

<http://dx.doi.org/10.7236/IIBC.2015.15.5.19>

IIBC 2015-5-3

## 저작물의 의미 기반 검색을 위한 온톨로지 적용 방안 연구

### A Study on Application Method of Ontologies for Efficient Semantic-based Search of Copyrighted Works

유민규\*, 김윤호\*\*

Min-Kyu Yoo\*, Yoon-Ho Kim\*\*

**요약** 한국저작권위원회 디지털저작권거래소는 저작물에 통합저작권관리번호(ICN)를 부여하고 이를 바탕으로 저작물 및 저작권이용허락계약에 대해 운영·관리하고 있으며, ICN이 부여된 저작물에 대해 통합검색 서비스를 제공하고 있다. 통합검색 서비스는 저작물의 ICN 메타데이터를 바탕으로 키워드 기반 검색 방법을 이용하여 메타데이터의 키워드 포함 여부에 따른 검색 결과를 제공한다. 하지만 키워드 기반 검색 방법은 이용자가 원하는 정확한 검색 결과를 제공하지 못하는 한계를 가지고 있다. 따라서 정확한 검색 결과를 제공할 수 있는 방법이 필요하다.

본 논문에서는 저작물에 대한 의미 기반 검색 방법의 적용 및 검색 결과의 높은 정확도를 제시하기 위해 ICN이 부여된 저작물의 ICN 메타데이터를 온톨로지로 확장하여 메타데이터 요소간의 관계 및 의미를 정의한 저작물 온톨로지를 구축하였다. 구축한 온톨로지에서 추론을 통한 의미 기반 검색의 가능함과 키워드 기반 검색 방법과의 정확도 비교를 위해 5가지 시나리오를 선정하였으며, 시나리오별 각 검색 방법의 결과 및 정확도의 비교 결과를 제시하였다.

**Abstract** Korea Digital Copyright Exchange of Korea Copyright Commission runs and manages copyrighted works and copyright license agreement, as well as assigns Integrated Copyright Number(ICN) to copyrighted works, also offers copyrighted works searching service. Keyword-based searching method based on ICN metadata provides results depending on the inclusion of metadata keyword. However, keyword-based searching method has limitations that cannot provide exact results that users need. Therefore, the essential method of offering exact result is needed. In this paper, we propose the semantic-based searching method about copyrighted works to increase the accuracy of searching results by extending ICN metadata to ontology, and implement the ontology of copyrighted works by defining relation and semantics about metadata elements. In order to compare semantic-based searching based on the ontology of copyrighted works with keyword-based searching method, the search results of each method according to 5 scenarios are presented and the accuracy of each method is compared.

**Key Words** : Integrated Copyright Number(ICN), Copyright ontology, Semantic-based search, keyword-based search

## 1. 서론

한국저작권위원회에서는 디지털저작권거래소의 운영을 통해 저작권이용허락계약 서비스의 관리 및 운영, 저

작권관리번호 체계를 마련하고 저작물 정보 및 저작물의 저작권 정보를 통합검색 서비스를 제공하고 있다. 디지털저작권거래소는 통합저작권관리번호(ICN: Integrated Copyright Number, 이하 ICN)를 발급하여 신뢰성 있는

\*정회원, 상명대학교 컴퓨터학과

\*\*정회원, 상명대학교 컴퓨터학과 (교신저자)

접수일자 : 2015년 9월 22일, 수정완료 2015년 10월 7일

계재확정일자 : 2015년 10월 9일

Received: 22 September, 2015 / Revised: 7 October, 2015 /

Accepted: 9 October, 2015

\*\*Corresponding Author: yhkim@smu.ac.kr

Dept. of Computer Science, Sangmyung University, Korea

저작물 정보 및 저작권 정보와 저작물을 이용하는 이용자의 저작권이용 편의성 및 온라인을 통한 온라인 저작권 이용허락계약 서비스를 제공함으로써, 저작물 유통에 있어 투명한 정산 및 분배를 목적으로 하고 있다<sup>11)</sup>. 디지털저작권거래소는 ICN을 부여한 저작물의 ICN 메타데이터를 기반으로 저작물 정보 및 저작권 정보에 대한 통합검색 서비스를 제공하고 있으며, 음악저작물, 어문저작물, 뉴스저작물 등 27개의 장르로 분류하여 검색 결과를 제공하고 있으며, 이용자는 디지털저작권거래소가 제공하는 통합검색 서비스를 통해 원하는 저작물 정보 및 저작권 정보를 얻을 수 있다.

그러나 디지털저작권거래소에서 운영하는 통합검색 서비스는 저작물의 ICN 메타데이터에 이용자가 입력한 키워드의 포함 여부에 따라 검색 결과를 제공하는 키워드 기반 검색 방법으로 저작물 정보 및 저작권 정보를 제공한다. 키워드 기반 검색 방법은 이용자가 입력한 키워드를 포함하고 있는 모든 대상을 검색의 결과로 제공하기 때문에 제공된 검색 결과 중 이용자가 원하지 않는 정보가 포함되어 있는 경우가 존재할 수 있다. 또한 이용자가 원하는 저작물 정보 및 저작권 정보가 키워드를 포함하고 있지 않은 경우, 대상이 존재함에도 불구하고 올바른 검색 결과를 제공하지 않으며 저작물의 메타데이터 중 변경된 내용과 누락이 존재할 경우 또한 이용자가 원하는 정확한 검색 결과를 제공하지 못하는 문제점이 존재한다.

