

# 시맨틱 웹

김은주 | 한국전산원 지식정보기술단 정보화표준부

웹기술은 팀버너스리가 1990년대 초에 최초로 월드 와이드웹을 창시한 이래로 꾸준히 발전해 왔으며, 2000년대에 들어서면서, 제2의 성장기를 맞이하고 있다. 기존의 웹이 양적성장을 목표로 하는 성장이었다면, 제2의 성장기는 시맨틱 웹, 신뢰의 웹 등에 기반한 질적인 성장을 의미한다. 새로운 웹의 패러다임으로 언급되기도 하는 시맨틱 웹은 사실 기존의 웹에 대한 개혁이라기 보다는 진화라고 표현하는 것이 옳을 것이다. 시맨틱 웹은 하이퍼텍스트 링크를 통한 가상공간에서의 위치적 연결을 넘어선 정보자원들 간의 의미적 연결을 컴퓨터가 이해할 수 있는 형태의 언어로 표현함으로써 가능해질 수 있다. 21세기의 웹은 궁극적으로 기계가 정보자원의 의미를 이해하고 이를 바탕으로 논리적 추론이 가능할 수 있게 됨으로써 기계들 사이에 커뮤니케이션이 가능할 수 있는 웹으로 발전하게 될 것이다.

시맨틱 웹을 한마디로 정의하자면 “컴퓨터가 정보의 의미를 이해하고 의미를 조작할 수 있는 웹”이라고 할 수 있다. 관계형 데이터베이스에서 ‘관계’가 ‘의미’를 뜻하듯이, 웹에서도 정보 리소스들 사이의 연관성을 잘 표현해준다면 웹에서 의미의 처리는 가능할 수 있는 것이다. 시맨틱 웹은 하이퍼텍스트 링크를 통한 가상공간에서의 위치적 연결을 넘어선 정보자원들 간의 의미적 연결을 컴퓨터가 이해할 수 있는 형태의 언어

로 표현함으로써 가능해질 수 있다. 21세기의 웹은 궁극적으로 기계가 정보자원의 의미를 이해하고 이를 바탕으로 논리적 추론이 가능할 수 있게 됨으로써 기계들 사이에 커뮤니케이션이 가능할 수 있는 웹으로 발전하게 될 것이다.

## 1. W3C에서의 시맨틱 웹 표준화

현재 시맨틱 웹에 관한 연구와 표준화 작업은 W3C를 중심으로 이루어져 오고 있다. 시맨틱 웹의 주요 기술로는 XML(eXtensible Markup Language), RDF(Resource Description Framework), OWL 등을 들 수 있다. 이들 기술 중 차세대 웹 아키텍처의 기반을 이루는 XML 기술을 제외한 W3C의 시맨틱 웹 관련 연구 및 표준화는 시맨틱 웹 Activity에서 다루어지고 있다. 시맨틱 웹 Activity내에는 RDF Core 작업반(Working Group)과 웹 온톨로지 작업반이 있으며, 주된 표준들은 이 두 작업반을 중심으로 산출되고 있다.

현재 시맨틱 웹 Activity의 활동기한은 2003년 3월을 기준으로 마쳐진 상태이며, 그간의 산출물의 완성도를 높이는 마무리 작업을 위하여 이 Activity의 활동기한을 2004년까지 연장하고자 하는 제안이 제출되

어 있는 상태이다. W3C는 또한 시맨틱 웹의 기반을 형성할 핵심 컴포넌트들을 개발하고 배포하기 위하여 노력하고 있다. 이러한 노력은 SWAD(W3C Semantic Web Advanced Development)를 통하여 진행중이며, 대표적으로 DAML 프로젝트 등을 들 수 있다.

## 2. XML 개요 및 표준화

웹에서 문서의 포맷을 정해주고 URI를 사용해 하이퍼텍스트 링크를 표시해 줄 수 있는 HTML 언어에서 태그는 출력형식에 상관하고 텍스트의 데이터 구조나 의미와는 무관하다. 이에 반해 1986년 Goldfarb에 의해 개발된 SGML은 문서의 출력형식이 아니라 문서의 구조를 기술하는 언어라 할 수 있다. 메타언어로서의 SGML은 문서와 데이터를 기술하기 위한 문법규칙의 집합으로서, HTML과 같은 언어의 태그들의 의미와 사용규칙을 정의해 주는 마크업 언어이다.

