

온라인 정보 자원을 위한 식별 체계 : URN, DOI, CID

이승재 정하경 최양희 권태경 주용완
서울대학교 컴퓨터공학부 한국인터넷진흥원
{sjlee, hkjung, yhchoi, tk}@mmlab.snu.ac.kr ywju@nida.or.kr

A Survey on Identifiers for Digital Content : URN, DOI, CID

Seungjae Hakyung Yanghee Taekyoung Yongwan
Lee Jung Choi Kwon Ju
School of Computer Science and Engineering National Internet
Seoul National University, Seoul, Korea Development
Agency of Korea
{sjlee, hkjung, yhchoi, tk}@mmlab.snu.ac.kr ywju@nida.or.kr

요 약

최근 인터넷 전자 상거래가 활발해지면서 온라인 정보 자원을 식별하기 위해 여러 기법들이 국내외에서 제시되었다. 이들 중 IETF에 의한 URN, International DOI Foundation(IDF)에 의한 DOI, 일본의 Content ID Forum(cIDf)에 의한 Content ID(CID)가 가장 주목받고 있다. URN과 DOI는 정보 자원을 위한 디지털 식별자로서, 자원의 내용과 위치 변화에 상관없이 자원을 식별하기 위한 기법이다. CID는 워터마킹 방법을 통해 디지털 콘텐츠에 삽입되는 코드로서, 콘텐츠 관리를 위한 메타데이터를 포함하는 방식이다. 이러한 기법들의 기본적인 목표는 유사하나 이들의 구체적인 기능에서 뚜렷한 차이를 보인다. 따라서, 본 논문에서는 URN, DOI, CID에 대하여 개별적으로 소개한 후, 확장성과 DRM(Digital Right Management)에 관한 고려, 기존 식별 체계와의 호환성 등의 관점에서 비교분석하고 각각의 장단점을 평가한다.

1. 도입

인터넷의 대중화로 인해 비디오, 이미지, 음악, 서적과 같은 디지털 콘텐츠들의 네트워크를 통한 유통이 촉진되고 있다. 인터넷 환경에서 URL로 대변되는 기존의 디지털 콘텐츠 식별체계는 식별자가 해당 콘텐츠의 물리적 위치를 표시하는 그 특성상 전자상거래 등 잦은 물리적 주소의 변경이 발생하는 상황에서는 적용하기 어렵다는 한계점을 지니고 있다. 따라서 이러한 점을 해결한 식별체계로서 여러 기법들이 국내외에서 제시되었다.

이러한 기법들 가운데 IETF에 의해 표준화된 URN(Uniform Resource Name), IDF(International DOI

Foundation)가 제안한 DOI(Digital Object Identifier), 일본의 cIDf(Content ID Forum)가 주도하고 있는 CID(Content ID)가 가장 대표적으로 주목받는 식별체계이다. URN과 DOI는 정보 자원을 위한 디지털 식별자로서, 자원의 내용과 위치 변화에 상관없이 자원을 식별하기 위한 기법이다. CID는 워터마킹 방법을 통해 디지털 콘텐츠에 삽입되는 코드로서, 콘텐츠 관리를 위한 메타데이터를 포함하는 방식이다. 이러한 기법들의 기본적인 목표는 유사하나 이들의 구체적인 기능에서 뚜렷한 차이를 보인다. 따라서 본 논문에서는 URN, DOI, CID에 대하여 개별적으로 소개한 후, 확장성과 DRM(Digital Right Management)에 관한 고려, 기존 식별 체계와의 호환성 등의 관점에서 서로 비교

분석하고 각각의 장단점을 평가하고자 한다.

2. Uniform Resource Name

2.1. URN의 목적

일반적으로 인터넷상의 정보 자원을 나타내는 데에는 URL(Uniform Resource Location)이 많이 쓰여 왔다. 이것은 정보 자원이 인터넷상에 존재하는 위치를 나타내기 때문에, 정보 자원이 삭제되거나 저장된 위치가 변경되면 기존의 URL로 정보 자원에 접근할 수 없게 되는 문제점을 갖고 있다. 또한, 한 자원이 여러 곳에 위치하는 경우나 자원에 대한 메타데이터를 체계적으로 관리할 필요가 있는 경우에도 URL은 적절한 해결책이 되지 못한다. 이러한 문제들을 해결하기 위해 정보 자원에 고유하고 영구적인 이름을 붙이는 URN 이름체계가 IETF에 의해 표준화되었다.

