제18회 한국정보처리학회 추계학술발표대회 논문집 제9권 제2호 (2002.11)

시맨틱 웹을 위한 온톨로지 시스템의 설계

정희준*, 유명한*, 이강찬**, 김성한**, 민재홍**, 정인정*
*고려대학교 전산학과
**한국전자통신 연구원 표준연구센터
e-mail:{joonny96, myong, chung}@korea.ac.kr
{chan, sh-kim, jhmin}@etri.re.kr

Design of Ontology System for the Semantic Web

Hee-Joon Chung*, Myong-Hwan Yoo*, Kang-Chang Lee**,
Sung-Han Kim**, Jae-Hong Min**, In-Jeong Chung*
*Dept of Computer Science, Korea University
**Protocol Engineering Center, Electronic and
Telecommunication Research Institute

요 약

컴퓨터 통신기술이 급속히 발달하고 인터넷이 광범위하게 사용됨에 따라서 웹은 다양하며 무한한 용량의 데이터 원천으로 부각되었다. 그러나 웹의 사용이 지수적으로 증가함에 따라 웹 상에서 사용자가 원하는 정보의 검색, 구성 및 통합하는 작업이 매우 어렵게 되었다. 이러한 문제점을 해결하기 위해 1990년대 말에 W3C(World Wide Web Consortium)에서는 차세대 웹으로써 시맨틱 웹을 제시하였다. 시맨틱 웹은 웹 데이터에 의미(meaning)을 부여하여 컴퓨터와 같은 기계가 이해할 수 있는 언어로 만들어 정보의 재사용성을 향상시키고, 컴퓨터가 정보를 이해할 수 있도록 하여 이기종 간의 상호운용성을 보장한다. 그러나, 현재 시맨틱 웹에 대한 시스템 아키텍처는 제안되어져 있지 않으며, 시맨틱 웹의 응용 분야 개발에 있어서 개별적인 도메인에 맞는 시스템의 아키텍처만이 제안 및 개발되어져 있는 실정이다. 본 논문에서는 시맨틱 웹의 전체적인 아키텍처에 앞서 시맨틱 웹의 중요 기술인 온톨로지에 대한 시스템 아키텍처를 제안한다. 이는 시맨틱 웹의 개발과 연구에 있어서 청사진을 제공하기 위해 필요하며, 시맨틱 웹의 응용 프로그램 개발에 있어서 효율적이고 유기적인 시스템 개발에 중요한 역할을 할 수 있다.

1. 서론

1990년대 중반에 이르러 인터넷이 광범위하게 사용됨에 따라 웹은 다양한 정보원으로써 무한한 저장 능력을 갖춘 정보의 바다로 부각되었다. 그러나, 정보가 급속하게 증가함에 따라 유용한 정보를 효율적으로 찾는 것이 점점 어렵게 되었다. 이와 같은 문제점은 현재의 웹 마크업 언어들이 표현을 위한 정보만을 표시하여 이기종 간의 정보 시스템과의 통합이 어렵고, 단일 기종 내에서도 디자인의 변화에 따라 정보도 함께 수정해야하며 정보의 의미전달이 취약하기 때문이다. 이러한 문제점을 해결하기 위해 1990년대 말에 W3C(World Wide Web Consortirum)에서는 차세대 웹으로써 시맨틱 웹을 제안하였다.

시맨틱 웹은 현재의 웹 환경의 문제점을 해결하고 자동화된 웹 서비스를 제공하며 컴퓨터의 지능적인 정보 처리가 가능토록 웹 문서 내에 지식표현을 위한 온톨로지를 삽입하고 지식간의 관계를 설정하며 추론규칙을 포함시킨다. 이를 통해서 이기종간의 상호운용성을 보장하고 사용자가 원하는 웹 서비스의 발견, 자동적인 웹 서비스의 실행과 동시에 웹 서비스들의 통합과 상호작용을 하여 사용자가 원하는 정보를 찾고 더 나아가 추론이 가능토록 한다[1],[2],[3].

시맨틱 웹에서 온톨로지는 특정 도메인에 맞는 지식을 개념화(conceptualization)하고 이를 명세화(specification)한다. 즉, 시맨틱 웹에서 온톨로지란 사람과 컴퓨터간의 공유되는 지식을 개념화한 구체적인 형식이며, 개념화와 개념화간의 관계를 표현하는 것으로 정의된다[4].

