



URI (Uniform Resource Identifier)

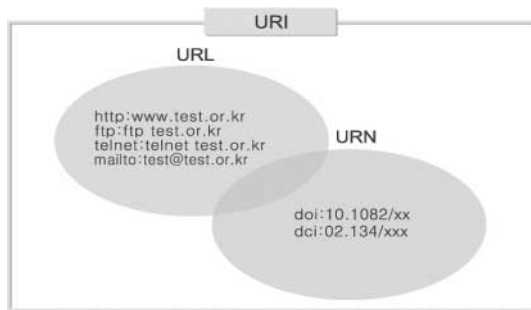
오상훈

TTA 데이터기술위원회 의장
한국데이터베이스진흥센터

I. 개요

본 고는 URI(Uniform Resource Identifier)의 개념 및 구조에 대한 간략한 정리를 시작으로 URI와 관련된 표준화방향과 내용을 소개하고자 한다.

URI는 인터넷상의 정보자원에 대한 식별체계로서 인터넷 주소체계인 URL과 고유이름 체계인 URN을 총칭하는 개념으로서 문서, 이미지, 음악파일, 동영상 등 다양한 정보자원에 대한 접근에 있어서 유일성을 부여하고 식별을 가능케 하는 관리체계를 의미하며 URI를 구성하는 URL과 URN은 상호보완적인 측면으로 사용된다.



〈그림 1〉 URI 개념도

① URL의 개념 : URL(Uniform Resource

Locators)이란, 인터넷상의 자원의 위치(Location)를 나타내는 문자열로서 자원의 물리적인 주소를 나타내며, 다양한 스키마를 통해 자원에 접근하는 수단을 제공하는 체계이다.

② URN의 개념 : URN(Uniform Resource Names)이란, 자원의 위치 또는 접근방법과는 무관하게 콘텐츠 자체에 영구적인 고유의 객체식별 기호를 부여함으로써 해당 정보자원으로 쉽게 접근하는 수단을 제공하는 체계이다.

II. URI 구문구조

URI는 프로토콜을 의미하는 「스키마」와 이에 대한 세부 사항을 나타내는 「설명부」로 이루어지는데, URL과 URN의 구문구조는 URI의 구문구조에 따르며, 각각의 스키마 정의 부분에 따라 그 의미가 달라진다.

스키마 명	스키마 설명부		
〈scheme〉	〈scheme-specific-part〉		
	〈authority〉	〈path〉	〈query〉

① URL 구문구조

요소	Protocol	Host	Port	Directory	File Name
기능	N/W 전송규약	질의 서버명	서버 내부망	세부 위치	자원 이름
예	http://	www.mic.go.kr	:8000/	jsp/	index.html

※ Protocol에는 ftp, http, gopher, mailto, news, nntp, telnet, wais 등의 규약이 있음. 브라우저는 URL 내의 웹서버 호스트 명을 네임서버에 질의하고 네임서버는 해당 웹서버 IP 주소값을 찾아 이용자의 브라우저로 전송한다.

② URN 구문구조

요소	URN	NID	NSS
기능	Label	URN의 취급방법 결정	식별 문자열
예	urn	doi	10.1082/abc

※ NID : Namespace Identifier / NSS : Namespace Specific String

□ URL과 URN의 비교

구분	URL	URN
체계	주소체계 (접근체계)	이름체계 (식별체계)
구문구조	Protocol + Domain+ Directory	NID + NSS
변환	<ul style="list-style-type: none"> • Domain ↔ IP Address • Inverse/Reverse Resolution 	<ul style="list-style-type: none"> • URN→URL, URN→URN • Multi-Resolution
부여방법	Top Down	Bottom UP
지속성	변형가능성	반영구적
Namespace	http, ftp, mailto, telnet...	doi, dci, cid, hdl....
확장 가능성	낮음	높음
공용성	높음	낮음

III. URN 기술 및 표준화동향

인터넷망이 고도화되어 갈수록 인터넷 상에서 유통되는 디지털화된 정보자원에 대한 중요성이 높아지고 있으며, URN은 이러한 디지털 객체에 대한 영구적인 식별을 통해 정보자원을 효율적으로 관리할 목적으로 고안되었다.

