

디지털 콘텐츠의 효율적 검색과 관리를 위한 UCI 식별체계의 온톨로지 적용

Applying Ontologies to UCI for the Efficient Search and Management of Digital Contents

하은옥(Eunok Ha)*, 김윤호(Yoonho Kim)**

초 록

디지털 콘텐츠 식별체계인 UCI(Universal Content Identifier)는 디지털 콘텐츠의 투명한 유통과정과 효율적 검색과 관리를 위해서 만든 URN(Uniform Resource Name)에 기반을 둔 식별체계이다. UCI 식별자를 부여받은 디지털콘텐츠는 사용자가 원하는 콘텐츠를 정확하게 전달하기 위해서는 다양한 메타데이터 정보를 필요로 한다. 그러나 UCI에서 제공하는 식별 메타데이터만으로는 콘텐츠에 대한 다양한 정보를 표현하기에는 부족하며, 정보의 보다 정확한 표현과 효율적 검색 및 관리를 위해서는 UCI에서 제공하는 메타데이터와 함께 메타데이터 내에 표현된 개념과 그 의미 관계를 정형화하고 명시적인 방법으로 정의하는 온톨로지를 필요로 한다.

본 논문에서는 UCI 식별체계의 메타데이터간 개념관계를 온톨로지로 확장하고 도메인 온톨로지를 설계함으로써 구축된 UCI 메타데이터 정보를 효율적으로 이용하여 의미 기반의 검색과 관리를 가능하게 하였으며, 다양한 질의어를 통하여 메타데이터만을 이용하는 UCI식별체계에 비하여 효율적인 검색과 관리가 가능함을 보였다.

ABSTRACT

UCI(Universal Content Identifier), a digital content identifier is a identification system based on URN(Uniform Resource Name) for the transparent distribution, efficient search and management of digital contents. Digital contents given a UCI number need metadata for serving a exact contents that users need. However, the only metadata provided by UCI are deficient to represent various information of digital contents, and ontologies which could standardize and define explicitly semantic relations between metadata (provided by UCI) are needed for better representation of information and efficient search and management of digital contents.

In this paper, we extended concept-relation between metadata with ontology and designed domain ontology enabling semantic-based search by expanding UCI meta data, and showed the more efficient search and management over the metadata by UCI through various queries.

키워드 : 디지털 콘텐츠 식별체계, UCI, 온톨로지
Digital Content Identifier, UCI, Ontology

* 상명대학교 대학원 컴퓨터과학과 박사과정

** 교신저자, 상명대학교 컴퓨터과학부 교수

2009년 06월 28일 접수, 2009년 08월 24일 심사완료 후 2009년 11월 04일 게재확정.

1. 서 론

인터넷과 정보 기술의 발전으로 급속하게 증가하는 정보에 대한 식별체계와 이를 지원하는 다양한 서비스가 요구되고 있다. 디지털 콘텐츠에 대한 식별체계는 콘텐츠의 무분별한 복제로부터 데이터를 보호할 수 있으며, 투명한 유통과정을 통해 지적 재산권을 보호할 수 있게 된다. 또한 콘텐츠의 유무를 용이하게 파악할 수 있어 자원의 재사용성을 증대시키고 불필요한 중복 생성을 방지할 수 있게 된다[1].

국내에서 사용되고 있는 디지털 콘텐츠 식별체계를 살펴보면 한국정보사회진흥원(구 전산원)의 UCI(Universal Content Identifier), 한국과학기술정보연구원의 KOI(Knowledge Object Identifier), 국립기술표준원의 SOI(Standard Object Identifier), 한국문화콘텐츠진흥원 COI(Content Object Identifier) 등이 있다. 그러나 2008년 정부조직 개편으로 구 정보통신부의 디지털 콘텐츠 관련업무가 문화부로 이관되면서 구 정보통신부의 UCI와 문화부의 COI가 통합되어 현재 국내 식별체계는 UCI로 통합되고 있는 추세이다. UCI는 2006년 6월 한국정보통신기술협회(TTA)에서 단체 표준으로 등록되었으며, 2009년 1월 한국정보통신표준(KISS)으로 정식 채택되어 정식 국가표준으로 인정받으면서 공공부분에서 민간영역으로 점차 확장되어 적용되고 있다. 이미 한국과학기술정보연구원에서도 KOI와 UCI 간의 연계 시스템을 구축하여 연계 서비스를 하고 있다[2].

UCI에서는 온라인과 오프라인 구분 없이

모든 분야의 자원에 대하여 식별체계를 부여할 수 있도록 함으로써 기존의 식별체계들과의 연계와 통합을 가능하게 하였다. 또한 구 문구조에서는 표현되지 않지만, 식별 메타데이터를 제공함으로써 정보의 검색 시 이용자는 메타데이터 기반의 검색방식을 활용할 수 있게 되어 원하는 자원을 쉽고 정확하게 찾을 수 있다. 메타데이터의 사용은 정보의 검색뿐 아니라 관리, 운영에 있어서도 큰 영향을 미친다.

디지털 콘텐츠를 검색하고 서비스하는데 있어서 사용자가 원하는 자료를 보다 정확하게 전달하기 위해서는 콘텐츠에 대한 정확하고 다양한 메타데이터 정보를 필요로 한다. 그러나 UCI에서 제공하는 식별 메타데이터만으로는 콘텐츠에 대한 다양한 정보를 표현하기에는 부족하다. 또한 등록기관별로 존재하는 응용메타데이터의 개별 관리로 인해서 정보의 공유가 어렵고, 중복 투자되는 비용이 발생하게 된다.

