

시맨틱 웹을 위한 마크업 언어들과 관련 기술들에 대한 연구

A Survey on the Markup Languages and
Related Technologies for the Semantic Web

유명환*, 정희준*, 민재홍**, 이강찬**, 정인정*

*고려대학교 전산학과 **한국전자통신연구원 표준연구센터

{myong, joonny96, chung}@korea.ac.kr, **{jhmin, chan}@etri.re.kr

요약

컴퓨터 통신기술이 급격히 발달하고 인터넷이 광범위하게 사용됨에 따라서, 웹은 다양하며 무한한 용량의 데이터 원천으로 부각되었다. 그리고 문법이 간단하고 사용하기 쉬운 HTML 및 XML과 같은 마크업 언어들로 다양한 데이터 원천이 웹 처리에 알맞게 표현될 수 있음에 따라서, 컴퓨터 전문가가 아닌 일반인들도 쉽고 편리하게 원하는 정보에 접근할 수 있게 되었다. 그러나 웹의 사용이 지속적으로 증가함에 따라 웹 상에서 원하는 유용한 데이터를 찾고 구성하며 통합하는 일들이 매우 어렵게 되고 있다. 이와 같은 웹의 문제점들은 지금까지 광범위하게 사용되고 있는 HTML, XML, RDF(S) 등과 같은 마크업 언어들이 사람의 이해와 처리를 위해 고안되어 웹 데이터의 문법형식은 충분히 표현할 수 있으나 의미를 표현하는 능력이 크게 미흡하기 때문이다. 이러한 웹의 심각한 문제점들을 해결하고 자동화된 웹 서비스를 실현시키기 위하여, 온톨로지 기반의 시맨틱 웹 기술이 1990년대 말에 소개되었다. 컴퓨터 기계가 정보를 처리할 수 있도록 목적된 시맨틱 웹은 전자상거래, 지식표현 및 관리, 지능형 정보검색 등의 다양한 분야에서 광범위하게 활용되고 있다. 본 논문에서 우리는 웹의 지금까지의 문제점들을 알아보고 시맨틱 웹을 위해 현재 개발 중인 마크업 언어 및 관련기술 등을 살펴본다. 마지막으로 시맨틱 웹을 효과적으로 구현하기 위한 환경을 연구·제안한다.

Abstract

With the rapid development of computer telecommunications technology and the widespread use of the internet, web has emerged as diverse and infinite data source. Also representation of web processing with grammatically simple and easily manageable mark-up languages such as HTML and XML, has made it possible not only for computer experts but also for the general public to access desired information easily and conveniently. However, the web's exponential growth rate makes it extremely difficult to find, organize, and integrate the desired useful data on the web. The reason is that currently widely used mark-up languages such as HTML, XML and RDF(S) are capable of expressing syntactical structure, while they lack ability to express the semantics (meaning) of the web data. Hence, today's web is designed primarily for human interpretation and use. In an effort to solve such crucial problems and to realize our vision of automatic web services, the semantic web, based on ontology, was introduced in late 1990's. The semantic web, aimed at machine-processible information, is widely used in such diverse application areas such as e-commerce, knowledge representation and management, and intelligent information retrieval, etc. In this thesis, we examine the web's existing problems, study the markup languages and their related technology which is currently being developed. Finally, we investigate and propose the environment to realize the semantic web effectively.

1. 서론

1990년대 중반에 이르러 인터넷이 광범위하게 사용됨에 따라 웹은 다양한 정보원으로

써 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 정보시스템과의 통합이 어렵고, 단일 기종 내에서도 디자인의 변화에 따라 정보도 함께 수정해야하며 정보의 의미 전달이 취약하기 때문이다. 이러한 문제점을 해결하기 위해 시맨틱 웹이 1990년대 말에 제안되었다.[1,3,4,6,7,8]

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 웹서비스의 발견, 자동적인 웹 서비스의 실행과 동시에 웹 서비스들의 통합과 상호작용을 하여 사용자가 원하는 정보를 찾고 더 나아가 추론이 가능토록 한다.

본 논문에서는 차세대 웹 환경인 시맨틱 웹을 위한 마크업 언어에 대한 연구로써 현재까지 개발되어 있는 혹은 개발 중인 여러 마크업 언어에 대해서 살펴보고 시맨틱 웹에 적합한 환경을 연구·제시한다.

