

# 온톨로지 기반 개인 미디어 메타데이터 자동 생성

최정화<sup>o</sup> 서희철<sup>+</sup> 박영택

송실대학교 컴퓨터학과, 한국전자통신연구원<sup>+</sup>

cjh79@aillab.ssu.ac.kr<sup>o</sup>, hcseo@etri.re.kr<sup>+</sup>, park@ssu.ac.kr

## Ontology-based Metadata Automated Generation for Personal Media

Junghwa Choi<sup>o</sup> HeeCheol Seo<sup>+</sup> Youngtack Park

School of Computing, Soongsil University

Electronics and Telecommunications Research Institute<sup>+</sup>

### 요 약

개인 디지털 콘텐츠 증가에 따른 개인 미디어의 관리를 위해 대량의 메타데이터를 자동으로 생성하는 연구가 반드시 필요하다. 본 논문에서는 온톨로지 기반의 추론을 이용하여 개인 미디어 메타데이터를 자동으로 생성하는 방법을 제안한다. 제안한 방법은 부족한 정보로부터 적합한 의미를 추출하여 메타데이터를 자동 생성하므로 콘텐츠관리의 어려운 문제점을 해결한다. 본 논문에서 제안하는 방법을 사용자가 메모를 부착하기만 하면, 온톨로지 기반 추론을 통해 메타데이터를 자동 생성하는 방법으로 다음과 같은 세 가지 기술과 특징을 갖는다. 첫째, 개인 미디어 온톨로지를 정의한다. 둘째, 미디어 메타데이터 표준을 정의한다. 미디어의 종류가 다르더라도 정의한 표준의 키워드만 추출할 수 있다면 미디어의 통합관리가 가능하다. 셋째, 메타데이터 자동 생성 기술을 연구한다. 단순히 온톨로지에 정의된 키워드의 의미만을 보지 않고, 온톨로지 기반의 추론엔진을 이용하여 사용자들 중심으로 관련 키워드의 관계를 고려한 메타데이터 생성의 정확성을 높인다. 이러한 기술을 기반으로 시맨틱 검색도 가능하며, 기존의 메타데이터 저작도구와 비교하여 보다 정확한 메타데이터 자동생성과 검색이 가능하다.

### 1. 서 론

개인 미디어는 디지털화된 인간 경험, 즉 개인이 일상을 보내면서 생성, 접근, 수집, 그리고 소장된 디지털 정보들이다. 예를 들면 이미지, 동영상, 텍스트 그리고 음악파일 등을 포함한다. 현재 컴퓨터 사용자들은 이와 같은 디지털 정보들을 web, e-mail, 검색엔진, 블로그 등을 통해 상호 공유함으로 개인콘텐츠가 폭발적으로 증가하고 있다. 이에 따른 개인 미디어 관리의 필요성이 대두되고 있으며, 이를 위해서 개인 미디어의 관리를 위해 메타데이터(metadata)를 자동 생성하는 연구가 반드시 필요하다. 하지만, 현재의 기술 수준으로는 이들의 상호 연관성 및 인덱싱 규칙을 발견하기 어려우며, 저장 시스템이 정보의 의미를 이해할 방법을 제시하지 못하고 있다 [1]. 즉 디지털화된 개인 경험을 저장 및 검색하기 어려움이 있다. 따라서 의미 수준에서 개인 콘텐츠를 분석하고 관리하는 연구가 절실히 필요하다.

본 논문에서는 형식(format)이 비슷한 개인 미디어에 대해 사용자가 메모만 부착하면 자동으로 메타데이터를 생성하는 것이 목적이다. 또한 모호성 해결 알고리즘을 제안하여 메모에 포함된 개념의 모호성을 해결하여 메타데이터 생성의 정확성을 향상시킨다.

사람이 편리하게 서로 다른 특성을 가진 디지털 정보를 찾거나 서비스를 받기 위해서는 사람의 일을 대신 할 수 있는 지능형 에이전트가 필요하다. 시맨틱 웹(semantic web)은 지능형 에이전트가 웹에 있는 정보의

의미를 이해하기 위한 공간을 제공한다. 시맨틱 웹 환경에서는 기존의 웹 공간과 이 웹에 있는 정보를 표현하는 메타 공간으로 구성된다. 이 메타 공간에 있는 정보는 웹 공간에 있는 문서의 의미를 표현하고 있으며, 지능형 에이전트가 메타공간에 있는 정보를 이해할 수 있도록 설계되어 있다. 따라서 지능형 에이전트가 이해할 수 있는 단순한 언어를 이용하여 문서의 내용을 메타데이터 형태로 표현하면 된다 [2, 3, 4].

메타데이터는 웹상의 문서나 정보의 의미를 표현하는데, 예를 들면, 그 문서의 주요 개념(concept), 저자, 저작 날짜 등이다. 그리고 사람이 생각한 시맨틱 개념을 지능형 에이전트가 이해할 수 있도록 표현한 것을 온톨로지(ontology)라고 한다. 온톨로지는 메타데이터를 표현하기 위한 해당 분야의 개념이며, 공유가 가능하다.