위와 같은 키워드 기반 검색 방법의 문제점을 해결하기 위해 본 논문은 ICN이 부여된 저작물의 ICN 메타데이터를 온톨로지로 확장하여 저작물 온톨로지를 구축하였다. 메타데이터를 온톨로지로 확장하여 메타데이터 요소간의 관계 및 의미를 정의하면 메타데이터가 변경되거나, 누락된 항목이 존재한다 하더라도 키워드 포함 여부에 따라 결과를 제공하는 키워드 기반 검색 방법과는 달

리, 정의된 메타데이터 요소간의 의미 및 관계를 기반으로 의미 기반 검색이 가능하고 이용자가 원하는 정확한 검색 결과를 얻을 수 있으며 메타데이터 요소간의 관계 및 의미를 통해 검색 결과에 대한 부가적인 정보를 얻을 수 있다는 장점을 가지고 있다.

본 논문의 구성은 다음과 같다. 2장에서 ICN에 대해 언급하고, ICN 메타데이터를 기반으로 저작물 온톨로지를 구축하는 과정을 기술한다. 3장에서는 2장의 결과로 구축한 저작물 온톨로지의 추론 과정을 Protégé 4.3<sup>12)</sup>을 통해 보였다. 또한 키워드 기반 검색 방법과 의미 기반 검색 방법의 한계점과 차이점을 분석하고 검색 결과의 정확도 비교를 위해 5가지 시나리오를 구성하여 시나리오에 대한 검색 결과의 비교를 통해 온톨로지 확장을 통한 저작물 정보 및 저작권 정보에 대해 의미 기반 검색의 가능성과 검색 결과의 높은 정확도를 제시하였다. 4장에서는 결론과 향후 연구방향을 제시하였다.

## II. 저작물 온톨로지

### 1. 통합저작권관리번호(ICN)

통합저작권관리번호는 저작물을 등록·관리하는 저작권신탁관리단체 및 권리관리기관에 따라 다르게 부여해 운영하던 저작물 번호를 통합하여 동일 저작물에 대해 유일한 하나의 식별코드를 부여해서 각각의 단체가 상호 운용이 가능하도록 하기 위해 저작물에 부여하는 통합된 고유번호이다. ICN은 저작물의 권리관리정보 식별을 위한 고유한 문자열 및 번호체계로 저작물의 생성, 유통, 소멸 등의 저작물 생명주기의 관리를 통해 정산과 분배를 목적으로 하고 있으며 한국저작권위원회 디지털저작권거래소를 통해 권리위탁관계, 저작물 이용허락계약에 관한 포괄적인 정보를 제공한다.

표 1. ICN 메타데이터 요소  
Table 1. Elements of ICN Metadata

ICN 메타데이터 요소	정의
원저작자 / 재산권자	저작물의 지적재산권과 직접적인 관계를 가진 권리자의 대한 정보
저작인접권자 / 배타적이용권자	저작물의 지적재산권과 간접적인 관계를 가진 권리자에 대한 정보
원저작자 예·필명	저작물의 지적저작에 기여한 자의 정보
제작형태	저작물이 창작된 형태
제작저작물 정보	저작물의 제작에 대한 정보
권리위탁	권리를 위탁한 기관에 대한 정보
이용계약(라이선스)	저작권 이용허락 계약 및 정산에 대한 정보

ICN 메타데이터는 저작권과 직접적인 연관을 가진 요소로 구성되어 있으며 표 1과 같이 기본요소를 정의하고 저작물의 장르에 따라 해당 요소에 대해 구체적으로 재정의할 수 있도록 하고 있다.[3][4]

## 2. 저작물 온톨로지 구축

온톨로지의 효과적인 구축을 위한 많은 방법론이 존재하며[5] 본 논문에서는 “Ontology Development 101[6]” 방법론에 따라 저작물 온톨로지를 구축하였다. “Ontology Development 101” 방법론에 따른 온톨로지 구축을 위한 첫 번째 단계는 도메인 온톨로지를 선정하는 것이다. ICN 메타데이터를 온톨로지화 확장하여 메타데이터 요소간의 관계 및 의미 정의를 바탕으로 저작물의 표현과 의미 기반 검색이 가능하도록 하고자하는 본 논문의 목적에 부합하도록 ICN이 부여된 저작물을 온톨로지의 도메인으로

선정하였다. 또한 온톨로지 구축을 위한 많은 언어가 존재하는데[7] 본 논문은 OWL(Web Ontology Language, 이하 OWL)[8] 언어를 사용한다. OWL 언어는 기계 및 에이전트가 처리할 수 있는 풍부한 어휘와 형식적 의미를 포함하고 있으며 XML, RDF, RDF Schema 보다 확장된 형태의 질의 처리가 가능하며 제약사항(Restriction)을 비롯한 표현을 통해 대상을 더욱 명확하게 표현할 수 있다. OWL 언어로 작성한 온톨로지는 명시하지 않은 사실에 대해 추론이 가능하며 이를 검증하기 위해 스텐포드 대학에서 개발한 자바 기반의 온톨로지 편집기인 Protégé 4.3을 이용한다. Protégé 4.3의 내장된 추론기는

온톨로지내의 직접적으로 명시하지 않은 관계 및 의미에 대한 추론 결과를 제공한다.

아래의 각 절은 “Ontology Development 101” 방법론의 온톨로지 구축 단계에 따른 저작물 온톨로지를 구축하는 과정에 대해서 기술하고 있다.