본 논문은 다음과 같은 구성으로 이루어져 있다. 2장에서는 현재 웹 환경의 마크업 언어와 그 문제점을 분석하고 지식을 표현하는 온톨로지와 시맨틱 웹의 연관성을 알아본다. 3장에서는 시맨틱 웹을 위한 마크업 언어의 개발 요구조건들을 설명하고 4장에서는 현재까지 개발되어 있는 마크업 언어의 특징을 분석설명하며 5장에서는 시맨틱 웹의 제

안 기저가 된 이론적 바탕에 대해 설명한다. 6장에서는 시맨틱 웹을 효과적으로 구현하기 위한 웹 환경의 인프라를 제시하며 7장에서는 시맨틱 웹의 활용 분야에 대한 설명을 하고 마지막으로 8장에서는 결론과 향후 연구 과제에 대해 언급한다.

2. 관련 연구

2.1. 현재 웹 환경의 문제점

현재의 웹은 HTML 기반의 문서들로써 표현위주의 웹이다. 표현위주 웹은 단일 기종 내에서도 디자인의 변화에 따라 정보도 함께 변해야하고 이는 이미 만들어 놓은 정보들의 재사용성을 떨어뜨린다. 또한 컴퓨터가 이해할 수 없어 사람에 의한 정보해석과 추출의 수고가 필요하다. 이러한 문제점을 해결하기 위해 에이전트, 자연언어처리나 XML과 같은 해결방안을 제시하였으나 웹이 텍스트뿐만 아니라 태그, 멀티미디어 자료 등과 같은 복합적인 구성으로 인하여 에이전트나 자연언어처리도 한계가 있으며 XML 역시 다음절에서 언급한 것과 같이 정보의 의미표현에 부적합하다. 이러한 정보의 표현력 부족은 더 나아가 정보교환을 어렵게 하여 다른 이기종 간의 정보 시스템과의 통합이 어렵고 상호운용성을 보장할 수 없으며 사용자가 원하는 자동화된 웹서비스를 제공하지 못 한다.

2.2. 기존의 마크업 언어

웹의 비약적인 성장은 문법이 단순하고 쉬운 HTML이라는 마크업 언어의 개발에서 비롯된 것이다. HTML은 텍스트, 이미지, 동영상 등을 보여줄 수 있으며 하이퍼링크를 통하여 다른 문서를 참조할 수 있는 장점이

있다. 그러나 HTML의 편리한 사용은 단점으로 작용하고 있다. HTML문서의 가장 큰 단점은 고정된 태그들만을 사용해야 한다는 점과 문서의 구조적인 정보를 표현할 수 없다는 점이다. HTML 문서는 데이터와 태그가 섞여있기 때문에 그 구조가 복잡하고 데이터의 변경이 용이하지 않다. 뿐만 아니라 한정된 태그로 인하여 이미 개발된 정보의 재사용성이 떨어지고 정보의 추출이 어렵다.[1,2,3]

이러한 HTML의 문제점을 극복하기 위해 XML이 개발되었다. XML은 표현과 데이터를 분리함으로써 데이터와 레이아웃을 수정이 용이하고 데이터에 의미를 줄 수 있다. 또한 사용자 정의의 임의적인 태그 정의가 가능하고 문서의 구조를 정의 해준다. 그러나 이러한 XML의 자율성은 임의적인 정보 구조와 태그의 사용으로 정확한 정보의 의미에 혼돈만을 가중시키는 단점이 있다. 또한 정보교환이나 비즈니스 프로세서 등과 같은 명확한 의미를 제공하기 위한 술어(terminology)를 제공해 주지 못한다. XML은 단지 문법 기술에 초점을 두어 개발된 언어으로써 정보들 간의 관계를 표현하기에 부족하고 의미가 상속되지 못한다.[11,12]

HTML이나 XML보다 발전된 형태의 마크업언어로는 RDF(S)가 있다. RDF(S)는 시맨틱 웹 마크업언어의 근간이 되는 언어으로써 XML보다 명시적이고 간단한 데이터 모델링으로 컴퓨터가 이해할 수 있도록 개발된 언어이다. RDF(S)는 XML 기반의 분산된 다양한 자원들을 기술하기 위한 것으로 상이한 메타 데이터를 효율적으로 교환하고 이해할 수 있는 정보 교환 수단을 제공한다. 또한 RDF(S)는 정보를 객체, 특성, 값으로 정의하고 정보의 재사용과 매칭이 용이하며 레어

어를 기반으로 하기 때문에 XML을 수용할 뿐만 아니라 XML의 변화에도 계속적으로 사용된다. 그러나 RDF(S)는 정보간의 관계나 동치성 등과 같은 연산·표현능력이 미약할 뿐만 아니라 추론을 위한 규칙표현을 해줄 수가 없다.[1,2,3] 이 문제점을 개선하기 위해 4장에서 언급한 OIL, DAML과 같은 언어가 개발되었다.

