

# 온톨로지 서브 그래프 검색 연구 동향

안후영\*, 박영호\*

\*숙명여자대학교 멀티미디어학과

e-mail : [hyahn85@sookmyung.ac.kr](mailto:hyahn85@sookmyung.ac.kr), [yhpark@sookmyung.ac.kr](mailto:yhpark@sookmyung.ac.kr)

## A Research Trend for Sub-graph Search on Ontologies

Hoo Young Ahn\*, Young-Ho Park\*

\*Dept. of Multimedia Science, Sookmyung Women's University

### 요 약

온톨로지는 특정 개념에 대한 부가정보 및 개념간의 관계를 기술하는 방법으로서 고차원의 웹 과 서비스를 실현하기 위한 시멘틱 웹, 그리고 지식관리 시스템을 비롯한 다양한 응용분야의 요구 와 관심이 증가하면서 그 중요성이 대두되고 있다. 온톨로지에서의 정보 검색은 특정 개념과 특정 관계를 가지는 데이터를 찾는 것이 주를 이루는데, 이러한 관계는 검색 대상에 대한 조상/ 후손을 모두 검색하여 찾아, 검색 결과를 반환해야 하기 때문에 정보를 처리하는 비용이 많은 비중을 차지 한다. 본 논문에서는 온톨로지의 효과적인 저장 및 검색을 위한 기존의 연구들의 동향을 살펴보고, 그 방법들에 대하여 정리하고, 보다 나은 방법에 대한 논의를 한다.

### 1. 서론

온톨로지는 철학의 ‘존재론’ 으로부터 그 기원을 찾아볼 수 있다. 고대 그리스의 존재론은 나를 포함한 이 세계가 어떻게 구성되어 있는가에 대한 문제 해결을 추구하였으며, 자연 법칙이나 실제 형태를 포함한 존재(being)에 관한 문제를 다룬다. 즉, 구성에 대한 정의를 하기 위해서는 용어에 대한 공통적인 표현이 필요하게 되며, 구성 개체간의 관계들에 대한 표현 및 의미를 부여할 수 있어야 했던 것이다. 존재론을 웹과 연관지어 살펴보면, URI(Uniform Resource Identifier)로 표현되는 자원은 모두 웹의 구성 개체가 되며, 구성 개체에 대한 정의 및 개체간의 관계들이 필요하게 된다.

그러나 현재의 HTML(Hypertext Markup Language) 기반의 웹은 시각적 효과를 위한 정보만을 컴퓨터에게 제공할 뿐 그 정보의 의미에 대한 정보는 제공하지 않는다. 따라서 W3C에서는 XML 및 RDF 를 기반으로 온톨로지 표현 언어를 설계하고 있다. 그러나 W3C 의 활동에서는 존재론에서 거론하는 모든 형상에 대한 표현은 실로 매우 어렵기 때문에 웹이라는 특정 분야 또는 각 응용 도메인 별의 국한된 범위 내에서의 표현만 가능하도록 웹 온톨로지 언어를 설계하고 있다. 초기의 온톨로지는 단순히 지식 공유와 재사용을 위해서 개발되었다. 최근에는 온톨로지의 개념이 지능적인 정보 통합, 협동 정보 시스템, 정보 검색, 전자 상거래, 지식 관리 같은 분야로 확대되고 있다. 또한 사람과 컴퓨터 사이에서 공유되고, 공통된 이해를 필요로 하기 때문에 온톨로지의 개념이 더욱 대중화되고 있다.

이러한 온톨로지는 그 종류와 크기가 매우 방대해

지며, 이에 대한 효과적인 검색 방법이 필요하다. 온톨로지를 효율적으로 저장 및 검색을 위한 다양한 방법들이 연구되어 왔으며, 온톨로지에 대한 시, 공간 효율적인 검색 방법에 대한 연구는 매우 활발히 진행되고 있다. 본 논문에서는 온톨로지에서의 서브 그래프 검색에 관련된 기존 연구를 살펴보고, 그 장, 단점을 분석한다.

