시맨틱 웹과 그 응용

The Semantic Web and Its Applications

양기철
목포대학교 정보공학부 멀티미디어 공학 전공

Yang Gi-Chul
Mokpo Univ.

요약

우리의 일상 생활의 일부가 되어버린 웹은 시맨틱 웹으로 한 번의 탈바꿈을 준비하고 있다. 웹상의 정보를 수집하고 처리하는데 더 이상의 인간의 전적인 개입을 요구하지 않겠다는 것이다. 시맨틱 웹을 통해서는 각종 자동화된 에이전트를 이용하여 효율적인 정보검색, 새로운 지식의 생성, 최적의 서비스 제공 등이 가능하게 된다. 본 논문 에서는 최근 연구가 활발히 진행되고 있는 시맨틱 웹과 그 응용분야에 대하여 알아본다.

Abstract

Web is a part of our daily life and now it is ready to become a new Web called Semantic Web. The Semantic Web does not require entire human process for collecting and processing the information on the Web. Automation of efficient information retrieval, new knowledge generation, exact web service providing and so on are become possible through the Semantic Web. In this paper, the main idea of the Semantic Web and its application areas are discussed.

I. 시맨틱 웹이란 무엇인가?

오늘날 많은 문서들이 웹(Web)을 통하여 유통되고 있다. 그러나 현재의 웹은 대부분은 HTML이나 XML로 작성되어 있어 기계에 의한 내용의 처리는 불가능하고 단지 보여주는데 그치고 있다. 하지만 인터넷은 사람만이 사용하는 것이 아니라 여러 가지 소프트웨어에 의해서 사용되기도 한다. 이러한 소프트웨어들은 정보검색, 정보추출, 데이터 마이닝 등 여러 가지 기능을 효율적으로 수행하는데 필수적인 소프트웨어이며, 앞으로는 이러한 소프트웨어에 의한 인터넷의 사용이 기하급수적으로 증가할 것이다.

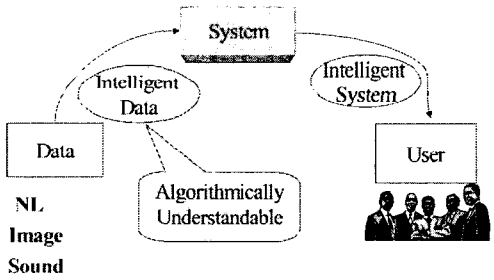
따라서 이러한 소프트웨어들이 내용(의미)을 잘 이해할 수 있는 형태의 웹 표현이 필요하다. 이를 위하여 국제조직인 W3C에서는 기계가 의미를 쉽게 이해할 수 있도록 웹을 작성할 수 있는 표현 형태인 Resource Description Framework (RDF)를 표준으로 제안하였다. 이는 온톨로지와 메타데이타를 이용하여 개념을 표현하며 Uniform Resource Identifier (URI)를 이용한 Name Space에 의하여 모호성을 해소하고 있다. 이처럼 기계가 의미를 파악할 수 있는 형태로 작성된 웹을 시맨틱 웹(Semantic Web)이라 한다[1].

지금까지는 우리는 자연어, 그림, 소리 등 다양한 미디어를 이용하여 저장된 수많은 정보를 저장하고 검색하여 이용하기 위하여 정보 시스템을 활용하였으며, 그 정보의 수가 보다 많고 다양하여 저감에 따라 정보 시스템을 보다 더 지능적으로 만들어 사용하려고 노력하였다. 하지만 아래 [그림 1]에 나타나 있는 것처럼 시맨틱 웹 방식은 정보 시스템을 지능적으로 만들기보다 데이터를 지능적으로 만드는 방식을 택한 것이다. 이렇게 하여 시맨틱 웹은 기계가 처리 가능한 웹 데이터를 생성하고 기존 정보시스템의 한

계를 극복할 수 있게 한 것이다.