2.3. 온톨로지

컴퓨터가 정보를 이해할 수 없는 현재 웹 환경의 문제점들을 해결하기 위하여 컴퓨터가 이해할 수 있는 지식을 웹 문서에 삽입해야 한다. 이를 위해 공통적이고 표준적인 지식표현 방법이 필요하다. 이러한 관점에서 온톨로지는 매우 중요하다.

온톨로지는 공유되는 데이터들의 개념화한 형식적이고 명백한 규정이다. 이는 특정 분야에서 사용되는 표준 어휘들의 모음이라고 할 수 있다. 에이전트의 수행을 위한 정보를 만들고 질의어를 이용한 정보검색을 위해서는 도메인의 개념화를 해야 한다. 즉, 도메인 내의 지식을 표현하고 의사소통 하기 위한 단어를 제공함으로써 이기종 간의 상호운용성을 보장하고 지식의 통일화를 시킨다.

온톨로지는 어휘 사전의 역할이외에 지식을 효과적으로 표현하기 위해 정보의 의미를 부여하고, 또 정보들간의 관계를 설정해 준다. 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.[1,5,16,18]

시맨틱 웹을 위한 마크업 언어는 웹 문서에 이러한 온톨로지를 삽입함으로써 웹 문서

에 정보의 의미성과 정보간의 관계를 설정해 주고 정보의 수집을 용이하게 하며 추론과 정보의 재조합과 자동적인 실행을 하기 위한 조건을 만들어 준다.

3. 시맨틱 웹을 위한 마크업 언어의 조건

효과적으로 시맨틱 웹을 구현하기 위해서는 도메인에 적합한 온톨로지의 구성과 더불어 온톨로지를 만들기 위한 마크업 언어의 연구가 필요하다. 본 장에서는 현재의 웹 환경을 수용하고 효율적인 시맨틱 웹을 구현하기 위한 마크업 언어의 조건에 대해서 설명한다.[2,3,7,13,16,19]

첫째, 온톨로지를 쉽게 이해할 수 있도록 프레임 시스템과 같은 사용자에게 직관적인 지식표현방법을 사용해야 한다.

둘째, 온톨로지를 공유할 수 있어야 한다. 이는 온톨로지의 목적과 결부되는 것으로 공유되고 분산된 웹 환경에서 공유되지 않은 온톨로지는 온톨로지의 역할을 수행할 수 없으며 재사용할 수 없다. 마크업 언어는 효과적으로 온톨로지를 구성함과 동시에 표준화된 온톨로지를 배포가 가능해야 한다.

셋째, 온톨로지의 변화를 수용해야 한다. 웹 환경이 급변하고 정보의 양이 많아짐에 따라 이에 대처하기 위해서 마크업 언어는 온톨로지를 수정할 수 있어야 한다. 이는 동시에 온톨로지의 확장성 역시 고려해야 한다. 이미 만들어진 온톨로지를 새로운 온톨로지에서도 사용함으로써 온톨로지의 확장을 꾀하고 또한, 기존의 온톨로지의 재사용성을 높일 수 있다.

마지막으로 기존의 웹 환경을 수용할 수 있어야 한다. 이는 현실적인 면에서 가장 중

요한 점이다. 기존에 만든 HTML 문서나 XML문서를 수용함으로써 새로운 웹의 개발의 부담을 줄이고 동시에 그 동안 축적된 노하우를 활용할 수 있다.

위 네 가지 사항은 시맨틱 웹을 효과적으로 실현시키기 위한 마크업 언어의 필요조건으로써 이는 시맨틱 웹의 발전과 잠재성을 이끌어 낼 수 있다.