정보의 보다 정확한 표현과 검색 및 관리를 위해서는 메타데이터와 함께 메타데이터 내에 표현된 개념과 그 의미 관계를 정형화하고 명시적인 방법으로 정의하는 온톨로지를 필요로 한다.

디지털 콘텐츠에 대한 정보를 온톨로지 기반으로 구축한다면, 작성한 메타데이터를 기반으로 온톨로지 언어인 OWL(Web Ontology Language)[3]이 제공하는 공리를 사용하여 추론함으로써 더 풍부한 메타데이터를 생성할 수 있게 되며, 이를 바탕으로 사용자가 원하는 검색 결과를 가져올 수 있다. 또한 등록기관, 시스템간 의미정보의 전달과 정보의 공

유가 가능해짐으로써 중복된 콘텐츠 관리비용의 감소효과를 가져온다.

그러나 기존의 UCI 식별체계 관련 연구들은 UCI의 적용과 유통 모델, 연계 시스템에 관련된 연구가 주를 이룬다. UCI의 메타데이터에 관한 확장이나 온톨로지 적용사례에 대한 연구가 아직 논의되고 있지 못하다.

따라서 본 논문에서는 디지털 콘텐츠에 대한 UCI 식별 메타데이터를 온톨로지로 확장하여 메타데이터의 의미정보, 메타데이터간의 개념관계를 나타내고, 존재하는 메타데이터를 기반으로 새로운 정보들을 추론함으로써 콘텐츠에 대한 정확한 정보검색을 가능하게 하며, UCI 등록기관간, 시스템간의 정보공유를 가능하게 하고자 한다.

이를 위해 먼저 본 논문에서는 온톨로지를 구축할 도메인을 선정하였다. 현재 디지털 콘텐츠의 실 거래가 가장 활성화되고 있는 곳이 음악 산업이라고 할 수 있어 도메인은 음악 온톨로지를 선정하였다. 선정한 도메인의 온톨로지는 Protégé3.4를 사용하여 구축하였으며, 추론을 수행하였다. 또한 온톨로지 확장시의 검색의 정확성과 효율성을 평가하기 위해 질의문을 작성하고, 작성된 질의문의 검색결과를 Protégé3.4의 화면으로 제시하였다.

본 논문의 구성은 제 1장은 서론을 나타냈으며, 제 2장에서는 UCI 식별체계를 온톨로지를 적용하여 확장하였고, 제 3장에서는 도메인 온톨로지인 음악 온톨로지를 구축하였다. 제 4장에서는 구축한 온톨로지를 기반으로 OWL 공리를 사용하여 추론이 이루어짐을 보이고, 추론된 내용을 바탕으로 질의를 선정하여 검색 결과를 나타냈다. 제 5장에서 결론을 도출하고 향후 연구 방향을 모색하였다.

2. UCI 식별체계의 확장

2.1 UCI 메타데이터

메타데이터는 정보 자원의 속성을 기술하는 데이터로서 흔히 데이터에 대한 데이터(data about data)로 정의할 수 있다. 이러한 메타데이터의 사용 목적은 크게 두 가지로 나눌 수 있다. 첫 번째는 데이터 자체를 표현하기 위한 목적이고 두 번째는 데이터를 검색하기 위한 목적이다. UCI 식별체계에서도 콘텐츠의 관리와 검색을 위해 식별 메타데이터를 제공한다. UCI 식별 메타데이터 요소는 <표 1>에 제시된 바와 같이 요소에 대한 container 1개를 포함하여 총 9개로 구성된다.

메타데이터를 구축하는 기본 목적은 정보의 소재를 파악하고, 정보를 해석하며, 데이터를 통합하기 위한 것이다. 그러나 메타데이터를 구축하는 것만으로는 전통적인 데이터베이스와 다를 것이 없으며 그 서비스는 한계를 가질 수밖에 없다[4]. 콘텐츠들의 메타데이터에 온톨로지를 적용함으로써 각각의 의미관계를 기반으로 의미의 확장 및 추론을 통하여 보다 정확한 서비스를 제공할 수 있게 된다.

2.2 UCI 메타데이터와 온톨로지

UCI 식별자는 온라인 자원과 실물 자원 즉 오프라인 자원에 대하여 식별 번호를 부여할 수 있다. 그러나 본 논문에서는 온라인에서 존재하는 디지털 콘텐츠에 대하여만 논의하기로 한다. 온라인상에서 존재하는 자원은 UCI 식별 번호가 부여되며 UCI 메타데이터로 표

〈표 1〉 UCI 식별 메타데이터 요소

	용어명		정의	빈도수	인코딩 스킴
	영문	한글			
1	UCI	UCI	한국정보사회진흥원에서 부여하는 고유한 자원 식별자	1..1	구문구조
2	identifier	기존식별자	UCI 이외 자원에 부여된 기존의 식별자	0..n	ISBN외 다수
3	title	자원명	알려져 있는 자원의 이름	1..n	
4	type	유형	자원의 주된 유형	1..1	uciType
5	mode	표현형태	디지털자원의 주된 감각적 혹은 지각적 표현 형태	0..1	uciMode
6	format	파일형식	디지털자원의 데이터 표현형식	0..1	uciFormat
7	contribution	기여	자원의 내용에 주된 책임을 가진 주체와 역할의 의미를 포괄하는 요소 ※ 기여자와 기여자역할의 상위요소로 container	1..n	
8	contributor	기여자	자원의 내용에 주된 책임을 가진 자의 이름 ※ contribution의 하위 요소	1..1	
9	contributorRole	기여자 역할	자원의 내용에 주된 책임을 가진 자의 역할 ※ contribution의 하위 요소	1..1	uciRole

주) 단, type이 디지털일 경우에는 mode와 format은 필수로 사용해야 함.