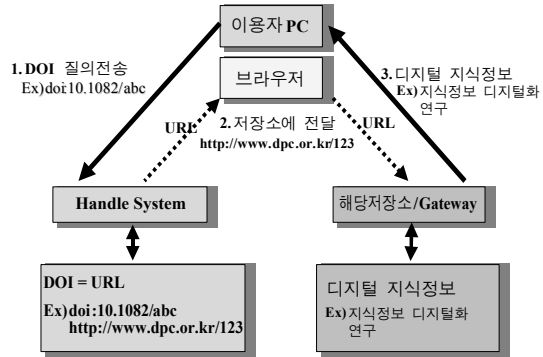
1994년 IETF는 디지털 객체에 대한 영구적인 식별방법으로서 URN에 대한 개념을 설정하였

고, 미국 IDF의 DOI와 일본 cIDF의 CID 등은 디지털 콘텐츠 식별기호에 대한 기술개발이 진행중에 있으며, 우리나라의 경우 정보통신부 주관 사업으로 미국 IDF의 DOI와 연동가능한 디지털콘텐츠 식별 시스템인 DCI가 개발되어 운영 준비중이다. 또한 URN 관련 기술에 대한 국제표준화는 IETF의 URN 워킹그룹에서 RFC를 통해 추진되고 있다.

- RFC 2141: URN Syntax

- RFC 1737: Functional Requirements for Uniform Resource Names
- RFC 2276: Architectural Principles of Uniform Resource Name Resolution

URN 기반에서 인터넷 정보자원에 대한 보다 효율적인 관리와 활용을 위해 그 자원 자체가 갖는 구체적인 속성 및 특성의 파악이 중요시 되면서 정보자원의 내용기술을 위한 표준 메타 데이터 형식인 Dublin Core, RDF, INDECS 등이 출현하고 있음.



<그림 2> DOI 작동과정

DOI(Digital Object Identifier) 란 ?

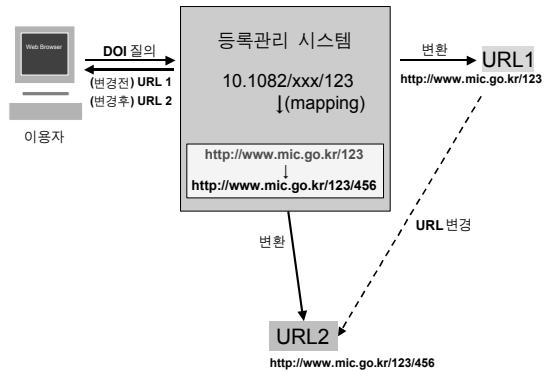
DOI란 URN 체계의 하나로서 디지털 자원에 대한 식별체계이며 1994년 CNRI에서 URN 체계의 핵심시스템으로 개발하여 운영중인 Handle System을 기반으로, 1997년 미국출판협회(AAP) 주도로 과학·기술·의학분야 전자저널(E-Journal)의 식별과 유통을 위해 고안된 시스템이다. DOI는 Prefix와 Suffix로 구성되는데, Prefix는 URL의 Domain과 같은 역할을 하고 Suffix는 URL의 디렉토리와 같은 역할을 하고 있다.

구조	Prefix	Suffix
예시	10.1082/	knowledge-002

- 10: “.”이하의 Prefix 번호를 할당해 주는 기관(등록 관리기관)
- 10(등록 관리기관)에서는 1082, 1083, 1084, ... 등을 할당
- “/”이하의 기호나 번호는 prefix를 할당받은 기관이 자체적으로 부여

이용자가 DOI를 클릭하면 메시지가 Handle System에 전송되어 DOI와 URL의 매핑 테이블을 참조, 해당 URL을 이용자의 인터넷 브라우저에 전송하고, 이용자는 정보자원 자체를 볼 수 있게 되거나 정보자원에 대한 정보나 접근 방법에 대한 정보를 획득.

디지털 정보자원이 다른 URL로 이동하거나 디렉토리가 변경되는 경우 변경내역이 등록관리 시스템에 기록됨으로써 변경 후 접근하는 이용자는 새로운 URL로 자동으로 안내되어진다.



<그림 3> DOI의 URL 변경관리

DOI는 2000년 미국의 표준화기구인 ANSI/NISO에서 국가표준으로 채택하였다.