시맨틱 웹은 처음에 Tim Berners-Lee에 의하여 그 아이디어가 제안 되었다[2]. Tim Berners-Lee가 생각한 미래의 웹은 HTML이나 XML을 이용하는 웹 이상의 것을 포함하고 있다. 이러한 것들을 나타내기 위해서는 별도의 데이터가 필요한데 이러한 메타데이터가 있음으로 해서 컴퓨터가 웹상의 정보를 이해하고 처리할 수 있게 된다. 이러한 메타데이터는 다음에 설명할 RDF(Resource Description Framework)를 이용하여 표현할 수 있으며, 웹상의 데이터를 보다 지능적으로 만든다.

Semantic Web Approach



▶▶ 그림 1. 시맨틱 웹의 접근 방식

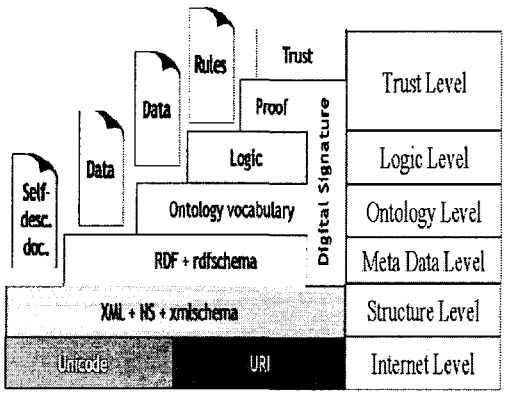
웹을 세대별로 구분하여보면 시맨틱 웹은 3세대 웹이라고 말할 수 있다. 1 세대 웹은 1990년대 초 HTML에 기반한 웹으로 문서가 있는 위치에 관계없이 인터넷을 이용해 어디에서든지 같은 형태로 문서를 볼수 있게 하였다. 따라서 웹의 표현(Style)이 중요했다. 2 세대 웹은 1990년대 후반 XML에 기반을 둔 웹으로 구조(Structure)를 중요시 했다. 따라서 같은 내용도 필요에 따라 표현이 다르게 보여 질 수 있다. 그리고 3 세대 웹인 시맨틱 웹은 기계가 콘텐츠의 의미를 이해할 수 있게 의미정보를 추가하여 웹 데이터를 확장하는데 역점을 두고 개발되고 있다[그림 2].

3 세대 웹	RDF, OWL 등에 기반을 두고 웹의 구조에서 의미를 분리시킴.
2 세대 웹	XML, XSL 등에 기반을 두고 웹의 표현(Presentation)에서 구조를 분리시킴.
1 세대 웹	HTML, PDF 등에 기반을 두고 위치(Location)에서 표현을 분리시킴.

▶▶ 그림 2. 세대별 웹의 비교

Ⅱ. 시맨틱 웹의 구조

시맨틱 웹은 [그림 3]과 같은 구조로 이루어진다. Internet Level에서 코드는 Unicode를 사용하고 URI(Uniform Resource Identifier)를 사용하여 객체를 지칭한다. 웹 내의 객체의 구조를 정의하는 Structure Level에서는 기본적으로 XML을 NS(Name Space)와 함께 사용한다. 그리고 객체간의 관계정의를 위하여 Meta Data Level에서 RDF(Resource Description Framework)를 사용한다. 각 객체와 개념들의 정의를 위해서 Ontology가 필요하고 보다 다양한 객체간의 관계 정의를 위하여 Logic이 필요하다. 또한 추론의 결과를 얻기 위해 Proof가 필요하며 추론에 사용한 데이터의 신뢰성검증을 위하여 Trust Level이 필요하다.



▶▶ 그림 3. W3C의 시맨틱 웹의 개념적 계층

Internet Level

인터넷상에서 문자나 기호를 표현하기 위해서는 코드체계가 필요한데 시맨틱 웹에서는 유니코드(Unicode)를 사용한다. 유니코드는 90,000개 이상의 문자와기호를 표현할 수 있어 한글 한자 등 다양한 국가의 언어를 표현할 수 있다. 하지만 한글의 경우 완성형 글자만 표현이 가능하다는 단점도 있다.