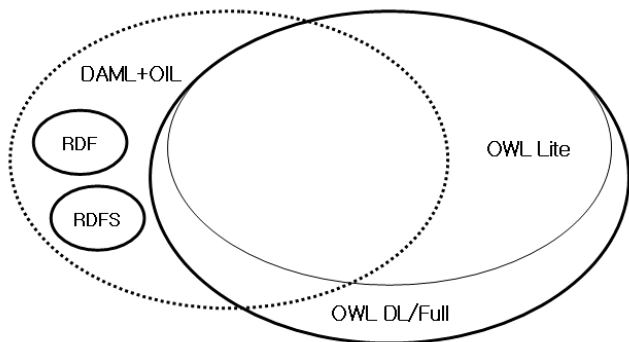
### 2. 관련 연구

온톨로지는 WWW[1] 에서 적용되는 시멘틱 웹을 가능하게 한다. 그리고 응용 프로그램 사이에서 웹 기반 지식 처리, 공유, 재사용 하는 것을 가능하게 하는 중요한 역할과 함께 사람과 응용 프로그램 시스템 사이에서 공통된 주제의 의사소통을 제공하여, 전자 상거래에서 구매자와 구입자간의 기계기반의 의사소통을 가능하도록 만들어 준다. 예로, 검색 엔진에서 찾으려고 하는 키워드와 의미적으로 유사하지만 구조적으로 다른 단어들을 가진 웹 페이지들을 찾아주는 역할을 한다.

시멘틱 웹을 위한 마크업 언어는 온톨로지를 표현하고 의미에 대한 정확한 정의와 데이터간의 논리적인 관계 설정 등의 기능을 가지고 있으며, 마크업 언어 기반의 에이전트를 이용하여 데이터간의 관계를 통한 정보추출과 추론이 가능해야 한다. 대표적인 인공지능 기반의 마크업 언어는 OIL(Ontology Inference Layer), DAML(DARPA Agent Markup Language), SHOE(Simple HTML Ontology Extensions)와 최근에 W3C 에서 개발중인 Notation 3(N3)가 있다.

OWL 은 문서에 포함된 정보를 어플리케이션을 이용하여 자동 처리하고자 할 때 활용하는 언어이다.

OWL 을 이용하면 임의의 어휘를 구성하는 용어 (term)의 의미와 용어들 간의 관계를 명시적으로 표현 할 수 있다. 이와 같이 용어와 용어들 간의 관계를 표현한 것을 온톨로지(Ontology)라 한다. OWL 은 XML, RDF, RDF-S 보다 더 많은 의미 표현 수단을 제공하므로, 웹 상에서 기계가 해석할 수 있는 콘텐츠를 작성하는데 있어 이들 언어보다 뛰어나다. OWL 은 DAML+OIL 웹 온톨로지 언어로부터 파생된 언어이다. OWL 에는 DAML+OIL 의 설계 및 활용 경험으로부터 습득된 지식이 반영되어 있다. 그림 1 은 온톨로지를 표현하는 언어 어휘 간의 포함 관계를 보인다.



(그림 1) 온톨로지 언어 어휘 간의 포함 관계

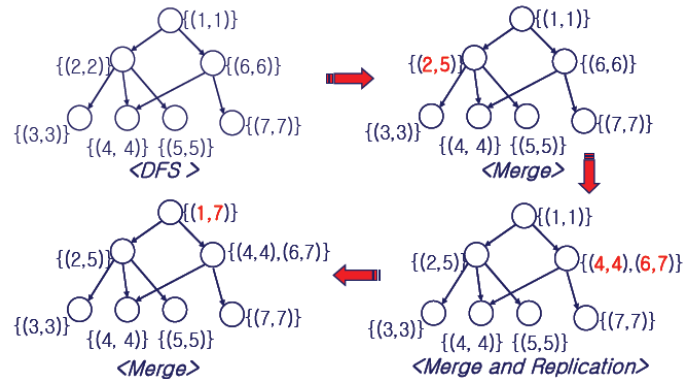
### 3. 본론

현재의 대표적인 온톨로지 저장소로서 특히 Jena[2]에서의 검색은 추론 엔진(reasoner)을 구현하여 제공하고 있으나 그 성능은 미약한 상태이고 특히 상/하위 개념 검색 질의를 처리하는 모듈인 transitive reasoner는 트랜지티브 리덕션(transitive reduction)을 통해 공간 효율성을 높이고 상/하위 개념 검색[3, 4]에 대해서는 메모리 캐쉬를 하는 단순한 방법을 사용하고 있다.

