

# 객체지향 시소리스를 RDF Schema로 변환하기 위한 매핑 구조의 설계 및 구현

양경아<sup>0</sup> 방선이 양재동  
전북대학교 컴퓨터정보학과  
{kayang<sup>0</sup>, sibang, jdyang}@ozzy.chonbuk.ac.kr

## Design of Mapping Structure and Implementation to transform Object-Oriented Thesaurus into RDF Schema

Kyoung-Ah Yang<sup>0</sup> Sun-Lee Bang Jae-Dong Yang  
Dept. of Computer Science, Chonbuk National University

### 요 약

본 논문에서는 객체지향 패러다임의 구조적 특성을 적용하여 만든 시소리스를 이용하여 웹 자원의 의미적 연관성을 RDF 형식으로 표현할 수 있는 RDF 문서 자동 생성기를 소개한다. 객체지향 시소리스를 RDF 형식으로 나타내기 위해 시소리스에 사용된 객체와 관계의 내용을 명세한 네임 스페이스를 정의하여 시맨틱 웹에서 메타데이터를 기술하는 표준이 되는 RDF와 RDF Schema 형식으로 매핑할 수 있다. 이러한 연구를 통해 객체지향 시소리스의 내용을 RDF 형식으로 표현함으로써 지식베이스(KnowledgeBase)라 불리는 온톨로지를 생성할 수 있으며, 구축된 온톨로지는 시맨틱 웹 상의 에이전트나 다른 어플리케이션에 응용될 수 있다.

### 1. 서 론

웹을 이용하는 사용자의 수는 지난 10년 동안 기하급수적으로 늘어났으며, 웹을 통해 접근할 수 있는 자료의 양도 그만큼 방대해지고 있다. 이제 급속도로 성장하여 거대해진 웹에서 사용자들은 원하는 정보를 추출하고 검색하는데 많은 시간과 노력을 들여야 한다.

이러한 문제를 해결하기 위해 많은 연구자들은 시맨틱 웹(Semantic Web)[1]이라 불리는 차세대 웹을 연구하기 시작했다. 시맨틱 웹은 기계로 하여금 정보가 지닌 의미를 충분히 이해함으로써 정보에 대한 지능적인 접근을 가능하도록 한다. 시맨틱 웹에서 기계가 정보의 의미는 물론 사람의 요구를 더 잘 이해하기 위해선 기계도 사람이 이해하는 만큼 각각의 사물이나 사건에 대한 공통된 특징을 파악할 수 있어야 한다. 이를 지원하기 위해 현재 여러 연구 단체에서 온톨로지(Ontology) 기술에 관한 연구를 하고 있으며, 그 결과로 지난 몇 년간 XOL, SHOE, DAML+OIL, OIL, RDF, RDF Schema 등의 온톨로지 언어가 개발되었다[2]. 특히 W3C의 연구단체들이 개발한 RDF[3]와 RDF Schema[4]는 웹 자원을 설명하는 메타데이터 스키마 표현을 위한 표준으로 자리 잡아 가고 있다. RDF 형식을 기반으로 온톨로지를 형성하는 시맨틱 웹이 현재의 웹을 대체할 방안으로 채택이 된다면, 현재 웹에 사용되는 기존의 기술들을 이 새로운 웹에 적용시키는 방식이 요구될 것이다.

본 연구에서는 시맨틱 웹에서 정보 검색을 하는데 있어 이미 존재하는 객체지향 시소리스[5]를 RDF 언어로 표현하여 온톨로지로 활용하기 위한 방안을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구에 대해 살펴보고, 3장에서는 객체지향 시소리스를 RDF Schema로 변환하기 위한 세 가지 단계를 설명한다. 4장에서는 객체지향 시소리스를 자동으로 RDF 문서로 생성하는 시스템의 구조에 대해 알아보고, 마지막으로 결론 및 향후 연구를 제시한다.

### 2. 관련 연구

시맨틱 웹에서 웹 지식을 구축하기 위해 또는 서로 다른 포맷을 가진 시소리스를 하나의 형식으로 단일화시키기 위한 목적으로 특정 시소리스를 RDF 형식으로 변환하는 프로젝트는 이전에도 있어왔다. ILRT(Institute for Learning and Research Technology)에서는 여러 시소리스의 개념적 관계들을 명세한 RDF Thesaurus Specification[6]을 지정하여 다양한 형식의 분류체제로 존재하는 시소리스들을 하나의 완전한 프레임으로 통합하는 방안을 제안하였다. 또한 LIMBER 프로젝트 팀에서는 TIF(Thesaurus Interchange Format) 형식을 설계하여 ISO에서 표준안으로 체결된 시소리스를 온톨로지 로 변환하기 위한 연구[7]를 현재 진행하고 있으며, 이미 존재하는 온톨로지와 시소리스 계층을 병합하여 RDF Schema를 작성하는 연구[8]도 다른 연구팀에 의해 수행되었다.