### 가. 클래스 및 클래스 계층 구조 정의

클래스는 온톨로지 구성 요소 중 개념(Concept)에 해당하며, 인스턴스(Instance)의 집합(Set)으로 표현한다. 클래스 및 클래스 계층 구조의 정의를 위해 사용하는 OWL 어휘로는 <owl:Class>, <rdfs:subClassOf>, <owl:equivalentClass> 등이 있다. 클래스 및 클래스 계층 구조 정의 시 고려해야할 점은 온톨로지를 구성하는 용어 중 인스턴스를 가지는 용어에 대해서만 클래스로 정의해야 한다는 점이다. 온톨로지를 구성하는 모든 용어를 클래스로 정의할 경우, 클래스간의 관계를 정의하는 속성의 수도 함께 증가하며, 많은 수의 속성을 통해 과도한 관계를 정의하면 오히려 관계의 명확성을 해칠 수 있다. 따라서 저작물 온톨로지는 표 1의 ICN 메타데이터 요소 중 실제 인스턴스가 존재하는 ‘원저작자/재산권자’, ‘저작인접권자/배타적이용권자’, ‘제작형태’, ‘권리위탁’만을 클래스로 정의하였다. ICN 메타데이터 요소 중 ‘원저작자/재산권자’는 클래스명 ‘원저작자/재산권자’로 정의하고 저작권자 및 재산권자의 역할을 표현하기 위해 ‘작곡’, ‘작사’, ‘제작’, ‘유통’ 등을 하위 클래스로 구성하였다. ‘저작인접권자/배타적이용권자’는 저작물을 이용하는 이용자를 표현하기 위해 클래스명 ‘이용자’로 정의

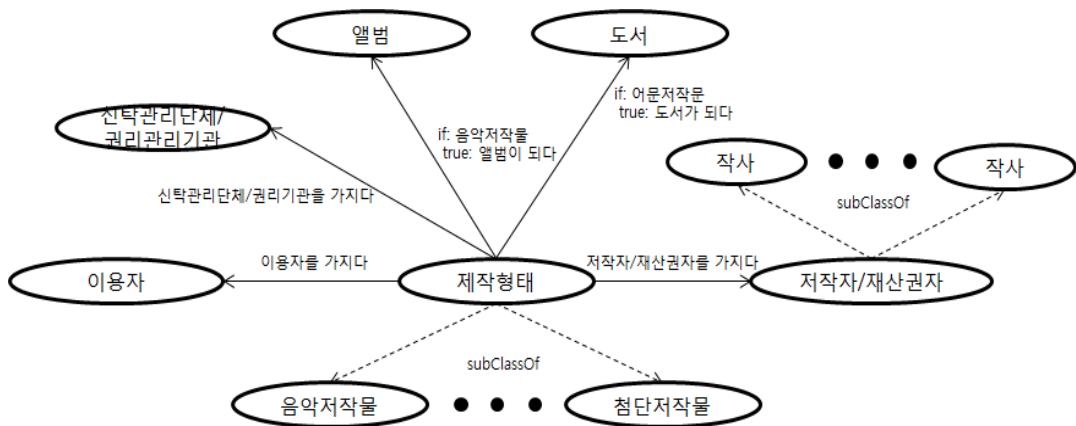


그림 1. 저작물 온톨로지  
 Fig. 1. Copyrighted Works Ontology

하였다. 저작물이 창작된 형태에 대한 정보를 담고 있는 ICN 메타데이터 요소인 ‘제작형태’는 메타데이터명 그대로 ‘제작형태’로 클래스명을 정의하였으며, 음악저작물, 어문저작물, 뉴스저작물 등 저작물의 제작형태를 하위 클래스로 구성하였다. 마지막으로 저작물의 권리를 위탁한 기관에 대한 정보를 담고 있는 ICN 메타데이터 요소인 ‘권리위탁’은 클래스명 ‘신탁관리단체/권리관리기관’으로 정의하였으며, 이 밖에 어문저작물 포함하는 ‘도서’, 음악저작물을 포함하는 ‘앨범’을 클래스로 정의하였다.

그림 1은 저작물 온톨로지를 그래프로 표현한 것이다. 그래프는 타원과 화살표로 구성되어 있으며, 타원은 저작물 온톨로지의 클래스이고 화살표는 클래스간의 관계를 표현하고 있다. 점선으로 된 화살표는 클래스의 상하 관계인 클래스 계층 구조를 표현하고 있으며, 실선으로 된 화살표는 객체형속성으로 클래스간의 관계를 표현하고 있다.

#### 나. 속성 정의

속성은 온톨로지 구성 요소 중 관계(Relation)에 해당한다. OWL 언어의 속성은 객체형속성(ObjectProperty), 데이터형태속성(DatatypeProperty) 2가지가 존재한다. 객체형속성은 클래스 및 인스턴스간의 관계만을 정의하며, 데이터형태속성은 인스턴스와 인스턴스가 가지는 값의 관계를 정의한다. 또한 속성은 온톨로지 구축 시 발생하는 오류의 최소화를 위해 정의역(Domain)과 치역(Range)을 결정하여 속성이 다른 용도로 사용되지 않도록 하며 제약사항을 추가할 수 있다. 객체형속성과 데이터형태속성은 각각 `<owl:ObjectProperty>`, `<owl:DatatypeProperty>`와 같은 OWL 어휘로 표현된다. 속성 역시 위 가. 에서 언급했듯이, 온톨로지 구축 시 관계의 명확성을 해칠 수 있는 과도한 수의 속성의 사용을 지양하고 속성의 정의역 및 치역, 제약사항을 이용하여 명확한 관계를 정의할 수 있도록 해야한다.

저작물 온톨로지에서는 객체형속성인 ‘~을 가지다’와 ‘~이 되다’를 사용하여 인스턴스간의 관계를 정의하고 있다. 예를 들어, ‘제작형태’ 클래스의 하위 클래스인 ‘음악저작물’ 클래스와 ‘원저작자/재산권자’ 클래스의 하위 클래스인 ‘작곡’, ‘작사’, ‘편곡’ 클래스와의 관계를 ‘작곡가를 가지다’, ‘작사가를 가지다’, ‘편곡가를 가지다’와 같은 객체형속성을 이용해 저작물의 저작자 및 재산권자와의 관계를 정의하고 있다. 다른 예로, 객체형속성 ‘라이선스를

가지다’를 이용해 해당 저작물과 이용자와의 관계를 정의하고 있다. 이 때, 객체형속성 ‘라이선스를 가지다’는 정의역으로 저작물의 타입(Type)인 ‘제작형태’ 클래스의 하위 클래스 중 하나와 ‘이용자’ 클래스를 각각 정의역과 치역으로 결정할 수 있다. 저작물 온톨로지에서 데이터 형태 속성은 ICN 메타데이터 요소와 메타데이터 요소가 가지는 값의 관계를 정의하고 있다. 예를 들면, ICN 메타데이터 요소 중 ‘제작 저작물 정보’에 해당하는 ‘디지털저작물여부’를 데이터형태속성으로 정의하고 “Y”, “N” 중 하나의 값만을 가질 수 있도록 제약사항을 추가하여 메타데이터 요소의 값이 누락되지 않도록 할 수 있다. 다른 예로, 저작물에 대해 이용자가 가지는 라이선스 세부정보인 계약기간, 서비스타입 등을 데이터형태속성으로 관계 및 의미를 정의하고 있다.