인터넷상에서 무엇에 대하여 기술하려면 그것을 인지하고 지칭할 수 있어야한다. 웹상에서는 이러한 인지와 지칭을 위하여 Identifier를 사용하며 일관된 체계를 유지하기 위하여 Uniform Identifier를 사용한다. 그런데 웹상의 모든 것들은 자원(Resource)이라 불리기 때문에 시맨틱 웹에서 사용하는 URI(Uniform Resource Identifier)라 한다. 따라서 URI를 사용하면 모호성이 없이 웹상의 모든 자원을 인지하고 지칭할 수 있다. 또한 웹상에서 모든 자원은 복사하여 가져오거나 중앙집중식으로 관리할 필요가 없고 필요할 때 링크를 이용하면 되기 때문에 URL(Uniform Resource Locator)이 아닌 URI(Uniform Resource Identifier)이다.

Structure Level

웹 상에서 객체들을 정의할 때 XML(eXtensible Markup Language)을 사용하면 다음과 같이 각 객체의 구조를 정의할 수 있다.

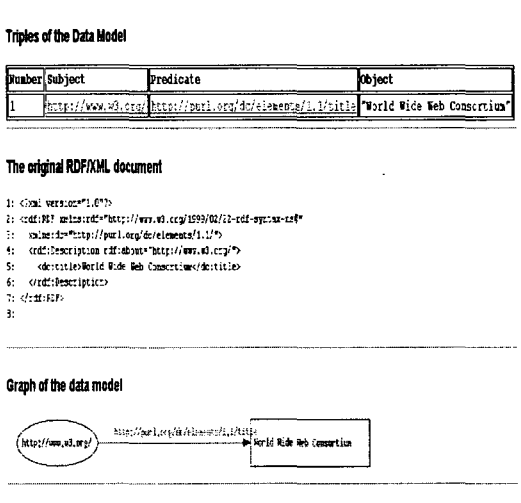
```
<ADDRESS>
    <NAME> TOM </NAME>
    <PHONE> 012 3456 </PHONE>
</ADDRESS>
```

하지만 XML로는 각 객체간의 관계를 정의할 수 없다. 또한 XML을 사용하면 사용하는 사람마다 다른 태그를 같은 용도로 사용할 수 있다. 하지만 XML을 사용하면 사람마다 서로 다른 형태의 태그를 사용하는 경우가 많으며 이는 태그의 의미를 이해하기 어렵게 만든다. 더욱이 컴퓨터는 이러한 태그의 의미를 이해할 수 없다.

Meta Data Level

컴퓨터가 문장의 의미를 이해하기 위해서는 추가적인 정보가 필요한데 이러한 정보가 Meta Data Level에 기술되고, 웹에서 의미기술을 위해 현재 가장 널리 알려진 언어가 RDF[4],[6]와 RDF Schma (RDFS)[3]이다. RDF는 [그림 4]에서 보는 것처럼 Subject, Predicate 그리고 Object라는 3가지의 요소로 이루어진다. 이는 [그림 4]의 맨 아래에 나타나있는 그래프 형태로도 표현이 되며 중간에 있는 XML 형태로 기술되기도 한다. 또한 Notation 3 (N3) 라는 표기법을 사용하기도하지만 N3로 기술된 내용은 사람이 읽기는 용이하지만 일반적인 XML이나 RDF 파서로는 파싱이 되지 않는다.

RDF는 간단한 관계형 모델로 데이터를 Object-Attribute-Value(또는 Subject-Predicate-Object)의 트리플 형태로 표현한다. 그리고 트리플의 각각은 리터럴 심볼이거나 URI이다. 예를 들어 "철수는 음악이야기의 저자이다"와 "음악이야기는 책이다"는 트리플로 다음과 같이 표현될 수 있다.



▶▶ 그림 4. RDF표현 형태

Ontology Level

온톨로지(Ontology)는 개념과 객체들을 계층적 관계로 정의하여 컴퓨터가 의미를 이해하는데 필요한 정보를 제공한다. [5]에 따르면 온톨로지는 다음과 같이 정의할 수 있다.

"An Ontology is a formal, explicit specification of a shared conceptualization."

위의 정의에서 보듯이 온톨로지는 'Formal'하고 'Explicit'하기 때문에 기계에 의한 개념의 자동 이해가 가능하고, 정보를 공유하는 특성을 갖는다.