현재 다양한 방면에서 기존의 구축된 시소리스를 온톨로지화하려는 연구가 진행 중이며, 본 논문도 그러한 연구의 일환으로 객체지향 시소리스를 시맨틱 웹에서 온톨로지 로 활용하기 위해 네임스페이스와 RDF Schema 매핑구조를 정의하였으며, 시스템을 설계하고 구현하였다.

3. 객체지향 시소러스를 RDF Schema로 변환

객체지향 시소러스는 개념과 개념의 속성값, 개념들 간의 상속적 계층구조와 연관관계를 표현함으로써 구축된다. 시맨틱 웹에서 이러한 시소러스를 사용하기 위해서는 시소러스를 표준언어인 RDF로 변환하여야 하며, 이를 위해선 다음 세 가지 단계가 필요하다. 첫째, 시소러스 구조를 RDF로 표현하기 위한 네임스페이스를 정의하고, 둘째, 관계 표현을 위한 RDF Schema를 정의한다. 마지막으로 파서를 이용해 시소러스의 구조를 정의한 매핑 구조에 따라 RDF 형식으로 출력한다.

3.1 RDF 변환을 위한 네임 스페이스 정의

RDF는 자원을 기술하기 위해 XML 모델과 구문을 사용하고, XML 네임스페이스를 통해 같은 용어를 다른 의미로 정의해 사용함으로써 발생될 수 있는 충돌을 피한다. RDF 네임스페이스는 다음과 같이 정의된다.

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
```

그러나, 객체지향 시소러스 내에서 사용되는 객체와 관계들은 RDF Schema에 정의되어 있지 않기 때문에 이들을 RDF Schema로 적절히 표현하기 위해서 다음과 같이 'thrsr' 네임스페이스를 정의하였다.

