

# Internet 정보자원의 식별 체계

정보자원을 식별하는 식별기호는 관련분야 내에서의와 관련분야간의 커뮤니케이션에 필수적인 도구가 된다. 이미 국제표준도서번호(ISBN)나 국제표준연속간행물번호(ISSN)는 도서와 연속간행물에 대한 식별기호로서 널리 사용되고 있으며 중요성을 인정받고 있다. 네트워크 환경이 일반화되고 인터넷상에 존재하는 정보자원의 식별이 중요한 이슈로 떠오르면서 URI, URN, URL, URC 등으로 이어지는 URx가 인터넷 정보자원의 식별과 검색을 위한 메타데이터로 활용되고 있다.

이재진/한국데이터베이스진흥센터 정책연구과

## 연 · 재 · 순 · 서

1. 메타데이터의 개요
2. DC(Dublin Core)
3. GILS(Government Information Locator Service)
4. IAFA Templates
5. MARC
6. PICS(Platform for Internet Content Selection)
7. RFC 1807
8. SOIF
9. TEI header
10. URx..... 이번호
11. Warwick Framework RDF (Resource Description Framework)
12. 메타데이터 향후 방향

년에서 1995년까지 IETF(Internet Engineering Task Force)의 URI WG(Working Group)이 운영되었고 URI WG에서는 각종 Internet-Draft와 RFC(Requests for Comments)를 제안하였다. 현재는 URN WG이 조직되어 활발하게 활동하고 있고 최근 Network WG에서는 URI에 관한 일반구문(Generic Syntax)을 최종 갱신하여 1998년 8월에 RFC 2396으로 제안하였다.

## URx의 개요

URI(Uniform Resource Identifier)는 인터넷상의 정보자원 식별을 위한 식별기호 체계로서 인터넷 주소 체계로 알려져 있는 URL(Uniform Resource Locator)과 고유 이름 부여 체계인 URN(Uniform Resource Name), 그리고 이들을 연결하기 위한 데이터 구조인 URC(Uniform Resource Characteristics)로 연계되어 이들을 URx로 총칭한다.

URx는 현재 사용되고 있는 인터넷상의 수많은 정보자원에 대한 식별 체계의 기본이며 이와 관련되어 많은 전문가들의 활발한 논의와 연구, 표준 제안 등이 이루어져 왔다. 즉, URI와 이와 관련된 여러 가지 문제를 해결하기 위해 1992

## URx의 구성

URx는 URI, URL, URN, URC 등 정보자원의 식별을 위한 체계 모두를 총칭한다.

### ■ URI

URI는 통일된(Uniform) 정보자원(Resource)의 식별기호(Identifier)란 개념이 함께 적용된 것이다. 통일성의 이점은 정보자원의 접근 방식이 다르더라도 서로 다른 유형의 정보자원 식별기호를 같은 방법으로 처리할 수 있고, 공통적인 문장 형식의 통일된 의미적 해석이 가능하고, 새로운 유형의 식별기호의 도입이 용이하다는 것이다.

또한 식별기호가 다른 환경에서도 재사용될 수 있도록 함

으로써 새로운 응용이나 프로토콜 상에서 광범위하게 사용될 수 있도록 한다. 정보자원은 문서, 이미지, 다운로드가 가능한 파일, 서비스, 전자 메일박스 등 모든 전자정보자원을 망라하며 이들 정보자원을 식별해 주는 것이 식별기호이다.

URI는 식별기호로서 제한된 구문의 문자열로 표현된다.

일반적으로 사용되는 URI의 예는 다음과 같다.