본 논문에서는 개인 PC에 저장된 이미지, 동영상 등의 동일한 특징 추출이 가능한 개인 미디어의 관리를 위해 색인, 분류, 공유, 검색 등의 과정을 통합적으로 지원하는 메타데이터 자동 생성 기술을 연구한다. 본 논문에서 제안하는 방법은 사용자가 미디어에 메모를 부착하기만 하면, 온톨로지 기반 추론을 통해 메타데이터를 자동 생성하는 방법으로 다음과 같이 세 가지 기술로 분류하고 특징을 설명한다. 첫째, 개인 미디어 온톨로지를 정의한다. 미디어의 공유 및 사용자의 미디어 입력 행위에 의해 도메인의 확장이 가능하다. 둘째, 미디어 메타데이터 표준을 정의한다. 미디어의 종류가 다르더라도 정의한 표준의 키워드만 추출할 수 있다면 미디어의 통합관리가 가능하다. 셋째, 메타데이터 자동 생성 기술을 연구한다.

단순히 온톨로지에 정의된 키워드의 의미만을 보지 않고, 온톨로지 기반의 추론엔진을 이용하여 사용자를 중심으로 관련 키워드의 관계를 고려한 메타데이터 생성의 정확성을 높인다. 이러한 기술을 기반으로 시맨틱 검색도 가능하며, 따라서 기존의 메타데이터 저작도구와 비교하여 단어의 모호성을 해결한 보다 정확한 메타데이터의 생성과 검색이 가능하다.

2. 관련 연구

시맨틱 웹이 출현한 이유는 지능형 에이전트가 웹의 정보를 이해할 수 있는 공간을 구축하기 위함인데, 이와 같은 공간이 결국은 메타데이터이다 [1, 3, 5]. 이러한 필요성으로 인하여 많은 연구들이 메타데이터를 생성하기 위해 시맨틱 웹 기술을 적용하여 제안하고 있다 [7, 8, 9, 11, 12, 13].

이러한 방법들을 간략히 살펴보면, Annotea [6]는 정보의 추출은 고려하지 않았지만, 웹 문서로부터 온톨로지에 정의된 개체(Entity)를 인식하여 RDF형식의 메타데이터를 자동 생성한다. 또한 메타데이터의 공유를 목적으로 검색을 제공한다. COHSE(Conceptual Open Hypermedia Services Environment) [7]는 RDF보다 어휘의 표현력이 풍부한 DAML+OIL 온톨로지 언어를 기반으로 웹 문서의 메타데이터를 자동 생성한다. KIM(Knowledge and Information Management) [8]은 RDF를 기반으로 모호성을 고려하여 웹문서에 메타데이터를 자동 생성한다. 모호한 단어인 경우 일반적으로 많이 쓰이는 개체로 의미를 부여한다. 단어들 간의 의미는 고려하지 않았으며 규칙기반의 정보추출을 제공하는 GATE[9] 시스템을 이용하여 자연어 처리를 목적인다. ALPHA[10]에서는 하나의 문장을 RDF 그래프로 표현하여 문장에 포함된 개념간의 관계를 추출한다. 하지만 특정 도메인 한정되어 범위가 작다. CREAM(Creating Relational, Annotationbased Metadata) [11]과 SemTag [12]은 온톨로지를 기반으로 텍스트의 모호성을 고려하였다. CREAM은 모호한 텍스트를 웹문서에 포함된 텍스트들과의 관계를 분석하여 메타데이터를 반자동으로 생성한다. SemTag는 모호한 텍스트에 근접한 텍스트를 추출하여 온톨로지 상에 포함되는 영역을 비교하여 모호성을 해결하였다. SemTag는 도메인의 범위가 큰 TAP [13] 온톨로지를 적용하였으며, KIM과 ALPHA와 같이 자연어처리를 목적으로 메타데이터의 자동 생성을 제안하였다.

이와 같이 메타데이터를 생성하는 방법은 크게 두 가지로 구분할 수 있다. 첫 번째 방법은 사용자가 온톨로지 기반의 저작 도구를 이용하여 수동과 반자동으로 메타데이터를 생성하는 방법이고 두 번째 방법은 사용자가 텍스트만 입력하면 온톨로지 기반의 생성도구가 자동으로 메타데이터를 생성해주는 방법이다. 두 방법은 메타데이터 생성의 정확성과 도메인의 규모에서 차이가 크게 난다.

3. 온톨로지 기반 개인 미디어 메타데이터 자동 생성

본 논문은 개인 미디어의 효율적인 관리를 위해 시맨틱 웹 기술을 적용하여 자동으로 메타데이터를 생성하는 방법을 제안한다. 제안한 방법은 개인 미디어 온톨로지 구축 기술과 미디어 메타데이터 표현 기술, 그리고 미디어 메타데이터 자동생성 기술로 구분된다. 그림 1은 본 논문에서 제안하는 개인 미디어 메타데이터 자동 생성의 전체적인 구조이다.