4. 시맨틱 웹을 위한 마크업 언어

4.1. OIL

OIL은 Ontology Interchange Language 또는 Ontology Inference Layer의 약어이다. OIL은 온톨로지를 표현함으로써 시맨틱 웹의 효과적인 구현을 위해 제안된 언어이다. OIL은 기술 로직들에 의해 추론 규칙을 삽입하고 프레임을 기반으로 하여 원시 정보를 온톨로지로 표현한다.[2,5,8]

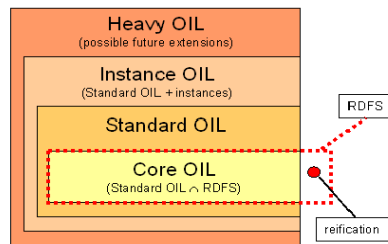


그림 1) OIL의 계층적 구조[2]

OIL은 개념을 표현하고 이를 상위 클래스와 슬롯으로써 나타내는 프레임 시스템의 모델링 원시 데이터를 통합시켰다. 또한 OIL은 잘 정의된 계층과 그 속성들을 나타내는 클래스 객체들으로써 슬롯을 표현한다. OIL에서의 기술로직은 개념의 정형화된 의미와 효과적인 추론을 지원하기 위해 사용된다. 또한 OIL은 XML의 문법을 사용한 RDF를 기저로 하여 개발되었고 이는 기존의 웹 언어

의 호환성을 위한 것이다.

OIL은 Core OIL, Standard OIL, Instance OIL, Heavy OIL로 각 계층을 이루고 있음으로써 OIL자체의 확장성을 고려하여 개발되었다.

4.2. DAML

DAML은 웹 문서에 존재하는 정보를 컴퓨터가 읽고 이해할 수 있도록 하기 위해서 XML 기술을 기반으로 하여 고안된 의미론적 언어이다. DAML의 프로젝트 리더인 짐 헨들러는 DAML을 문맥 정의 언어라고 정의한다. 서로 관련 있는 웹 문서라도 서로 다른 의미를 사용하기 때문에 발생하는 의미론적 장애를 해결하기 위한 수단으로 온톨로지와의 결합 방법을 제공한다는 점에서 HTML이나 XML과 같은 메타 태그와 구별된다.[8,14,15]

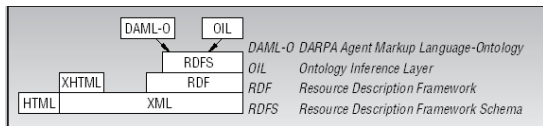


그림 2) 마크업 언어의 계층 구조[4]

DAML 이전에 앞서 OIL을 개발하였으나 이는 온톨로지를 표현하기에 부족한 점을 보완할 새로운 언어의 필요성 대두로 인하여 제안되었다. 이로 인하여 OIL이 RDF나 RDFS가 가지고 있었던 관계보다 의미적으로 더욱 밀접하게 결합된 DAML+OIL이라는 마크업 언어가 연구되었고 이는 RDF 문서로 작성된다.[4,10]

4.3. SHOE

SHOE는 Simple Html Ontology Extensions의 약어로서 웹에서 사용되는 HTML문서에 의미적인 정보를 추가하고

HTML의 확장태그로써 사용되기 위한 목적으로 만들어졌다. 즉, 웹 문서에서 컴퓨터가 이해할 수 있는 의미적 지식을 제공하기 위한 방법으로써 HTML을 확장시킨 것이다. 이는 앞서 설명한 OIL이나 DAML과 같은 언어와는 다르게 표현위주의 마크업 언어인 HTML을 주 대상으로 하고 있으며 HTML 문서를 그대로 이용이 가능하다.

SHOE는 HTML 문서 내에 SHOE 태그를 사용하여 해당 HTML 문서에 적합한 온톨로지를 구성하고 인스턴스를 생성한다 이렇게 HTML문서에 SHOE를 첨가함으로써 후에 웹 로봇이나 지능형 에이전트가 지식을 쉽게 수집할 수 있도록 한다. 즉 SHOE는 HTML 문서에 SHOE 태그를 삽입시킴으로써 사람의 관점의 정보 표현을 위한 문서인 동시에 컴퓨터가 이해할 수 있는 지식을 문서 내에 삽입시킨다.[8,17]

4.4. 기타 마크업 언어

앞의 세가지 마크업 언어는 시맨틱 웹의 마크업 언어로써 주류를 이루는 것으로써 그 외에도 몇가지의 마크업 언어들이 있다.