이러한 온톨로지는 사용 목적이나 응용분야 등 에 따라 다양한 형태로 만들어 질 수 있다[Fensel01]. 잘 알려진 온톨로지로는 WordNet, RosettaNet, CYC 등 이 있다. 또한 온톨로지를 구현할 때 사용하는 언어도 여러 가지가 있는데 OIL(Ontology nference Layer), DAML(DARPA Agent Markup Language) 그리고 OIL과 DAML을 합한 DAML+OIL 등이 있다. OIL과 DAML은 Description Logic에 기반을 둔 논리적 언어이다. DAML+OIL은 논리와 객체지향 개념 그리고 RDF(S)와 XML같은 웹 언어의 특성들이 합성되었다. 따라서 DAML+OIL과 같은 언어를 사용하면 웹 상의 메타데이타를 더 쉽게 작성할 수 있다. 또한, OWL(Web Ontology Language)은 W3C (http://www.w3.org/2001 /sw/WebOnt/) 에서 제안한 것으로 DAML+OIL에 이어 온톨로지 언어의 표준으로 자리잡고 있다.

Logic & Trust Level

온톨로지는 개념과 객체들을 정의하지만 이들을 어떻게 사용해야 되는지는 알려주지 않는다. 예를 들어, 컴퓨터를 5대 사야 되는데 한 회사에서 3eomff 사고 다른 회사에서 2대를 주문했으면 더 이상 주문이 필요없다는 사실을 사용자에게 알려 주어야한다. 이러한 일은 온톨로지만 있어가지고 해결되는 것이 아니다. 이러한 결론을 도출하기 위해서는 절차적인 행동과 논리적인 추론이 있어야 가능하며, 이러한 작업들이 시맨틱 웹의 Logic Level에서 이루어진다.

또한, 논리적 추론의 결과가 참이다는 사실을 증명할 수 있어야 된다(Proof Level). 이때 추론 결과가 참이라는 것은 전제로 이용하는 웹데이타가 참이라는 가정하에서 가능하다. 하지만 또 하나의 문제는 공개된 수많은 웹상의 자료들을 모두 믿을 수 있는가 하는 것이다. 따라서 웹상의 자료의 신뢰도를 측정하고 보장할 수 있는 방법이 필요하다. 이러한 작업은 시맨틱 웹의 Trust Level에서 이루어진다. 한 가지 방법은 웹 사용자가 전자서명을 이용하여 각 사이트에 신뢰도를 평가하여 기록하도록 하는 것이다.

Ⅲ. 시맨틱 웹의 응용 분야

메타데이타나 온톨로지 사용 기술같은 시맨틱 웹 관련 기술은 다음과 같은 여러 분야에서 응용될 수 있다.

- 정보검색: 웹 사용에 있어서 가장 큰 문제중의 하나가 사용자가 필요한 정보나 서비스를 빠리 찾기가 어렵다는 것이다. 이는 인터넷 상의 정보가 너무 많을 뿐만 아니라 구조화 되어있지 않기 때문이다. 따라서 기존의 키워드를 이용한 방식보다 내용기반 정보검색이 가능하도록 하여야 한다. 이러한 시도를 한곳으로 Karlsruhe University AIFB Institute의 Semantic Portal[7 있다.
- 지식관리: 최근 들어 기업이나 조직에서 자신들이 보유하고 있는 지식관리에 관심이 크다. 그러나 각 기관이 보유하고 있는 비정규적인 문서는 경쟁력있고 효율적인 지식관리를 어렵게 만든다. 하지만 시맨틱 웹 기술을 사용하면 다음과 같은 장점이 있다.

- 키워드가 아닌 내용기반 정보검색이 가능하다.
- 열거된 정보를 단순히 보여 주는 것이 아니라 원하는 정보를 조직하여 만들어 낼 수 있다.
- 이형의 시스템간 문서의 이동이나 교환이 쉽다. 따라서 서로 다른 병원의 데이터나 서로 다른 회사의 카달로그, 박물관 자료의 통합 등이 가능해진다.