현되는 부가정보를 가진다. 해당 자원은 식별 번호와 부가 정보를 통해 사용자에게 서비스 된다. 서비스 될 때, 사용자는 메타데이터에서 제공하는 항목에 대해서만 검색이 가능하다. 본 논문은 UCI 메타데이터를 온톨로지 개념을 사용하여 확장함으로써 보다 다양하고 정확한 정보를 제공하고자 한다.

이 장에서는 UCI 메타데이터에 대응하는 온

톨로지를 구축한다. 메타데이터 항목 중에 ContributorRole의 허용 값에 해당하는 용어(가수, 각색자, 감독, 개발사 등)는 모두 클래스로 구축한다. 이 클래스들은 음악 온톨로지의 행위자 클래스의 하위 클래스와 같은 객체를 가리키는 경우 owl:sameAs를 사용하여 동치임을 나타낸다. <표 2>는 UCI 메타데이터 항목에 해당하는 속성을 나타낸 표이다.

〈표 2〉 UCI 메타데이터의 온톨로지 속성

UCI 메타데이터	속성명	정의역	치역
UCI	UCI	owl:Class	rdfs:Literal
Identifier	기존식별자	owl:Class	rdfs:Literal
Title	자원명	owl:Class	rdfs:Literal
Type	유형	owl:Class	rdfs:Literal
Mode	표현형태	owl:Class	rdfs:Literal
Format	파일형식	owl:Class	rdfs:Literal
Contribution	기여	owl:Class	owl:Class

속성은 오브젝트 타입(owl:ObjectProperty)의 속성과 데이터 타입이 속성(DatatypeProperty)으로 구성된다.

속성 중에서 데이터 값을 가지는 데이터 타입의 속성들을 RDF 추상 구문을 사용하여 표기하면 다음과 같다.

```
DatatypeProperty(uci:Type range(oneOf
    ("Digital" "Physical")))
DatatypeProperty(uci:Mode range(oneOf
```

("Visual" "Audio" "AudioVisual"))

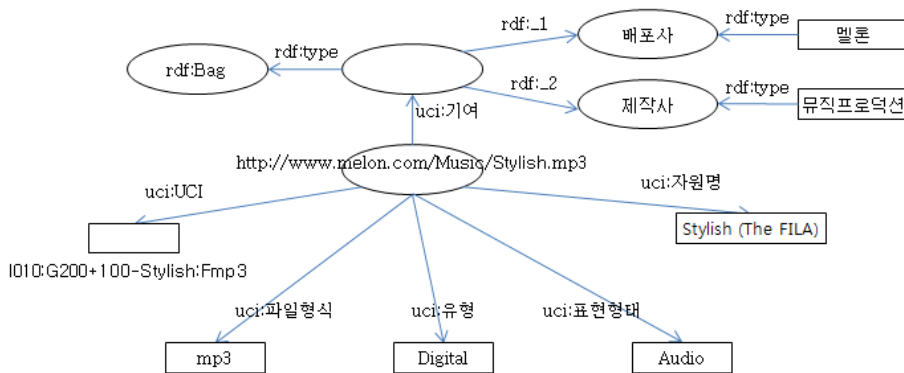
```
DatatypeProperty(uci:Format range(oneOf
    ("aif" "aiff" "asf" "asp" "au" ... "zip")))
```

예를 들어, Type은 Digital과 Physical이라는 문자열을 속성 값으로 가지며, 이를 OWL을 사용하여 표기한 것이 <표 3>이다.

<그림 1>은 위에서 작성한 UCI 메타데이터 온톨로지의 클래스와 속성을 사용하여 Stylish.mp3에 대한 정보를 그래프로 표현한

<표 3> OWL로 작성한 Type 속성

```
<owl:DatatypeProperty rdf:ID = "Type">
  <rdfs:range>
    <owl:DataRange>
      <owl:oneOf>
        <rdf:List>
          <rdf:first> Digital </rdf:first>
          <rdf:rest>
            <rdf:List>
              <rdf:first> Physical </rdf:first>
              <rdf:rest rdf:resource = "&rdf:nil"/>
            </rdf:List>
          </rdf:rest>
        </rdf:List>
      </owl:oneOf>
    </owl:DataRange>
  </rdfs:range>
</owl:DatatypeProperty>
```



<그림 1> 그래프 예제

예이다.

3. 온톨로지 구축

3.1 음악 온톨로지 구축

UCI는 다양한 분야와 형식의 콘텐츠에 부여할 수 있기 때문에 어떠한 자원에도 적용이 가능하다. 본 논문에서는 UCI 메타데이터의 확장을 위해 음악이라는 특정 도메인을 선정하여 온톨로지를 구축하였다.