#### 다. 인스턴스 생성

인스턴스는 클래스를 구성하는 실례이며, ‘개체’라는 이름으로도 사용할 수 있다. 인스턴스를 생성하고, 클래스 지정 및 속성을 이용해 실제 대상을 표현할 수 있다. 이 때, 인스턴스를 포함하고 있는 클래스는 인스턴스의 타입이 되며 인스턴스는 다른 인스턴스와의 속성을 이용한 관계 정의가 가능하다. 예를 들면, 저작물 온톨로지서 ‘음악저작물’ 클래스에 속하는 국내가요 ‘거리에서’ 인스턴스의 타입은 ‘음악저작물’이 되며, 위 2.2에서 정의한 객체형속성 및 데이터형태속성을 이용하여 ICN 메타데이터 요소간의 관계 정의를 통해 실제 음악저작물인 ‘거리에서’를 표현할 수 있다.

ICN 메타데이터 요소 중 ‘이용계약(라이선스)’에 대해 분석해본 결과, 저작물에 대해 이용자별로 각기 다른 라이선스 정보를 가지는 것을 확인할 수 있었다. 특이사항으로는 동일한 장르에 대한 라이선스 정보 중 같은 서비스타입을 가지고 있다면 해당 장르의 모든 저작물에 대해 이용자가 가지는 라이선스 정보가 모두 동일함을 알 수 있었다. 표 2는 ICN 메타데이터 요소 ‘이용계약(라이선스)’의 분석결과를 바탕으로 ‘이용자’ 클래스에 속하는 인스턴스가 라이선스 정보 중 서비스타입으로 ‘음악감상형’을 가질 때 ‘이용자’ 클래스의 인스턴스가 가지는 음악저작물의 라이선스를 표현하고 있다.

표 2. '엠에프유' 라이선스 정보  
 Table 2. License Information of "MFU"

```

<owl:NamedIndividual rdf:about="엠에프유">
  <rdf:type rdf:resource="이용자"/>
  <계약기간 rdf:datatype="xsd:string">2015-06-01-2024-12-31</계약기간>
  <자동연장여부 rdf:datatype="xsd:string">Y</자동연장여부>
  <홈페이지 rdf:datatype="xsd:anyURI">www.mfu.co.kr</홈페이지>
  <서비스타입 rdf:datatype="xsd:string">음악감상형</서비스타입>
  <이용자가_되다 rdf:resource="거리에서"/>
  ...
  ...
  <이용자가_되다 red:resource="세 사람(with 성시경)"/>
</owl:NamedIndividual>
    
```

### III. 추론 과정 및 검색 결과 분석

#### 1. 추론 과정 분석

의미 기반의 검색을 가능하게 하는 온톨로지 특징 중 하나는 명시하지 않은 사실에 대한 추론이 가능하다는 점이다. 이러한 추론은 공리를 통해 이루어지는데, 온톨로지 구축 시 공리는 어휘와 개념의 관계 및 의미를 정의하고 온톨로지에 포함된 어휘와 개념을 사용하여 추론을 가능하게 한다[9][10]. 2장에서 저작물 온톨로지를 구축하기 위해 OWL 언어를 사용했으므로 OWL 언어의 공리[11]를 이용해 저작물 온톨로지내에서 추론이 가능하며, 추론을 통해 직접적으로 명시하지 않은 관계 및 의미를 얻을 수 있어 저작물에 대한 의미 기반 검색이 가능하다. 아래의 예는 저작물 온톨로지에서 OWL 공리 중 속성 공리 "inverseOf"를 통해 Protégé 4.3을 이용한 추론 과정 및 결과를 보이고 있다. 그림 2는 앞서 예시로 등장한 '음악저작물' 클래스의 인스턴스 '거리에서'에 대해 OWL 언어로 표현한 객체형속성만을 Protégé 4.3의 객체형속성 기술 탭을 통해 보이고 있다.



그림 2. "거리에서" 객체형속성 기술 탭  
 Fig. 2. ObjectProperty assertion tab of "거리에서"

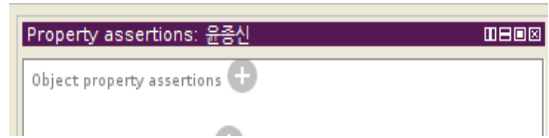


그림 3. 추론 전 '윤종신' 객체형속성 기술 탭  
 Fig. 3. ObjectProperty assertion tab of "윤종신" before reasoning

그림 2를 통해 인스턴스 '거리에서'의 가수, 작곡가, 작사가 등의 저작자 및 재산권자, 라이선스를 가지는 이용자, 신탁관리를 맡고 있는 신탁관리단체, 앨범과의 관계를 알 수 있다. 그림 3은 '거리에서'의 작곡가와 작사가의 역할을 가지고 있는 '원저작자/재산권자' 클래스의 하위 클래스인 '작사', '작곡' 클래스의 인스턴스 '윤종신'의 객체형속성을 기술한 Protégé 4.3의 객체형속성 기술 탭이다. 저작물 온톨로지 구축 시 인스턴스 '윤종신'에 대해 어떠한 객체형속성도 기술하지 않았기 때문에 인스턴스의 객체형속성 기술 탭을 통해 아무런 내용도 확인할 수 없음을 알 수 있다.