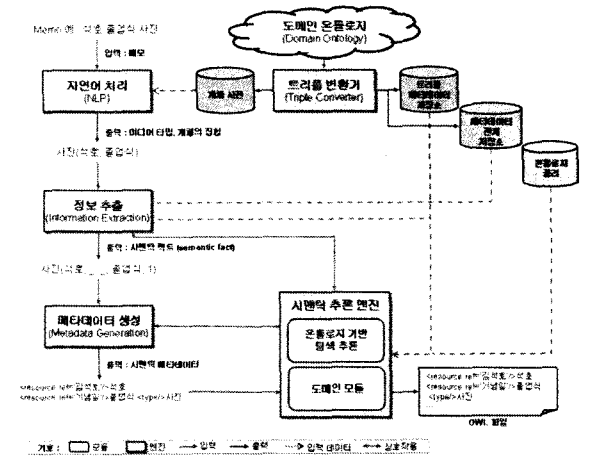


그림 1. 개인 미디어 메타데이터 자동 생성 시스템 구조

시스템 구조를 살펴보면, 먼저, 사용자가 미디어 관리 시스템에 삽입한 미디어에 부착한 메모가 입력된다. 자연어 처리 엔진에서 개체 사전에 저장된 개체 목록을 통해 메모에 포함된 개체와 미디어 타입을 구분하여 추출한다. 개체 사전은 온톨로지를 기반으로 트리플(triple) 변환기를 통해 추출된 개체 목록이다. 다음으로, 정보 추출 엔진에서는 자연어 처리엔진에서 추출한 개체의 모호성을 고려하여 시맨틱 팩트(fact)를 생성한다. 시맨틱 팩트는 본 논문에서 제시한 미디어 표준 형식이며, 미디어의 특징정보를 포함한다. 마지막으로, 메타데이터 생성 엔진은 정보 추출엔진에서 생성한 시맨틱 팩트에 의미를 부여하여 시맨틱 메타데이터를 생성한다. 정보 추출 엔진과 메타데이터 생성 엔진은 시맨틱 추론 엔진을 이용하여 온톨로지에 공리로 표현된 개체간의 관계와 규칙기반의 도메인 정보의 의미를 고려한다. 시스템 구조에 따른 메타데이터 생성을 통해 사용자에게는 생성된 OWL 파일의 내용을 텍스트 형태로 사용자 이해가 쉽게 제공한다.

3.1 개인 미디어 온톨로지 구축

개인 미디어 온톨로지는 개인 미디어에서 추출할 수 있는 키워드의 의미를 명확히 정의한다. 온톨로지는 개체간의 의미적 연관성을 이용하여 지능형 에이전트가 이해할 수 있는 문서를 만들어 시맨틱 웹의 구현을 가능하게 한다 [14]. 본 논문에서의 온톨로지는 사용자의 프로파일 정보를 포함하고 있어서 사용자 중심의 의미 분석을 통해 메타데이터 생성의 정확성을 높인다.

제안하는 개인 미디어 온톨로지는 키워드의 특징에 따라 핵심 온톨로지, 확장 온톨로지, 그리고 도메인 온톨로지 로 분류한다. 그림 2는 본 논문에서 제안하는 개인 미디어 온톨로지이다.

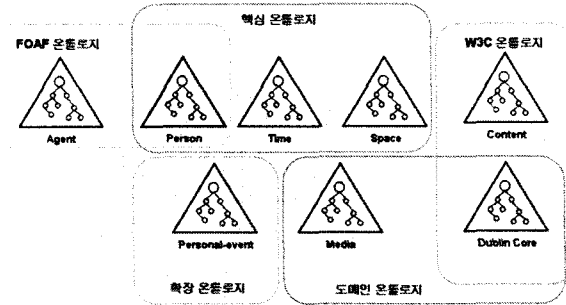


그림 2. 개인 미디어 온톨로지

본 논문에서는 지능형 에이전트의 처리를 위하여 온톨로지는 온톨로지 구축도구를 이용하여 OWL의 형태로 구현한다. OWL은 온톨로지 언어 중 가장 많은 의미 관계를 표현 할 수 있으며 RDF를 기반으로 한다. RDF는 레코드(record)를 하나의 기술 단위로 취급해온 기존의 방식과 달리 자원, 속성, 속성의 값을 하나의 단위로 취급하는 트리플 개념을 이용해서 에이전트의 처리가 가능하도록 각각의 자원, 속성, 속성 값들을 URI를 통해서 고유 식별자를 부여하고 상호간의 의미 관계를 정의한다 [15]. 본 논문의 온톨로지는 이러한 의미 관계를 통해 메모에 포함되지 않은 의미를 추출하여 메타데이터 생성을 가능하게 한다. 또한 온톨로지 표현에 있어서 OWL에서 제공하는 subClassOf와 sameAs 등 50여개의 공리를 사용하여 도메인 지식을 계층구조로 표현한다.