현재 디지털 콘텐츠의 실 거래가 가장 활성화 되고 있는 곳이 음악 산업이라 할 수 있다. 국내 디지털음악 시장은 2003년 이후 연평균 성장률 8.1%를 기록하며 2008년 3,427억원의 시장 규모를 형성한 것으로 조사됐다. 이는 2007년에 비해 8.1%증가한 결과이다[5]. 음악 시장은 과거 음반 산업 중심의 구조에서 디지털화, 압축 기술, 네트워크 기술 등의 IT 관련 기술이 음악 상품에 접목되면서 오프라인 음반시장에서 디지털 음원 시장으로 패러다임의 변화가 이루어지고 있다[6].

이처럼 디지털 콘텐츠의 유통이 활발히 이루어지는 음악 분야에 온톨로지를 적용함으로써 디지털 콘텐츠 산업에 새로운 서비스와 이익을 창출하게 되며, UCI 보급에도 영향을 미칠 수 있을 것이라 생각한다.

반면, 음악 콘텐츠는 가장 활발하게 콘텐츠 거래가 일어나는 부분이지만 불법 복제와 불법 유통이 제일 심한 콘텐츠이기도 하다. 따라서 음악 온톨로지의 구축을 통해 다양한 저작자와 저작 정보를 제공한다면 저작권보호에도 기여할 수 있다고 생각되며 저작권과

온톨로지와의 관계를 연구하는데도 도움이 될 것이라 생각된다.

본 논문에서는 웹 온톨로지 언어 중 OWL을 사용하여 온톨로지를 구축하였다. OWL은 기계 또는 에이전트가 처리할 수 있는 풍부한 어휘와 형식적 의미를 포함하고 있어서 클래스간의 관계(교집합, 합집합), 동치성(Equality), 비동치성, 차수(Cardinality), 속성간의 역관계 등의 세부 관계를 기술할 수 있다. 또한 XML, RDF, RDF 스키마를 이용한 질의 유형 보다 확장한 형태의 다양한 질의 처리가 가능하며, 온톨로지의 다양한 표현에 대한 보다 많은 제약사항을 기술하기 때문에 클래스와 구성원간의 관계를 온톨로지로 정의되지 않는 사실들까지 논리적인 추론이 가능하다.

온톨로지의 효율적인 설계를 위해 스탠포드 대학의 Noy and McGuinness에 의한 "Ontology Development 101"[7]의 방법론에 따라 구축하였다. 이를 참고로 하여 온톨로지를 작성 시에 일반적인 순서를 살펴보면 다음과 같다.

- 가. 기존 어휘의 재사용을 고려한다.
- 나. 중요한 용어를 나열한다.
- 다. 중요한 용어들을 클래스와 클래스 계층 구조로 정의한다.
- 라. 클래스들 간의 관계를 속성으로 정의한다.
- 마. 인스턴스를 기술한다.

3.1.1 클래스와 클래스 계층 구조 정의

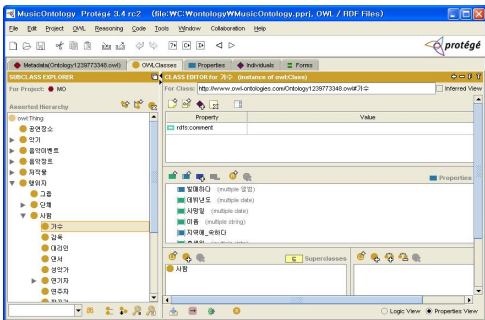
온톨로지 구축 시 중요한 용어를 명사와 명사절을 사용하여 정의한 뒤, 클래스와 각 클래스의 동의어들을 나열한다. 동의어는 owl :

sameAs 어휘를 사용하여 표현하며 표현 규칙은 <표 4>와 같다.

<표 4> 동의어 표현 예

```
owl:sameAs("공연" : "상연")
owl:sameAs("가요" : "대중가요")
owl:sameAs("클래식" : "고전음악")
owl:sameAs("콘서트" : "음악회")
owl:sameAs("리사이틀" : "발표회" : "연주회")
owl:sameAs("오페라" : "가극")
```

정의된 클래스를 바탕으로 클래스들 간의 계층 관계를 기술하며, 클래스의 계층 구조는 is-a관계로서 파악할 수 있다. 즉, 서브클래스 A는 상위클래스 B의 일종이라는 관계가 성립한다. <그림 2>는 음악 온톨로지의 클래스와 클래스들 간의 계층 관계를 Protégé3.4 [8]를 사용하여 구축한 그림이다. Protégé는 스탠포드 대학에서 개발한 자바 기반의 온톨로지 편집기로 본 논문에서는 버전 3.4를 이용하였다.



<그림 2> 음악 온톨로지 클래스와 클래스 계층 구조

<표 5>는 음악 온톨로지에서 클래스와 클래스 계층 구조를 OWL의 <owl:Class>, <rdfs:subClassOf> 어휘를 사용하여 표현한 예이다.