```
<rdf:RDF
xmlns:thrsr="http://jdblab.chonbuk.ac.kr/project/Thrsr#">
```

객체는 RDF 태그의 Class로 선언하고 관계 정보들은 RDF 태그 내에 Property로 선언한다. 각 Property는 시소러스 스키마에서 정의한 'Concept' 클래스를 리소스(resource)로 하는 'domain'과 'range'의 Constraint Property를 가진다.

3.2 객체지향 시소러스의 RDF Schema 정의

객체지향 시소러스를 RDF 표현 방식으로 나타내기 위한 매핑구조를 표로 나타내면 다음과 같다.

객체지향 시소러스	RDF Schema	
객체	객체(용어)	<rdfs:Class>
관계	subclass of	<rdfs:subClassOf>
	synonym of	<thrsr:synonymOf>
	part of	<thrsr:partOf>
	association of	<thrsr:associationOf>
	instance of	<thrsr:instanceOf>

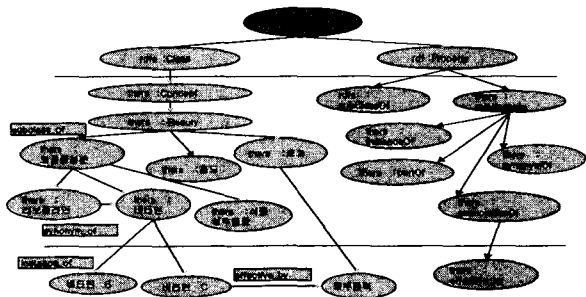
[표 1] 객체지향시소러스의 RDF Schema 매핑구조

우선 객체지향 시소러스의 객체는 RDF 형식으로 변환하기 위해 분류나 유형에 대응되는 클래스(Class)로 표현될 수 있다. RDF Schema의 rdfs:Class를 이용해 시소러스 개념노드를 클래스로 정의하고 이 클래스는 Concept과 Instance로 구분이 된다.

객체지향 시소러스에서 개념들 사이의 연관성을 나타

내는 관계 중 대부분은 RDF Schema에서 지원하지 않으므로 'thrsr' 네임스페이스를 지정하여 그 의미를 정의하고 이를 사용할 수 있다. 객체지향 시소러스에 쓰이는 관계들은 'subclass of', 'synonym of', 'part of', 'association of', 'instance of' 등 크게 다섯 가지로 나누어 볼 수 있다. 개념들 간의 상속적 관계를 나타내는 'subclass of'는 RDF Schema에 정의된 rdfs:subClassOf Property를 사용한다. 'subclass of'를 이용해 상위 클래스와 하위 클래스를 계층적으로 지정할 수 있으며, 한 클래스는 복수 개의 클래스의 하위 클래스가 될 수도 있다. 'synonym of', 'part of', 'association of', 'instance of' 등의 관계는 RDF Schema에 지정되어 있지 않으므로, 이들을 표현하기 위해 rdf:Property를 상속받은 thrsr:relationship을 정의하고 나머지 관계들을 thrsr:relationship의 하위 Property로 명세한다. 구체적으로 이를 살펴보면 우선, 시소러스 상에서 클래스가 다른 클래스와 동의어 관계임을 나타내는 'synonym of' 관계는 thrsr:synonymOf Property로 나타낸다. 특정 클래스가 계층 내에서 한 레벨 위의 클래스의 부분임을 나타내는 'part of' 관계는 thrsr:partOf Property로 나타내며, 클래스와 클래스간의 연관관계에 있다는 것을 의미하는 'associationOf'는 thrsr:associationOf Property로 명세한다. 마지막으로 개념과 객체간의 관계를 나타내는 'instance of'는 thrsr:instanceOf Property로 정의한다. 다음은 객체지향 시소러스의 관계와 RDF Schema의 매핑 구조를 표로 나타낸 것이다.

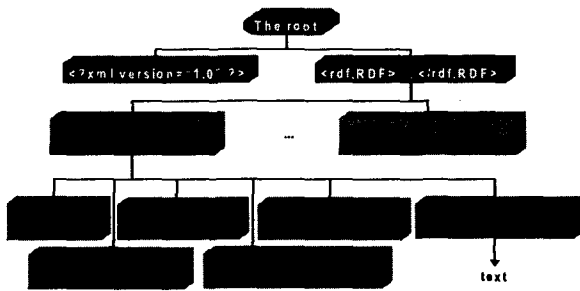
객체지향 시소러스의 객체와 관계를 RDF Schema로 매핑시킴으로써 시소러스의 내용을 RDF 언어로 표현할 수 있다. 다음은 객체지향 시소러스 중 화장품에 관련된 어휘를 구조화한 코스메틱 시소러스를 예로 든 것이다. 코스메틱 시소러스는 최상위 개념으로 'Beauty'라는 용어(Term)가 있고, 이는 '화장품성분', '용도', '효과' 등의 용어를 자식 노드로 가진다. 'Beauty'와 자식 노드들은 'subClassOf' 관계를 가지게 되며, '화장품성분'은 '용도'와 '효과'와 'associationOf' 관계에 있다. 또한 '화장품성분'의 자식 노드인 '비타민'은 '비타민B'와 '비타민C'를, '효과'는 '피부탄력'을 각각 Instance로 가지며 '비타민 C'는 '피부탄력'에 효과가 있다는 관계를 정의한다. 코스메틱 시소러스를 RDF Schema 양식으로 나타내면 [그림 1]과 같다.



[그림 1] 코스메틱 시소러스를 RDF Schema로 표현

### 3.3. RDF 출력 과정

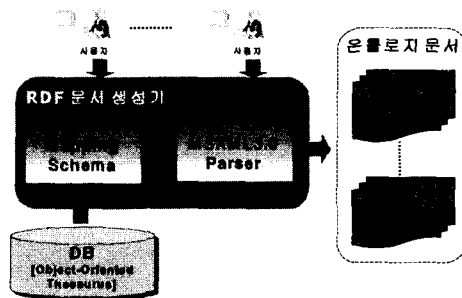
