

트리를 이용한 효율적인 온톨로지 병합 알고리즘

김영태*, 임재현*, 공헌택*, 김치수*
*공주대학교 컴퓨터공학과
zerot@kongju.ac.kr

Efficient Ontology Merging Algorithm Using Tree

Young-Tae Kim*, Jae-Hyun Lim*, Heon-Tag Kong*, Chi-Su Kim*
*Dept of Computer Engineering, KongJu National University

요약

일반적으로 온톨로지는 관심 있는 특정 도메인에 대해 생성되므로 넓은 문제 영역에서 단일 온톨로지가 응답하는 것을 기대할 수 없다. 이것은 큰 작업에서 상이한 온톨로지의 데이터를 사용할 필요가 있음을 의미한다. 또한 대부분의 온톨로지는 한 사람이나 작은 그룹에 의해 개발되고, 그래프가 아닌 트리로 생각할 수 있다.

본 논문에서는 트리를 이용하여 온톨로지를 구축하고, 이해하고, 처리하는 효율적인 방법을 보이고자 한다. 온톨로지 병합 오퍼레이션의 정의를 위해 온톨로지의 구조, 구성 요소, 표현을 자세하게 정의하고, 두 온톨로지를 병합하는 방법에 대한 세부 사항을 보인다.

1. 서론

협력 가능성에 대한 기대는 온톨로지 개발의 중요한 동기이다. 온톨로지는 일반적으로 관심 있는 특정 도메인에 대해 생성된다. 따라서 넓은 문제 영역에서 단일 온톨로지가 응답하는 것을 기대할 수 없다. 함축적으로 이것은 큰 작업에서 상이한 온톨로지의 데이터를 사용할 필요가 있음을 의미한다. 다른 말로 하면 상이한 온톨로지가 중요한 정보를 처리하기 위해 협력한다는 가정이 된다.

정보 통합은 온톨로지의 중요한 목표이다. 온톨로지는 데이터 웨어하우스나 연합 데이터베이스 같은 이종의 소스와 데이터베이스로부터 정보를 통합하는데 사용될 수 있다. 구조적인 이종, 의미적인 이종 둘 다 적당한 온톨로지를 사용하여 극복할 수 있다.

현재 소프트웨어 개발에서 중요한 문구는 “재사용”이다. 온톨로지 세계에서 재사용 가능한 온톨로지 라이브러리를 생성할 수 있기를 원한다. 이 라이브러리는 잘 분리되고 기능적으로 응집력이 있는 작고, 독립적인 온톨로지를 포함한다. 좀 더 크고 보다 강력한 온톨로지를 구축하는 것은 기존의 온톨로지를 재사용하는 것만으로도 가능하다.

실제 유용하고 강력한 온톨로지 시스템은 부 온톨로지의 결합과 병합에 의해 끊임없이 스스로를 적응시킨다. 이런 지능 시스템은 새로운 데이터, 새로운

데이터 타입, 새로운 요구와 질의 타입, 또한 새로운 사용자를 처리할 수 있는 능력이 있다. 온톨로지 개발과 사용에 대한 보다 강력한 접근방법이 연구되고 있다.

본 논문에서는 효율적인 데이터 통합을 위하여 트리 구조를 이용한 온톨로지 병합 방법에 대하여 논하고자 한다.

2. 관련연구

2.1 XML, RDF, OWL

시멘틱 웹 구축을 위한 중요한 기술인 XML과 RDF는 이미 적절하다. XML은 사용자 커뮤니티의 정도, 도구 지원의 가능성, 기술 의존적인 비즈니스 모델들의 생존 능력의 관점에서 온톨로지 보다 훨씬 성숙하다. 온톨로지가 불확실성 감소를 위해 사용되는 반면, XML은 복잡성 감소를 위해 사용된다는 것이 관찰되었다[1].

XML은 온라인 데이터 교환을 위한 표준으로 만들어졌고, 상호운용과 데이터 통합을 증진시킨다. XML은 자기 서술 데이터를 생성하고, 월드 와이드 웹, 인터넷, 그리고 어디에서든 형식과 자료를 공유하기 위한 유연한 방법이다[2].