### 3.1.1 핵심 온톨로지

핵심 온톨로지로는 도메인에 독립적으로 사용될 수 있는 온톨로지로서 사람의 특징을 추출하여 개체와 개체간의 의미 관계를 정의한 Person 온톨로지와 시간의 특징을 정의한 Time 온톨로지, 그리고 장소에 따른 특징을 정의한 Space 온톨로지가 포함된다. Person 온톨로지는 W3C의 FOAF[16] 온톨로지를 기반으로 확장해서 구축하였다. W3C의 FOAF 온톨로지는 사람에 관계되는 특징을 추출하여 65개의 클래스와 속성으로 용어를 정의하였다. 클래스의 예로는 Agent, Person, Organization, Group등 있으며, 이 중 Person의 속성 예로는 사람의 이름을 의미하는 firstName, 사람과 관련된 이미지를 표현하기 위한 knows등이 있다. 이 용어들은 지능형 에이전트가 처리할 수 있게끔 OWL 온톨로지 언어로 정의된다. 제안한 Person 온톨로지에서는 FOAF의 Person과 Group 클래스에 포함된 일부 속성을 참조하여 확장하여 구축한다. Time 온톨로지는 시간에 관련된 계절과 기간의 특징을 정의한다. 계절은 봄, 여름, 가을 등을 정의하고, 기간은 낮과 밤의 하루를 기준으로 한 정보와 올해, 작년 말 등의 일년을 기준으로 한 정보의 의미 관계를

표현한다. Space 온톨로지는 국외, 국내, 실내, 야외로 나누어 지역 및 특정 장소를 표현한다. 핵심 온톨로지는 다른 도메인에서도 공유 및 재사용이 가능하다.

### 3.1.2 확장 온톨로지

확장 온톨로지는 본 논문의 도메인에서만 사용될 수 있는 메모에 포함될 수 있는 특징 데이터를 추출한 온톨로지로서 사용자의 행위에 따른 확장이 가능하다. Personal-event라는 클래스가 포함되는데 미디어 정보에 포함될 수 있는 일상생활의 이벤트들의 특징을 추출하여 의미를 표현한다. Personal-event 클래스는 휴가, 여행, 기념일 등을 서브클래스를 가지며, 상위 클래스인 Personal-event가 포함하는 이벤트들의 공통 속성을 상속받는다. 휴가의 예로는 여름휴가, 겨울휴가 등이 있으며, 여행의 예로는 국외와 국내로 나누어 미국과 제주도, 인천 등이 포함된다. 기념일의 예로는 신혼여행, 결혼식, 생일 등이 있다. 확장 온톨로지는 공유 및 재사용보다는 사용자가 생성하는 메모에 포함된 특징 있는 정보들의 의미를 부여한다.

### 3.1.3 도메인 온톨로지

도메인 온톨로지는 본 논문의 도메인인 미디어 관련 특징 정보를 정의한다. W3C의 Dublin Core 온톨로지를 확장하여 표현하였으며, 개인 미디어 메타데이터 생성 시 포함되는 미디어의 특징을 포함한다. DublinCore 클래스로부터 사용하는 속성으로는 미디어의 타입, 작성자, 제목, 날짜 등이 있으며, 이 속성을 상속받은 Media 온톨로지는 본 연구에서 제안하는 미디어 메타데이터 표준 형식의 속성을 포함한다. 미디어 메타데이터 표준 형식의 속성은 다음 장에서 다룬다. 도메인 온톨로지는 핵심 온톨로지와 확장 온톨로지서 정의한 개체를 기반으로 의미가 부여되며 같은 특징을 추출할 수 있는 미디어 정보에 대해 재사용이 가능하다. 물체가 존재하는 환경에 대한 기본적인 인지가 가능하다.

## 3.2 개인 미디어 메타데이터 표현

본 절에서는 개인 미디어 정보의 특징을 추출하여 메타데이터를 생성할 수 있는 표준 형식을 설명한다. 메타데이터 표준 형식에 대해 설명하고 단어의 의미가 모호한 예에 들어 모호성 해결 알고리즘의 필요성을 언급한다. 마지막으로 모호성 해결 알고리즘에 대해 설명한다.

### 3.2.1 개인 미디어 메타데이터 표준 형식

메타데이터 표준 형식은 그림 3과 같이 사용자가 미디어 데이터에 대해 입력한 메모에 대해 자연어 처리를 통해 생성된 팩트를 미디어 메타데이터 표준 형식의 시맨틱 팩트로 정형화 한다.