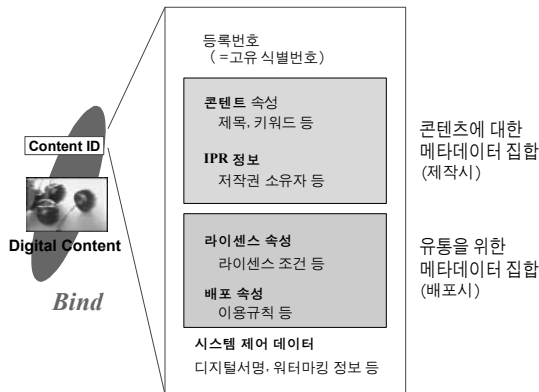
- Syntax for the Digital Object Identifier -- ANSI/NISO Z39.84, 2000; 디지털환경에서 지식정보 자원의 식별을 위해 이용되는 DOI의 정의, 형식 및 특징 등을 정의.

CID(Content Identifier) 란 ?

일본의 cIDf(Content ID Forum)에서 콘텐츠

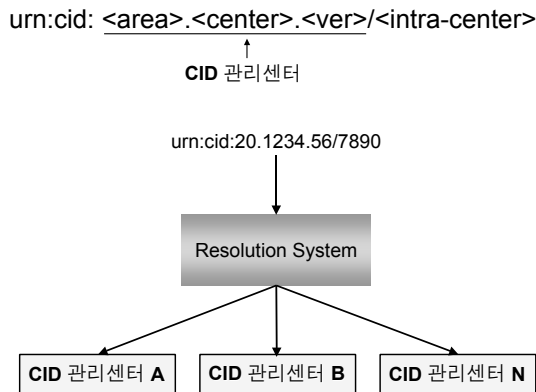
및 유통속성을 기술하기 위해 개발한 식별기호 및 메타데이터 관리시스템이며 CID의 구성은 아래와 같은 구조로 표현.

- 등록번호 + 콘텐츠 메타데이터 + 유통메타 데이터 + 시스템제어 데이터



〈그림 4〉 CID의 구조

CID의 변환은 자체적인 변환시스템 (Resolution System)을 이용하여 CID 관리센터 번호로 변환하는 기능을 한다.



〈그림 5〉 CID의 URN 구문구조와 변환

작년 3월 cIDf는 spec.1.0을 발표하고 국제 표준화기구(ISO/IEC)를 통해 표준화 추진중

- 콘텐츠 유통을 위한 참조모형

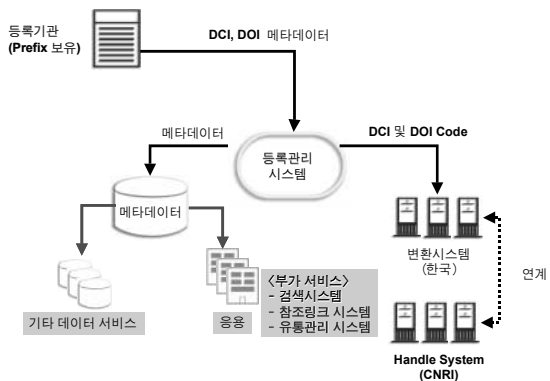
- Content ID Format(XML/RDF, 상세사항 기술)
- Content ID Profile(지적재산권 DB 정보, 디스크립터, 고유 ID)
- 저작권 보호기술(2계층 워터마킹, 디지털서명, 감시도구(Net Watcher))
- Content ID 관리센터에서의 과정 및 절차

DCI(Digital Content Identification System) 란 ?

디지털 콘텐츠 식별 시스템(DCI)은 URN 기반 식별체계로서 디지털콘텐츠 등록관리 시스템, 검색 시스템, 유통관리 시스템, 참조링크 시스템, 변환 시스템으로 구성되었으며, IDF의 DOI 및 cIDf의 CID와 연동되었다. DCI의 구분 구조는 Prefix(100.xxx/), Suffix로 구성되었다.

구조	NID	Prefix	Suffix
예시	dci:	100.1234/	internet-002

- ※ 100.1234는 등록 기관번호를 의미
- ※ “/”이하의 기호나 번호는 등록기관에서 자체적으로 부여



〈그림 6〉 DCI의 작동과정

메타데이터(METADATA) 란 ?