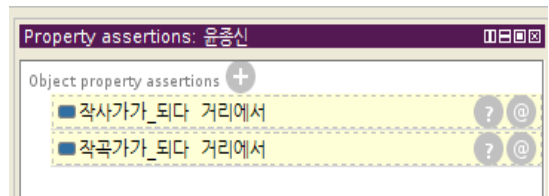


그림 4. 추론 후 '윤종신' 객체형속성 기술 탭  
 Fig. 4. ObjectProperty assertion tab of "윤종신" After reasoning

그림 4는 Protégé 4.3의 내장된 추론기가 OWL 공리를 이용해 추론한 결과이다. 그림 4와 같은 결과를 얻을 수 있는 이유는 저작물 온톨로지 구축 시 객체형속성인 '작사가를 가지다'와 '작곡가를 가지다'를 정의할 때, 그 역의 관계를 의미하는 '작곡가가 되다'와 '작사가가 되다'를 정의했기 때문에 추론기가 속성 공리인 "inverseOf"를 바탕으로 인스턴스 '윤종신'과 '거리에서'간의 명시하지 않은 사실에 대해 추론의 결과를 제공한다. 위의 추론 과정 및 결과를 통해 저작물 온톨로지 구축 시 정의한 클래스 및 속성 공리를 이용해 명시하지 않은 사실에 대해서 의미 및 관계가 정의됨을 알 수 있다. 따라서 의미 기반 검색이 추론을 포함할 때, 검색에 대한 결과를 정확하게 제시함을 가정할 수 있으며 이는 다음 절의 추론을 포함하는 시나리오를 통해 증명한다.

## 2. 검색 결과 분석

이 절에서는 앞 절의 추론을 통해 명시하지 않은 사실에 대해서도 의미 및 관계가 정의됨을 바탕으로 저작물 온톨로지의 의미 기반 검색의 가능성을 제시한다. 또한 키워드 기반 검색 방법의 검색 결과와의 정확도를 비교한다. 이를 위해 5가지 시나리오를 선정하였다. 비교를 위한 키워드 기반 검색 시스템은 한국저작권위원회 디지털저작권거래소의 저작권정보제공 통합검색 서비스를 이용한다. 디지털저작권거래소의 저작권정보제공 통합검색 서비스는 상세검색을 통해 저작물 및 저작권 정보에 대해서 키워드 및 키워드의 조합인 키워드의 합집합(AND)과 교집합(OR)을 이용한 검색이 가능하다.

저작물 온톨로지서 의미 기반 검색 결과를 얻기 위해서는 온톨로지 질의 언어를 사용해야 한다. OWL 언어로 작성된 온톨로지는 RDF 트리플(RDF Triple)로 변환되는 특성이 있기 때문에 RDF를 위해 개발된 질의 언어인 SPARQL<sup>[12]</sup>의 적용이 가능하다. 따라서 시나리오를 SPARQL 질의로 변환하여 의미 기반 검색의 결과를 얻을 수 있다.

키워드 기반 검색 방법의 경우, 대부분 단일 키워드 또는 키워드의 조합을 검색의 입력으로 이용한다. 단일 키워드를 검색의 입력으로 사용하는 경우, 이용자가 원하는 검색 결과를 포함하고 있으나, 메타데이터에 입력한 키워드를 포함하고 있는 모든 대상을 검색의 결과로 제공하기 때문에 검색 결과에 대한 정확도가 떨어진다. 한편 키워드 조합 중 키워드의 교집합을 검색의 입력으로 사용하는 경우 단일 키워드를 사용한 검색에 비해 적은 검색 결과를 제공하기 때문에 정확도가 일시적으로 상승하는 효과를 볼 수 있다. 반면, 키워드의 합집합을 검색의 입력으로 사용하는 경우에는 메타데이터에 입력한 키워드를 포함하고 있는 모든 대상을 검색의 결과로 제공하기 때문에 단일 키워드를 입력으로 사용한 검색 결과에 비해 정확도가 감소한다. 또한 조합을 통해 이용자의 요구와는 전혀 관련이 없는 결과가 제공될 수 있으며 대상이 실제 존재함에도 메타데이터가 키워드를 포함하고 있지 않아 검색의 결과로 제공되지 않을 수 있다.

메타데이터의 온톨로지 확장을 통한 의미 기반 검색 방법은 이용자의 입력에 대해 의미 기반으로 검색하기 때문에 제공하는 검색 결과가 키워드 기반 검색 방법의 검색 결과에 비해 높은 정확도를 보인다. 키워드 기반 검색 방법과 의미 기반 검색 방법의 검색 결과에 대한 정확

도의 비교를 위해 아래의 5가지 시나리오를 선정하였다<sup>[13]</sup>. 시나리오를 SPARQL 질의로 작성할 때 질의를 통해 저작물 온톨로지에 대해 클래스와 속성의 관계 및 계층 구조, 추론을 모두 포함하여 의미 기반 검색이 가능하도록 하였다.

### 시나리오 1: 저작권자, 저작권자의 통합 저작권ID 검색

시나리오 1은 저작물과 저작권자가 가지는 개체형속성을 통한 관계를 기반으로 저작권자를 검색하고, 저작권자가 가지는 통합저작권자ID를 데이터형태속성을 통한 관계를 기반으로 검색한다.

### 시나리오 2: 서비스타입별 이용자 라이선스 정보 검색

시나리오 2는 저작물과 이용자간의 관계를 추론을 통해 이용자를 검색하고, 이용자와 데이터형태속성으로 관계를 가지는 이용자의 라이선스 정보를 검색한다.