- ftp://ftp.is.co.za/rfc/rfc1808.txt  
→ 파일 전송 프로토콜 서비스(FTP)를 위한 ftp 스키마
- gopher://spinaltap.micro.umn.edu/00/Weather/California/Los%20Angeles  
→ Gopher와 Gopher+ 프로토콜 서비스를 위한 gopher 스키마
- http://www.math.uio.no/faq/compression-faq/part1.html  
→ 하이퍼텍스트 전송 프로토콜 서비스를 위한 http 스키마
- mailto:mduerst@ifi.unizh.ch  
→ 전자우편주소를 위한 mailto 스키마
- news:comp.infosystems.www.servers.unix  
→ USENET 뉴스그룹과 아티클을 위한 news 스키마
- telnet://melvyl.ucop.edu/  
→ TELNET 프로토콜을 통해 상호 서비스를 하기 위한 telnet 스키마

URI는 소장 위치, 이름, 혹은 이 둘 모두로 분류할 수 있다. 즉, URL은 URI의 하부집합으로서 정보자원의 이름이나 기타 속성으로 자원을 식별한다기 보다는 주요 접근 메커니즘(네트워크상의 위치)으로 정보자원을 식별하는 것이다. 이와는 달리 URI의 다른 하부집합인 URN은 고유하고 영

구적으로 남을 수 있는 정보자원의 이름을 나타낸다. 이러한 URI의 기본적인 관계는 <그림 1>과 같이 나타낼 수 있다.

### ■ URL

URL(Uniform Resource Locator)은 인터넷상의 정보자원에 접근하는 방법을 명시적으로 지시하는 것으로서 정보자원의 위치(location)를 표현하는 간단한 문자열로 구성된다. URL을 표현하는 구문은 일반적으로 다음과 같이, 사용되는 스키마 이름(<scheme>)과 콜론(:), 그리고 스키마에 따라 해석되는 문자열 부분(<scheme-specific-part>)으로 구성된다.

<scheme>:<scheme-specific-part>

스키마는 해당 정보자원을 위해 사용되는 프로토콜이나 호스트명, 포트번호, 디렉토리, 파일명 등을 의미하며 ftp, http, gopher, mailto, news, nntp, telnet, wais, file 등이 포함된다. ftp, http, file 스키마의 경우, 계층 구조를 이루는 문자열이 뒤따를 수 있으며 각 계층은 '/'로 구분된다. URL의 예는 다음과 같다.

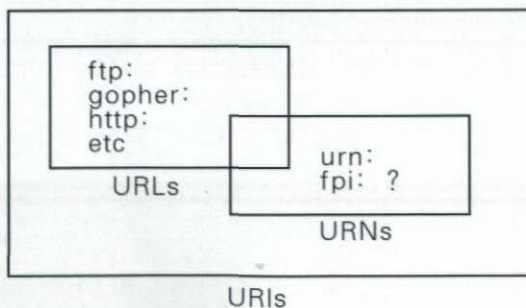
<http://www.acme.edu.au:8080/projects/oil.html>

URL은 인터넷상의 정보자원을 검색하는 수단으로서 널리 활용되고 있다. 그러나 URL에 대응되는 정보가 없거나 더 이상 이용할 수 없게 되는 경우에는 검색 수단으로서 기능을 상실하게 된다. 이러한 문제점을 해결하고자 제안된 것이 URN과 URC이다.

### ■ URN

URN(Uniform Resource Name)은 정보자원의 소장 위치, 프로토콜, 호스트 등과는 상관없이 고유의 기호로 표현한 것으로서 영구적으로 남을 수 있는 이름이며 다른 이름 체계를 URN으로 매핑하기 쉽도록 설계되었다. 따라서 URN을 표현하는 구문은 기존의 ISBN, ISSN 식별기호와 관련을 맺으며 이들 기호를 활용하기도 한다.

URN은 일반적으로 레이블인 URN, URN을 어떻게 다룰 것인가를 결정하는데 사용하는 NID(Namespace



<그림 1> URI의 관계

Identifier), 그리고 이름을 식별하는 문자열인 NSS(Namespace Specific String)로 구성되며 다음과 같은 구조를 갖는다.

