MDR 기반의 구조화 된 데이터를 이용한 웹 온톨로지 학습 및 확장 모델

정혜진*, † 백두권**, † 정동원*

*군산대학교 정보통계학과
**고려대학교 컴퓨터학과

Web Ontology Learning and Population Model using Structured Data Based on MDR

Hyejin Jeong*, Doo-Kwon Baik**, Dongwon Jeong*

*Dept. of Informatics and Statistics, Kunsan National University
**Dept. of Computer Science & Engineering, Korea University

E-mail : {xhyejin86x, djeong}@kunsan.ac.kr, baikdk@korea.ac.kr

요 약

기존의 웹을 확장한 시맨틱 웹의 등장으로 웹 온톨로지의 구축이 중요시 되고 있다. 이로 인하여 현재 웹 온톨로지의 관리 및 활용을 위한 편집기, 웹 온톨로지 기술언어, 저장소 및 추론 엔진 등 다양한 기술 및 시스템들이 개발되어 웹 온톨로지의 구축이 용이해졌다. 이제는 구축된 웹 온톨로지를 응용 시스템에 활용하기 위한 웹 온톨로지 클래스에 대한 인스턴스를 풍부하게 할 수 있는 웹 온톨로지의 확장에 대한 연구가 요구된다. 웹 온톨로지의 확장을 위해서는 먼저 웹 온톨로지를 보다 정확하게 정의해야 하며 웹 온톨로지를 보다 풍부하게 확장할 수 있는 방법이 개발되어야 한다. 웹 온톨로지의 보다 정확한 정의를 위해서는 표준화 된 공통 개념을 이용하여 웹 온톨로지 스키마를 생성해야하며 이를 기반으로 한 웹 온톨로지 간 상호운용성 향상되어야 한다. 따라서 이 논문에서는 표준화 된 공통 개념을 관리하는 메타데이터 레지스트리(Metadata Registry)를 기반으로 구조화 된 데이터를 이용한 웹 온톨로지의 학습 및 확장 모델을 제안한다. 또한, 제안 모델을 위한 프로토타입을 구현하고 제안 모델의 평가에 대하여 기술한다.

1. 서론

웹 정보에 잘 정의된 의미를 부여하여 컴퓨터도 쉽게 문서의 의미를 해석할 수 있도록 하여 정보의 검색·해석·통합 등의 업무를 자동화하기 위한 시맨틱 웹을 구현하기 위해서는 웹 온톨로지의 구축이 필수적으로 요구된다[1-2]. 현재 온톨로지의 관리 및 활용을 위한 다양한 편집기(Protege, OntoStudio), 웹 온톨로지 기술 언어(RDF, RDF Schema, OWL), 저장소 및 추론 엔진(Jena, Sesame, Bossam)등이 개발되었으며 보다 풍부한 정보를 제공하기 위한 웹 온톨로지의 확장에 대한 연구가 요구되고 있다. 이를 위해서는 먼저 웹 온톨로지의 보다 정확한 정의가 요구된다. 웹 온톨로지에 대한 보다 정확한 정의 및 확장을 위해서 요구되는 사항은 다음과 같다.[3-7]

- 보다 정확한 웹 온톨로지 정의를 위한 일관성 있고 보편적인 개념을 이용한 웹 온톨로지 스키마 생성과 이를 기반으로 한 웹 온톨로지 간 상호운용성의 향상
- 보다 풍부한 정보 제공을 위한 웹 온톨로지 학습 및 확장 방법의 개발

이 논문에서는 앞서 언급한 웹 온톨로지의 확장을 위해 요구되는 사항들을 충족시키기 위하여 메타데이터 레지스트리(MDR, Metadata Registry)를 기반으로 한 웹 온톨로지의 스키마 생성과 MDR 기반의 구조화 된 데이터를 이용한 웹 온톨로지 확장 모델을 제안한다.

MDR은 ISO/IEC에서 국제 표준(ISO/IEC 11179)으로서 표준 개념, 즉 일관성 있는 의미 정보를 관

본 연구는 2008년 정부(교육과학기술부)의 재원으로 한국학술진흥재단 연구과제(과제번호: KRF-2008-314-D00485)의 연구비지원에 의하여 연구되었음.

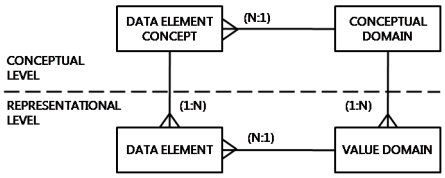
† Co-Corresponding Authors

리함으로써 이를 기반으로 생성된 데이터베이스 간 의미 공유 및 교환이 용이하다[10]. 따라서 MDR 내의 표준 개념을 이용하여 보다 정확한 의미 교환이 가능한 웹 온톨로지 생성이 가능하다.