2.2. URN의 문법

```
<urn> := "urn:" <NID> ":" <NSS>
<NID> = 명명기관의 이름
<NSS> = 각 명명기관에서 정의하는 형식의 문자열
(예: 'urn:ietf:rfc:3187'
      'urn:isbn:0451450523')
```

[그림 1] URN의 문법

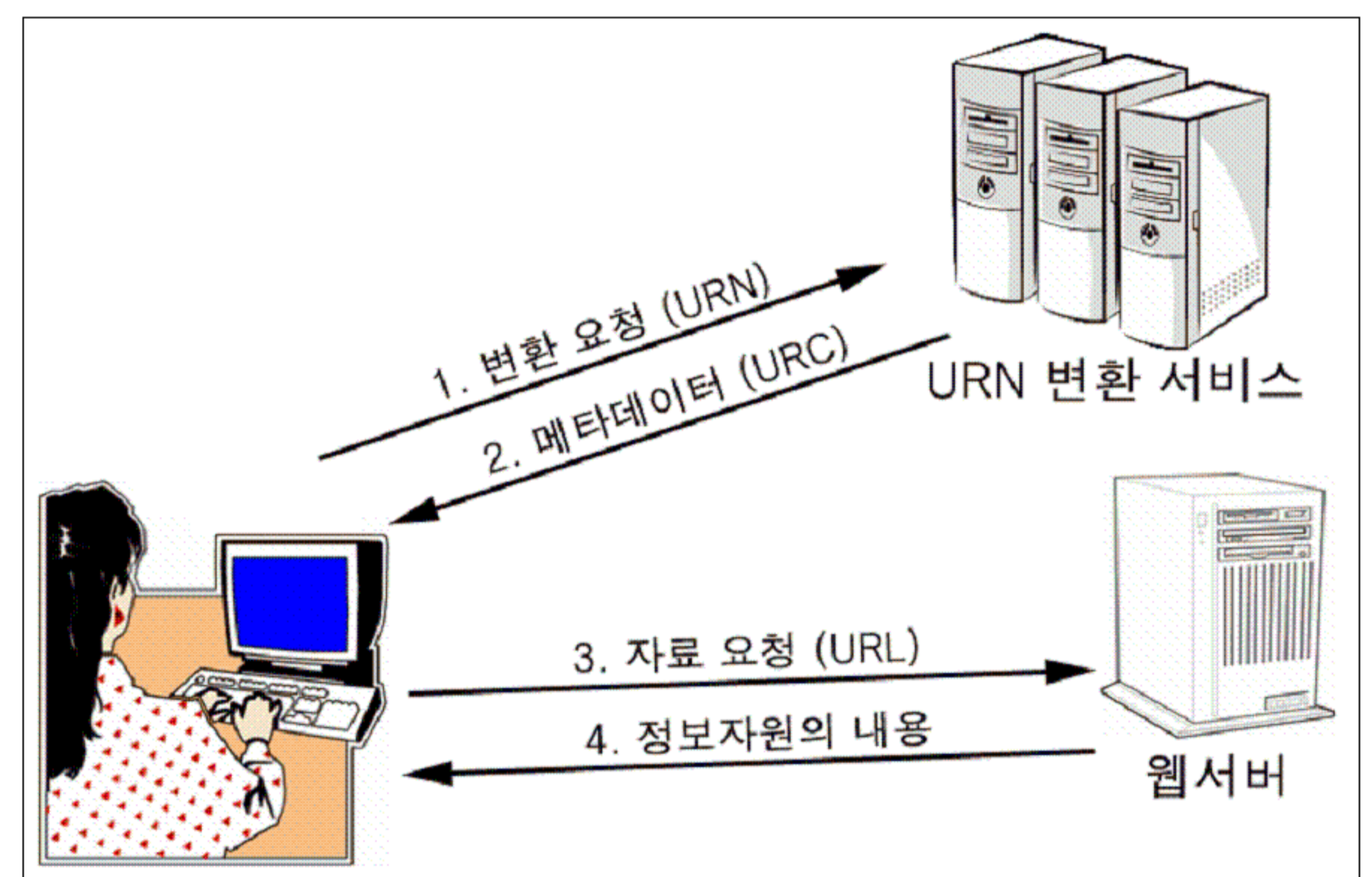
URN은 해당 식별자가 URN임을 나타내는 'urn:' 문자열과 NID(Namespace Identifier), NSS(Namespace Specific String)의 세 부분으로 구분할 수 있다. NID는 명명기관의 이름이다. 각 명명기관은 자신의 NID를 등록함으로써 URN 이름 공간의 일부를 할당받고, 자신이 관리하는 정보 자원들에 대해 NSS를 명명할 권한을 가진다. NSS는 사람이 읽고 쓸 수 있는 문자로 구성되어야 한다는 규칙 이외에 형식에 대한 다른 제한은 없다.