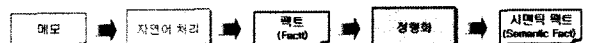


그림 3. 미디어 표준 형식 생성 과정

본 논문에서 제안하는 미디어 메타데이터 표준 형식은 표 1과 같다. 시맨틱 팩트는 Media<sub>i</sub>(W3E) 형식으로 구

성된다. *Media*는 미디어의 타입을 의미하며, 사진과 동영상에 포함된다. *W3E*는 사진과 동영상의 의미를 포함할 수 있는 특징 정보들이다. 이 특징 정보는 60명의 사용자로부터 미디어 파일에 부착할 메모를 조사하여, 메모의 의미를 해석할 수 있는 특징들을 추출한 것이다. *Who*는 미디어 파일에 포함된 사람을 명시한다. *Where*는 미디어 파일을 찍은 장소를 의미하고 *When*은 미디어 파일이 생성된 때를 의미한다. *Event*는 미디어 파일이 어떤 상황이 일어났을 때 생성되었는지를 나타내며, *Persons*는 미디어 파일에 포함된 사람 수를 의미한다.

표 1. 미디어 메타데이터 표준 형식

미디어 메타데이터 표준 형식
<i>Semantic fact, sf</i> = <i>Media</i> ; <i>(W3E)</i> ,
<i>W3E</i> = { <i>Who, Where, When, Event, Persons</i> }.
<i>Media</i> = {사진, 동영상}

그림 4는 하나의 예를 통한 시맨틱 팩트의 생성 과정이다.

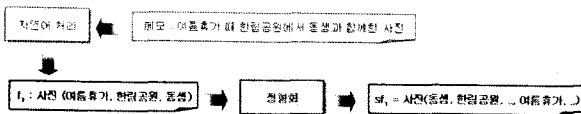


그림 4. 메모의 미디어 메타데이터 표준 형식으로의 생성 과정

자연어 처리 엔진은 “여름휴가 때 한림공원에서 동생과 함께”라는 메모가 입력되면 지능형 에이전트가 처리할 수 있게끔 사진(여름휴가, 한림공원, 동생)의 팩트를 생성한다. 사진은 미디어의 타입을 의미하며, 괄호안의 단어는 개체의 집합이다. 자연어 처리 엔진은 본 연구의 범위에 포함되지 않으며, 팩트는 온톨로지에 정의된 개체들을 기반으로 생성한 개체 사진을 통해 추출된 결과로 본다.

### 3.3 미디어 메타데이터 자동 생성

본 절에서는 본 논문에서의 온톨로지를 기반으로 추론 엔진을 이용하여 메타데이터를 자동 생성하는 방법을 설명한다. 그림 5는 시맨틱 추론 엔진의 개념도이다. 본 논문에서는 OWL로 작성된 온톨로지를 지능형 에이전트가 처리할 수 없기 때문에 트리플 변환기를 통해 트리플 형식으로 변환한다. 기존의 RDF와 DAML+OIL에서 정의된 공리는 일차논리(First Order Logic) 표현방식에 따르고 있다. 즉 추론엔진을 통해 일차논리형식의 추론을 하기 위해서는 온톨로지를 <property> <subject> <object> 형태의 트리플 형식으로 변환해야 한다. 본 시맨틱 검색 시스템의 트리플 변환기는 Jena의 RDF 파서를 사용하여 구현하였다 [17, 18]. 이와 같은 방식으로 온톨로지 스키마와 인스턴스의 개체들을 트리플 변환기를 통하여 PSO형태의 트리플로 자동 변환하고 다시 시맨틱 추론 엔진의 형태에 맞게 술어 형태로 변환 후 추론 엔진의 작업메모리에 넣어 온톨로지 OWL의 공리와

메타데이터를 생성하고자 하는 시맨틱 팩트를 바탕으로 구성된 규칙을 적용하여 전방향 추론을 수행한다.

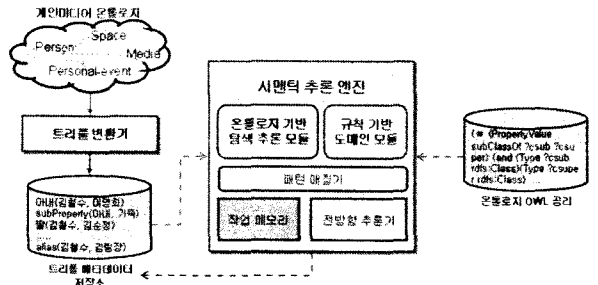


그림 5. 시맨틱 추론방식 개념도

그림 5의 시맨틱 추론 방식을 통해 시맨틱 팩트의 의미를 부여하여 메타데이터를 생성하게 된다. 본 논문에서는 사용자가 입력한 메모의 키워드에 정확하게 매칭되는 부분이 트리플 메타데이터 저장소에 없을 경우에 시맨틱 추론 원리를 사용하여 추론을 한다. 본 논문에서는 시맨틱 추론을 수행하기 위해서 두 단계의 추론 방법을 거친다. 첫 번째는 온톨로지 기반의 탐색 추론 방법으로 OLW 공리를 이용하여 온톨로지 상에 표현된 개체 간의 관계를 통해 추론을 하는 방법이다. 두 번째는 규칙 기반 도메인 추론 방법으로 도메인 온톨로지를 기반으로 사용자의 메모와 메타데이터의 속성으로 표현된 부분이 정확하게 매칭 되는 부분을 규칙을 통하여 검색하는 방법이다.