또한 MDR을 기반으로 생성된 데이터인 구조화된 데이터를 이용한 기 생성된 온톨로지의 확장이 용이하다. 즉, MDR의 표준 데이터 요소를 이용하여 정의한 실제 데이터베이스의 데이터(인스턴스)를 기존에 생성한 온톨로지에 추가함으로써 보다 풍부한 정보 제공 및 서비스 개발이 가능함을 의미한다. 이 논문에서는 이러한 MDR의 특징과 장점을 이용하여, 표준화 된 공통 개념인 MDR을 기반으로 하여 생성된 웹 온톨로지 생성 및 확장을 위한 새로운 모델을 제안한다.

2. 관련연구

MDR(Metadata Registry)은 ISO/IEC JTC 1/SC 32에 의해 개발된 국제 표준(ISO/IEC 11179)으로서, 데이터 의미의 표현, 등록, 관리, 교환, 공유를 주목적으로 한다.



[그림 1] ISO/IEC 11179 MDR 모델

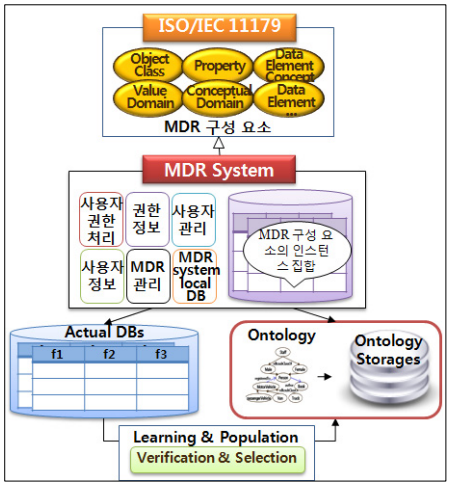
ISO/IEC 11179 표준문서는 식별, 명명, 정의, 분류 등에 중점을 두어 교환 및 공유 될 데이터의 의미, 표현방식 등에 대하여 이용자의 이해를 돕기 위한 메타데이터의 등록과 유지를 위한 지침을 제공하여 레지스트리에 등록된 정보의 상호운용성 확보에 중요한 기반을 제공한다[8].

MDR은 구현 내용의 재사용과 공유를 용이하게 하기 위하여 메타 모델의 형태로 기술되어 있으며 레지스트리의 기능은 관리와 식별(Administration and Identification), 명명과 정의(Naming and Definition), 분류(Classification), 데이터 요소 개념(Data Element Concept), 개념 및 값 영역(Conceptual and Value Domain), 데이터 요소(Data Element)로 나뉜다. 또한, MDR에 의해 관리되는 데이터의 기본 단위는 데이터 요소(Data Element)이며, 객체 클래스(Object Class), 특성(Property), 표현(Representation)으로 구성된다[9].

[그림 1]은 ISO/IEC 11179 MDR 모델로서 MDR에서 가장 중요한 개념적 모델의 요소인 개념 영역, 값의 영역, 데이터 요소 개념, 데이터 요소(Data Element) 간의 관계를 보여준다[10].

3. 제안 모델

[그림 2]는 제안 모델의 전체적인 개념도이다.

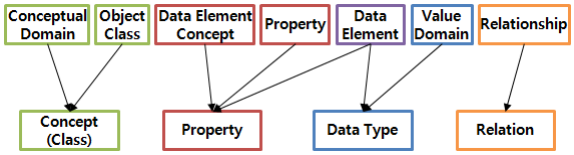


[그림 2] 전체적인 제안 모델의 개념도

제안 모델은 웹 온톨로지를 보다 정확하게 정의하고 확장하는 모델로서, 일관성 있고 보편적인 개념, 즉 MDR의 표준화 된 공통 개념을 이용하여 웹 온톨로지 스키마를 생성하여 웹 온톨로지를 보다 정확하게 정의한다. 또한, MDR의 표준화 된 공통 개념들을 이용하여 데이터베이스 스키마를 정의한 실제 데이터베이스의 인스턴스를 정의된 웹 온톨로지에 추가함으로써 보다 풍부하게 웹 온톨로지를 확장한다.

이를 위하여 ISO/IEC 11179 표준문서의 지침에 따라서 MDR의 구성 요소인 Object Class, Property, Data Element Concept, Value Domain, Conceptual Domain, Data Element 등을 MDR 시스템에 등록하고 관리한다.