2.3. URN의 특징

- **전 세계적인 유일성**: 하나의 URN은 세계 어디에서나 동일한 자원을 가리키는 의미로 쓰인다. NID가 각 명명기관에 대해 IANA에 전 세계적으로 유일하게 등록되고, 각 명명기관은 자신이 관리하는 이름공간에 대해 유일성을 제공함으로써 모든 URN은 전 세계적으로 유일함을 보장받는다.
- **영구성(Persistency)**: URN은 그것을 명명하고 관리하는 기관의 수명이나 대응되는 정보 자원의 수명에 관계없

이 영구적으로 유지된다. 즉, 한번 한 자원에 할당된 식별자는 더 이상 그 자원이 존재하지 않게 되더라도 다른 자원을 식별하기 위한 용도로 재할당되는 일이 없다.

- **양적 확장성(Scalability)**: URN은 앞으로 매우 긴 시간이 지나더라도 정보 자원을 식별하는 데 문제가 없다. 이는 URN의 길이에 제한이 없기 때문에 가능한 일인데, IPv4가 주소 공간 부족 문제로 새로운 프로토콜을 만들어야 했던 것과 대조되는 특징이다.
- **구조적 확장성(Extensibility)**: URN 스펙에서는 길이나 구조 등 NSS의 형식을 제한하지 않고 각 명명기관이 책임과 권한을 갖고 자유롭게 체계를 만들어 URN을 생성할 수 있도록 함으로써, 앞으로 URN 체계가 확장되더라도 하위호환성을 유지할 수 있게 하였다.
- **기존 식별 체계 지원**: NSS의 형식이 제한되지 않았다는 특징 때문에, 기존 식별 체계를 URN의 하위 이름공간으로 쉽게 흡수할 수 있다는 장점이 있다. 예를 들어 IETF의 인터넷 표준 문서인 RFC#3187은 'urn:ietf:rfc:3187'과 같이 쓸 수 있고, 출판물 분류 번호로 널리 쓰이고 있는 ISBN도 'urn:isbn:0451450523'과 같이 쓸 수 있다. 즉, URN의 하위 이름공간으로 편입된 기존 식별자 체계를 그에 해당하는 URN으로 변환하기 위해서는 복잡한 과정이 필요 없고 단지 기존 식별자에 URN 접두사와 각 체계에 해당하는 NID를 붙이면 된다.



[그림 2] URN 변환 메커니즘의 예

- **변환가능성(Resolvability)**: URN을 URL 등의 다른 이름공간에 있는 식별자로 변환하는 메커니즘이 있어야 한다. URN은 정보 자원의 이름일 뿐이고 자체적으로 자원의 위치 등 다른 정보를 갖고 있지 않으므로, URN으로부터 정보 자원의 내용이나 메타데이터를 얻어내기 위해서는 변환 메커니즘이 필수적이다. 이러한 변환 서비스를 RDS(Resolution Discovery Service)라고 한다. 이 서비스는 DNS에 기반을 둔 구조로 구성되어 있는 거대한 서버 네트워크를 통해 제공되며, URN을 기본키(Primary Key)로 삼고 각 URN에 해당하는 메타데이터를 돌려주는, 전

세계에 분산된 데이터베이스라고 볼 수 있다.