W3C에서 1998년도에 새로운 웹 표준언어로 채택된 XML(Extensible Markup Language)은 새로운 태그를 정의하여 사용할 수 있는 언어이다.

XML은 SGML의 축소판이라고 볼 수 있지만 웹 상에서 문서의 이동이 더욱 용이하고 웹에 적합한 형태로 만들어졌다. 즉, XML은 웹에서 데이터를 사용하기 위한 마크업 언어이다. XML은 데이터를 기술하기 위해 고안되었고 매우 다양한 응용 프로그램으로부터 구조화된 데이터를 전달할 수 있다. XML 태그들은 XML에서 미리 정의되어 있지 않기 때문에 필요한 태그는 사용자가 직접 정의해야만 한다. 이처럼 확장성과 유연성이라는 장점에도 불구하고 XML을 시맨틱 웹 표준 언어로 사용하는 데에는 한계가 있다. XML에서 데이터의 의미를 정의해 주는 태그는 사용자가 이

해할 수 있는 차원이기 기계가 자동적으로 그 의미를 이해하는 차원은 아니다. 또한 태그들 사이의 의미적 연관성을 표현할 수 있는 방법은 매우 제한적이다.

## 3. RDF(Resource Description Framework) 개요

최근 W3C에서 제안된 RDF(Resource Description Framework)와 같이 시맨틱 웹을 구현하기 위한 기술에 있어서 핵심적 요소 두 가지를 들라면 메타데이터와 지식표현이라 하겠다. 이들은 서로 중복되는 부분이 많은 개념이지만 서로 구분하여 설명 하겠다.

일반적인 의미로 메타데이터는 “데이터에 관한 데이터”로 정의할 수 있는데, 메타데이터는 그 자체가 또한 데이터이다. 예를 들어 데이터베이스에 저장되어 있는 데이터의 형식에 대한 내용을 기술하는 데이터가 메타데이터이다. 웹에서의 메타데이터의 목적은 정보검색의 처리과정을 줄여주고, 사용자가 원하지 않는 데이터를 미리 걸러주며, 관련성이 많은 정보의 발견 가능성을 높여줌으로써 정보검색을 향상시키기 위한 것이다. 메타데이터의 예로서 도서관의 목록카드를 들 수 있는데, 개별 도서가 데이터라면 목록카드는 도서에 관한 주제별, 저자별, 연도별 분류를 가능하게 해주는 메타데이터로서 검색을 용이하게 해준다. 지도의 범례도 일종의 메타데이터라 할 수 있는데 지도에 표시된 여러 상징들에 대한 설명을 해줌으로써 지도 검색을 용이하게 해준다. HTML 페이지 내의 <META> 태그도 문서를 설명해주는 메타데이터이다. 메타데이터에는 웹 페이지의 작성자와 버전 및 내용에 대한 목록정보를 표시할 수 있으며, 웹 페이지 상호간의 관련성을 나타낼 수 있고, 또 배포권 및 프라이버시 코드 등과

같은 사회적 정보도 표시할 수 있다.

메타데이터는 관계성에 관한 것이다. 관계성이란 정보 리소스에 대하여 기술(description)하는 것을 의미한다. 그런데 메타데이터로 데이터를 기술할 때 몇 가지 정해진 방식에 제한 받지는 않는다. 예를 들어 도서관의 책을 분류하는데 쓰이는 도서목록카드의 주제나 저자뿐만 아니라 책의 물리적 크기나 발행연도에 대한 정보도 중요할 때가 있다. 이렇게 분류나 검색을 위한 메타데이터는 사용자의 관점과 사용목적에 따라 달라질 수 있다. 따라서 메타데이터는 정보검색을 위한 사용자 프로파일로 쓰일 수도 있다.