$\langle \text{urn} \rangle ::= \text{"urn:"} \langle \text{NID} \rangle \text{"::"} \langle \text{NSS} \rangle$

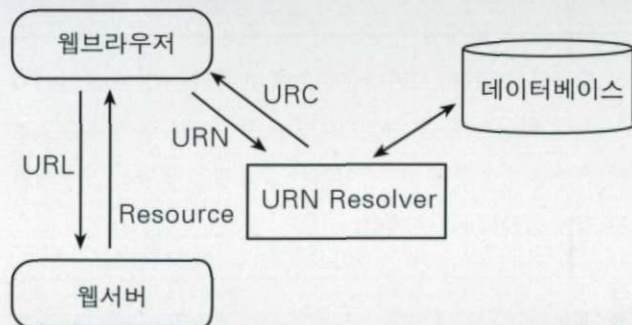
“”로 표시된 부분은 반드시 나타나야 하는 필수요소이다. URN의 예는 다음과 같다.

`urn:isbn:123456789X`

`urn:inet:dstc.edu.au:tr0088`

`urn:telecom:61733654310`

URN은 정보자원에 대한 영구적인 이름체계로서 소장 위치에 상관없이 사용할 수 있도록 설계된 것이다. 따라서 실제로 정보자원을 검색하고자 할 때에는 해당 URL에 대한 정보를 가진 데이터 구조를 필요로 하게 된다. 이를 위해 URC가 개발되었고, URC는 정보자원에 대해 기술하는 메타데이터 형식으로서 데이터베이스로 저장된다. 이들 URN, URC 그리고 URL의 관계는 <그림 2>와 같이 나타낼 수 있다.



<그림 2> URN, URC, URL의 연결

정보자원의 탐색시 URN이 브라우저에서 제공되면 URN의 이름체계를 분석·변환하는 URN Resolver를 통해 URC가 수록되어 있는 데이터베이스를 검색하여 해당되는 URC를 브라우저로 제공하게 된다. 브라우저는 이 URC에 포함되어 있는 URL을 이용하여 서버를 탐색하게 되며 그 결과를 이용자에게 제공한다.

## ■ URC

URC(Uniform Resource Characteristics)는 정보자원에 대한 정보를 기술한 것으로서 원정보에 나타나지 않은 정보를 수록하므로 일종의 메타데이터라 할 수 있다. 이때 기술되는 메타데이터에는 식별기호로서 URI 뿐만 아니라 정보원에 대한 정보로서 서명, 저자, 주제 등도 포함될 수 있다. 이외에도 URC는 그들 자신에 대한 메타데이터를 필요로 하게 되는데, 이러한 메타데이터 집합의 예는 다음과 같다.

Title: *The User Interface of URNs and URCs*

Author: *Renato Iannella*

Subject: *Uniform Resource Names and Characteristics,...*

Identifier: <http://www.dstc.edu.au/RDU/reports/AUUG96WWW/index.html>

urn: *inet:dstc.edu.au:tr0088*

URC-Type: *DublinCore*

URC-Date-Created: *199501011200Z*

URC-Date-Modified: *199501011230Z*

URC-Created-By: *Mary Smith (smith@dstc.edu.au)*

기술 예에서 보는 바와 같이 URC는 구문 형식으로서 속성-값쌍(attribute-value pair)의 형식을 취하거나 SGML 형식을 취하게 된다. 이러한 메타데이터 집합 가운데 URC에 대한 정보를 알려주는 메타데이터는 클라이언트가 URC를 파싱하여, 사용된 메타데이터 집합을 확인하고, 이와 관련된 속성-값쌍을 사용할 수 있도록 한다. SGML의 형식을 취한 기술 예는 다음과 같다.