MDR 시스템에 표준화 된 공통 개념들이 등록되면 등록된 표준화 된 공통 개념을 이용하여 웹 온톨로지 스키마를 생성하게 된다. 이때 표준화 된 공통 개념과 웹 온톨로지의 구성요소의 사상이 이루어져야 하며, [그림 3]의 관계에 의하여 정의한다[11].



[그림 3] MDR과 웹 온톨로지 구성 요소 간 관계

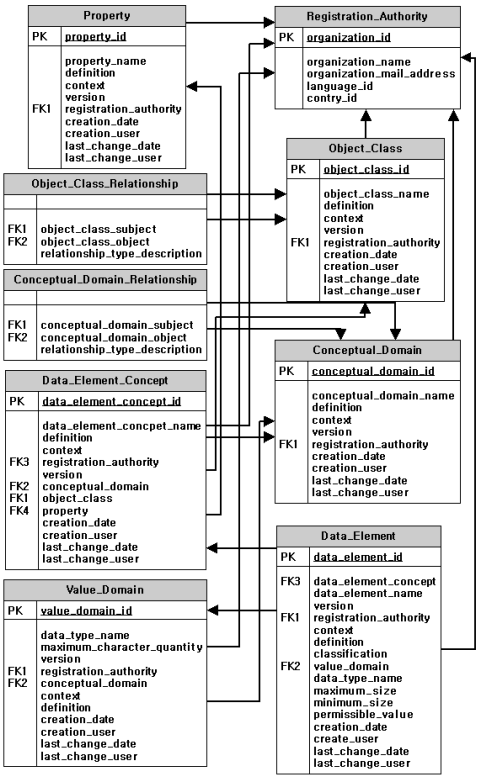
또한, 실제 데이터베이스에서는 MDR의 구성 요소 중에서도 가장 기본이 되는 데이터 요소(Data Element)를 이용하여 데이터베이스 스키마 필드를 정의해야 한다. 이는 이 논문의 제약 사항이다. MDR의 표준화 된 공통 개념에는 데이터 요소(Data Element) 이외에도 많은 개념이 있지만, 데이터의 형식(값의 영역, 데이터 타입, 표현 클래스 등)을 포함하는 데이터 요소(Data Element)만을 이용하여 정의한다고 가정한다. 다시 말해, 이 논

문에서는 실제 같은 필드지만 데이터 타입이 다른 경우가 있을 수 있으나 이러한 경우는 배제한다.

실제 데이터베이스의 인스턴스를 이용하여 웹 온톨로지를 확장하는 과정에서는 동일한 공통 개념을 온톨로지와 실제 데이터베이스 스키마 필드에서 사용함에 있어서 Rule 기반의 학습이 이루어진다. 이를 통해 얻어진 결과는 정확성을 위하여 개발자의 확인과 검증 작업을 필요로 한다. 개발자의 확인과 검증 작업이 완료되면 웹 온톨로지의 확장 과정이 종료된다.

4. 구현 및 평가

4.1 MDR 저장 모델

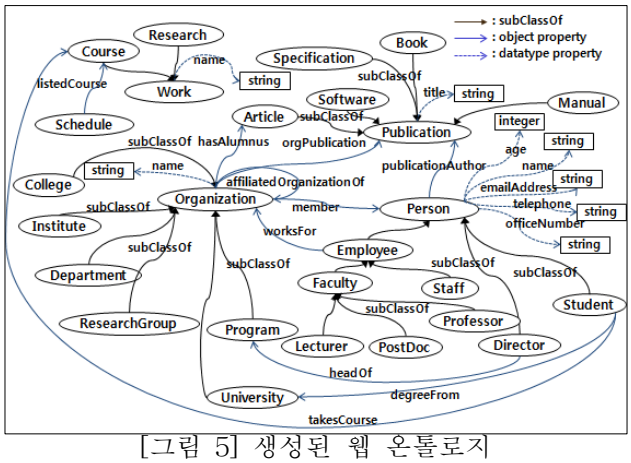


[그림 4] MDR 저장 모델

제안 모델은 MDR을 기반으로 하여 확장을 하기 때문에 먼저 MDR이 구축되어 있어야 한다. 따라서 ISO/IEC 11179 MDR 모델을 참고로 하여 간단한 MDR을 구축하기 위한 MDR 저장 모델을 정의한다. [그림 4]는 MDR 저장 모델을 표현한 것이다.