본질적으로 RDF는 명확히 설명적인 정보를 제공하

기 위한 규칙 집합이다. 이를테면 간단한 설명을 하기 위한 표준 방법이다.

OWL의 주요 개념은 결정 절차를 따르는 온톨로지의 효율적인 표현을 가능하게 하는 것이다. 이것은 논리적으로 일관되는지 살펴보기 위해, 또는 온톨로지에 특별한 개념이 포함되어 있는지 결정하기 위해 온톨로지를 검사한다[3].

2.2 온톨로지와 시멘틱 웹

시멘틱 웹은 의미 있는 웹 페이지의 내용 구조를 가져오며, 페이지를 돌아다니며 사용자를 위해 복잡한 작업을 용이하게 할 수 있는 소프트웨어 에이전트 환경을 만든다[4].

전형적인 시멘틱 웹 애플리케이션의 경우, 구조의 개발을 위해 다음과 같은 특징들을 고려해야 한다.

- 1) 데이터는 지역적으로 분산되어 있다.
- 2) 데이터는 다양한 소유주가 있다.
- 3) 데이터는 이중이며 실세계의 데이터이다.
- 4) 애플리케이션은 열린 세계이어야 한다. 이를테면 정보는 절대로 완벽하지 않다.
- 5) 애플리케이션은 반드시 어떤 형식적인 데이터 의미 기술을 사용한다.

따라서 온톨로지가 시멘틱 웹을 실현 가능한 목표로 만들고, 시멘틱 웹이 월드 와이드 웹의 풍부한 잠재성을 가져올 수 있다는 것을 알 수 있다. 온톨로지는 질적으로 새로운 서비스 수준을 제공할 잠재성이 있다. 이를테면 증명, 검증, 차이 분석과 같은 서비스는 이전에는 구현되지 않았다[5]. 온톨로지는 정보를 정의할 뿐만 아니라 풍부한 표현과 특성 추론을 추가한다.

2.3 온톨로지의 관계, 트리, 그래프

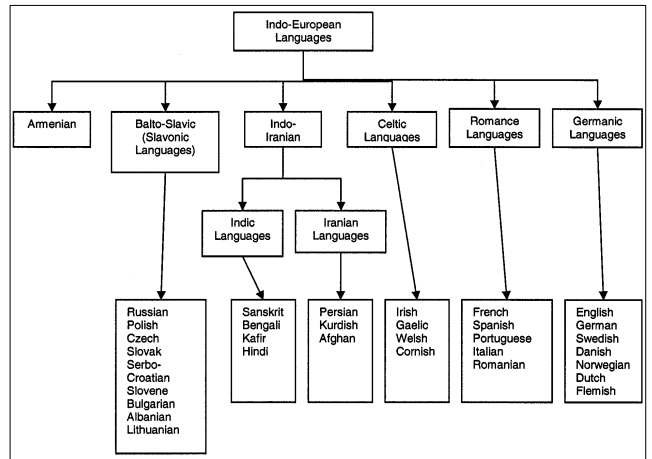
온톨로지는 개념화를 기반으로 한 결과로서 두 가지 다른 타입의 데이터의 내부 구조로 표현될 수 있다. 두 가지 타입은 레벨 중심 온톨로지와 관계 중심 온톨로지이다.

레벨 중심 온톨로지는 일반적으로 명시적으로 분류된 레벨의 객체와 정보로 형성된다. 이런 레벨 혹은 계층은 잘 정의되고 명확히 이해된다. 높은 계층은 보다 일반적인 정보로 표현되고, 낮은 계층은 보다 상세한 정보로 표현된다.