`ftp://ftp.math.gatech.edu/pub/zsh`

*The Z-shell, a command interpreter for many UNIX systems which is freely available to anyone with FTP access. Zsh is more powerful than every other common shell (sh, ksh, csh, tcsh and bash) put together. The maintainer is Richard Coleman, zsh@math.gatech.edu*

=====

`ftp://ftp.sterling.com/zsh`

A mirror site in the US

```

<urc>
<urn>urn: x-dns-2:shells.unix.computing.subjects.int:zsh
  </urn>
<author>Coleman, Richard</author>
<author type="email">zsh@math.gatech.edu</author>
<title>The Z-shell</title>
<subject scheme="abstract">
A command interpreter for many UNIX systems
which is freely available to anyone with FTP access. Zsh is
more powerful than every other common shell (sh, ksh,
csh, tcsh and bash) put together.
</subject>
<instance>
<coverage>Canonical distribution site</coverage>
<url>ftp://ftp.math.gatech.edu/pub/zsh</url>
</instance>

```

URC를 사용하는 또하나의 장점은 웹 색인기(예, Lycos, Altavista)가 전체 문서 자체가 아닌 메타데이터만을 색인함으로써 색인된 데이터의 품질을 높이고 자원 탐색 도구를 통한 정보검색의 효율을 높여준다는 것이다.

URC가 다루는 정보자원은 매우 다양하기 때문에 하나의 기술 요소집합이 이들 모든 형식의 자원을 기술하는데 적합하기란 불가능하다. 따라서 URC는 URC 하부타입(Subtype)을 규정하여 서로 다른 기술 스키마를 생성할 수 있도록 하고 있다. URC 하부타입은 Dublin Core나 IAFA 템플릿의 요소를 기반으로 하고 있으며, Dublin Core에 기반한 하부타입에서 지정한 필수요소는 다음과 같다.

- URC : 그룹화된 구조
- URL : URC로 기술하고 있는 정보자원의 판(버전)을 나타내기 위해 URL을 사용한다.
- Location : 실제 정보자원의 위치를 나타내는 문자열로서 URL과 같은 특정 위치 표현 메커니즘을 식별할 수 있는 체계의 속성을 제공한다.
- URN : URC 요소가 기술하고 있는 정보자원의 식별기호

이다. <URL>은 <URC>요소에 기술되어 있는 정보자원의 위치를 명시하는데 사용되며, <location scheme="URL">은 관련 정보자원의 위치를 명세하는데 사용된다.

- Identifier : 정보자원에 대한 유일한 식별기호이다. <URN>은 URC 요소에 기술되어 있는 정보원을 참조하는데 사용되며, <location scheme="URN">은 다른 정보자원을 참조할 때 사용된다.
- Format : 정보자원을 데이터 형식을 기술하는데 사용된다. 초기값은 MIME Content-type으로 알려져 있는 IMT(Internet Media Type)이다.

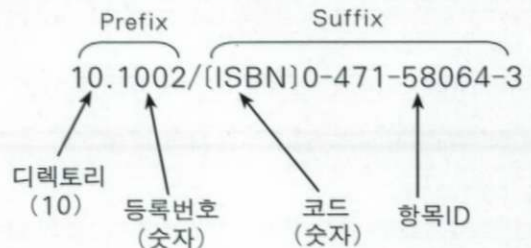
이밖에 추가적인 요소로는 저자(Author), 서명(Title), 주제(Subject), 발행자(Publisher), Date(발행일), 다른 기관/인물(Other Agent), 객체 타입(Object Type), 형식(Form), 관계(Relation), 출처(Source), 언어(Language), 범위(Coverage), URC-Creation (URC 생성일), URC-Source(URC 정보원), URC-Change (URC 수정일) 등이 포함된다.

### 관련 연구

URx는 현재 인터넷상의 정보자원의 식별과 검색을 위한 기본 수단이 되고 있고 이를 응용한 식별기호 체계가 상당수 연구되고 있고 실제로 통용되고 있다. 대표적으로 URN의 원리를 응용한 식별기호 체계인 DOI 시스템과 URL과 관련된 PURL을 들 수 있다.