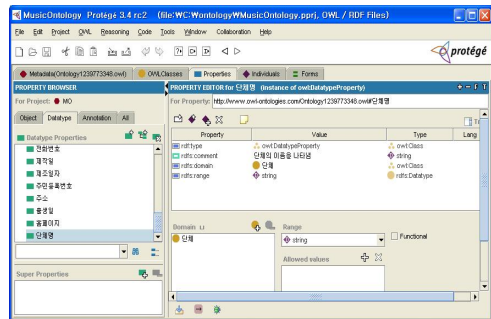
<표 5> OWL로 표현한 클래스와 클래스 계층구조

```
<owl:Class rdf:ID = "행위자"/>
<owl:Class rdf:ID = "사람">
  <rdfs:subClassOf rdf:resource = "#행위자"/>
</owl:Class>
<owl:Class rdf:ID="단체">
  <rdfs:subClassOf rdf:resource = "#행위자"/>
</owl:Class>
<owl:Class rdf:ID = "작곡가">
  <rdfs:subClassOf rdf:resource="#사람"/>
</owl:Class>
```

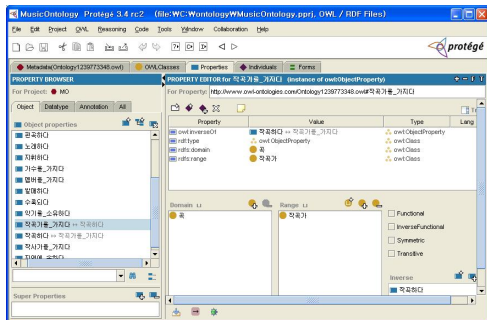
3.1.2 속성 정의

클래스 간의 관계 및 특징을 나타내는 속성을 정의한다. 이때, 상위 클래스에 공통적인 속성은 하위 클래스보다 상위 클래스에 정의하며, 속성은 오브젝트 타입(owl:ObjectProperty)의 속성과 데이터 타입이 속성(DatatypeProperty)으로 구성된다.

속성을 정의할 때는 속성의 주어부분이 되는 클래스를 정의역, 목적어가 되는 클래스를 치역이라고 부르며 <rdfs:domain>과 <rdfs:range>를 사용하여 지정한다. <그림 3>은 데이터 타입의 속성을 나타내며 <그림 4>는 오브젝트 타입 속성을 나타낸다.



<그림 3> 음악 도메인 데이터 타입 속성 정의



〈그림 4〉 음악 도메인 오브젝트 타입 속성 정의

속성의 제약사항은 온톨로지를 구축할 때 발생하는 오류를 최소화 시킬 뿐 아니라 향후 추론 규칙을 구성하는데 영향을 미친다. <표 6>는 속성의 제약사항을 기술 한 예이다.

‘공연하다’라는 속성은 정의역을 행위자 클래스를 가지며, 치역을 음악이벤트 클래스를 가진다. ‘작곡가를_가지다’라는 속성의 경우, 저작물인 곡은 하나 이상의 작곡가를 가지기 때문에 반드시 1개의 작곡가가 존재해야 함을 제약조건을 통해 명시하였다.

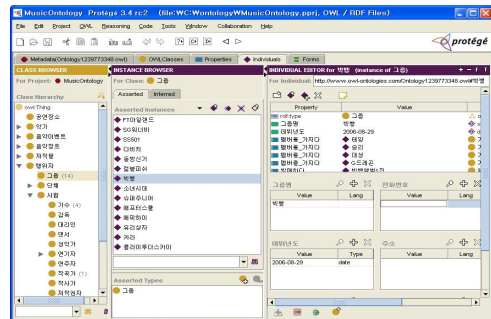
〈표 6〉 속성 정의 예

```

<owl:ObjectProperty rdf:ID = "공연하다">
  <rdfs:domain rdf:resource = "#행위자"/>
  <rdfs:range rdf:resource = "#음악이벤트"/>
  <rdfs:subPropertyOf rdf:resource
    = "#예술가를_묘사하다"/>
</owl:ObjectProperty>
<owl:Class rdf:ID = "곡">
  <owl:Restriction>
    <owl:onProperty rdf:resource
      = "#작곡가를_가지다"/>
    <owl:minCardinality> 1
      </owl:minCardinality>
  </owl:Restriction>
</owl:Class>
  
```

3.1.3 인스턴스 기술

클래스와 속성을 모두 정의한 후 인스턴스를 기술한다. <그림 5>는 Protégé3.4를 이용하여 인스턴스를 작성한 화면이다. 가운데 창에 인스턴스를 입력하면 그에 해당하는 속성들을 기술 할 수 있는 창이 오른쪽에 나타난다.



〈그림 5〉 인스턴스 작성

<표 7>은 OWL로 작성된 인스턴스의 예이다. 가요 그룹 빅뱅이 부른 ‘Dirty Cash’ 곡에 대한 정보를 나타내며, 가상의 UCI 식별 번호(I010:G200+100-DirtyCash:Fmp3)를 갖는 MP3형태로 존재함을 기술하였다.

〈표 7〉 인스턴스 예

```

<곡 rdf:ID = "Dirty_Cash">
  <uci:UCI rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"> I010:G200+100-DirtyCash:Fmp3 </uci:UCI>
  <곡제목 rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"> Dirty Cash </곡제목>
  <uci:type> Digital </uci:type>
  <uci:mode>Audio</uci:mode>
  <uci:format>mp3</uci:format>
  <수록되다 rdf:resource = "#빅뱅앨범1집"/>
  <작곡가를_가지다 rdf:resource = "#G드레곤"/>
</곡>
  
```


4. 추론 및 검색 결과 분석

4.1 OWL 공리를 이용한 추론

공리란 하나의 이론에서 증명 없이 바르다고 하는 명제, 즉 조건 없이 전제된 명제이다. 온톨로지 구축 시 공리는 다음 두 가지 역할을 할 수 있다.