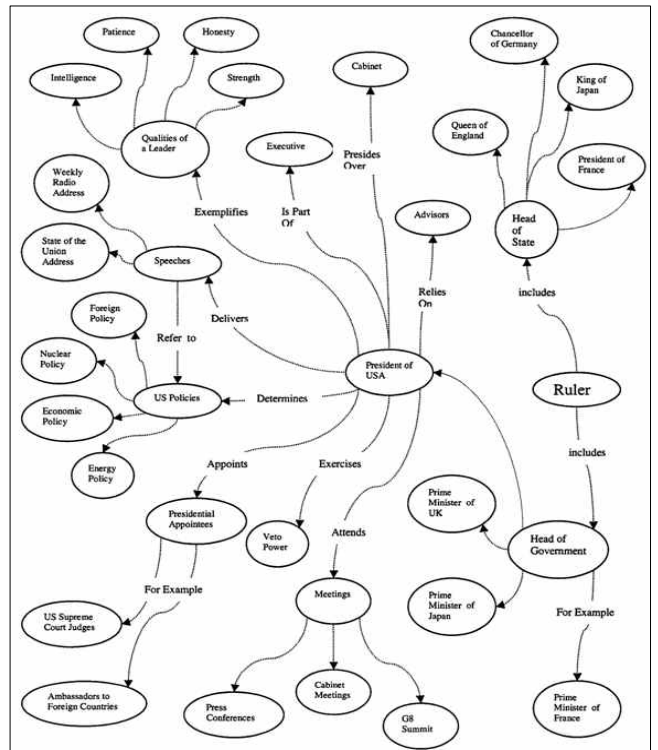
한편, 관계 중심 온톨로지는 시스템내의 가장 흥미로운 객체를 이해하기 위해 탐색한다. 그 다음 이해하기 위해 노력하고, 객체, 활동(프로세스), 객체에

속한 관계를 표현한다. 이 상호작용은 관계를 가장 사실적으로 나타내는 형식으로 묘사된다.

다음 [그림 1]과 [그림 2]는 레벨 중심 온톨로지와 관계 중심 온톨로지의 예를 보여준다. [그림 2]에서 보이는 몇 가지 일반적인 관계는 “For example”, “Unifies”, “Describes”, “Refers to”, “Depend on”, “Is part of”, “Reports to”, “Passes data to”, “Must be done before”등이며, 일부는 Tim Berners-Lee가 설명했다[6].



[그림 1] 레벨 중심 온톨로지의 예

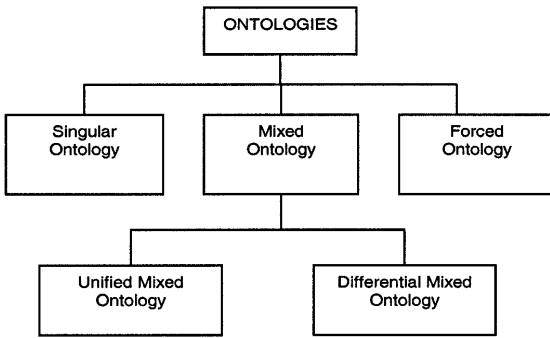


[그림 2] 관계 중심 온톨로지의 예

3. 온톨로지의 트리 병합

3.1 온톨로지의 종류

온톨로지는 본질적으로 계층적인 분류 시스템으로 볼 수 있다. [그림 3]에 보인 것처럼 내부 구조를 기반으로 세 종류의 온톨로지-단일 온톨로지, 혼합 온톨로지, 강제 온톨로지, 강제 온톨로지로 나눌 수 있다.



[그림 3] 내부 구조를 기반으로 한 온톨로지 종류

3.2 온톨로지 병합 오퍼레이션

두 온톨로지 O1과 O2의 병합은 매우 복잡한 프로세스이거나 매우 간단하고 단순한 프로세스이다. 병합 프로세스는 기본적인 병합 오퍼레이션의 모음으로 이루어진다. 다양한 기본적인 병합 오퍼레이션은 다음과 같다.

3.2.1 Siblinging

O1의 루트 노드와 O2의 루트 노드는 새로운 온톨로지 O3의 형제관계가 된다.

3.2.2 Childing

온톨로지 O2의 루트 노드가 온톨로지 O1의 어떤 노드의 자식 노드가 된다. 명확히 온톨로지 O2 전체가 온톨로지 O1에 포함된다. 그러므로 O1이 주요한 온톨로지이고, O2가 보조적인 온톨로지이다.

3.2.3 Aggregating

온톨로지 O2의 노드는 O1의 노드로 합해진다. O1과 O2가 같은 하부구조를 가지므로 결과 온톨로지 O3가 된다.