◦ 전자상거래: 인터넷의 발달로 새로운 형태의 상거래 시스템이 형성되었다. 기업과 기업간 전자상거래인 B2B나 기업과 개인간 전자상거래인 B2C를 위해 시맨틱 웹 기술은 유용하게 사용될 것이다. 전자상거래를 위한 국제표준으로 다음과 같은 것들이 시도 되었다.

- ebXML: ebXML(Electronic Business XML)은 기업들의 전자상거래시 사용할 수 있는 XML 실행의 표준화를 시도하였다.
- SOAP: SOAP(Simple Object Access Protocol)는 HTTP 프로토콜에 기반한 표준 프로토콜로 시스템간에 XML을 이용한 정보교환이 가능하다. 이 밖에도 에이전트간 통신에 사용되는 언어로 KQML(Knowledge Query and Manipulation Language), FIFA(Foundation for Intelligent Physical Agents) 그리고 OMG(Object Management Group) 등 이 있다.
- Ontology: 전자상거래에 사용되는 온톨로지 중 RosettaNet은 IT와 전자제품 분야에서 가장 잘 알려진 표준이다.

위에서 본 것처럼 여러 분야에서 시맨틱 웹이 응용될 수 있으며 실제로 사용자가 시맨틱 웹을 이용하기 위해서는 웹 서비스를 활용하여야 한다. 시맨틱 웹 기반 웹 서비스의 활용에는 자동 웹 서비스 검색, 자동 웹 서비스 실행, 자동 웹 서비스 구성과 상호운용, 자동 웹 서비스 실행 모니터링 등 다양한 작업이 가능하다.

자동 웹 서비스 검색은 주어진 제약 사항을 만족시키는 서비스를 제공하는 웹 서비스의 위치를 자동으로 찾아내 주는 것이다. 즉, 사용자가 원하는 웹 서비스를 제공하는 사이트의 위치를 사용자가 일일이 찾아다니지 않아도 자동으로 찾을 수 있는 것이다. 자동 웹 서비스 실행이란 컴퓨터나 에이전트가 찾아낸 웹 서비스를 자동으로 실행할 수 있게 해 주는 것이다. 예를 들어 현재는 웹상에서 치과병원에 예약을 하려고 하면 사용자가 직접 원하는 병원의 홈페이지에 접속하여 필요한 정보를 입력하고 실행버튼을 눌러 예약을 한다. 이러한 작업을 컴퓨터가 자동으로 해주는 것이다. 자동 웹 서비스 구성과 상호운용이란 하나의 목적을 이루는데 여러 가지 웹 서비스가 필요할 경우 필요한기능이다. 기존의 웹에서는 이러한 작업을 사용자가 따로따로 수행하여 그 결과들을 조합하여야 가능하지만 시맨틱 웹상에서는 이러한 통합 작업이 자동으로 실행될 수 있다. 자동 웹 서비스 실행 모니터링 기능은 각각의 웹 서비스가 실행되는데 필요한 시간이나 진행 상황 등에 대한 정보를 알려줄 수 있는 기능이다. 이와 같이 시맨틱 웹상에서는 여러 가지 웹 서비스가 자동으로 실행 가능하게 되어 기존의 웹에서는 할 수 없었던 일들의 자동 수행이 가능하게 된다.

▌참 고 문 헌▐

[1] Berners-Lee, T., J. Hendler & O. Lassila(2001). The semantic web. Scientific American, 284(5):3443, May.

[2] Tim Berners-Lee, Information Management: A Proposal CERN March 1989, May 1990

[3] Brickley D. & Guha R., Resource Description Framework(RDF) Schema Specification 1.0, W3C Candidate Recommandation2000-03-27, February 2000.

[4] Hjelm J., Creating the Semantic Web with RDF. John Wiley & Sons, New York, 2001.

[5] Gruber T.R., A Translation Approach to Potable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, 1993.

[6] Lassila, O. & Swick R., Resource Description Framework(RDF): Model and Syntax Specification. Technical Report, W3C, 1999.
[7] Maedche, A., et al., Semantic Portal - the SEAL approach. Technical report, Institute AIFB, Universitv of Karlsruhe. Germanv. 2001.