4.2 웹 온톨로지 생성

[그림 5]는 MDR 저장 모델의 인스턴스를 이용하여 앞서 기술한 MDR과 웹 온톨로지 구성요소 간 관계에 따라 정의한 웹 온토롤지를 그래프 형태로 표현한 것이다. 생성된 웹 온톨로지는 표준 웹 온톨로지 언어인 OWL로 기술하였으며 저장소 및 추론엔진 역할을 하는 Jena를 통하여 현재 대부분의 웹 온톨로지 저장소인 관계형 데이터베이스에 저장된다. 관계형 데이터베이스는 현재 가장 널리 사용되고 있으며 다른 관계형 데이터베이스 관리 시스템에 비해 뛰어난 안정성과 성능을 제공하는 오라클을 이용하였다. 웹 온톨로지가 저장된 테이블은 트리플 구조(subject, predicate, object)이다.



[그림 5] 생성된 웹 온톨로지

4.3 웹 온톨로지 확장

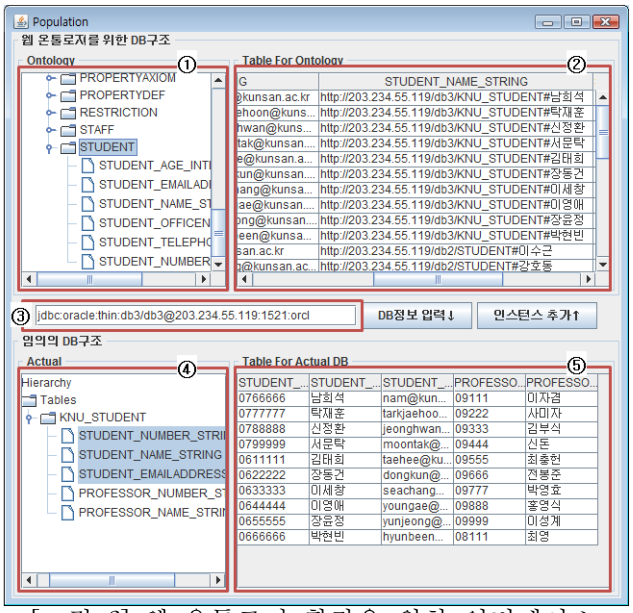
실제 데이터베이스의 인스턴스를 이용하여 정의된 웹 온톨로지를 확장하기 위해서 다른 데이터베이스로부터 가져온 인스턴스들을 저장하기 위하여 [그림 5]에서 정의한 웹 온톨로지 중 하위클래스를 갖지 않는 클래스를 중심으로 별도의 테이블을 생성하였다. 또한 임의의 DB 3개를 정의하였다.

- DB1 : 클래스 Student의 모든 속성의 값만 모두 가지고 있는 DB
- DB2 : 클래스 Student의 일부 속성 값만 모두 가지고 있는 DB
- DB2 : 클래스 Student의 일부 속성 값과 클래스 Professor에 해당하는 일부 속성 값을 가지고 있는 DB

[그림 6]은 웹 온톨로지 확장을 위한 인터페이스를 보여준다. 그림에서 ①은 확장할 온톨로지에 대한 구조를 보여주는 부분으로, 클래스 정보와 프로퍼티의 정보를 보여준다. ②는 해당되는 클래스의 인스턴스들을 보여주는 부분이다. 또한 ③은 실제 임의의 데이터베이스에 접속하기 위한 정보를 입력하는 부분이다. ④는 임의의 데이터베이스의 테이블 정보를 보여주는 부분이며 ⑤는 테이블에 해당하는 인스턴스들을 보여준다. ①에서 해당되는 클래스와 프로퍼티를 선택한 후 ③에서 실제 임의의 데이터베이스의 정보를 입력하고 DB정보 입력 버튼을 누른 후 ④에서 해당되는 테이블과 필드를 선택한 후 인스턴스 추가 버튼을 누르게 되면 [그림 6]에서 보는 바와 같이 인스턴스가 추가된 것을 볼 수 있다.

여기에서 실제 임의의 데이터베이스의 인스턴스

들을 웹 온톨로지에 추가할 때, 데이터들이 데이터베이스의 인스턴스이기 때문에 URI를 'http://DB의 IP 정보/DB 사용자명/테이블명'의 형태로 기술하였다.



[그림 6] 웹 온톨로지 확장을 위한 인터페이스

4.4 평가

현재 MDR기반의 구조화 된 데이터를 이용하여 웹 온톨로지를 확장하는 연구가 아직 이루어지지 않았기 때문에 비교·평가가 어렵다. 따라서 제안 모델의 장점과 한계성을 중심으로 기술한다.