### 시나리오 3: 앨범 및 도서에 포함된 모든 음악저작물 및 어문저작물 검색

앨범명 및 도서명에 대해서 추론을 통해 앨범 및 도서에 포함된 음악저작물 및 어문저작물을 검색한다.

### 시나리오 4: 다른 표현으로 같은 대상을 나타내고 있는 정보 검색

대상과의 객체형속성을 통한 관계를 기반으로 다르게 표현하고 있는 인스턴스에 대해 검색한다.

### 시나리오 5: 저작물에 유일하게 부여된 메타데이터 요소의 값을 이용한 저작물 및 저작물 정보 검색

데이터형태속성으로 관계가 정의된 저작물의 유일한 메타데이터 요소의 값을 통해 클래스 및 클래스 계층구조, 추론을 이용한 저작물을 검색한다.

키워드 기반 검색 방법과 의미 기반 검색 방법의 검색 결과를 비교하기 위해 시나리오의 대상을 실제 존재하는 저작물을 대상으로 하였다. 저작물 온톨로지 구성 요소를 통해 실제 저작물을 완전하게 표현할 수 있다. 시나리오 1, 2는 음악저작물로 분류할 수 있는 국내가요 가수 다비치의 '8282'를 선정하였다. 시나리오 3을 위해서는 '저 꽃이 불편하다'라는 시집을 선정하였다. 시집은 하나의 도서에 여러 어문저작물을 포함하고 있기 때문에 시나리오를 위한 좋은 사례로 판단하였다. 시나리오 4를 위해서는 가수 성시경의 앨범 "THE BALLADS"를 선정하였다. 앨범의 경우 앨범 발매 당시 제작과 유통을 담당하는 제작사 및 유통사에 대한 정보를 포함하고 있는데, 이

후 제작 및 유통사의 상호가 변경된 경우가 발생할 수 있다. 하지만 변경된 정보가 저작물의 메타데이터에 적용되지 않았을 때 같은 대상에 대해서 다르게 표현하고 있는 경우가 존재하여 시나리오를 위한 적절한 예로 판단하였다. 시나리오 5는 ICN, 국제표준식별체계(UCI: Universal Content Identifier, 이하 UCI)[14] 등의 고유식별번호를 포함하고 있는 공공저작물을 대상으로 선정하였다.

표 3. 시나리오 질의 구성  
 Table 3. Scenario Query

번호	SPARQL 질의
1	<pre>SELECT ?곡제목 ?작곡가 ?작곡가통합저작권자ID       ?작사가 ?작사가통합저작권자ID       ?편곡가 ?편곡가통합저작권자ID WHERE { ?곡제목:음원명을_가지다:8282.   ?곡제목:작곡가를_가지다 ?작곡가.   ?작곡가:통합저작권자ID ?작곡가통합저작권자ID.   ?곡제목:작사가를_가지다 ?작사가.   ?작사가:통합저작권자ID ?작사가통합저작권자ID.   ?곡제목:편곡자를_가지다 ?편곡가.   ?편곡가:통합저작권자ID ?편곡가통합저작권자ID. }</pre>
2	<pre>SELECT ?노래 ?이용자 ?서비스타입 ?계약기간 ?자동연장여부 WHERE { ?노래:음원명을_가지다:8282.   ?이용자:이용자가_되다 ?노래.   ?이용자:서비스타입 ?서비스타입.   ?이용자:계약기간 ?계약기간.   ?이용자:자동연장여부 ?자동연장여부.   ?이용자:홈페이지 ?홈페이지. }</pre>
3	<pre>SELECT ?도서 ?어문저작물 WHERE { ?도서:도서명을_가지다:저쪽이불편하다.   ?어문저작물:도서가_되다 ?도서. }</pre>
4	<pre>SELECT ?앨범명 ?제작자 WHERE { ?앨범명:앨범명을_가지다:THE_BALLADS.   ?앨범명:제작자를_가지다 ?제작자. }</pre>
5	<pre>SELECT ?공공저작물 ?UCI ?원리관리기관 ?저작재산권 ?CCL ?OGL WHERE { ?공공저작물:ICN ?ICN.217=CM010220110086793**xsd:string.   ?공공저작물:UCI ?UCI.   ?공공저작물:신탁관리기관/원리기관을_가지다 ?원리관리기관.   ?공공저작물:저작재산권 ?저작재산권.   ?공공저작물:CCL정보 ?CCL.   ?공공저작물:OGL정보 ?OGL. }</pre>

각 시나리오를 SPARQL 질의로 변경하면 표 3과 같다. 표 4는 키워드 기반 검색 방법을 이용하여 각 시나리오에 대한 단일 키워드를 검색의 입력으로 사용했을 경우 나온 검색 결과이며, 표 5는 시나리오에 대해 SPARQL 질의로 작성한 의미 기반 검색 방법의 검색 결과이다. 표 4를 통해 시나리오 3을 제외한 나머지 모든 시나리오에 대한 키워드 기반 검색 결과는 이용자가 원하는 정보를 제공하지만 원하지 않는 검색 결과 또한 포함하고 있는 경우와 이용자가 입력한 키워드에 대해 어

떠한 결과도 제공하지 않는 경우가 존재함을 확인할 수 있다. 논문에 제시하지는 않았지만 각 시나리오에 대한 키워드의 조합을 검색의 입력으로 사용한 결과, 키워드의 교집합이 검색의 입력일 경우 입력한 키워드 조합에 대해 어떠한 검색 결과도 제공하지 않았으며, 키워드의 합집합을 검색의 입력으로 사용할 경우 키워드 조합을 구성하는 단일 키워드의 검색 결과로 제공되는 모든 검색 결과를 제공함을 알 수 있었다. 표 6은 키워드 기반 검색 방법과 의미 기반 검색 방법의 검색 결과의 정확도 비교를 위해 두 검색 방법의 검색 결과 중 시나리오별 가장 높은 정확도를 비교한 것이다. 정확도 P를 측정하는 방법은 전체 검색 결과를 A, 제공된 검색 결과 중 적합한 검색 결과를 B라고 하면 식 (1)과 같다[15].