3. Digital Object Identifier

3.1. DOI의 목적

DOI(Digital Object Identifier)는 그 정의상 ‘디지털 네트워크의 개체를 위한 이름’이다. 새로운 식별 체계의 필요성을 느낀 미국출판협회가 온라인 정보 자원을 위한 식별 체계의 개발을 제안하여 DOI의 개발이 이루어졌고, DOI를 관리하기 위한 기관인 IDF(International DOI Foundation)를 만들어서 오늘날에 이르고 있다.

2.2. DOI의 문법

```
<doi> := "doi:" <prefix> "/" <suffix>
<prefix> = "10." 등록기관의 번호
<suffix> = 각 등록기관에서 정하는 형식의 문자열
(예: '10.1000/123abc')
```

[그림 3] DOI의 문법

DOI의 구조는 겉모습만 다를 뿐 실질적인 구조는 URN과 사실상 동일하다. ‘/’를 기준으로 하여 앞쪽에는 등록기관(URN의 명명기관과 같은 역할을 한다)의 번호가, 뒤쪽에는 각 등록기관에서 정하는 형식의 문자열이 들어간다. ‘/’ 뒷부분은 등록기관이 자유롭게 형식을 정할 수 있다.

2.3. DOI의 특징

DOI는 URN과 거의 동일한 개념으로 만들어진 것이기 때문에 식별자 문법 구조도 거의 비슷하고 그 특징에 있어서도 유사한 모습을 보인다. 예를 들어, 각 등록기관에서 자유롭게 식별자 접미사의 형식을 정할 수 있게 했기 때문에 ISBN 등의 기존 식별자 체계를 변경 없이 그대로 흡수하거나 새로운 구조를 제안하여 사용하는 데 적합하다는 점은 URN에서 동일하게 나타나는 특성이다.

차이점도 존재한다. URN에서는 명명기관의 이름을 한눈에 알아볼 수 있는 몇 개의 영문자로 정하고 있는 반면에, DOI에서는 등록기관을 나타내기 위한 형식은 숫자여야 하며 숫자를 통해 어떠한 뜻도 유추할 수 없어야 한다고 규정하고 있다. 이는 등록기관이 소멸하더라도 그 기관이 관리하던 DOI를 번호의 변경 없이 다른 등록기관에서 관리하기 용이하게 하기 위함이다.

또한, 필요할 경우에는 식별하고자 하는 개체의 일부분

에 대해서도 식별할 수 있을 것을 기능적 요구사항으로 두고 있다. 예를 들어, 하나의 논문집, 그 안의 논문, 그 안의 본문, 그 안의 테이블을 모두 별개의 DOI로 식별하는 것이 가능하다[2].

2.4. URN과 DOI

URN과 DOI는 만들어진 목적은 같으나 상호 호환되지 않기 때문에¹⁾ 서로 대립되는 표준이라 볼 수 있다. 이들은 표준화를 추진한 기관에서부터 분명히 대비되며, 현재도 서로 다른 길을 걷고 있다.

URN은 IETF에 의한 인터넷 표준이다. IETF가 1997년의 URN 문법에 대한 RFC를 시작으로 하여 2002년에 urn 관련 표준화 활동을 완료한 뒤로, 실제로 URN을 사용할 수 있는 변환 서비스의 구현 등 URN을 전 세계적으로 널리 보급하는 일을 주도하고 있는 단체는 없는 실정이다.