첫째, 어휘와 개념의 의미 정의를 정확하게 표현하는 것

둘째, 온톨로지에 포함된 어휘와 개념을 사용하여 추론하는 것

OWL 언어의 기술 논리를 설명할 수 있는 공리는 <표 8>과 같다[9].

<표 8> OWL 공리

공리	기술논리(DL)구문
subClassOf	$C_1 \sqsubseteq C_2$
equivalentClass	$C_1 \equiv \dots \equiv C_n$
subPropertyOf	$P_1 \sqsubseteq P_2$
equivalentProperty	$P_1 \equiv \dots \equiv P_n$
disjointWith	$C_i \sqsubseteq \neg C_j$
inverseOf	$P_1 \equiv P_2^-$
DifferentIndividuals	$\{O_1\} \neq \{O_n\}$
SameIndividual	$\{O_1\} = \{O_n\}$
Transitive	$P^+ \sqsubseteq P$
Symmetric	$P \equiv P^-$

OWL로 작성된 온톨로지는 OWL에서 공리를 사용하여 온톨로지 구축 시에 직접적으로 명시하지 않은 내용들을 추론해 낼 수 있다.

이행속성(Transitive Property)은 속성 P가 존재할 때, $P(x, y) \wedge P(y, z)$ 이라면 $P(x, z)$ 가 성립한다는 관계를 추론해 낼 수 있다. 이행

속성을 가진 속성 P는 OWL에서 owl:TransitiveProperty로 나타낸다. 예를 들어, 속하다(제1합창단, 합창단)이고 속하다(합창단, 서울시향)일 때, 제 1합창단은 서울시향에 속한다는 사실을 추론할 수 있다. 예를 OWL로 표현하면 아래와 같다.

```
<owl:TransitiveProperty rdf:ID = "속하다">
  <rdfs:domain rdf:resource = "#연주단체"/>
  <rdfs:range rdf:resource = "#연주단체"/>
</owl:TransitiveProperty>
```

대칭속성(Symmetric Property)은 속성 P가 존재할 때, $P(x, y)$ 이면 $P(y, x)$ 가 성립함을 추론해 낼 수 있다. 대칭속성을 가진 속성 P는 OWL에서 owl:SymmetricProperty로 나타낸다. 예를 들어, 협주하다라는 속성은 주어와 목적어를 바꾸더라도 같은 관계가 성립하며 속성의 정의역과 치역은 같은 클래스가 된다. 예를 OWL로 표현하면 아래와 같다.

```
<owl:SymmetricProperty rdf:ID = "협주하다">
  <rdfs:domain rdf:resource = "#악기"/>
  <rdfs:range rdf:resource = "#악기"/>
</owl:TransitiveProperty>
```

온톨로지를 기반으로 하면 단순한 메타데이터만으로 데이터에 대한 정보를 표현할 때보다 추론을 통하여 더 풍부하고 다양한 정보를 제공할 수 있다.

4.2 검색 결과 분석

본 장에서는 지금까지 구축한 온톨로지를

바탕으로, 질의를 작성하고 질의에 대한 검색의 결과를 나타냄으로써 UCI 식별 메타데이터만으로 검색했을 경우와 온톨로지를 적용했을 경우의 차이를 보인다.

온톨로지 쿼리 언어로는 SPARQL[10]을 사용하기로 한다. OWL로 작성된 온톨로지는 RDF트리플(RDF Triples)로 변환되기 때문에 RDF를 위한 질의의 언어로서 개발된 SPARQL을 OWL 웹 온톨로지에 적용할 수 있다. 이는 웹 온톨로지가 그래프 모델이면서 트리플 집합이라는 측면에서 SPARQL을 이용한 OWL에 대한 검색 연산이 가능하다[11].

4.2.1 질의 정의

음악 도메인에서 온톨로지를 적용하여 디지털 콘텐츠에 대한 내용 검색 시에 질의의 내용을 다음과 같이 선정하였다. 질의는 클래스와 속성들의 관계와 계층구조, 추론을 포함하여 총 7개로 구성되어 있으며 'LUBM : A Benchmark for OWL Knowledge Base Systems [12]'를 참고하여 작성하였다. 다음 질의들은 정확한 결과분석을 위해 입력의 크기와 클래스의 계층구조, 속성의 계층 구조 등을 다양하게 포함하여 작성하였다.

(질의 1) 한국음악저작권협회에 등록된 노래 검색

이 질의는 큰 입력과 높은 선택성을 지닌다. 클래스들 간의 계층구조를 포함하지 않으며 하나의 클래스와 하나의 속성에만 관련이 있다.

(질의 2) 그룹 활동을 하는 가수가 발매한 단독 싱글 앨범 검색

이 질의는 3개의 클래스와 3개의 속성을 필요로 하며 질의 1에 비해 복잡성이 증가했다.

(질의 3) 가수별로 발매한 장르별 앨범 검색

이 질의는 음악 장르 클래스에 관련되어 광범위한 클래스 계층 구조를 대상으로 검색이 이루어진다.

(질의 4) 연주 단체에 소속된 연주자들의 이름, 전화번호 검색

이 질의는 작은 입력과 높은 선택성을 지니며, 하나의 클래스에 대한 여러 속성에 대한 질의이다.

(질의 5) 특정 소속사에 속한 멤버 검색

소속사에 속한 사람 클래스들과 그 하위 클래스들의 검색을 포함하며, '멤버다'라는 속성의 하위 속성들을 포함한다.