$$P = \frac{B}{A} \times 100 \quad (1)$$

시나리오를 대상으로 키워드 기반 검색 결과 중 시나리오 3을 제외한 모든 검색 결과에 대해서 낮은 정확도를 확인할 수 있다. 키워드 기반 검색 결과의 정확도가 낮은 이유는 키워드의 교집합을 검색의 입력으로 이용할 경우, 정확도가 높아질 것이라는 가정과는 달리 어떠한 결과도 제공되지 않았으며, 키워드의 합집합을 검색의 입력으로 이용한 경우, 단일 키워드의 검색 결과로 제공되는 모든 검색 결과의 합을 제공함으로써, 식 (1)에 대입하면 전체 검색 결과는 증가하나 원하는 검색 결과의 수는 정해져 있기 때문에 정확도가 감소한다. 반면, 의미 기반 검색의 경우 높은 정확도를 보이는 이유는 검색의 입력으로 사용한 키워드의 관계 및 의미를 기반으로 원하는 정보를 검색 결과로 제공하기 때문이다.

하지만, 높은 정확도의 의미 기반 검색 결과를 얻기 위해서는 2개 이상의 메타데이터 요소가 필요하다는 가정이 필요하다. 단일 메타데이터 요소로 의미 기반 검색 방법을 이용한다면 키워드 기반 검색 방법의 검색 결과와 차이점을 찾을 수 없다. 위의 실험에서 알 수 있듯이 의미 기반 검색 방법의 경우 메타데이터 요소를 추가함으로써 정확도의 증가를 보이는 반면, 키워드 기반 검색 방법은 단일 키워드를 사용하는 경우에 비해 키워드의 조합을 검색의 입력으로 사용하기 위해 키워드의 수를 증가시킨다고 하여도 정확도의 증감을 예측할 수 없을 뿐만 아니라 조합할 수 있는 키워드의 수도 제한적이다. 따라서 메타데이터의 온톨로지 확장을 통한 의미 기반 검



표 4. 시나리오에 대한 키워드 기반 검색 결과  
Table 4. Result of Keyword-based-search about scenario

	입력 키워드	전체 검색 결과 (건)	적합한 검색 결과 (건)
시나리오 1	강은경	29	1
	김도훈	126	1
	이현승	47	1
	19910807000010000	.	.
	19980701000010000	.	.
시나리오 2	19912110000100000	.	.
	애플뮤	.	.
시나리오 3	음악감상형	.	.
	저꽃이불편하다	49	49
시나리오 4	서울음반	46	0
	로엔엔터테인먼트	4,745	0
시나리오 5	ICN.217=CM010220110086793	.	.
	G700:RA101+R100-pc0521000039	.	.

색이 가능하며, 메타데이터 요소가 검색의 입력이 될 경우 키워드 기반 검색 방법에 비해 의미 기반 검색 방법이 더욱 정확한 결과를 제공할 뿐만 아니라 결과에 대한 부가적인 정보의 제공이 가능함을 알 수 있다.

표 5. 시나리오 질의에 대한 결과  
Table 5. Result of Semantic-based-search about scenario

번호	질의 결과						
	곡제목	작곡가	작사가	편곡자	작곡가통합저작권자ID	작사가통합저작권자ID	편곡자통합저작권자ID
1	8282	김도훈	강은경	이현승	"19980701000010000"	"19910807000010000"	"19991211000010000"
	8282	이현승	강은경	이현승	"19991211000010000"	"19910807000010000"	"19991211000010000"
2	노래	이용자	서비스타입	계약기간	출처까지		
	8282	애플뮤	"음악감상형"~<http://n*2015-06-01~2024-12-31*~"www.mfu.co.kr"~<http://www.w3.org/2001/XMLSchema#anyURI>				
3	8282	주식회사 로엔엔터테인먼트	"연하를 이용한 서비스(발송"2014-12-15~2015-12-15*~"LISTEN"~<http://www.w3.org/2001/XMLSchema#anyURI>				
	8282	KB5미디어	"온라인거실 및 미니라이브"2014-03-01~2015-02-28*~"스마트누리/온라인서비스(동요"~<http://www.w3.org/2001/XMLSchema#anyURI>				
3	도서			어문저작물			
	저꽃이불편하다				웹미션에서		
	저꽃이불편하다				비수구미에서		
	저꽃이불편하다				그 여자		
	저꽃이불편하다				연평도의 말		
	저꽃이불편하다				나에게 묻는다		
	저꽃이불편하다				게, 한마리, 가고, 있다		
	저꽃이불편하다				말		
	저꽃이불편하다				길		
	저꽃이불편하다				길 위에서		
	저꽃이불편하다				해창에서		
	저꽃이불편하다				눈, 산		
	저꽃이불편하다				빛속에서		
	저꽃이불편하다				봄빛		
	저꽃이불편하다				겨울비		
저꽃이불편하다				물, 땅			
저꽃이불편하다				강가에와서			
저꽃이불편하다				결경			
저꽃이불편하다				行旅			
저꽃이불편하다				내가, 만난, 뒤			
4	발행명			저작자			
	THE BALLADS				서울음반		
5	공공저작물		UCI	권리관리기관	저작재산권	CCL	OGL
	김병서호적표	"G700:RA101+R100-pc0521000039"		국립춘천박물관	"복제배포전송권"	"CCL 해당없음"	"OGL 해당없음"



표 6. 키워드 기반 검색과 의미 기반 검색 결과의 정확도(%) 비교

Table 6. Comparison of precision of Keyword-based search result and Semantic-based search result

	키워드 기반 검색	의미 기반 검색
시나리오 1	4.16	100
시나리오 2	0	100
시나리오 3	100	100
시나리오 4	0	100
시나리오 5	0	100

#### IV. 결 론

본 논문은 이용자가 원하는 저작물 정보 검색에 대해 키워드 기반 검색 방법이 아닌 의미 기반 검색이 가능하도록 하기 위해 ICN이 부여된 저작물이 가지고 있는 ICN 메타데이터를 온톨로지에 적용하는 방법을 제시하였다.