DAML+OIL은 DAML과 OIL의 특징들이 합쳐진 것으로써 이 마크업 언어는 컴퓨터가 이해할 수 있고 의미를 잘 정의함과 동시에 자동적인 추론 메커니즘을 제공한다. 또한 에이전트간의 통신을 위한 방법을 제공한다.[12,15]

DAML+L은 DAML의 확장된 형태로써 표현과 추론을 위한 제약조건의 복잡함을 줄인 것이다.[15]

DAML-S는 DAML+OIL을 이용한 시맨틱 웹서비스를 위한 마크업 언어로 DAML-S는 자동적인 웹 서비스 검색, 실행, 구성과 상호 운용성을 가능하게 한다.[15]

XOL은 XML기반의 온톨로지 교환언어로써 바이오 인포메틱스 분야에서 온톨로지의 교환을 위해 제안된 것이다.[9,21]

시맨틱 웹을 제안한 팀버너스리에 의해 만들어진 N3(Notation3)는 사람이 좀더 편하게 사용하기 위해 개발된 언어로써 SWAG를 제공하여 RDF로의 쉬운 전환을 할 수 있다. N3는 RDF와 로직을 표현하는 새로운 표시법에 대해 임의로 이름을 붙인다. 이는 결과값을 위한 RDF 데이터 모델로 매핑되는 같은 언어 안에서 데이터와 로직의 표현을 최적으로 하기 위한 목적이다.[20]

5. 형식 도구

컴퓨터가 자동적으로 정보의 추출, 추론, 처리 등을 수행하는 시맨틱 웹의 목적을 실현하기 위해서는 정형화된 도구가 필요하다. 시맨틱 웹을 위한 형식 도구는 크게 지식표현과 웹 서비스 두 측면으로 나눌 수 있다.

5.1. 지식표현을 위한 형식 도구

지식표현은 기존의 인공지능 분야에서 주로 연구된 학문분야로써 시맨틱 웹의 주요 정보 단위 요소인 온톨로지의 표현하여 지식을 컴퓨터나 에이전트에게 효과적으로 전달하기 위한 형식 도구들이다.

시맨틱 웹에서 지식표현을 위한 형식 도구는 주로 도메인 영역 내에서 나타나는 온톨로지의 특성을 규정하는 그 특징이 있다. 시맨틱 네트워크와 프레임 시스템은 주로 도메인 영역내의 개념들을 규정하여 온톨로지로서 표현하고 기술로직이나 혼 클로즈 로직은 주로 온톨로지간의 관계를 설정하고 추론을 위한 규칙 혹은 명제를 정의하는 형식 도구이다.

5.2. 웹 서비스를 위한 형식 도구

웹 서비스를 위한 형식적 도구는 컴퓨터나 에이전트가 작업을 수행하는 과정을 수학적 모델로써 표현한 것이다.

객체 지향 패러다임은 지식표현을 위한 온톨로지의 모델링에 그 초점을 두고 온톨로지 설계자의 입장의 도구로써 온톨로지의 설계와 정보 전달을 하여 효율적인 시맨틱 웹의 기능을 수행하도록 돕는다.[11]

기능성에 초점을 둔 형식 도구는 사용자에게 제공할 정보 혹은 온톨로지들의 상태변화를 나타내기 위해 오토마타의 FA(Finite Automaton)을 사용하고 다중 링크를 이용하여 관련정보의 참조를 사용자나 에이전트에게 알려주기 위한 것으로 PFA(Parallel FA)를 사용한다.[22]

또한 페트리 넷은 웹 서비스에서 사용자의 트랜잭션이나 행위에 초점을 두고 웹 서비스의 변화를 나타내기 위한 형식 도구이다. 페트리 넷은 시맨틱 웹에서 기본적인 과정(atomic process)을 수행하는 방법과 조합과정을 수행하는 방법에 대해서 좋은 형식 도구를 제공한다. 즉, 일련의 정보와 지식들을 선조건(precondition)으로 하여 최종적으로 수행할 웹 서비스의 계산 메커니즘을 제공하고 sequence, split, split and join, if-then-else, repeat-condition과 같은 기능을 제공함으로써 조합과정을 수행하는 도구로 사용된다. 또한 활성화조건, 상태 전이의 조건, 온톨로지의 추론규칙이 페트리 넷의 트랜지션으로 작용하여 관련 정보들의 참조를 위한 모델링 도구로써도 사용된다.[10,22]

6. 시맨틱 웹의 효과적 구축을 위한 인프라

분산된 정보의 수집, 처리와 이기종 간의 상호운용성을 보장하기 위해서는 시맨틱 웹의 표준화가 필요하다. 표준화는 온톨로지, 마크업 언어, 추론 및 정보 수집을 위한 API로 세 가지 관점에서 볼 수 있다.