#### 3.3.1 메타데이터 구축

3.3.1에서는 사용자가 미디어에 부착한 메모에 대해 자동으로 메타데이터를 생성하는 과정을 예로써 설명한다. 사용자가 “여름휴가 때 한림공원에서 동생과 함께”라는 메모를 입력하면, 자연어 처리 엔진(NLP)은 사진(여름휴가, 한림공원, 동생)라는 팩트를 생성한다. 이 팩트는 모호성을 해결한 알고리즘과 시맨틱 추론 엔진을 이용하여 시맨틱 팩트를 생성하고, 다음과 같은 과정으로 메타데이터를 생성한다.

- (1) 우선, 팩트에서 미디어 타입으로 구분된 ‘사진’을 추출하여 온톨로지 상의 Media 클래스의 type속성 값으로 입력한다. Media 온톨로지는 DublinCore 온톨로지의 하위 클래스로 존재하므로 DublinCore 온톨로지의 속성들을 상속받는다.
- (2) 다음으로 메모에서 추출한 특징들을 시맨틱 팩트의 속성인 Who, Where, When, Event, Persons에 매핑(mapping)하면서 메타데이터를 생성한다. 본 논문의 개인 미디어 온톨로지는 사용자의 프로필 정보를 포함하며, 추론엔진은 메타데이터 생성 시 사용자를 중심으로 의미를 부여하는 것이 특징이다. 따라서 우선, 메모를 생성한 사용자를 인식하고 사용자를 중심으로 분석한다. 이 예의 사용자는 Person 클래스의 인스턴스로 존재하는 ‘이영희’이다. Person은 속성으로 ‘동생’을 가지며 ‘이영희’의 동생은 ‘이순희’로 정의되어 있다. Media 클래스에

는 *Who, Where, When, Event, Persons* 시맨틱 팩트의 속성이 포함되어 있다.

(2)-① 속성 *Event*: 사진(여름휴가, 한림공원, 동생) 팩트의 첫 번째 속성인 '여름휴가'는 온톨로지 추론을 통해 Media클래스의 event속성 값으로 매핑 된다. Media클래스의 속성 event는 레인지로 Personal-event의 인스턴스를 가지도록 정의되어 있으며, Personal-event클래스의 인스턴스로 '여름휴가'가 존재한다. 따라서 시맨틱 팩트의 *Event*속성의 의미는 '여름휴가'로 부여된다.

(2)-② 속성 *Where*: 사진(여름휴가, 한림공원, 동생) 팩트의 두 번째 속성인 '한림공원'은 온톨로지 추론을 통해 Media클래스의 where속성 값으로 매핑 된다. Media클래스의 속성 where는 레인지로 Space클래스의 인스턴스를 가지도록 정의되어 있으며, Space클래스의 하위 클래스로 국내, 국내클래스의 하위 클래스로 제주도클래스에 관광영소라는 속성의 값으로 '한림공원'이 존재한다. 따라서 시맨틱 팩트의 *Where*속성의 의미는 '제주도', '한림공원'으로 부여된다.

(2)-③ 속성 *Who*: 사진(여름휴가, 한림공원, 동생) 팩트의 세 번째 속성인 '동생'은 온톨로지 추론을 통해 '이영희'의 동생으로 매핑 된다. Media클래스의 who속성은 레인지로 Person을 가리킨다. 따라서 시맨틱 팩트의 속성 *Who*는 '이영희', '이순희'로 의미가 정의된다. 속성 *Who*에 '이영희'가 포함되는 이유는 (2)-⑤에서 설명한다.

(2)-①, ②, ③단계에 따라 시맨틱 팩트의 *Who, Where, Event*속성의 값이 부여되어 사진([이영희, 이순희], [제주도, 한림공원], \_, 여름휴가, \_)의 시맨틱 팩트가 생성된다. 팩트에 존재하지 않아서 아직 의미가 부여되지 않은 *When*과 *Persons*의 의미부여는 다음과 같다.

(2)-④ 속성 *When*: 사진(동생, 한림공원, 여름휴가) 팩트에서 *When*에 대한 키워드를 추출해 낼 수 없다. 따라서 미디어에 기본적으로 존재하는 메타데이터를 기반으로 온톨로지 기반 탐색 추론 모듈 통해 *When*에 해당하는 의미를 부여하게 된다. 미디어는 종류, 크기, 생성 날짜 등의 미디어 생성 시 기본적으로 부여되는 메타데이터가 있다. 또한 Media클래스의 속성 *when*은 레인지로 Time클래스의 인스턴스를 가지도록 정의되어 있으며, Time클래스의 하위 클래스로 Season과 Period클래스가 존재한다. 미디어의 기본 메타데이터 중 생성날짜를 기반으로 속성 *When*의 값을 추론한다. 따라서 시맨틱 팩트의 속성 *When*은 미디어의 생성날짜에 대해 온톨로지를 기반으로 추론하여 '올해'와 '여름'에 찍은 미디어라는 값을 부여한다.