시맨틱 웹에서 메타데이터 사용의 다양성은 중앙집중 방식을 지양하고 웹의 분산성을 증진하려는 웹의 기본적 정신과 관련이 있다. 예를 들어, RDF에서는 정보 리소스들간의 관계성이 웹 문서의 저자에 따라 다르게 정의될 수 있다. 즉, 웹에서 어휘의 의미를 표준화하여 어휘 사용을 한 방향으로 경직시키는 것이 아니라, 누구든지 자유롭게 어휘의 정의에 대한 메타데이터를 웹에 공개할 수 있는 것이다. 따라서 언제든지 공개된 메타데이터를 원하는 사람은 전체 혹은 부분적으로 사용할 수 있다. 마치 자연어에서 어휘의 의미가 진화하는 것과 같다.

### 3.1 RDF Model and Syntax

#### (1) 개요

RDF(Resource Description Framework)는 메타데이터의 기술과 교환을 위한 프레임워크로 W3C에서 개발한 시맨틱 웹의 핵심 기술이다. RDF는 다양한 메타데이터 사이의 연결을 위해 의미(semantics), 구조(structure) 및 구문(syntax)에 대한 공통적인 규칙을 지원한다. 이러한 메카니즘을 통해 기계가 이해할 수

있는 정보자원을 교환하는 메타데이터 사이의 상호운용성을 지원한다. RDF는 다양한 메타데이터의 의미를 정의하지 않고 메타데이터에 필요한 데이터 요소를 정의해 사용할 수 있게 만든다.

RDF는 지식 표현(Knowledge Representation) 분야의 영향을 받아 웹 자원을 기계가 이해할 수 있고, 자동적으로 기술할 수 있도록 설계되었다. RDF의 설계 목적은 기계가 접근가능한 웹 자원에 의미(semantics)를 부여할 수 있도록 함으로써 지능형 웹 에이전트가 처리할 수 있는 지식기반 정보를 구축하는데 있다. 다음은 RDF 관련 표준들의 웹 사이트들이다.

- Resource Description Framework(RDF) Model and Syntax Specification
  - 1999년 2월 22일 발표(W3C Recommendation)
  - <http://www.w3c.org/TR/1999/REC-rdf-syntax-19990222>
- RDF Model Theory
  - 2002년 4월 29일 발표(Working Draft)
  - <http://www.w3c.org/TR/rdf-mt/>
- RDF/XML Syntax Specification(Revised)
  - 2002년 3월 25일
  - <http://www.w3c.org/TR/rdf-syntax-grammar/>
- RDF Test Cases
  - 2002년 4월 29일
  - <http://www.w3c.org/TR/rdf-testcases/>
- RDF Primer
  - 2002년 4월 26일
  - <http://www.w3c.org/TR/rdf-primer/>

#### (2) RDF 데이터 모델

RDF 데이터 모델은 정보자원의 속성과 정보자원들 사이의 관계를 표현하며, 메타데이터의 교환과 통합을 지원하기 위한 모델이다. RDF 데이터 모델로 기술되는 내용들은 구문적으로 다를 수 있지만 RDF 파서에 의해 동일한 의미로 해석된다.

RDF 데이터 모델은 자원(resource), 속성타입(property type) 및 속성 값(value)으로 구성된다. 자원은 RDF 데이터 모델에서 기술되는 URI를 갖는 모든 객체를 말한다. 하나의 자원은 여러 속성 유형과 속성 값을 가질 수 있다. 속성 타입은 자원의 속성 명을 의미하며, 속성 값은 속성 타입에 해당하는 값으로 문자열이나 자연어로 기술할 수 있다. RDF 데이터 모델은 노드와 아크 그래프를 이용해 메타데이터를 정의하기 위한 추상적이고 개념적인 틀을 제공한다.

그러나 노드와 아크 그래프는 메타데이터를 시각화하기 위해 유용한 반면 웹에서 응용 애플리케이션 사이의 메타데이터를 공유하기 위한 명확한 방법을 제공하지 못한다. 따라서 기계가 처리할 수 있도록 표현하기 위해 RDF는 XML을 사용하여 구문을 표현할 수 있다. RDF는 객체 지향적 접근의 지식 표현방식을 취하며, 세 개의 요소로 이루어진 구조를 기본으로 한다. 즉, 개체(object)-속성(attribute)-값(value)을 갖고 있으며, A(O,V)로 표현된다. 그러나 일반적인 객체 지향적 혹은 프레임 방식과는 달리, RDF는 객체 중심이 아니라 속성 중심(property-centered)의 구조를 갖고 있다.