온톨로지는 시스템간의 공유와 공통적인 이해가 주목적이다. 온톨로지는 시맨틱 웹에서 중요한 역할을 하는 것으로 명확한 온톨로지의 구성과 각각의 문제 도메인 영역에 적합한 온톨로지의 표준화가 있어야 될 것이다.[1,16,18]

다음으로 마크업 언어이다. 마크업 언어는 온톨로지를 표현하고 의미의 정확한 정의와 데이터간의 논리적 관계의 설정 등의 기능을 가지고 있다. 이러한 마크업 언어 기반의 에이전트는 데이터간의 관계를 통해 정보 추출과 추론이 가능하다. 현재는 W3C를 주축으로 하여 OIL과 DAML의 표준화에 노력을 하고 있다.[2,13,15]

마지막으로 API는 에이전트에 중요한 역할을 한다. 에이전트는 분산된 소스로부터 정보를 수집하고 또한 다른 에이전트들을 사용함으로써 사용자의 제약 사항과 선호 조건에 맞는 정보를 제공해 주어야 한다. 이러한 작업을 수행하는 에이전트를 위해서는 에이전트간의 일관된 API가 필요하다.[6,19]

그 외에 시맨틱 웹을 효과적으로 구현하기 위한 웹 환경의 인프라에 대한 연구이다. 시맨틱 웹의 정보 조합과 추론의 원론적인 문제뿐만 아니라 분산환경, 보안, 예외처리 등과 같은 문제가 발생한다.[8,10,19]

시맨틱 웹 환경은 정보 추출을 위해 한 정보의 원천에 동시에 여러 에이전트들이 질의를 수행할 경우 다중처리를 위한 트랜잭션 관리나 교착상태(Deadlock)문제 해결해야 한다. 또한 효과적인 질의를 위해서는 질의의

계획이나 최적화가 필요하고 에이전트에서 여러 개의 정보 원천에 접근하기 때문에 분산 스케줄링 역시 필요하다. 사용자의 제약 조건과 선호조건에 적합한 사용자의 질의나 에이전트의 질의와 같은 유동적인 요청으로 인한 예외상황이 많이 발생한다. 그러므로 예외처리를 통하여 시스템을 견고하게 해야 한다. 시맨틱 웹은 여러 정보의 원천을 돌아다니고 사용자의 제약 사항이나 선호 조건을 비롯하여 시스템의 지역 정보에 접근하기 때문에 보안 분야의 연구가 이루어져야 한다.

7. 활용분야와 사례

시맨틱 웹은 여러 분야에서 이용될 수 있다. 첫째, 전자 상거래이다. 현재의 시스템은 단순정보제공이나 래퍼(wrapper)에 의한 상품정보검색으로 인해 그 효율성이 떨어지므로 시맨틱 웹을 적용할 경우 사용자의 요구 조건과 제약조건에 만족하는 전자상거래가 가능하다.[1,2] 둘째, 지식 관리 분야와 E-비즈니스 분야이다. 시스템의 거대화, 광역화로 인한 비효율성과 관리와 지식 표현이 어렵다. 온톨로지를 사용하여 문서의 구조나 의미를 정의해 주어 정보의 의미를 부여하고 이를 통해 지능형 검색이나, 지식 표현을 해 줄 수 있다. 또한 각 회사간의 정보교환을 위해 표준화된 방법을 제공하여 이를 통해 광범위한 비즈니스 분야 적용이 가능하며 서로 다른 비즈니스 분야간의 온톨로지 변환을 통하여 비즈니스의 통합을 꾀할 수 있다.[1,2,7,24] 마지막으로 생물 정보학 분야이다. 현재 유전자를 나타낼 수 있는 온톨로지를 구축을 연구가 진행중이다. 시맨틱 웹의 기술을 사용하여 유전자 검색과 정보의 분석 등에 이용할 수 있다.[9,23]