(2)-⑤ 속성 *Persons*: 본 논문에서 사용하는 자연언어 처리 엔진은 메모에 포함된 '와', '과', '함께'라는 조사와 팩트의 속성 *Who*에 해당하는 값을 분석하여 미디어에 포함되는 사람의 수를 인식할 수 있다. 따라서 시맨틱 팩트의 *Persons*에 해당하는 값은 '둘'로 매핑 된다.

위의 과정을 통해 생성된 미디어 메타데이터 표준 형식은 사진([이영희, 이순희], [제주도, 한림공원], ['올해', '여름'], 여름휴가, '둘')라는 시맨틱 팩트를 생성한다. 그림 6은 (2)-①에서 (2)-⑤의 과정을 그림으로 표현한 것이다.

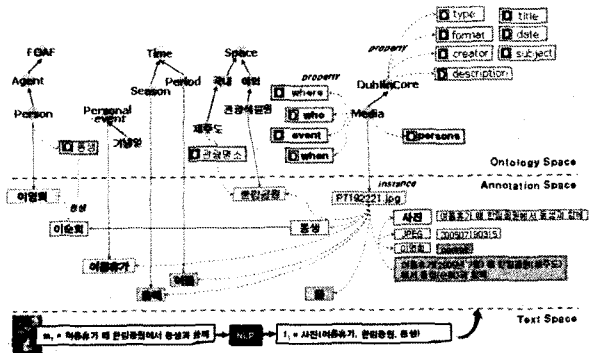


그림 6. 메타데이터 구축 과정

### 3.3.2 미디어 메타데이터 생성 예

그림 7은 3.3.1의 메타데이터 생성과정을 통해 생성된 개인 미디어 메타데이터이다. 왼쪽 상자는 메타데이터 생성과정을 통해 생성된 OWL 인스턴스이며, 오른쪽 상자는 사람의 인식을 위해 표기한 메타데이터이다. 본 논문에서 제시한 개인 미디어 표준 형식의 속성들, 즉 *Who, Where, When, Event, Persons*가 메타데이터 안에 표현된 것을 볼 수 있다.

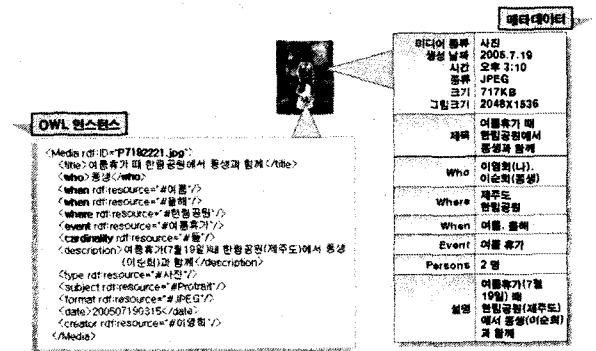


그림 7. 개인 미디어 메타데이터

### 3.3.3 온톨로지 기반 추론을 이용한 검색

본 절에서는 본 논문에서 제안하는 메타데이터 자동 생성 방법과 같은 방법으로 시맨틱 추론을 하였을 때, 3.3.2에서 생성된 메타데이터의 이미지를 찾는 과정을 설명한다. 이 과정을 통해 본 논문에서 제안하는 방법의 우수성을 증명한다.

그림 8은 온톨로지를 기반으로 추론하는 예이다. 3.3.1에서 제시한 예의 "여름휴가 때 한림공원에서 동생과 함께"라는 메모를 통해 생성된 메타데이터의 미디어 파일을 "올해 제주도에서 가족과 함께한 여행 사진"이라는 질의(query)가 들어오자 검색할 수 있음을 나타낸다. 우선 그림 8의 예제 질의는 자연어 처리 엔진으로부터 사진(올해, 제주도, 가족, 여행)이라는 팩트를 제공받고, 본 논문에서 제안하는 메타데이터 생성방법과 같은 과정

으로 사진(가족, 제주도, 올해, 여행, 가족 함께)의 시맨틱 팩트를 생성한다.

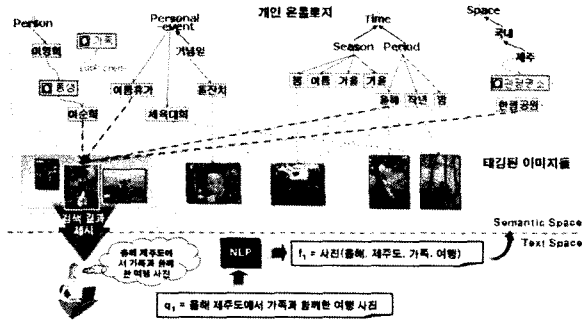


그림 8. 온톨로지 기반 추론 시맨틱 검색

본 논문에서 제안한 개인 미디어 온톨로지의 Person 클래스의 동생 속성은 가족 속성과 subPropertyOf 관계를 갖는다. 따라서 동생 속성의 값은 모두 가족 속성의 값으로 매핑 된다. 그리고 Space 클래스의 한림공원은 제주도의 인스턴스이다. 따라서 시맨틱 추론 엔진은 "올해 제주도에서 가족과 함께한 여행 사진"이라는 질의가 들어와도 "여행휴가 때 한림공원에서 동생과 함께"라는 메모로 생성한 미디어 파일을 검색 할 수 있다.