RDF의 간단한 표현방식에 대해 예를 들면 다음과 같다.

```
author( 'http://www.ontology.or.kr/id1971',
        'http://www.dku.edu/hgkim' )
title( 'http://www.ontology.or.kr/id1971', '시맨틱웹 이해하기' )
```

```
name( 'http://www.dku.edu/hgkim', '김홍기' )
```

[http://www.ontology.or.kr/id1971]과 같은 URI로 표현되는 리소스(resource) 혹은 개체는 이것이 갖고 있는 author, title, name과 같은 여러 속성들을 통해 그 의미가 표현될 수 있다.

### (3) RDF 구문

XML을 이용해 RDF를 표현하기 위해서는 연속 구문(serialization syntax)과 축약 구문(abbreviated syntax)을 통해 처리할 수 있다. 연속 구문은 특성값을 엘리먼트로 기술하는 것이고, 축약 구문은 특성값을 엘리먼트의 속성값으로 기술하는 방법을 말한다. RDF 데이터 모델과 파서에 의한 해석은 두 구문에서 동일하다. 위에 제시된 RDF 데이터 모델을 XML을 이용하여 표현하면 다음과 같다.

```
<rdf:Description about=" http://www.ontology.or.kr/id1971">
  <author>http://www.dku.edu/hgkim</author>
</rdf:Description>
```

### (4) RDF와 XML 비교

XML은 콘텐츠와 표현을 분리함으로써 HTML의 한계를 해결했고, 웹 검색엔진 사용시 검색의 정확성 문제는 RDF를 사용하여 효율성을 높일 수 있다. RDF가 데이터 모델을 표현하기 위해 XML을 이용하지만, RDF 데이터 모델과 XML 데이터 모델에는 상이한 점이 존재한다.

표 4-1과 같이 XML 데이터 모델은 객체의 배열된 순서가 중요한 트리 지향 모델의 텍스트 마크업

(markup) 모델이고, RDF 데이터 모델은 객체 사이의 관계를 정의하는 데이터 모델이다. XML은 트리 형태의 문서로 고정되게 설계되어 노드가 문서에 포함되어 인덱스되기 때문에 메타데이터를 표현하기 위한 유연성이 부족하다. 반면 RDF에서 노드는 인덱스되지 않고 URI를 갖는 자원이기 때문에 메타데이터를 표현하기 위한 유연성을 지원할 수 있다.

XML 스키마는 생성가능한 XML 문서의 집합을 제약하는 문법적 해석이 주된 기능인 반면, RDF는 의미 해석에 이용된다. 다시 말해 XML 스키마는 XML 문서의 구조를 모델링하여 사용하는 데이터 모델링 언어이고, RDF는 지식을 모델링하기 위한 메타데이터 모델링 언어이다. 따라서 메타데이터가 데이터로 인코딩될 때 XML 구문은 매우 유용하지만, 순수한 XML로 메타데이터를 모델링할 때는 의미적 상호운용성에 한계가 있다. 즉, 웹 정보자원을 표현하는데 있어 XML이 구문적 계층에서 데이터의 구조를 정의하는데 이용되고, RDF가 의미적 계층에서 지식표현을 위한 기능을 하는 것이다. 따라서 두 언어는 배타적이기보다 상호 보완적인 특성을 갖고 있다.

표. RDF와 XML

구분	XML	RDF
데이터 모델	순서화된 트리지향 모델	객체의 관계 지향 모델
스키마	문법적 해석	의미 해석

## 3.2 RDF Schema

### (1) 개요

RDF 데이터 모델 명세서는 특성들을 선언하고, 특성과 다른 자원 사이의 관계를 정의하기 위한 어떤 메카니즘도 제공하지 않는다. 이런 문제는 RDF 스키마 명세서에서 구체화한 RDF 스키마의 역할이다. 다시

말해 RDF/XML이 XML 구문을 이용하여 메타데이터를 표현하고, RDF 스키마는 메타데이터의 무결성을 보장하기 위해 구문에 관계없이 메타데이터의 자원과 특성에 대한 유효성을 검증한다. RDF 스키마는 자원의 특성을 기술하기 위해 사용될 수 있는 자원의 집합이고, 웹 자원을 기술하기 위해 이용된 메타데이터의 구조를 해당 애플리케이션에서 사용되는 특정한 어휘로 정의한다.