3.2.4 Data Appending

온톨로지 O1과 O2는 정확히 같은 구조를 갖는다. 단말 노드의 데이터는 구조 변화 없이 연관되어야 한다.

3.2.5 Node Generation

노드 생성은 병합 프로세스의 중간 단계이다. 노드 생성 오퍼레이션은 온톨로지의 노드로부터 정보를 추출하고 온톨로지에 새로운 노드를 생성한다.

새로 생성된 노드는 레벨에 의존적이다. 노드 생성은 세 가지 유형이 있다.

- 1) 루트 노드 생성 : 추가되는 새로운 노드가 온톨로지의 루트 노드가 된다.
- 2) 단말 노드 생성 : 추가되는 새로운 노드가 온톨로지의 단말 노드가 된다.
- 3) 중간 노드 생성 : 추가되는 새로운 노드가 온톨로지의 중간 노드가 된다.

4. 효율적인 온톨로지 병합

두 온톨로지 O1과 O2의 병합 알고리즘은 도메인 호환성이 있다고 가정한다. 두 온톨로지가 같은 도메인에 속한다면 도메인 호환성이 있다고 한다. 두 경우에 대해 온톨로지 병합 실험을 수행하고자 한다. 첫 번째 경우는 동등한 레벨의 온톨로지이고, 두 번째 경우는 주 온톨로지와 보조적인 온톨로지로서 이루어진 경우이다.

4.1 동등 레벨 온톨로지의 병합

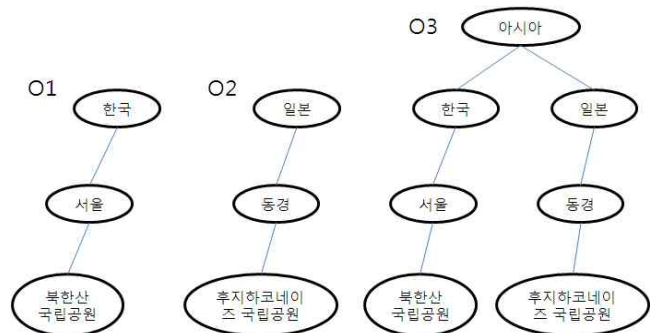
두 온톨로지 O1과 O2가 다음과 같으면 동등 레벨이라고 한다.

- 1) 깊이 동치 : O1의 레벨 수 = O2의 레벨 수 = n
 - 2) 내용 동치 : O1.레벨_i ≡₃ O2.레벨_i
- 여기서 표기법 ≡₃ 는 두 레벨의 3 레벨 동치를 의미한다.

두 레벨이 다음과 같으면 3 레벨 동치라고 한다.

- 1) 레벨 호환성이 있다.
- 2) 부모 레벨이 레벨 호환성이 있다.
- 3) 자식 레벨이 호환성이 있다.

루트가 동일한 경우와 루트가 동일하지 않지만 유사한 경우를 고려해 볼 수 있고, [그림 4]는 루트가 동일하지 않지만 유사한 경우의 병합 예이다.



[그림 4] 루트가 동일하지 않지만 유사한 경우의 병합 예

4.2 주 온톨로지와 보조적인 온톨로지의 병합

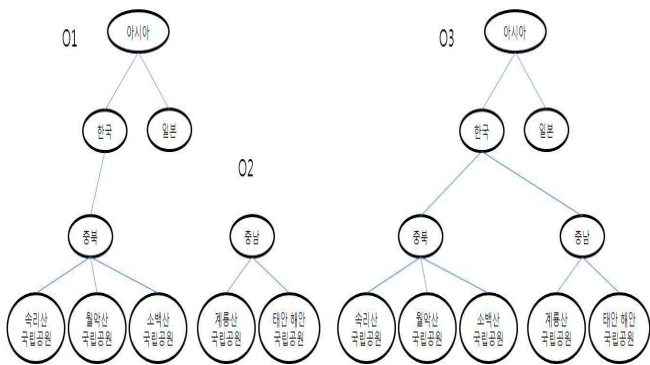
두 온톨로지 O1과 O2에서 O2의 R-L 경로 유형 집합이 O1의 R-L 경로 유형 집합의 부분집합이라면 O1은 주된 온톨로지라 하고, O2는 보조적인 온톨로지라 한다.

