연관된 주제별로 자동으로 분류해주는 웹 콘텐츠 자동분류 시스템을 구현할 수 있다. 또한 DDC-DOI번호표기방법은 일반 문헌콘텐츠뿐만 아니라 이미지나 사운드, 동영상과 같은 다양한 디지털콘텐츠에 대해서도 동일하게 적용할 수 있다. 새로운 DOI구성체계인 DDC-DOI의 장점들을 열거해보면 다음과 같다.

#### ① 보편성

전세계 공통적으로 사람 뿐만 아니라 기계인 컴퓨터도 이해, 활용할 수 있는 보편적인 의미전달방식

#### ② 용이성

기존의 전통적인 도서관분류의 노하우를 그대로 DOI번호 부여시에도 적용가능

#### ③ 편리성

기존의 서지정보를 디지털화 할 경우, 이미 부여된 도서관 분류기호를 토대로 자동으로 DOI번호부여가 가능함

#### ④ 확장성

DDC-DOI의 구성항목에 기존의 타 등록기관에서 사용중인 DOI 부여방식을 그대로 수용함으로써, 기존의 타 등록기관에서 사용중인 DOI번호부여 방식과 쉽게 혼용이 가능함(참조<표 3>).

## 5 결 론

의미없는 식별자인 DOI의 불편함을 해소하고자, DOI식별자 자체만으로 해당 디지털 콘텐츠에 대한 주제의미전달이 가능한 새로운 DOI구성체계인 DDC-DOI식별체계를 개발하였다. 새로운 방식의 DOI식별자인 DDC-DOI식별체계는 기존의 각 DOI등록기관마다 의미체계의 구성방식이 각기 서로 다르기 때문에 보편적인 의미전달이 불가능하다는 문제점을 해결하였으며, 기존의 타 등록기관에서 사용중인 DOI번호부여 방식과 쉽게 혼용이 가능함에 따라 뛰어난 확장성을 가지고 있다. 따라서 기존의 정보제공기관은 DDC-DOI식별자를 활용하여 전세계적으로 통용될 수 있는 의미 참조 연계서비스(semantic-based reference linking service)를 제공할 수 있다.

DDC-DOI식별체계는 여러 가지 용도로 다양한 분야에 활용되어질 수 있는데, DDC기반 주제별 웹 정보검색, 자동화된 디지털 콘텐츠 분류시스템, 지능형 웹 정보검색 에이전트 개발 등 다양한 분야에 활용이 기대되어 진다. 이는 DDC-DOI식별자를 통해 인간 뿐만 아니라 기계인 컴퓨터가 스스로 대상 콘텐츠에 대한 의미정보를 해독할 수 있기 때문이며, 기존 도서관 분류서비스를 기반으로 하는 다양한 네트워크 정보시스템들과의 연동이 가능하다. 앞으로의 연구과제는 실제 DOI서비스의 운영을 통해, DDC-DOI식별체계에 대한 확장 및 수정이 계속 진행되어야 할 것이다.

## 참 고 문 헌

- 김세정, 안계성. 2000. DOI 구분 식별 코드 개발. 『제7회 한국정보관리학회 학술대회』, 2000년 8월 17-18일. [서울. 이화여자대학교].
- 정상원, 오상훈. 2000. 디지털 콘텐츠 식별 시스템(DOI)의 구축. 『제7회 한국정보관리학회 학술대회』, 2000년 8월 17-18일. [서울. 이화여자대학교].
- 조소연, 조현주. 2000. DOI 메타데이터의 역할 및 개발 현황. 『제7회 한국정보관리학회 학술대회』, 2000년 8월 17-18일. [서울. 이화여자대학교].
- 한혜영. 2000. 『국내 학술지 논문의 DOI기반 연계시스템 구축에 관한 연구』. 박사학위논문, 이화여자대학교 대학원, 문헌정보학과.
- IDF. 2000. "DOI Handbook." [online]. [cited 2000.12.3]. <<http://www.doi.org>>.
- Kahle, Brewster. 1997. "Archiving the internet." *Scientific American*. [online]. [cited 2000.12.3]. <[http://www.archive.org/sciam\\_article.html](http://www.archive.org/sciam_article.html)>.
- Larry Lanno. 2000. "CrossRef Technical Introduction and Overview." Corporation for National Research Initiatives. [online]. [cited 2000.12.3]. <<http://www.crossref.org>>.
- NISO Press, 1999. "Syntax for the Digital Object Identifier." [online]. [cited 2000.12.3]. <<http://www.niso.org/z3984.html>>.
- Norman Paskin, 1997. "Information Identifiers." *Learned Publishing*, 10(2): 135-156.
- Norman Paskin, 2000. "The DOI Handbook", International DOI Foundation. [online]. [cited 2000.12.3]. <[http://www.doi.org/handbook\\_2000/index.html](http://www.doi.org/handbook_2000/index.html)>.
- Priscilla Caplan and William Y. Arms. 1999. "Reference Linking for Journal Articles." *D-Lib Magazine*, 5(7/8). [online]. [cited 2000.12.3]. <<http://www.dlib.org/dlib/july99/caplan/07caplan.html>>.
- Paskin, Norman. 1998. "The digital object identifier initiative: Current position and review forward." IDF. [online]. [cited 2000.12.3]. <[http://www.doi.org/about\\_the\\_doi.html](http://www.doi.org/about_the_doi.html)>.
- Sun, Sam X and Lannom, Larry. 1999. "Handle System Overview." Corporation For National Research Initiatives. [online]. [cited 2000.12.3]. <<http://www.ietf.org/internet-drafts/draft-sun-handle-system-03.txt>>.