그러나, 현재 시맨틱 웹에 대한 각 주요 사항 및 요소들의 유기적인 관계를 하나로 나타내는 시맨틱 웹에 대한

전체적인 시스템 아키텍처는 제안되어 있지 않다. 본 논문에서는 전체적인 시맨틱 웹의 전체적인 아키텍처에 앞서, 시맨틱 웹의 가장 중요한 개념인 온톨로지에 대한 시스템 아키텍처를 제안한다. 제안된 시스템 아키텍처는 시맨틱 웹과 온톨로지에 대한 유기적인 관계를 나타내며 시맨틱 웹에서 온톨로지가 어떻게 특정 도메인의 지식을 개념화 하는지를 잘 표현하고 있으며 온톨로지가 특정 도메인의 지식을 개념화함으로써 시맨틱 웹의 전체적인 도메인에 알맞은 시스템 구조를 가지고 있다.

본 논문의 구성은 2장에서 관련연구로 시맨틱 웹과 시맨틱 웹의 중요 기술인 온톨로지, RDF와 RDFS에 대해서 알아보고, 3장에서 온톨로지 시스템의 설계와 각 구성요소를 제안한다. 마지막으로 4장에서 결론 및 향후 연구 방향에 대해서 언급한다.

2. 관련 연구

2.1 시맨틱 웹

1990년대 후반에 W3C에서는 차세대 웹으로써 시맨틱 웹을 제안하였다. 시맨틱 웹을 간단히 정의하자면, 시맨틱 웹은 웹 데이터에 의미(semantics)를 부여하여 컴퓨터 기계가 이해할 수 있는 언어로 만들어 컴퓨터에 의해 처리될 수 있도록 고안된 차세대 인터넷이다. 인터넷을 제안하였던 Tim Berners Lee에 의해 시맨틱 웹의 계층적 구조는 [그림1]과 같다[11].

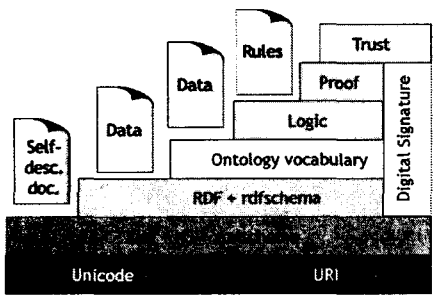

그림 1 . 시맨틱 웹 아키텍처[11]

시맨틱 웹은 W3C에서 현재 표준안과 표준안으로 상정을 한 기술들을 사용하여 기존의 인터넷과 상호운용성 및 호환성을 보장하고 있다. XML과 이에 관련된 네임 스페이스(name space), URI(Uniform Resource Identifier) 기술은 시맨틱 웹에서 기본 기술로써 사용을 되고 있다. 또한, 현재 표준 추천안으로 상정된 RDF와 RDFS는 시맨틱 웹에서 의미(meaning)을 표현하기 위한 언어로써 사용을 하고 있으며, 온톨로지는 이기종 환경에서 도메인의 지식을 공유를 위하여 사용하고 있다. 마지막으로, 자동적이고 지능적인 기능을 수행하기 위한 에이전트는 사용자에게 편리성과 지능적인 기능의 수행을 한다.

2.2 온톨로지

시맨틱 웹의 중요 기술인 온톨로지는 지식 표현을 위한 수단으로써 인공지능 분야에서 오랜 시간동안 연구를 하고 있는 분야이다. 온톨로지란 공유되는 데이터들의 개념화한 형식적이고 명백한 규정이며, 이는 특정분양에서 사용되는 표준 어휘들의 모음이라고 할 수 있다. 즉, 온톨로지는 도메인 내의 지식을 개념화하고 이를 명세화(specification)하는 것으로써 정의된다.

온톨로지는 어휘 사전의 역할이외에 지식을 효과적으로 표현하기 위해 정보의 의미를 부여하고 ,또 정보간의 관계를 설정한다. 즉, 온톨로지는 광범위한 도메인에 적용이 가능하도록 표준을 제시함으로써 웹 문서에 나타난 지식을 표현, 공유와 재사용을 그 목적으로 하고 있다. 더 나아가 시맨틱 웹의 목적인 자동적인 실행과 추론을 하기 위해 그 중요성이 부각되고 있다.