DOI는 최초에 산업계(미국출판협회)에서 그 필요성을 제기하였고 IDF에 의해 활발하게 추진되어 미국 표준이 되었고, 현재는 ISO 표준으로 채택되었다(ISO TC46/SC9). DOI 식별자 변환은 현재 웹을 통하여 바로 수행해볼 수 있는데, 이는 공개된 변환 서비스가 없는 URN과는 대조적이다. 그런데 IDF는 URN을 강하게 의식하고 있는 것으로 보인다. URN의 변환 체계가 아직 널리 보급되어 있지 않고, DOI가 URN의 하위 공간으로 등록될 경우 불필요한 관리계층이 추가된다는 점 때문에 DOI를 공식적인 URN으로 등록하지 않았다고 말하고 있다[6]. 또한 인터넷 표준 초안으로 ‘The “doi” URI Scheme for the DOI’[4]를 제출하여 DOI가 URN과 동일한 지위를 얻고자 시도하였으나 이것이 RFC로 확정되지는 않았다.

URN이 국제적인 인터넷 표준이기는 하나 미국 산업계가 DOI를 주도하고 있다는 점에서, 이변이 없는 한 미국의 온라인 정보 자원 식별자는 앞으로 DOI가 실질적인 표준으로 사용될 것으로 보인다. 이 두 가지 표준은 서로 호환성이 없지만, URN을 사용하는 시스템과 DOI를 사용하는 시스템 모두와 상호작용하기 위해서는, 하나의 등록기관이 자체적으로 식별자 형식을 정한 뒤 URN과 DOI 모두의 하부 이름공간으로 등록하는 것도 가능한 해결책이다. 예를 들어, urn:foo:1234-5678과 doi:10.573/1234-5678을 같은 기관에서 관리하며 동일한 자원으로 매핑하는 것이다.

국내에서는 한국전산원이 URN 기반의 식별 체계인 UCI(Universal Content Identifier)를 국가사업으로 진행하고

1) 두 식별 체계의 특성상 URN이 DOI 하로, 혹은 DOI가 URN 하로 편입되는 것은 가능하다. 그러나 두 체계의 이름공간은 서로 별개의 기관이 관리하므로 두 기관의 긴밀한 협의가 없으면 동등한 위치에서 호환성을 유지하는 것은 불가능하다.

있다. 현재 개발은 완료된 단계이고 시범사업을 수행하고 있다.

4. Content ID

cIDf(Content ID Forum)는 저작권을 보호하면서 콘텐츠 유통을 촉진할 수 있는 프레임워크를 만들기 위해 1999년 8월 동경대와 NTT를 비롯한 7개 회사에서 조직한 기관이며, 이 결과로 2003년 8월 관련 기술문서인 cIDf Specification 2.0이 발표되었다.

4.1. Content ID의 정의와 속성

cIDf는 디지털 콘텐츠의 유통을 위한 고유 식별체계로서 유일한 식별자인 Content ID(CID)를 정의하였다. CID는 워터마킹과 같은 특수 기술을 이용하여 각각의 디지털 콘텐츠에 삽입하기 위한 고유한 저작권 코드일 뿐 아니라, 콘텐츠를 관리할 수 있는 완전한 메타데이터(속성)를 포함하고 있다. 이를 통해 디지털 콘텐츠의 저작권을 명확히 하여 저작자는 콘텐츠의 판매와 재이용에 따른 합당한 보상을 받고 콘텐츠의 유통 기록을 유지하기 위한 바코드 시스템을 운영하기 위해 구상된 개념이다.

CID는 저작자의 신청에 따라 부여되는 고유번호와 내용 개요, 권리 번호, 로열티 배포 등을 표시하는 속성으로 구성된다. [표 1]은 CID의 포맷을 보여준다.