DAMLDining은 식당의 각종 정보를 DAML로 온톨로지를 구성함으로써 에이전트를 통해 다양한 방법의 질의를 할 수 있도록 하였고 Teknowledge Corp는 채권, 금융 정보를 온톨로지로 구성하여 인텍싱과 에이전트를 위한 DAML 사용방법을 논의되고 있다. 또한 SHOE를 개발한 메릴랜드 대학에서는 컴퓨터학과의 온톨로지를 구성하여 예시으로써 배포하고 있다. 생물정보학 분야에서는 분자생물학을 표현한 TAMBIS 온톨로지[23]를 OIL로 표현하기 위한 연구가 진행 중이다. OAA(Open Agent Architecture), CongnIT, Expose, SHOE Search, DAML Crawler[8,15,17] 등의 경우 웹 문서에 마크업 언어로 표현된 온톨로지들을 검색할 수 있는 에이전트나 웹 로봇을 제공하고 있으며 ITTalks[24]의 경우 사용자가 요구하는 분야, 시간, 장소, 제약 사항 등의 사용자 프로파일을 사용하여 사용자의 스케줄, 패턴, 프로파일 자동 수정 기능을 함으로써 사용자가 원하는 미팅 스케줄이나 토론을 할 수 있는 장을 만들어 준다. 웹 문서에 시맨틱 웹 마크업 언어를 기입하기 위해 The Knowledge Annotator, Oiled, Protege2000, OntoEdit, Ontomat와 같은 편집기가 제공되고 있으며 이는 웹 상에서 직접 실행할 수 있거나 다운로드받아 사용하도록 배포되고 있다. HyperDAML, DAML Validator, PIQ, ConsVISor, Closed World Model, DAML/XSLT Adapter와 같은 마크업 언어의 전환, 유효성 검사, 질의 도구들이 개발되어 있으며 온톨로지 설계에 UML 설계를 사용할 수 있도록 Rational Rose나 Visio의 경우 관련 도구들을 발표하였다. 이외에도 DAML Viewer, PalmDAML과 같은 뷰어가 개발 배포되고 있다. PalmDAML의 경우 컴

퓨팅 환경이 빈약한 모바일 기기에서도 시맨틱 웹을 사용할 수 있으며 이기종의 다양한 정보접근의 방법의 가능성을 보여주고 있다. 시맨틱 웹의 관련 API로는 스탠포드 대학에서 JAVA로 개발한 RDF API를 발표하여 Protege2000, OntoEdit 등과 같은 편집기에서 쓰이고 있을 뿐만 아니라 소프트웨어의 효율적인 개발을 위해 개발된 MS사의 .NET에서도 RDF.NET을 발표하여 OIL과 DAML의 기반이 되는 RDF 모델을 처리하고 파싱하기 위한 프레임워크를 발표하였다.[17]

8. 결론 및 향후 연구 과제

시맨틱 웹은 현재의 웹 환경의 문제점을 해결하고 자동화되고 추론과 정보의 재조합이 가능한 웹 환경의 구축을 목적으로 한다. 이에 본 논문에서 시맨틱 웹을 위한 마크업 언어와 각종 기술에 대한 이해를 돕고 효과적인 시맨틱 웹 환경을 구축하고자 하였다.

시맨틱 웹은 태동기으로써 발전 가능성과 그 잠재능력이 무한하다. 이에 대하여 효율적인 시맨틱 웹을 구축하기 위해서는 좀더 정형화되고 효율적인 형식 도구의 연구와 온톨로지를 표현하고 추론 규칙을 정의하기 위한 마크업 언어의 연구가 활발히 이루어져야 할 것이다. 또한 시맨틱 웹에서 이러한 문서 정보를 수집하고 처리하기 위해서는 에이전트의 역할이 중요하고 지금껏 연구가 진행되어 온 에이전트의 연구를 시맨틱 웹 분야로의 활용을 모색해야 할 것이다.

시맨틱 웹은 주로 지식표현과 추론분야를 연구하는 인공지능과 로직 프로그래밍 분야로 연구가 진행되어오고 있다. 시맨틱 웹의 효율성과 가능성을 재고하기 위해서는 이의

에도 에이전트, 분산처리, 보안등과 같이 오랫동안 연구가 되어 왔고 많은 노하우를 축적한 기타 여러 학문 분야의 접근이 필요하다. 이와 함께 기존의 웹 환경을 수용할 수 있도록 기존의 웹 데이터들의 통합과 시맨틱 웹으로의 전환이 필요하다.