2.3 RDF와 RDFS

RDF와 RDFS는 현재 W3C에서 표준 추천으로 상정되어 있으며, 시맨틱 웹에서 의미의 표현과 온톨로지의 표현 및 모델링에서 사용을 한다. 즉, RDF는 시맨틱 웹을 구성하는 중요한 구성요소 중에 하나이다. RDF는 XML 기반으로 만들어진 마크업 언어로써 웹상의 분산된 다양한 자원들의 기술함과 동시에 그 의미를 표현하기 위해 개발된 언어이다. 웹 상의 자원을 메타 데이터로 표현하고 이를 위한 표준화된 방법을 제공하며 웹 자원의 효율적인 관리와 상호운영성을 위해 개발되었다[7].

RDF 문서에 나오는 자원들은 URI로 나타낸다. 즉, RDF 문서에서는 웹 문서, 문서의 일부, 문서를 이루는 요소들을 직접적으로 나타내지 않고 URI를 통해 나타낸다. RDF에서 나타난 데이터 모델은 객체(Object), 속성(Attribute), 값(Value)으로써 표현을 하며 OVA 관계라고도 한다. 이는 방향성과 라벨이 있는 그래프로써 모델의 인스턴스를 표현한다[4],[7].

RDFS는 자원의 특성을 기술하기 위해 사용될 수 있는 자원들의 집합을 나타내고 웹 자원을 기술하기 위해 이용된 메타 데이터의 구조를 해당 어플리케이션에서 사용되는 특정한 어휘로 정의한다. 또한 RDFS는 데이터의 메타 데이터의 무결성을 보증하기 위해 구문에 관계없는 메타 데이터의 자원과 특성에 대한 유효성을 검증하는 역할도 한다[4],[7].

3. 시맨틱 웹을 위한 온톨로지 시스템

3.1 시스템의 개요

시맨틱 웹에서 온톨로지는 시맨틱 웹이 가능하게 하는 중요 기술이다[1],[2],[6]. 그림 1에서와 같이 온톨로지는 RDF와 RDFS의 레이어 위에 존재하며 온톨로지는 그 상위 계층인 논리(logic)의 근간을 이루고 있다. 즉, 온톨로지는 도메인 내의 지식을 공유하고 추론을 가능하게 하는 역할을 담당하는 역할을 한다[1],[3].

현재 온톨로지에 관한 연구는 온톨로지 저장소(repository) 및 온톨로지의 생성과 온톨로지 버저닝(versioning)및 주석(annotation) 처리와 온톨로지 학습과 통합등 각 분야별로 연구가 이루어지고 있다.

그러나, 시맨틱 웹에서 사용하기 위한 온톨로지의 전체적인 시스템은 아직 제안되어지지 않았다. 전체적인 시스템이 제안되어져 있지 않은 이유는 시맨틱 웹에서 온톨로지가 도메인의 지식을 개념화함으로써, 도메인에 맞는 시스템의 아키텍처를 지녀야하기 때문이다. 본 논문에서는 시맨틱 웹에 대한 전체적인 시스템 아키텍처에서 온톨로지에 대한 시스템의 아키텍처를 제안한다.

이러한 시스템의 아키텍처는 시맨틱 웹의 연구와 개발에 있어서 전체적인 시스템의 흐름과 조직, 개발등에 있어서 로드맵(roadmap)의 역할을 하며, 또한 시맨틱 웹의 개발에 있어서 효율적이고 유기적인 설계 및 개발에 있어서 필요하다.

시맨틱 웹의 전체적인 시스템의 아키텍처는 그림 2와 같다.

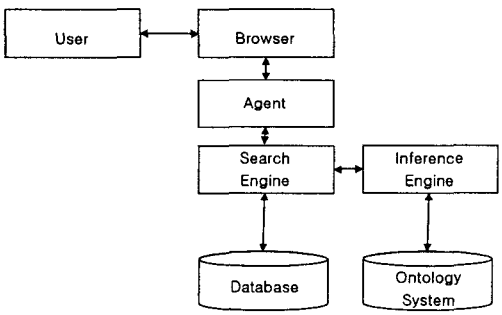

그림 2 시맨틱 웹의 시스템 아키텍처

시맨틱 웹의 시스템 아키텍처는 그림 2와 같은 구조를 가지고 있으며, 이러한 시스템은 현재의 시맨틱 웹의 응용 프로그램들의 구조와 동일한 구조를 가지고 있다[1].