구체적으로 살펴보면, Jena 시스템은 RDF 그래프에 대한 추상화된 온톨로지 모델을 유지하는데 이때 중간에 Reasoner 에 의해 확장된 그래프를 모델로 가지게 된다. 이 확장은 Jena 의 Reasoner 패키지이고 그 중 transitive reasoning 에 관여하는 것이 transitive reasoner 이다. Jena 에서는 트랜지티브 리덕션을 사용하여 저장 공간의 효율성을 높이고 있지만, 검색의 상/하위 개념을 일반적인 그래프 탐색 기법을 사용하여, 모든 노드를 방문 한 후, 트랜지티브 클로저를 구하여 검색 결과를 반환 하기 때문에 온톨로지가 방대해 질수록 디스크 I/O 가 증가하여 검색의 성능이 저하되는 단점을 가진다. Jena 의 검색 방법은 저장 공간의 효율성은 있으나, 검색 시간의 성능이 매우 저하되는 단점을 가진다.

[5]는 Jena 의 단점을 해결하고자 온톨로지에 그래프 레이블링 기법을 사용하여, 저장 공간의 효율성을 도모하고, 그래프 레이블의 범위 비교를 통한 조상/후손 관계 검색을 통하여 검색 시간 성능의 향상을 도모한 연구이다. 위의 연구에서는 XML 의 구간 기

반 레이블링 기법을 온톨로지에 적용하여 레이블링 하고 있으며, 이러한 레이블링을 통해 그래프 탐색을 하지 않고도 레이블 정보만 보고 두 노드 사이의 연결 가능성(Reachability)을 결정할 수 있도록 하였다. 그림 2 는 구간 기반 레이블링 과정을 보인다.



(그림 2) 구간 기반 레이블링 과정

### 4. 결론

본 논문에서는 다양한 종류와 방대한 양의 온톨로지에서의 효과적인 검색을 위한 기존의 연구들의 동향을 정리하고, 그에 대한 장,단점을 분석하였다.

온톨로지에서의 정보에 대한 접근은 특정 개념과 특정 관계를 가지는 데이터를 찾는 것이 주를 이루는데, 이러한 관계에서는 검색 대상에 대한 상/하위개념에 대한 검색까지 확장되는 관계이기 때문에 특정 개념에 대한 검색이 재귀 호출의 형태로서 그 처리 비용 또한 매우 크다. 따라서, 이러한 질의에 처리를 지원하기 위하여, 공간 효율적이고, 시간 효율적인 검색 방법들이 많이 연구 되고 있다. 향후, 보다 지능화된 시스템을 위하여 온톨로지를 통한 검색은 더욱 다양한 분야에서 활용될 것이며, 그 검색 성능 향상을 위한 연구들을 더욱 활발히 진행 될 것으로 사료된다.

### 참고문헌

- [1] W3 Consortium (W3C): <http://www.w3c.org>
- [2] K. Wilkinson, C. Sayers, H. Kuno, D. Reynolds. "Efficient RDF Storage and Retrieval in Jena2", SWDB 2003.
- [3] J. Broekstra, A. Kampman, F.V. Harmelen. „Sesame: An Architecture for Storing and Querying RDF Data and Schema Information”, Semantics for the WWW, 2001
- [4] I. Horrocks, "Reasoning with Expressive Description Logics", Theory and Practice Proceedings of DADE-02, 2002. Springer-Verlab Lecture Notes in Artificial Intelligence.
- [5] 김종남, 정준원, 민경섭, 김형주, "온톨로지에서의 레이블링을 이용한 효율적인 트랜지티브 클로저," 정보과학회논문지: 데이터베이스, 제 32 권, 제 5 호, 2005