4. 결론 및 향후 연구

본 논문은 온톨로지 기반의 추론을 이용하여 개인 미디어 메타데이터 자동 생성 방법을 제안한다. 제안한 방법은 미디어 메타데이터 표준 형식의 속성을 추출하고, 추출된 속성 값은 규칙 기반 추론엔진을 이용해 의미를 부여하여 메타데이터로 생성된다. 본 논문에서는 사용자에 적합한 메타데이터를 생성하기 위해 온톨로지에 사용자 프로파일 정보를 포함하여 사용자의 특징을 중심으로 단어의 의미를 분석하는 추론 기법을 제안한다. 메타데이터를 생성하는 기존 방법들을 살펴보면, 특정 도메인에 대한 온톨로지를 기반으로 정의된 개념에 대해서만 메타데이터를 생성하여 명확히 개념과 일치하지 않는 정보에 대해서는 정확한 메타데이터를 생성할 수 없다. 그러나 제안하는 방법에서는 온톨로지 기반 추론 엔진을 통해 온톨로지 개념과 일치하지 않더라도 관련된 의미의 개념을 추출하여 정확한 메타데이터를 생성 할 수 있다. 본 논문은 특징이 같은 이미지와 동영상 미디어에 대해 사용자가 메모만 부착하면, 온톨로지 기반 추론을 이용하여 메타데이터를 자동 생성한다. 이 기술을 기반으로 시맨틱 검색도 가능하며, 자연어 처리 엔진은 연구하지 않았다.

참고문헌

[1] MyLifeBits: Fulfilling the Memex Vision  
 [2] Tim Berners-Lee, James Hendler, Ora Lassila, "The Semantic Web," Scientific American, May 2001.  
 [3] Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster, Spinning the Semantic Web, MIT Press, 2003.

[4] Grigoris Antoniou and Frank van Harmelen, A Semantic Web Primer, MIT Press, 2004.  
 [5] Dieter Fensel, Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, 2nd edition, Springer, 2004.  
 [6] J. Kahan and M. Koivunen and E. Prud'Hommeaux and R. Swick, "Annotea: Open RDF Infrastructure for Shared Web Annotations. In Proc. of the WWW10 International Conference. Hong Kong, 2001.  
 [7] S. Bechhofer and C. Goble, "Towards Annotation Using DAML+OIL. First International," Conference on Knowledge Capture (K-CAP 2001). Workshop on Semantic Markup and Annotation. Victoria, B.C., Canada. October 2001.  
 [8] Borislav Popov, Atanas Kiryakov, Damyan Ognyanoff, Dimitar Manov, Angel Kirilov, "KIM - a smantic platform for information extraction and retrieval," Journal of Natural Language Engineering, Vol. 10, Issue 3-4, pp. 375-392, 2004.  
 [9] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications". Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia, 2002.  
 [10] Y. Li, L. Zhang, Y. Yu, "Learning to Generate Semantic Annotation for Domain Specific Sentences," In: K-CAP 2001 Workshop on Knowledge Markup & Semantic Annotation, October 21, 2001.  
 [11] S. Handschuh and S. Staab and A. Maedche, "CREAM- Creating relational metadata with a component-based, ontology-driven annotation framework," First International Conference on Knowledge Capture (K-CAP 2001), Victoria B.C., October 2001.  
 [12] Stephen Dill, Nadav Eiron, David Gibson, Daniel Gruhl, R. Guha, Anant Jhingran, Tapas Kanungo, Sridhar Rajagopalan, Andrew Tomkins, John A. Tomlin, and Jason Y. Zien, "Semtag and seeker: Bootstrapping the semantic web via automated semantic annotation," in Intl.Conf. WWW2003, May 2003. Budapest, Hungary.  
 [13] R.Guha and R. McCool. Tap: Towards a web of data. <http://tap.stanford.edu/>.  
 [14] Thomas R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," Stanford Knowledge Systems Laboratory, August 1993.  
 [15] Smith, M., Welty, C., Deborah, L. and McGuinness, D. 2004. OWL Web Ontology Language Guide. W3C Recommendation, <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>  
 [16] "FOAF Vocabulary Specification" Namespace Document 27 July 2005, <http://xmlns.com/foaf/0.1/>  
 [17] Frank Manola, Eric Miller, "RDF Primer," W3C Working Draft 23 January 2003.  
 [18] Fikes, Richard, Jessica Jenkins, and Qing Zhou. "Including Domain-Specific Reasoners with Reusable Ontologies." Proceedings of the 2003 International Conference on Information and Knowledge Engineering, LasVegas, Nevada, USA, June 23-26, 2003.