- RDF Vocabulary Description Language 1.0  
- 2002년 4월 30일  
- <http://www.w3c.org/TR/rdf-schema/>
- Resource Description Framework(RDF) Schema Specification 1.0  
- 2000년 3월 27일 발표  
- <http://www.w3.org/TR/2000/CR-rdf-schema> - 20000327/

### (2) RDF 스키마와 XML 스키마

스키마란 사람이나 기계가 읽을 수 있는 형태로 정보를 전달할 수 있도록 특별한 구문의 의미와 제약조건을 구체화한 것이고, RDF 스키마는 자원과 특성의 집합을 네임스페이스로 간단히 표현한 것이다. RDF 스키마와 XML 스키마는 기능적으로 유사한 특성을 갖고 있지만 상이한 특성도 존재한다. 다음은 두 스키마 언어의 유사한 특성을 열거한 것이다.

- RDF 스키마의 Range 제약조건과 XML 스키마 type 제약조건
- RDF 스키마의 domain 제약조건과 XML 스키마의 type과 엘리먼트의 콘텐츠 모델 정의
- 특정 도메인을 위한 목록 리스트나 통제화된 어휘 값의 정의
- RDF 스키마의 comment와 XML 스키마의

annotation

비록 두 스키마 언어가 기능적으로 공통점이 있지만, 구조적으로 갖고 있는 특성은 매우 다르다. RDF 스키마가 타입 계층, 클래스와 프로퍼티 관계 등 사람이 읽을 수 있는 기술 등을 정의하는 기능을 통해 풍부한 의미적 정보를 제공하고, 구조적 제약, 데이터 타입을 위한 제약조건에 대한 부분을 정의하기 위해서는 한계가 있다. 반면 XML 스키마는 명확하게 구조적, 데이터 타입 제약사항을 지원하지만 메타데이터의 도메인 사이의 의미를 표현하기 위한 동적이고 유연한 매핑을 하기에는 한계가 있다. 따라서 제약조건 사이의 비일관성이나 중복성을 제거하기 위한 각각의 스키마 언어의 책임 영역에 대한 경계를 명확히 정의해야 할 필요가 있다.

#### 4. DAML+OIL/OWL

##### (1) 개요

DAML+OIL은 미국의 DARPA<sup>1)</sup>와 유럽의 IST (Information Society Technologies)가 공동으로 개발한 언어로 DAML의 문법적 특성을 상속받고 있다. DARPA 프로젝트가 XML에서 의미적 상호운용성을 지원하는 반면 DAML+OIL은 OIL의 같은 객체들을 공유하고 있다. DAML+OIL은 RDF(S)를 기반으로 표현될 수 있으며, DAML-ONT라 불리는 초기 스펙을 대체하고 있다.

OWL 웹 온톨로지 언어는 웹에 표현된 콘텐츠를 이해하기 위해 요구되는 애플리케이션이 사용할 수 있는

언어를 제공하기 위해 W3C 웹 온톨로지 워킹그룹<sup>2)</sup>에 의해 설계된 온톨로지의 공유와 출판을 위한 시맨틱 마크업 언어이다. OWL은 DAML+OIL 웹 온톨로지 언어를 기반으로 하고 있다. OWL의 지식 베이스는 RDF/XML 구문 명세에서 정의된 RDF 트리플의 집합으로 구성되었으며, OWL에 정의한 어휘를 사용하여 트리플을 위한 특정한 의미를 기술한다. 다음의 문서들은 OWL과 관련된 구문과 언어의 특징을 설명하고 있다.