본 논문에서는 시맨틱 웹의 전체적인 아키텍처에서 온톨로지 시스템에 대한 시스템 아키텍처를 제안한다. 그림 3은 온톨로지 시스템의 아키텍처를 나타내고 있다.

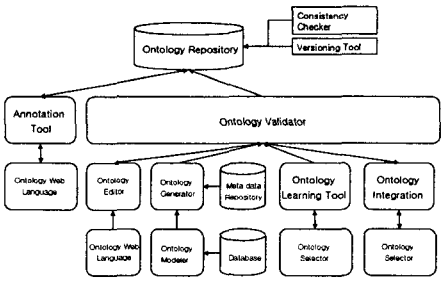

그림 3 온톨로지 시스템

온톨로지 시스템은 온톨로지 저장소와 온톨로지 생성을 하기 위한 온톨로지 생성기 부분, 온톨로지를 수정하기 위한 온톨로지 편집기 부분 및 온톨로지 학습과 온톨로지 통합을 하기 위한 구성 요소를 가지고 있으며, 온톨로지에 사용자가 주석 첨부를 하기 위한 온톨로지 주석기가 존재한다. 또한, 온톨로지 저장소에는 온톨로지가 생성이나 수정이 될 때에 이를 버저닝을 하기 위한 온톨로지 버저닝 구성요소와 온톨로지 무결성 검사기가 온톨로지의 오류를 체크하게 된다.

3.2 시스템의 구성요소

온톨로지 시스템은 온톨로지의 생성, 수정 및 저장을 하기 위한 온톨로지 시스템이다. 온톨로지 생성과 수정 및 통합 과정에서 발생 할 수 있는 오류를 검사를 하기 위한 온톨로지 유효성 검사기(validator)는 각 구성요소에서 작성된 온톨로지를 검사한다.

3.2.1 온톨로지 저장소(repository)

온톨로지 저장소는 생성과 수정된 온톨로지를 저장하는 역할을 한다. 온톨로지 저장시의 오류의 검사와 같은 도메인에서 새로이 작성된 온톨로지와 수정된 온톨로지가 기존의 온톨로지와 충돌이 발생하지 않도록 버전을 기록하는 버저닝 기술이 적용되어야 한다.

또한, 자체적으로 온톨로지를 분류하기 위한 온톨로지 분류기(classification tool)나 온톨로지 클러스터링 처리기가 내부적으로 적용이 되어 온톨로지의 효율적인 관리를 할 수 있도록 하여야 한다.

3.2.2 온톨로지 생성 컴포넌트

온톨로지 생성 컴포넌트는 2개의 구성요소를 가지고 있다. 먼저 온톨로지 모델러(ontology modeler)를 가지고 있으며, 도메인의 데이터 베이스와 연동이 되도록 하고, 도메인 내의 지식을 개념화하는 기능을 수행한다. 온톨로지 모델러에는 UML이나 RDF와 같은 도구를 사용을 한다. UML은 온톨로지 모델링을 할 때에 관계의 설정과 구조를 제작자 및 사용자가 알기 쉬운 방법이며[8], RDF는 개념화를 위한 모델링 방안을 제공한다[4].

다른 구성요소인 온톨로지 생성기는 메타 데이터 저장소와 연동을 하여 온톨로지 모델러가 생성한 개념화를 명세화하는 작업을 한다. 이때에 온톨로지를 생성하기 위해서 웹 온톨로지 언어인 OIL[9], DAML[9], SHOE[9]와 RDF[4][7]와 RDFS[4][7]등을 사용하여 온톨로지를 생성한다.

3.2.3 온톨로지 학습(learning) 컴포넌트

온톨로지의 학습은 생성과 통합을 위한 구성요소이다. 온톨로지 생성 컴포넌트와 통합 컴포넌트의 기능을 모두 가지고 있다. 그러나, 두 컴포넌트와는 다른 점은 기존의 온톨로지에서 기계학습의 방법을 사용하여 새로운 온톨로지를 생성하는 것과 통합시에 데이터마이닝 알고리즘은

적용하여 보다 자동화된 기능을 수행을 하는 것이다.