(질의 6) 작곡가별 작곡한 곡 검색

이 질의는 OWL의 inverseOf로 표현한 속성들의 역관계를 이용하여 결과를 검색한다. '작곡한다' 라는 속성은 '작곡 되다'라는 속성과 역관계를 가지는 속성으로 정의되었다.

(질의 7) 어떤 히트곡을 갖는 가수의 곡 검색

이 질의는 속성의 동치성과 관련이 있다. 히트 곡 속성은 대표곡, 인기곡이라는 속성과 같은 인스턴스를 가짐을 정의하였다.

<표 11>은 위의 질의들을 온톨로지 쿼리 언어인 SPARQL로 표현한 것이다.

4.2.2 검색 결과

<그림 6>은 Protégé3.4를 이용하여 위의 4.2.1의 질의를 검색했을 때 각각의 결과를 나타낸다.

〈표 11〉 SPARQL 질의문

질의 번호	질의 내용	질의 번호	질의 내용
질의 1	SELECT ?X WHERE { ?X rdf:type ?곡. ?X : 등록되다 : 한국음악저작권협회.}	질의 4	SELECT ?X ?Y WHERE { ?X : 멤버이다 : 서울시립교향악단. ?X : 전화번호 ?Y }
질의 2	SELECT ?X ?Y ?Z WHERE { ?X rdf:type ?그룹. ?X rdf:type ?가수. ?Y rdf:type ?앨범. ?Y : 멤버이다 ?X. ?Y : 발매하다 ?Z.}	질의 5	SELECT ?X WHERE { ?X rdf:type ?그룹. ?X : 멤버이다 : DSP_미디어.}
질의 3	SELECT ?Y ?Z WHERE { ?X rdf:type ?가수. ?Y rdf:type ?음악장르. ?Z rdf:type ?앨범. ?Z : 음악장르를_가지다 ?Y. ?X : 발매하다 ?Z.} ORDER BY? 음악장르	질의 6	SELECT ?X WHERE { ?X rdf:type ?곡. ?X rdf:type ?작곡가. ?X : 작곡하다 ?Y.}
		질의 7	SELECT ?X ?Y WHERE { ?X rdf:type ?가수. ?Y rdf:type ?곡. ?X : 히트곡 : 라라라. ?Y : 가수를_가지다 ?X.}

디지털 콘텐츠에 대한 메타데이터만으로 작성한 경우, 식별 메타데이터에서 제공하는 항목에 대한 정보만을 확인 할 수 있다. 위의 질의 중에서 (질의 6)은 식별 메타데이터만으로 검색이 가능하지만, 그 외 6개의 질의들은 식별 메타데이터만으로는 검색이 불가능하다.

그러나 온톨로지를 이용한 경우에는 식별 메타데이터를 기반으로 메타데이터의 의미정보, 개념관계까지 포함함으로써 더욱 상세한 정보 검색이 가능해진다.

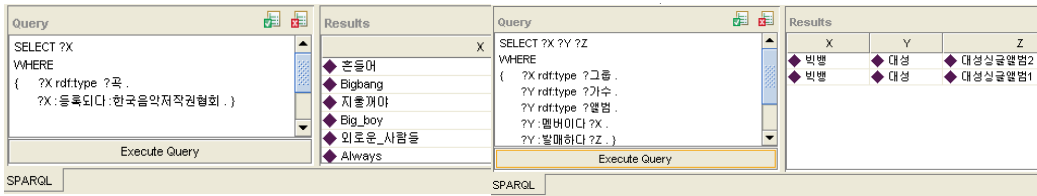
음악 콘텐츠를 검색할 때, 온톨로지를 사용함으로써 제목과 가수, 저작권자 정보로만 정보를 제공하는 방식에서 각 음악 콘텐츠의 작곡자와 작사가뿐만 아니라 관련된 단체와

사람에 대한 정보를 제공할 수 있으며, 장르, 히트곡, 공연 정보 등을 포함한 다양한 정보들을 제공할 수 있게 된다.

온톨로지를 바탕으로 보다 풍부한 의미 기반의 정보검색 서비스의 제공이 가능해진다. 또한 기존의 구축되어 존재하는 다양한 온톨로지와의 통합과 공유를 통해 더욱 다양하고 정확한 정보들을 검색하는 것이 가능해진다.

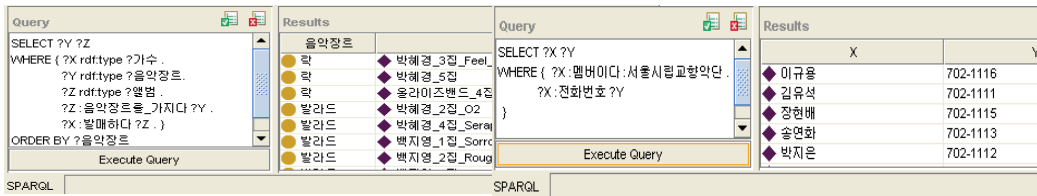
5. 결 론

본 논문에서는 디지털 콘텐츠의 효율적인 관리와 검색을 위해 UCI 식별 체계와 온톨로



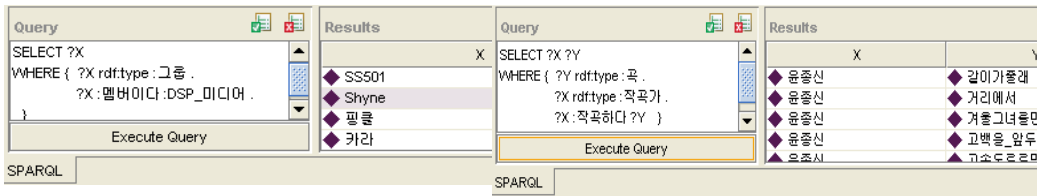
(질의 1)