참고 문헌

- [1] McIlraith. S.A., Son. T.C., Honglei Zeng, Semantic Web services , IEEE Intelligent Systems, Vol.16 Issue: 2 , Page(s): 46-53, 2001
- [2] Fensel. D., van Harmelen. F., Horrocks.I., McGuinness. D.L., Patel-Schneider. P.F., OIL: an ontology infrastructure for the Semantic Web , IEEE Intelligent Systems, Vol. 16 Issue: 2, Page(s): 38-45, 2001
- [3] Decker. S., Melnik. S., van Harmelen. F., Fensel. D., Klein. M., Broekstra. J., Erdmann. M., Horrocks. I., The Semantic Web: the roles of XML and RDF , IEEE Internet Computing , Vol. 4 Issue: 5, Page(s): 63-73, 2000
- [4] Lassila. O., van Harmelen. F., Horrocks.I., Hendler. J., McGuinness. D.L., The semantic Web and its languages , IEEE Intelligent Systems, Vol.15, Issue: 6, Page(s): 67 -73, 2000
- [5] D. Fensel, I. Horrocks, F. Van Harmelen, S.Decker, M. Erdmann, M. Klein, OIL in a Nutshell , Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000), 2000
- [6] Hendler, J. Agents and the Semantic Web. IEEE Intelligent Systems Vol.16 Issue: 2, Page(s): 30 -37, 2001
- [7] O. Lassila, D. McGuinness. The Role of Frame-Based Representation on the Semantic Web. Linkoping Electronic Articles in Computer and Information Science, ISSN 1401-9841, Vol. 6. 2001
- [8] Jeff Heflin. Towards the Semantic Web: Knowledge Representation in a Dynamic Distributed Environment. Ph.D. Thesis, University of Maryland, College Park, 2001
- [9] Peter D. Karp., An ontology for biological function based on molecular interactions ,Bioinformatics Vol. 6 Page(s) 269~285, 2000
- [10] Srinu Narayanan, Sheila McIlraith Simulation, Verification and Automated Composition of Web Services. WWW2002 May. Honolulu, Hawaii, 2002
- [11] Stephen Cranefield. UML and the Semantic Web. Proceedings of the International Semantic Web Working Symposium (SWWS), 2001
- [12] Yolanda Gil, Varun Ratnakar. A Comparison of (Semantic) Markup Languages. Proceedings of the 15 th International FLAIRS Conference. 2002
- [13] <http://www.w3.org/TR/webont-req>
- [14] <http://www.w3.org/TR/2001/NOTE-daml+oil-reference20011218>
- [15] <http://www.daml.org/>
- [16] <http://www.semanticweb.org/knowmarkup.html>
- [17] <http://www.cs.umd.edu/projects/plus/>

SHOE/

[18] <http://www.cs.umd.edu/users/hendler/ontologies.html>

[19] <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>

[20] <http://www.w3.org/DesignIssues/Notation3.html>

[21] <http://www.ai.sri.com/pkarp/xol/>

[22] http://www5conf.inria.fr/fich_html/papers/P19/Overview.html

[23] Robert Stevens, Carole Goble, Ian Horrocks, Sean Bechhofer, Building a bioinformatics ontology using OIL, IEEE Transaction on information technology in biomedicine Vol. 6, no2, June 2002

[24] ITtalks: a case study in the Semantic Web and DAML+OIL, IEEE Intelligent Systems, Vol. 17 Issue: 1 , Jan.-Feb. 2002

유명환

고려대학교 자연과학대학 전산학과 졸업
2003년 2월:
고려대학교 자연과학대학 전산학과 대학원 졸업

정희준

고려대학교 자연과학대학 전산학과 졸업
2003년 2월:
고려대학교 자연과학대학 전산학과 대학원 졸업

민재홍

1974년 ~ 1978년:

고려대학교 산업공학과(공학사)

1988년 ~ 1990년:

고려대학교 경영대학원(경영정보석사)

1999년 ~ 2001년:

고려대학교 자연과학대학 전산학과 (박사수료, 인공지능)

1978년 ~ 현재:

한국전자통신연구원 표준연구센터
표준기반연구팀장

관심분야:

데이터베이스, 정보검색

이강찬

충남대학교 컴퓨터공학과(학사)

충남대학교 컴퓨터공학과(석사)

충남대학교 컴퓨터공학과(박사)

현재:

한국전자통신연구원 표준연구센터
표준기반연구팀 선임연구원

관심 분야:

데이터베이스, 정보통합, XML,
미디어이터, 시맨틱 웹

정인정

서울대학교 계산통계학과(이학사)

한국과학기술원 전산학과(공학석사)

1989:

University of Iowa 전산학과(Ph.D.)

현재:

고려대학교 자연과학대학 전산학과 교수