콘텐츠ID 센터관리 번호	콘텐츠 속성	권리 소유권 속성	권리 계약속성	배포 속성	사용료 배포 속성	자유영역
---------------------	-----------	-----------------	------------	-------	--------------	------

[표 1] CID의 포맷

콘텐츠ID센터관리번호는 콘텐츠의 패키지를 식별할 수 있는 고유번호를 나타내는 속성으로서 버전번호, 타입번호, 그룹번호, 센터번호, 센터내부관리번호로 구성된다. 콘텐츠 속성은 패키지에 포함된 디지털 콘텐츠를 기술하는 속성으로서 창작자의 이름, 제목, 초록 등과 같은 정보 등이 포함되며, 권리소유권속성은 콘텐츠 아이템의 권리 소유자(들)를 식별하는 속성이다. 콘텐츠속성과 권리소유권속성은 콘텐츠의 생성시점에 정의된다. 권리계약속성은 권리의 허락, 위임, 양도에 관한 정보를, 배포속성은 라이선스의 범위를 기술하며, 사용료배포속성은 콘텐츠의 이용으로 인한 로열티나 사용료가 누구에게 어떻게 배포되어야 하는지를 기술한다. 권리계약속성, 배포속성, 사용료배포속성은 배포 이전에 정의된다. 마지막으로 CID에는 자유속성을 두어 새로

운 어플리케이션에 대해 확장이 가능하도록 하고 있다. 여기서 주목할 점은 CID에는 정적인 속성과 동적인 속성이 모두 포함되어 있다는 것이다. 콘텐츠 유통의 촉진이라는 CID의 목적을 위해서는 유통과 배포의 조건이 다른 동일 콘텐츠의 사본에 다른 ID를 할당하는 것이 중요하며, 이러한 관점이 반영된 것으로 이해할 수 있다.