제안 모델은 MDR을 이용하여 구조화 된 데이터를 기반으로 웹 온톨로지 스키마를 정의하였기 때문에 개념의 일관성을 향상시키며 보다 정확한 의미를 갖는 웹 온톨로지를 생성할 수 있다. 또한, 제안 모델은 학습 및 확장을 통하여 인스턴스를 수집하여 보다 풍부한 온톨로지를 생성할 수 있다.

따라서 직접 정보를 수집하여 인스턴스를 추가하는 것에 대비하여 보다 풍부한 정보를 수집할 수 있으며 보다 풍부한 정보를 수집하는 데에 드는 비용과 시간을 절감할 수 있다.

그러나 제안 모델은 실제 데이터베이스 스키마 필드를 정의할 때 MDR의 표준화 된 공통 개념을 이용하여 정의해야하며 실제 데이터베이스 스키마 필드를 정의할 때 MDR의 구성 요소 중 데이터 요소(Data Element)의 인스턴스만 이용해야 한다는 제약 사항이 있다. 또한, MDR 시스템이 구축되어 있어야 하며 표준화 된 공통 개념이 등록되어 있지 않을 경우 등록해야하는 번거로움이 있다.

5. 결론 및 향후 연구

이 논문에서는 MDR의 표준화 된 공통 개념을 이용하여 웹 온톨로지를 생성하고 MDR의 표준화 된 공통 개념을 이용하여 실제 데이터베이스 스키마 필드를 정의한 데이터베이스의 인스턴스를 기존에 생성한 웹 온톨로지에 추가함으로써 보다 풍부하고 정확한 정보를 제공하기 위한 웹 온톨로지 확장 모델을 제안하였다. 또한 제안 모델에 대한 프로토타입을 구현하여 실제 웹 온톨로지를 확장 하는 방법을 보였다. 그러나 이 논문에서는 MDR의 구성 요소 중 데이터 타입과 데이터 형식 등을 포함하고 있는 데이터 요소(Data Element)만을 이용하여 실제 데이터베이스의 스키마 필드를 정의한다는 제약 조건과 MDR에 등록된 표준화 된 공통 개념만을 이용한다는 전제 조건이 있었다.

따라서 (1) 실제 데이터베이스 스키마 필드를 정의할 때, 다른 구성 요소들을 부분적으로 사용했을 경우, (2) 실제 데이터베이스 스키마 필드를 MDR에 등록된 요소를 사용하지 않았지만 MDR에 등록된 요소 중 비슷한 개념이 있을 경우 학습을 하여 웹 온톨로지를 확장할 수 있는 방법에 대한 연구가 향후 요구된다.

[참고문헌]

[1] Tim Berners-Lee, "The World Wide Web: Past, Present and Future," IEEE, IEEE Computer Magazine, Vol. 29, No. 10 pp. 69-77, October 1996.

[2] Asuncion Gomez-Perez and Oscar Corcho, "Ontology Languages for the Semantic Web," IEEE Computer Society, IEEE Intelligent Systems, Vol. 17, No. 1, pp. 54-60, January/February 2002.

[3] World Wide Web Consortium. http://www.w3.org/, 2008.

[4] W3C, Resource Description Framework, http://www.w3.org/RDF/, 2008.

[5] W3C, Web Ontology Language, http://www.w3.org/2004/OWL/, 2008.

[6] Jena Semantic Web Framework, http://jena.sourceforge.net/, 2008.

[7] The Protege Ontology Editor and Knowledge Acquisition System, http://protege.stanford.edu/, 2009.

[8] 오삼균, "시맨틱 웹 기반 메타데이터 레지스트리 설계에 관한 연구", 한국도서관·정보학회, 한국도서관정보학회 하계학술발표회, pp. 7-36, 2005. 06.

[9] 백두권, "데이터 표준화와 메타데이터 레지스트리(MDR:MetaData Registry)", 한국정보통신기술협회, TTA 저널, 제2000권 제71호, 2000. 10.

[10] ISO/IEC JTC 1/SC 32/WG 2 Web site, http://metadata-standards.org/, 2008.

[11] 정혜진, 정동원, "구조화된 데이터 기반의 웹 온톨로지 학습 및 확장 모델 설계", 한국컴퓨터정보학회, 제39차 2008 동계학술대회논문집, 제16권, 제2호, pp.329-334, 2009. 01.

[12] 정동원, 김정동, 손지성, 김장원, 백두권, "메타데이터 레지스트리 기반 웹 온톨로지 생성에 관한 연구", 한국컴퓨터정보학회, 제39차 2008 동계학술대회 논문집, 제16권, 제2호, pp. 19-24, 2009. 01.