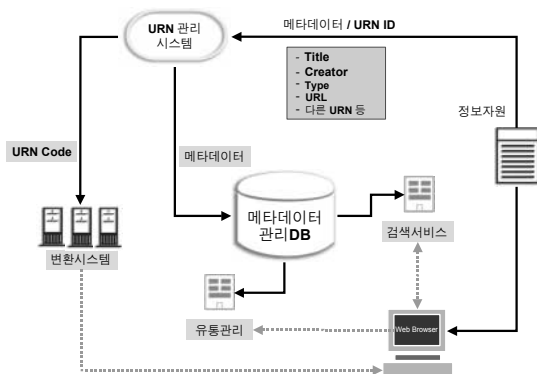
메타데이터(metadata)란 데이터에 대한 구조

화된 데이터, 자원에 대해 기술하고 있는 정보 (이름, 주제, 위치, 특성 등)의 집합체; 해당 자원을 기술하는 데 필요한 속성 또는 요소들의 집합체를 의미한다.

메타데이터를 네트워크 상의 정보자원과 연결시킴으로써, 자원에 대한 색인을 가능하게 하며 또, 그 자원 자체에 대한 접근과는 별도의 요약정보에 대한 접근을 가능케 함으로써, 정보 자원의 검색, 적합한 정보의 확인, 질의처리, 데이터 교환, 자원의 기술, 요소의 정의 확인 등에 활용된다. 정보검색에 있어서도 웹 에이전트가 전체 웹 정보자원이 아닌 메타데이터만을 색인 함으로써 색인된 데이터의 품질을 높이고 탐색 도구를 통한 정보검색의 효율을 높일 수 있다.

정보시스템에서 메타데이터 공유의 필요성에 대한 인식이 확산되면서, 정보시스템간 데이터 교환시의 데이터 의미해석을 명확히 하고 데이터 해석을 자동화할 수 있는 환경을 제공하기 위한 메타데이터 레지스트리(Metadata Registry)가 구축되고 있다. 메타데이터의 표준화와 관련하여, Dublin Core 메타데이터 요소는 2001년 5월 미국의 정보화 관련 국가표준기구인 NISO에서 표준으로 채택되었음(NISO Z39.85 : The Dublin Core Metadata Element Set, 2001). 메타데이터 레지스트리는 국제 표준화기구인 ISO/IEC의 JTC1 SC32에서 ISO/IEC 11179 시리즈를 통해 표준화를 추진이며 RDF는 인터넷 관련 표준화기구인 W3C에서 표준화활동 등이 추진되고 있다.

※ RDF(Resource Description Framework)란 인터넷에서 상이한 메타데이터 간의 의미, 구문, 구조에 대한 공통적인 규칙을 지원하여 웹 상에 존재하는 상이한 메타데이터간의 상호운용성을 확보하기 위한 프레임워크



〈그림 7〉 URN 체계에서 메타데이터의 역할

메타데이터는 인터넷 정보자원의 기술 (Description) 및 관리에 이용되는 Dublin Core 메타데이터 요소가 대표적이며, 상이한 메타데이터간의 의미공유 및 데이터 교환을 위한 프레임워크로 RDF(Resource Description Framework), 메타데이터 레지스트리, 전자상거래에 활용되는 INDECS(Interoperability of Data in E-Commerce Systems) 프레임워크가 있다. 한편, 다양한 메타데이터 형식이 출현하고 각종

IV. 마침말

국내 지식정보 자원에 대한 원활한 접근, 이용과 유통활성화를 위해 인터넷 정보자원의 관리체계를 보다 체계적이고 적극적인 방법으로 고도화하여 다양한 접근체계의 개발과 정보자원 유통기반의 제공을 위한 체계적인 기술개발 및 관리가 필요합니다. 또한 국제적인 표준화 트랙을 추적 반영하는 동시에 국내에 개별적으로 개발되고 있는 정보자원의 효율적인 관리체계와 관련된 기술분야의 표준화를 마련해야 하는 토대를 위해 URI포럼은 다양한 분야의 전문가와 산업계의 참여를 유도할 수 있는 창구의 역할을 수행할 것으로 기대합니다. TTA