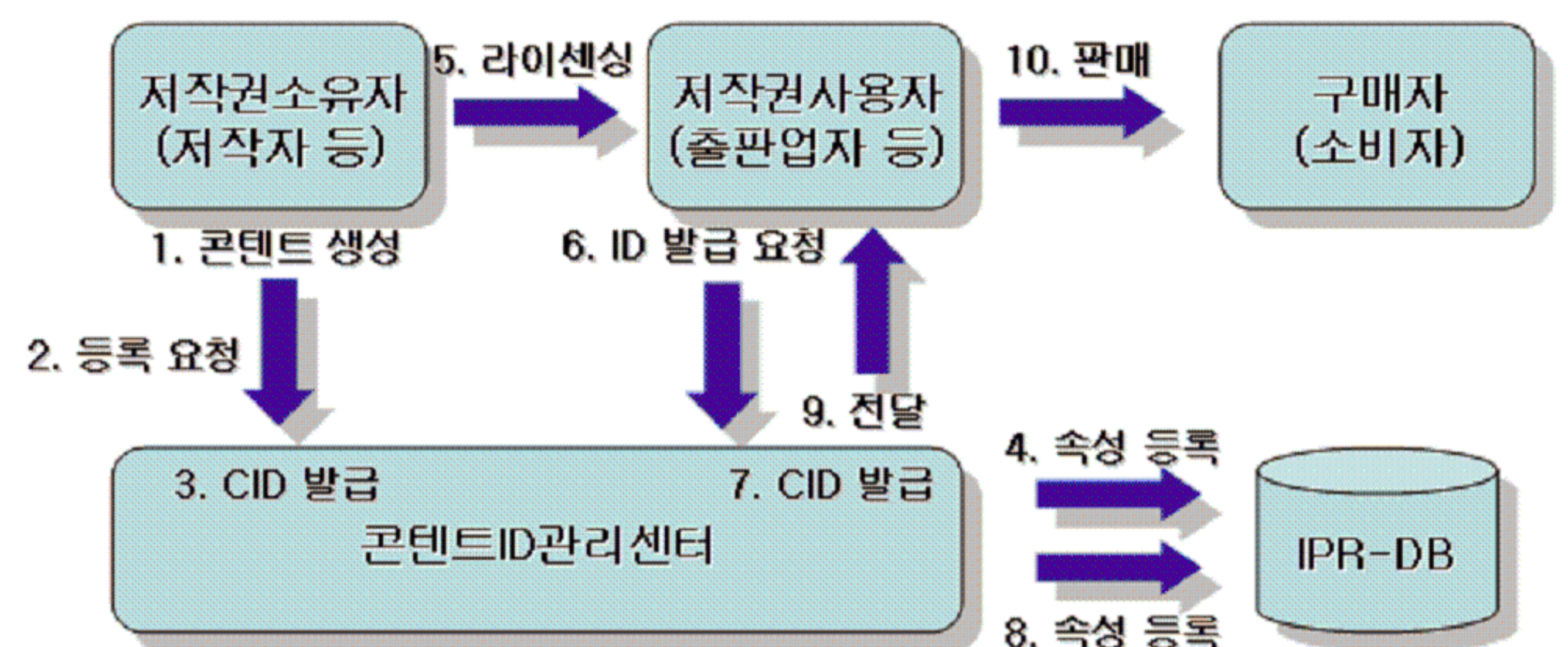
4.2. Content ID 발행 체계

콘텐츠ID관리센터는 각 콘텐츠에 CID를 부여하며, 저작권 정보를 저장하는 IPR-DB(Intellectual Property Rights Databases)에 부여한 CID를 등록하여 관리한다. IPR-DB에는 콘텐츠ID관리번호를 통해 접근 및 갱신을 할 수 있도록 되어있다. 한 지역에는 복수개의 콘텐츠ID관리센터가 존재할 수 있으나, 이들은 모두 단일 RA(Registration Authority)에 의해 관리된다. RA는 콘텐츠ID관리센터를 인가하면서 이들에 센터번호를 부여한다.

CID는 여러 센터에서 발행하지만, RA에서 관리하는 변환시스템이 각 CID를 이를 발행한 센터의 IPR-DB로 연결하는 것이 가능하기 때문에, 각 속성들은 분산 환경에서 질의될 수 있다.

4.3. Content ID 발행 흐름

그림은 CID를 활용한 콘텐츠 유통을 위한 처리과정의 한 예를 나타낸 것이다.



[그림 4] 콘텐츠 처리과정의 예

먼저, ①저작자가 콘텐츠를 생성하고, ②콘텐츠ID관리센터에 ID를 발급해줄 것을 요청한다. 센터는 ③콘텐츠에 관한 ID를 발급하고, ④IPR-DB에 속성을 등록한다. 그 후, ⑤저작권자와 저작권이용자가 저작권 이용에 관해 계약을 맺고, ⑥저작권이용자가 센터에 ID를 발급해줄 것을 요청한다. ⑦센터는 ID를 발급해주고, ⑧콘텐츠의 배포속성을 IPR-DB에 등록한다. 그 후 ⑨센터는 저작권이용자에게 워터마킹된 콘텐츠를 전달하며, ⑩저작권이용자가 콘텐츠를 소비자에게 판매한다.

5. URN, DOI, CID의 비교

	URN	DOI	CID
추진 기관	IETF	IDF	ciDf
주된 식별 대상	문서	문서	멀티미디어
DRM에 관한 고려	없음	없음	있음
개별 인스턴스 구별	불가능	불가능	가능
사람이 읽고 쓸 수 있음	가능	가능	불가능
확장성	있음	있음	제한됨
기존 식별체계와 호환성	있음	있음	없음
부분 객체 식별 고려	없음	있음	없음

[표 2] URN, DOI, CID의 특징과 장단점

URN, DOI, CID의 특징을 나열해 보면, URN과 DOI끼리는 서로 유사하고 CID와는 상당히 다름을 알 수 있다. 먼저, 식별 대상으로서 URN과 DOI는 주로 문서를 대상으로 하는 반면, CID는 워터마킹이 용이한 음악이나 비디오 같은 멀티미디어 자원을 주 대상으로 하여 식별 체계를 디자인하고 사용 시나리오를 만들었다. 물론, 각각을 다른 대상에 적용할 수 없는 것은 아니다.