온톨로지의 도메인은 ICN을 부여한 저작물로 선정하였으며, 도메인에 대한 중요 용어는 ICN 메타데이터를 바탕으로 사용하였다. “Ontology Development 101” 방법론과 OWL 언어를 사용하여 저작물 온톨로지를 구축하였으며, OWL 언어로 작성된 온톨로지의 공리를 이용한 추론의 과정 및 결과를 Protégé 4.3의 내장된 추론기를 통해 보였다. 또한 저작물 온톨로지에서 추론을 포함한 의미 기반 검색의 가능성 및 키워드 기반 검색 방법의 검색 결과를 비교하기 위해 5가지 시나리오를 선정하였다. 시나리오가 의미 기반 검색이 가능하도록 온톨로지질의 언어인 SPARQL 질의를 이용하였다. 메타데이터의 온톨로지 확장을 통해 의미 기반 검색의 가능성과 키워드 기반 검색 방법의 검색 결과 및 검색 결과의 정확도 비교를 통해 의미 기반 검색 방법이 키워드 기반 검색 방법에 비해 보다 정확한 검색 결과를 제공함과 제공된 결과에 대해 부가적인 정보를 제공할 수 있음을 알 수 있다.

향후 연구로는 단일 메타데이터 요소로 의미 기반 검색의 정확도를 높이기 위해, 저작물을 식별할 수 있는 식별메타데이터를 제공하는 UCI를 온톨로지 확장과 저작물 온톨로지와 통합<sup>16)</sup>하여 저작물에 대한 UCI 식별 메타데이터와 ICN 메타데이터가 상호보완된 저작물의 메타데이터를 통해 보다 정확한 저작물 및 저작권 정보를 제공할 수 있는 방안을 제시하고자 한다.

#### References

- [1] Korea Digital Copyright Exchange: www.kdce.or.kr.
- [2] Protégé 4.3: protege.stanford.edu/.
- [3] Sang-Hoon Oh, Jung-Jea Lee, “A Study on the Development of the Digital Content Service Scenario, An Integrated Structure of ICN and UCI Identifier System.”, The Journal of Korea Society for Information Management, vol.26 no.3, pp.317-333, 2009.
- [4] Yoon-Ho Kim, Seong-Hwan Cho, “A Study on the Linkage and Integration of UCI (Universal Content Identifier) between ICN (Integrated Copyright Number)”, The Journal of The Institute of Internet, Broadcasting and Communication (JIIBC), vol.14 no.5, pp27-33, 2014.
- [5] Eun-Kyong Kim, Young-Joon Nam, “The Comparative Study on the Methodologies of Building Ontology toward Semantic Web”, Journal of Information Management, vol.35 no.2 p57-85, 2004.
- [6] Natalya F. Noy, Deborah L. McGuinness, “Ontology development 101: A guide to creation your first ontology”, <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>.
- [7] Do-Heon Jeong, “A Study on the Ontology Language and Application Systems for the Semantic Web”, Journal of Information Management, vol.34 no3 pp88-109, 2003.
- [8] OWL: <http://www.w3.org/TR/owl-features>.
- [9] EunOk Ha, Yoonho Kim, “Applying Ontologies to UCI for the Efficient Search and Management of

Digital Contents”, The Journal of Society for e-Business Studies, vol.14 no.4 pp215-228, 2009.11.

- [10] Hyeon-geun Park, “Reasoning on Production of Rules and Facts Using OWL-based Semantic Web Ontology”, Chung-Ang University, Ph.D. thesis, 2004.12.
- [11] HyenMin Kang, “A Study on the Description of Personal Name Access Point Control Ontology Using Axiom Definition”, Journal of the Korean society for library and information science, vol.46 no.2 pp157-174, 2012.05.
- [12] SPARQL, [www.w3.org/TR/rdf-sparql-query](http://www.w3.org/TR/rdf-sparql-query)
- [13] Yuanbo Guo, zhengxiang Pan, Jeff Hflin, “LUBM: A Benchmark for OWL Knowledge Base Systems”, Journal of Web Semantics, vol.3, 2005.
- [14] Universal Content Identifier, [www.uci.or.kr](http://www.uci.or.kr).
- [15] Jagendra Singh, Dr. Aditi Sharan, “A Comparative Study between Keyword and Semantic Based Search Engines”, International Conference on Cloud, Big Data and Trust, 2013.
- [16] In K. Lee, Dosam Hwang, “An Ontology Integration Process”, Annual Conference on Human and Language Technology, 19th, 2007.

## 저자 소개

### 유 민 규(정회원)



- 2014년 : 상명대학교 컴퓨터과학부 (학사)
- 2014년 ~ 현재 : 상명대학교 대학원 컴퓨터과학과 석사과정

<주관심분야 : 분산시스템, 저작권보호기술, 디지털콘텐츠>

### 김 윤 호(정회원)



- 1985년 : 서울대학교 계산통계학과 (학사)
- 1987년 : 서울대학교 대학원 계산통계학과 (이학석사)
- 1996년 서울대학교 대학원 전산과학 박사
- 1997년 ~ 현재 : 상명대학교 소프트웨어대학 컴퓨터과학부 교수

<주관심분야 : 분산시스템, 저작권보호기술, 디지털콘텐츠>

※ 본 논문은 상명대학교 교내연구비를 지원받아 연구되었습니다.  
(This research project was supported by Research Fund from Sangmyung University)