(질의 2)



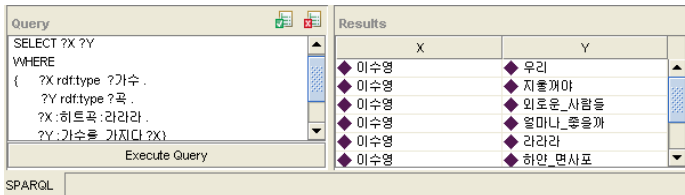
(질의 3)

(질의 4)



(질의 5)

(질의 6)



(질의 7)

〈그림 6〉 질의 검색 결과

지의 적용에 대하여 제안하였다. UCI 식별 체계를 적용할 때 식별 메타데이터 만으로는 상세한 정보를 제공하기 어렵기 때문에 사용자가 원하는 정확한 정보를 제공하기 위해서는 메타데이터와 메타데이터간의 관계를 온톨로지로 구축하였다.

UCI 식별 메타데이터를 온톨로지로 확장하여 적용하기 위하여 도메인 온톨로지를 선정하였으며, 도메인 온톨로지는 본 논문의 목

적에 가장 적합한 음악 온톨로지를 구축하였다. 구축한 온톨로지는 여러 사용자에게 재사용될 수 있다는 장점을 가지며, 온톨로지로 작성된 디지털 콘텐츠에 대한 정보는 UCI 기관간, 시스템간 공유가 가능하므로 콘텐츠 관리 비용을 절감하는 효과를 얻을 수 있다.

작성한 온톨로지를 바탕으로 OWL의 공리를 이용하여 추론해냄으로써 새로운 온톨로지의 생성이 가능함을 보였고, 구축된 디지털

콘텐츠의 온톨로지의 검색의 정확성을 보이기 위해 다양한 입력의 크기와 클래스의 계층구조, 속성의 계층구조를 포함하는 질의를 선정하였다.

검색 결과 메타데이터만으로 디지털 콘텐츠를 검색할 경우, 메타데이터에서 제공하는 항목 정보 위주의 검색만 가능하며, 콘텐츠들 간의 관계와 세부 정보들을 검색하는 것이 어렵지만 온톨로지를 사용함으로써 보다 다양하고 정확한 정보의 검색이 가능함을 알 수 있었다.

본 논문에서는 디지털 콘텐츠 관리에 시맨틱 기술인 온톨로지를 접목하여 보다 효율적이며 의미적 적합성을 기반으로 한 검색을 가능하도록 하였다. 보다 개선 할 점은 저작권 보호를 위하여 저작권에 대한 온톨로지 부분이 더욱 강화되어야 한다는 것이다. 이점은 향후에 연구하기로 한다.

참 고 문 헌

- [1] 한국전산원, "UCI 명세서(version 2.2)", 2007. 12.
- [2] 신동구, 김재수, 윤정모, 권이남, 전성진, 정택영, "식별체계 간 연계시스템 구축에 관한 연구", 한국정보과학회 2005년도 가을 학술발표논문집, 제32권, 제2호(1), 2005.
- [3] Web Ontology Language, <http://www.w3.org/2004/OWL/>.
- [4] 이정연, 김정민, 최석두, 김이겸, "기초학문자료 메타데이터 설계 분석 및 온톨로지 적용 방안 연구", 한국문헌정보학회지, 제41권, 제2호, 2007, pp. 291-316.
- [5] 한국소프트웨어진흥원, "2008년도 국내 디지털콘텐츠산업 시장조사 보고서", 2008.
- [6] KOCCA, "문화콘텐츠 산업별 유통구조 분석", 2006.
- [7] Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101 : A Guide to Creating Your First Ontology," 2001-03, <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>.
- [8] <http://protege.stanford.edu>.
- [9] 이혜원, "음악 자원을 대상으로 한 이벤트 중심 ABC 온톨로지 확장 모형에 관한 연구", 2007, 박사논문.
- [10] Eric Prud'hommeaux, "Andy Seaborne, SPARQL Query Language for RDF," 2006.
- [11] 정동원, 백두권, "온톨로지 추론 모델에 독립적인 SPARQL 추론 질의 처리를 위한 재작성 알고리즘", 한국정보과학회논문지, 제35권, 제6호, 2008, pp. 505-517.
- [12] Yuanbo Guo, Zhengxiang Pan, and Jeff Heflin, "LUBM : A Benchmark for OWL Knowledge Base Systems," Journal of Web Semantics, Vol. 3, 2005.

저 자 소개



하은옥
2000년
2003년
현재
관심분야

(E-mail : arumlily@smu.ac.kr)
상명대학교 소프트웨어학과 (학사)
상명대학교 컴퓨터학과 (석사)
상명대학교 컴퓨터학과 (박사과정)
시맨틱웹, 온톨로지



김운호
1985년
1987년
1996년
1997년~현재

(E-mail : yhkim@smu.ac.kr)
서울대학교 계산통계학과 (학사)
서울대학교 계산통계학과 계산학전공 (석사)
서울대학교 전산과학전공 (박사)
상명대학교 소프트웨어대학 컴퓨터과학부교수