CID는 URN이나 DOI와는 달리 태생적으로 디지털 콘텐츠의 유통을 촉진하면서 동시에 저작권자의 저작물을 보호하는 것을 그 목적으로 하기에, DRM(Digital Right Management)의 적용을 고려하여 설계된 것이 두드러진 차이점이다. 일례로서 고유 식별자로만 구성된 URN, DOI와 달리 CID는 고유 식별자 외에도 권리, 유통, 배포 등에 관련된 속성들을 포함하고 있으며, CID가 워터마킹 기술을 이용하여 콘텐츠에 포함되어 소비자에게 전해지게 된다. 또한, 배포조건이 동일한 종류의 콘텐츠를 식별하는 기준이 될 수 있다는 점에서 타 식별체계와 차별화된다. URN과 DOI의 경우 같은 콘텐츠의 동일한 사본이 존재하는 경우 이들은 같은 식별자가 주어지지만, CID는 동일한 사본일지라도 배포 조건(배포자, 라이선스 조건 등)이 서로 다른 콘텐츠를 상이한 콘텐츠로 취급하게 된다. 그러나 CID는 DRM의 적용을 위해 다양한 정보를 식별자 내에 포함하여 콘텐츠 자체에 워터마킹하기 때문에, 사람이 CID를 쉽게 읽고 쓸 수 없다. 반면에 URN과 DOI는 식별자를 사람이 읽고 쓸 수 있게 설계되었기 때문에, 사람이 쉽게 옮겨 적거나 웹 문서에서 하이퍼링크를 걸 수 있다.

URN과 DOI는 개별 명명기관 혹은 등록기관이 자신이 소유한 이름공간에 속한 식별자 구조의 의미를 스스로 정의할 수 있도록 완전히 위임하였기 때문에 높은 확장성을 갖는다. 또한 이런 특징을 이용하여 기존에 존재하던 식별체계도 그대로 흡수해서 이용할 수 있다는 장점이 있다.

그러나 각 기관이 식별자 구조의 의미를 정하기 때문에, 각각의 구조에 관한 사전 정보 없이 모든 URN 혹은 DOI 식별자로부터 그 의미를 해석해내는 일반적인 방법은 존재하지 않는다. 기능적인 면에서 DOI가 URN과 다른 점은, 식별하고자 하는 개체의 부분에 대해서도 개별적으로 식별할 수 있다는 점이다. 비록 URN이 현재 그러한 기능에 대한 고려가 없다고는 하나, 높은 확장성을 이용하여 URN의 하부 기관은 그에 관한 기능을 추가할 수 있을 것이다.

6. 결론

본 논문에서는 URN, DOI, CID 등을 소개하고 서로 비교해 보았다. 온라인 정보 자원을 식별하기 위한 목적으로 만들어졌다는 점에서 URN과 DOI, CID는 한 범주에 있으나 각자 뚜렷한 차이점이 존재한다. URN과 DOI는 정보 자원의 이름이라는 개념이며, 개별 사본을 구분할 수 없다는 점에서 책을 구분하는 기존 식별 체계인 ISBN에 비유할 수 있다. CID는 정보 자원의 개별 사본마다 부여되므로, 상품마다 부착하는 RFID 태그에 비유할 수 있다. 반면에 URN과 DOI는 식별자가 정보 자원 자체에 삽입되지 않기 때문에, DRM의 적용을 위해서는 디지털 워터마킹과 관련된 다른 기법의 도움을 받아야 한다.

참고문헌

- [1] RFC 2141 : URN Syntax, R. Moats, 1997.
- [2] The DOI Handbook Edition 4.2.0, International DOI Foundation, Inc.(<http://www.doi.org>), 2005.
- [3] URN Namespaces, <http://www.iana.org/assignments/urn-namespaces>
- [4] The "doi" URI Scheme for the Digital Object Identifier (DOI), <http://www.ietf.org/internet-drafts/draft-paskin-doi-uri-04.txt>, 2003.
- [5] RFC 4179 : Using Universal Content Identifier (UCI) as Uniform Resource Names (URN), S.Kang, 2005.
- [6] Factsheets, International DOI Foundation, Inc. (<http://www.doi.org/factsheets.html>), 2005.
- [7] ciDf Specification 2.0, 2003.
- [8] Hideki Sakamoto, Overview of the Content ID Forum, Content Id Forum Secretariat, 2001.
- [9] Content ID 표준화(기술) 추적보고서, 안계성, URI 포럼, 2001.
- [10] UCI Homepage, <http://www.uci.or.kr/>