온톨로지 학습 컴포넌트는 온톨로지 선택기(selector)에 의해서 관리자가 온톨로지를 선택하고, 온톨로지 학습기(learning tool)가 관리자가 원하는 기능을 수행하는 것이다.

3.2.4 온톨로지 통합(integration) 컴포넌트

온톨로지 통합 컴포넌트 도메인 내의 유사한 온톨로지를 통합하고, 새로운 온톨로지를 생성하는 작업을 한다. 이는 같은 도메인 내에서의 온톨로지의 통합과 다른 도메인에서의 유사한 부분을 통합하는 역할을 담당한다.

이를 위해서 온톨로지 학습 컴포넌트와 같이 온톨로지 선택기가 온톨로지를 선택하고, 선택한 온톨로지의 통합을 수행하는 기능을 수행한다. 일련의 작업은 완전 자동화된 작업이 아니라 반자동의 기능을 수행한다.

4. 결론 및 향후과제

시맨틱 웹은 현재 새로이 제안되었으며, 이에 관련된 기술과 시스템들이 시험적으로 개발되고 있다. 시맨틱 웹은 웹 문서내에 온톨로지를 삽입함으로써 이상적인 기능과 서비스를 제공한다. 이를 위해서 본 논문에서는 시맨틱 웹에서 사용이 가능한 온톨로지 시스템의 설계를 제안하였으며, 이 시스템은 온톨로지 시스템에 필요한 기능들을 제시하고 있다.

이러한 시스템의 아키텍처는 향후 시맨틱 웹의 연구 및 개발 시에 전체적인 흐름을 이해하고 참조할 수 있는 역할을 할 수 있다.

이 시스템을 위해서는 온톨로지 응용에 대한 시나리오의 개발과 아울러 중요 컴포넌트인 온톨로지 생성을 위한 휴리스틱한 방법의 연구, 온톨로지 학습을 위한 기계학습 및 자연언어 처리 연구가 이루어져야 한다. 또한, 온톨로지 모델링을 할 때에 도메인 내의 지식을 개념화를 자동적으로 추출하기 위한 방안도 함께 연구가 되어야 한다. 아울러, 시스템의 재사용과 확장성을 위하여 소프트웨어 공학 측면에서의 연구도 필요하다.

참고문헌

[1] McIlraith.S.A, Son T.C, Honglei Zeng, "Semantic web services", IEEE Intelligent Systems, Volume:16 Issue:2, page(s):46-53, March-April, 2001

[2] J. Hendler. Agents and the Semantic Web. IEEE Intelligent Systems. Vol.16 Issue 2, p.30-37 March-April 2001.

[3] Introduction to Ontologies on the Semantic Web http://www.cs.umd.edu/users/hendler/ ontologies.html

[4] S. Decker, P. Mitra. S. Melnik. Framework for the semantic Web: an RDF tutorial. IEEE Internet Computing , Vol. 4 Issue 6 p.68-73, Nov.-Dec. 2000

[5] T.R.Gruber, "Towards Principles for the Design of Ontologies Used for Knowledge Sharing", International Workshop on Formal Ontology, Padova, Italy, 1993

[6] Berners-Lee. Tim, Heldler. J and Lassila. O, "The Semantic We", Scientific American, May 2001, 28-37.

[7] Decker.S, Melinik. S., van Harmelen.F., Fensel. D., Klein. M., Broekstra. J., Derdmann. M., Horrocks. I., "The Semantic Web: the roles of XML and RDF", IEEE Internet Computing, Volume: 4 Issue: 5, page(s) 63-73, Sep-Oct. 2000.

[8] S. Cranefield, S. Haustein, M. Purvis UML-Based Ontology Modelling for Software Agnets Proceedings of the Workshop on Ontologies in Agent Systems, 5th International Conference on Autonomous Agents. 2001

[9] Gomez-Perez. A, Corcho.O, "Ontology languages for the semantic web", IEEE Intelligent Systems, Volume:17 Issue:1, Jan-Feb, page(s):54-60, 2002

[10] Maedche. A., Staab. S., "Ontology learning for the Semantic Web", IEEE Intelligent Systems, Volume:16 Issue: 2, page(s) 72-79, March-April 2001

[11] World Wide Web Consortium, http://www.w3c.org