RDF는 XML 구문형식을 기반으로 하기 때문에 객체지향 시소러스를 RDF 문서로 출력하기 위해서 MSXML 3.0 parser를 이용한다. RDF 문서 출력 트리 구조는 [그림 2]와 같다. 먼저, RDF 문서의 Document Element로 <rdf:RDF>를 설정한다. DB에 저장되어 있는 객체지향 시소러스 중, RDF 문서로 출력하려는 시소러스의 노드에 해당하는 Concept들에 대해 각각 <rdf:Description> Element로 출력하고, 해당 Concept에 대해 'subclassOf', 'partOf', 'associationOf', 'instanceOf' 관계를 가지는 Concept과 'synonymOf'에 해당하는 Literal을 Child Element로 출력해준다.



[그림 2] RDF 문서 출력 트리 구조도

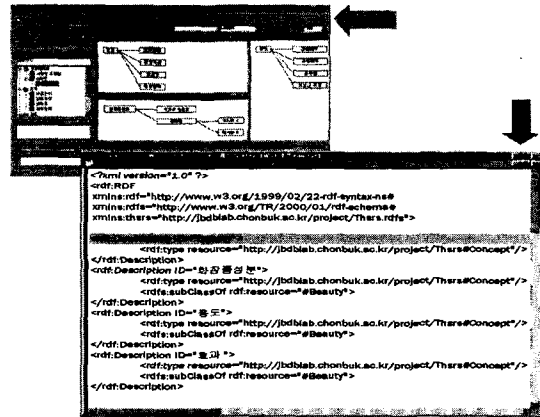
### 4. 시스템 구조 및 예

본 논문이 제안한 RDF 문서 생성기의 구조는 [그림 3]과 같다. DB 내에 시소러스에 관련된 정보를 저장하고 있고, 사용자가 변환하기 원하는 시소러스 도메인을 선택하면 정의된 매핑 스키마에 의해 시소러스가 RDF 스키마 구문으로 대응되고, 이를 출력 구조에 따라 XML parser가 문서로 출력을 하게 된다. 시스템 구현을 위해 Visual C++ 6.0을 이용하여 시스템을 작성하였으며, 객체지향 시소러스를 저장하기 위한 데이터베이스로 Oracle을 사용하였다. 또한 시소러스를 정의한 스키마는 MSXML 3.0 parser를 이용해 출력한다.



[그림 3] RDF 문서 생성기 구조도

다음은 객체지향 시소러스 편집기를 통해 본 코스메틱 시소러스와 RDF 문서 생성기를 이용하여 코스메틱 시소러스를 RDF Schema로 매핑한 후 RDF 문서로 출력한 예를 보인 것이다.



[그림 4] 코스메틱 시소러스와 RDF 문서 변환 결과

### 5. 결론 및 향후연구

본 논문에서는 시소러스를 이용하여 시맨틱 웹에서 온톨로지를 표현하고자 객체지향 시소러스를 RDF로 변환하는 네임스페이스와 매핑구조를 정의하였다. 이를 기반으로 객체지향 시소러스를 자동으로 RDF 문서로 변환하는 시스템을 구현하였다.

시소러스를 변환하여 RDF 형식을 기반으로 한 온톨로지를 구현한다면, 이를 여러 분야에 활용할 수 있다. 특히 시맨틱 웹 상에서 에이전트와 생성된 온톨로지를 연동하여 활용하는 방안에 대한 연구가 요구된다. 또한 변환의 대상이 되는 시소러스의 범위가 객체지향 시소러스에 국한되지 않고 범용적 시소러스로 확대되기 위한 연구도 수행되어야 할 것이다.

### 참고문헌

1. Tim Berners-Lee, James Hendler and Ora Lassila, "The Semantic Web", Science American, May 2001.
2. Asuncion Gomez-Perez and Oscar Corcho, "Ontology Languages for the Semantic Web", IEEE Intelligent System, Vol. 7 Issue 1, 2002.
3. O.Lassila and R. Swick, "Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation", WWW Consortium, 1999.
4. D.Brickey and R.V.Guha, "Resource Description Framework (RDF) Schema Specification 1.0, W3C Candidate Recommendation", World Wide Web Consortium, Boston, 2000.
5. 최재훈, 김기현, 양재동, "객체기반 시소러스 시스템의 설계 및 구현: 반자동화 방식의 구축, 추상화 방식의 개념 브라우징 및 질의기반 참조" 정보과학회 논문지(데이터베이스), Vol. 27, No. 1, pp.64-78, March, 2000.
6. Phil Cross et al., "RDF Thesaurus Specification(draft)", Institute for Learning and Research Technology, Irt.org/discovery/2001/01/rdf-thes
7. B.M.Mattuews et al., "Thesaurus Interchange Format in RDF", Semantic Web Conference 2002.
8. B.Amann, I.Fundulaki, "Integrating Ontologies and Thesauri to Build RDF Schema", International Journal of Digital Libraries, 1999.