### ■ DOI

DOI(Digital Object Identifier) 시스템은 미국출판자



<그림 3> DOI의 구성

협회(AAP)에서 처음 제안한 것으로서 디지털 저작물의 식별 체계로 개발된 것이다. 현재는 국제 DOI 재단에 의해 운영되고 있다. DOI는 디지털 저작물의 새로운 식별기호로서 출판계를 비롯한 인터넷 정보자원 관리 영역에서 큰 주목을 받고 있다.

DOI 시스템은 크게 식별기호, 디렉토리, 그리고 데이터베이스의 세부분으로 구성되어 있다. 식별기호는 접두사(prefix)와 접미사(suffix)로 이루어진다. 접두사는 디렉토리 관리자에 의해 부여되며, 디렉토리 관리자가 부여한 '10'의 접두어를 사용하며 DOI를 등록하려는 출판사를 지시하는 등록번호로 구성된다. 접미사는 출판사가 부여하는 개별 자원(항목)에 대한 고유 식별기호 체계로서 흔히 SICI(Serial Item and Contribution Identifier)나 PII(Publisher Item Identifier)와 같은 기존 식별기호 체계 표준을 따른다. DOI의 식별기호 구성은 <그림 3>과 같다.

DOI의 디렉토리는 출판사에 의해 유지되는 모든 DOI 번호와 서버의 주소를 관리하여 이용자와 자원 소유자간을 연결해 주는 일종의 매개체로서의 역할을 수행한다. 그리고 출판사 혹은 자원 소유자는 실제 정보자원의 내용을 담고 있는 데이터베이스를 보유하여 실제 정보자원의 요청시 데이터베이스에서 정보자원을 가져온다.

#### ■ PURL

PURL(Persistent Uniform Resource Locator)은 OCLC에서 개발한 식별기호 체계로서 URL의 단점을 보완하고자 URN이 제안되는 과정에서 탄생한 것이다. PURL의 구조는 URL과 유사하며 크게 프로토콜, Resolver 주소, 자원 이름으로 구성된다.

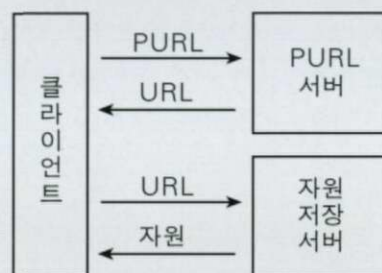
Http://purl.oclc.org/OCLC/PURL/FAQ

프로토콜      Resolver주소      자원 이름

<그림 4> PURL의 구조

PURL의 가장 큰 특징은 기능적으로는 URL과 같지만

인터넷 정보자원의 소장 위치를 직접적으로 가리키는 대신 매개 역할을 하는 변환 서비스를 가리킨다는 것이다. 즉, PURL 변환 서비스는 실제 URL과 PURL을 연관시켜 클라이언트에게 URL을 전달해 준다. 이러한 PURL 시스템의 구조는 다음 <그림 5>와 같이 나타낼 수 있다. Resolver는 PURL과 고유의 URL을 연결시켜 주고, 관리 기능은 PURLs의 생성과 연관된 URLs의 수정을 용이하게 해 준다.



<그림 5> PURL 시스템 구조

#### 결론

URx는 급증하는 인터넷 정보자원의 식별과 검색을 위한 도구로서 매우 유용하게 사용되고 있다. 특히, URL과 같이 단순한 위치 정보만으로 식별기호로서의 역할을 충분히 소화하지 못한 한계 상황에서 URN과 URC와 같은 체계의 개발은 수시로 사라지고 변경되는 인터넷 자원의 속성이 갖는 문제점을 최대한 극복할 수 있도록 하였다.

또한 이러한 체계 개발을 기반으로 한 관련 분야의 연구 확대와 이용 증가는 URx의 중요성을 다시 한번 강조하는 계기가 되고 있다. 이제는 국내에서도 이와 관련하여 URN 이름부여 체계나 URC 메타데이터 형식 개발 등이 시급히 수행되어 전세계의 인터넷 환경 추세에 부응해 나가야 할 것이다. ☞