우월성의 세 가지 종류를 정의할 수 있다.

1) 온톨로지의 단순 우월성 : 두 온톨로지 O1과 O2에서 O2의 R-L경로 유형 집합이 O1의 R-L 경로 유형 집합의 지속적인 부분집합이라면 O1은 주된 온톨로지이고, O2는 보조적인 온톨로지이며, O1이 O2에 대해 단순 우월성(혹은 평범함 우월성)을 보인다고 한다. [그림 5]는 단순 우월성을 갖는 경우의 병합 예이다.

2) 온톨로지의 전체 우월성 : 두 온톨로지 O1과 O2에서 O2의 R-L경로 유형 집합이 O1의 R-L 경로 유형 집합의 전체적인 부분집합이라면 O1은 주된 온톨로지이고, O2는 보조적인 온톨로지이며, O1이 O2에 대해 전체 우월성을 보인다고 한다.

3) 온톨로지의 부분 우월성 : 두 온톨로지 O1과 O2에서 O2의 R-L경로 유형 집합이 O1의 R-L 경로 유형 집합의 부분적인 부분집합이라면 O1은 주된 온톨로지이고, O2는 보조적인 온톨로지이며, O1이 O2에 대해 부분적인 우월성을 나타낸다고 한다.



[그림 5] 단순 우월성을 갖는 경우의 병합 예

5. 결론 및 향후연구

개발자-사용자 양분의 해소는 시멘틱 웹의 개발과 수용의 주요 목표이다. 주요 논쟁의 하나는 도메인 전문가가 일반적으로 온톨로지를 개발하기 때문에 온톨로지를 설계하는 과학자와 기술자가 온톨로지

사용자보다 양질의 유능한 온톨로지를 사용한다는 것이다. 따라서 온톨로지가 일반 사용자에게 유용하지 않다고 볼 수 있다. 온톨로지 병합 전략과 기술의 매우 큰 부분은 온톨로지에 대한 개발자와 사용자간의 차이가 증가된다는 것이다. 온톨로지는 정의와 규칙을 포함하기 때문에 데이터가 시멘틱 웹에 공유되면 많은 사용자로부터 많은 정의, 설명, 규칙 집합을 얻는다. 온톨로지 병합은 모든 사용자가 이해하고 재사용하는 통합된 온톨로지를 생성한다. 따라서 온톨로지는 모든 것을 공유 가능하고, 서로 다른 사람의 지식으로부터 이익을 얻을 수 있는 세상을 촉진한다.

본 논문에서는 온톨로지 병합에 대한 이론, 공리, 형식화, 가이드라인의 큰 틀을 제공한다. 본 논문의 알고리즘과 정의는 특정 도메인에 한정되지 않는다. 이것은 일반적이고, 크기 조절 가능하고, 효율적이다.

향후 연구로는 의학 등과 같은 특정 도메인의 두 정보 시스템을 병합하는 실제적인 능력을 갖도록 하기 위해 온톨로지 병합 알고리즘의 포괄적인 시스템을 구축하는 것이다.

참고문헌

[1] C. Holsapple, K Joshi, "A Collaborative Approach to Ontology Design", Comm. of the ACM, vol 45, No 2, pp. 42-47, February 2002

[2] David Hunter 등 저, 박형일, 위장환 공역, Beginning XML(3판), 지앤선, 2005

[3] Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, "The Semantic Web Revisited", IEEE Intelligent Systems vol 21, no 3, pp 96-101, July 2006.

[4] Tim Berners-Lee, James Hendler, Ora Lassila, "The Semantic Web", Scientific American, May, 2001

[5] L. Stojanovic, J. Schneider, A. Maedche, S. Libischer, R. Studer, T. Lumpp, A. Abecker, G. Breiter, J Dinger, "The Role of Ontologies in Autonomic Computing Systems", IBM Systems Journal, vol 43, no 3, pp 598-616, 2004

[6] Tim Berners-Lee, "Information Management: A Proposal", CERN, March 1989