- Feature synopsis for OWL Lite and OWL
  - 2002년 7월 29일 발표
  - <http://www.w3.org/TR/2002/WD-owl-features-20020729/>
- OWL Web Ontology Language Reference
  - 2002년 7월 29일 발표
  - [http://www.w3.org/TR/2002/WD-owl-features-20020729](http://www.w3.org/TR/2002/WD-owl-features-20020729/)
- OWL Web Ontology Language Abstract Syntax
  - 2002년 7월 29일 발표

##### (2) DAML + OIL과의 차이점

OWL은 DAML+OIL의 언어적 특성을 상속받았지만 다음과 같은 특성은 서로 상이한 점을 갖고 있다.

- 네임스페이스 수정 - <http://www.w3.org/2002/07/owl>
- RDF 코어 워킹그룹에 의해 오류가 수정된 공식적인 RDF 구문을 사용
  - `rdfs:subClassOf`와 `rdf:subPropertyOf`를 이용하여 클래스와 프로퍼티의 순환적 삽입가능

1) [www.daml.org](http://www.daml.org)

2) W3C Web Ontology Working Group, <http://www.w3c.org/2001/sw/WebOnt/>


- 복수의 rdfs:domain과 rdfs:range 특성을 지원
- 컬렉션 타입 지원 : rdf:parseType="Collection"
- 새로운 프로퍼티타입(owl:SymmetricProperty) 추가 : 대칭관계 특성표현
- DAML+OIL의 몇 가지 프로퍼티는 이름이 변경되었음

DAML+OIL	OWL
daml:hasClass	owl:someValueFrom
daml:toClass	owl:allValuesFrom
daml:UnambiguousProperty	owl:InverseFunctionalProperty
daml:UniqueProperty	owl:FunctionalProperty

## 5. 결론

시맨틱 웹은 무질서하게 연결되어 있는 현재의 웹에 관계성을 더해줌으로써, 새로운 웹의 활용을 낳을 수

있는 차세대 기술로 평가되고 있다. 물론 웹에서의 정보들 사이의 관계성은 개별 데이터베이스와는 규모 면에서 비교가 안될 정도로 복잡하고 다양할 수 있다. 시맨틱 웹은 문서의 각 부분을 컴퓨터가 이해할 수 있는 형식으로 기술(description)할 수만 있다면 복잡하게 얽혀져 있는 정보 리소스들 사이의 의미적 연관성으로 인해 인터넷을 통하여 다양한 정보를 보다 효과적으로 활용할 수 있게 해준다.

시맨틱 웹에 대한 연구 및 표준화를 주도하고 있는 W3C의 시맨틱 웹 Activity는 활동기한이 이미 마쳐진 상태이나, 이를 계승하는 다음 Activity의 활동은 기존의 시맨틱 웹에 관한 성과를 보다 가시화할 수 있는 것이 될 것으로 기대된다. 또한, 웹 서비스 등을 포함한 발전하는 웹 기술에서도 시맨틱 웹을 이용하는 시나리오를 포함시키고 있는 만큼, 이러한 기술이 현실화됨에 따라 시맨틱 웹기술이 현재의 웹구조에 실현될 가능성은 점점 높아지고 있으며, 이에 부응하여 국외뿐만 아니라 국내 연구들도 함께 활성화될 것으로 예상된다. 

### 통계로 본 정보통신 발달사

우정총국이 설립된 지 119년, 최초의 전화통화가 이뤄진 지 107년이 지난 2003년 현재 우체국은 전국에 3709개, 전화가입자는 2349만명으로 늘어났다. 그러나 인터넷의 발달로 우편물의 증가속도가 느려지기 시작했고 전화도 이동전화로 대세가 넘어가고 있다. 초고속인터넷 이용자는 2627만명으로 인구의 절반을 훌쩍 넘었고 가입가구도 1040만가구에 달한다. 전자상거래도 177조원 규모로 올라섰다. 지난 10년새 일어난 일이다. 이동전화 가입자는 3234만명에 달한다. 이는 산업으로 연결됐다. IT업체는 2만2127개로 늘어났고 정보통신 생산액이 189조원, 수출은 463억달러로 늘어나 연간 167억달러의 흑자를 안기는 '한국경제의 1등공신'으로 자리매김했다.