

## Semantic Web 소개

김 병 학 | bhkim@kins.re.kr

한국인 지력연구소 | 출신 선임연구원

	I. 서	론
II. Semantic Web	소개	
III. Semantic Web	구조	
IV. 결	론	

SW이란 컴퓨터가 스스로 웹상의 자료를 처리할 수 있도록 하기 위한 차세대 지능형 웹을 말한다. 현재 웹을 무시하고 전혀 새로운 웹을 만들고자 하는 시도가 아니고 현재까지 검증된 기술에 추가적인 기능을 부가하여 추론 등의 기능을 구현하고자 하는 시도이다.

### I. 서 론

인터넷 웹이 생활의 일부로 급속히 파고들어 사람들의 생활 패턴을 변화시키고 있으며 앞으로는 더욱더 삶의 일부로 자리잡을 것이다. 웹이 성공적으로 발전할 수 있었던 요인은 초기부터 자료 구성 및 이용이 단순했기 때문이다.

그러나 최근들어 멀티미디어 정보 증가, 이용환경 다변화, 정보 및 웹 사용자의 기하급수적인 증가 등으로 인하여 초기 웹이 발전하는데 한계에 이른 느낌이 든다. 더구나 자료는 대부분 사람들 위주로 작성되어 있기 때문에 컴퓨터가 자동으로 부가정보를 만들어 낸다는 것은 매우 어려운 일이고, 검색연결을 통하여 정보를 검색하면 엉뚱한 정보가 나오는 경우도 있다.

이런 웹 환경 변화에 대처하기 위한 새로운 웹 기술로 Semantic Web(이하 SW)에 대한 관심이 고조되고 있다. SW은 웹상에 존재하는 자료에 의미를 부가하고 사람이 관여하지 않아도 컴퓨터가 자동으로 처리를 할 수 있는 차세대 지능적인 웹을 말한다. SW은 W3C의 Director인 Tim Berners-Lee가 수년 전부터 주장해 온 비전으로 W3C에서 그 중요성을 인정하여 2001년

2월에 공식적으로 Semantic Web Activity를 착수하도록 했다. W3C의 Metadata Activity도 이어받은 새로운 Activity의 목적은 웹상에 존재하는 각종 자원들을 최대한 활용할 수 있는 환경을 만드는 것이다.

외국에서는 이미 SW에 대한 연구결과로 몇몇 시스템들이 출현되고 있으나 아직은 초기단계에 있는 경우가 대부분이다. 따라서 본고에서는 SW에 대한 기술적인 내용보다는 Tim Berners-Lee가 소개한 SW의 계층적인 구조에 대해 간단하게 소개하고자 한다.

### II. Semantic Web 소개

SW이란 컴퓨터가 스스로 웹상의 자료를 처리할 수 있도록 하기 위한 차세대 지능형 웹을 말한다. 현재 웹을 무시하고 전혀 새로운 웹을 만들고자 하는 시도가 아니고 현재까지 검증된 기술에 추가적인 기능을 부가하여 추론 등의 기능을 구현하고자 하는 시도이다.

SW이 제공할 것으로 예상되는 기능을 나열해 보자. 당연히 기존 웹처럼 단어를 식별해서 관련된 사이트나 문서를 찾아주는 것은 기본이고, 새롭게 구성될 문서에 사물간의 관계를 명확히 기술하여 정확하고 의미 있는

정보를 제공한다. 또 컴퓨터가 스스로 자료를 처리하고, 통합하는데 필요한 정보들을 포함함으로써 사람들은 더욱 편리하게 컴퓨터의 도움을 받으며 일을 할 수 있게 될 것이다. 사용자의 취향을 분석한 후 가장 적합한 정보를 제공하기 위해 관련된 웹 페이지를 스스로 검색할 수도 있으며 다른 웹을 찾아다니면서 필요한 정보들을 자발적으로 제시해주는 기능도 갖게 될 것이다.

또한 현재의 웹보다 더욱 지능적인 웹 서비스가 가능하고 의미적으로 더욱 강력해진 검색엔진을 만들 수 있고 몇 개의 독립된 사이트로부터 발견한 정보들을 통합하여 어떤 결론을 도출할 수도 있을 것이다.

위와 같은 다양한 기능이 가능하게 되면 현재 우리가 일하는 방식을 근본적으로 변경해야 할 것이다. 예를 들어 소프트웨어를 만든다는 것은 웹상에서 적당한 키워드들을 찾는 일과 이러한 키워드들을 연결하기 위한 분석을 기술하는 일이 될 것이다. 새로운 소프트웨어는 또다시 웹상에 등록되어 다른 프로그램에게 재사용 될 수 있을 것이다.

이와 같은 막강한 정보처리가능이 구현되려면 컴퓨터가 웹상에 구축되어 있는 문서들을 이해하여 스스로 정보를 수집하고 추론할 수 있어야 한다. 이같은 기술은 XML과 RDF가 개발되면서 현실화되고 있다.

SW의 구현에 필요한 기반 기술로는 하이퍼텍스트, Unicode, URI, XML/Schema, RDF/Schema 등이 있는데 이러한 기술들을 계층적으로 나누어 살펴보자.

### III. Semantic Web 구조

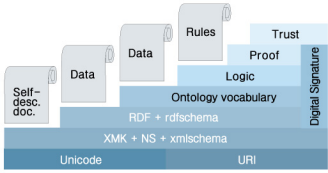
SW의 구조는 아래 그림과 같다. 다음은 각 계층에 대해 설명이다.

#### 1. Unicode 계층

문자 인코딩을 위한 국제 표준인 유니코드는 65536 개의 문자를 포함할 수 있는 2바이트 문자를 지원하므로 이를 지원하는 SW에서는 국제적인 데이터 교환을 원활히 할 수 있을 것이다.

#### 2. URI(Universal Resource Identifier) 계층

URI는 현재도 웹의 기초를 이루는 개념으로서 인터넷에 존재하는 각종 객체의 명칭, 위치 등에 대한 표현 방법을 일컫는 말이다. URI는 URL(Uniform Resource Locator)과 URN(Universal Resource Name) 등을



(그림 1) Semantic Web의 계층적 구조

포함하는 개념이다. URL은 인터넷에 존재하는 수많은 정보자원의 위치를 정확하고 편리하게 표현하기 위한 방법으로 일반적인 주소, 자료에 접근할 프로토콜, 접속할 호스트 이름, 자료 파일 경로 및 파일 이름 등으로 구성된다. 또 URL은 전세계 정보를 찾는데 있어서 보편적인 방식을 제시해 주고 웹상에 다른 문서를 연결할 경우에 있어서도 중요한 역할을 한다.

URI의 특징을 살펴보자. 어떤 것이라도 URI를 가지면 웹에 올릴 수 있고 어떤 것이든지 URI를 가질 수 있다. URI는 분산적인 특성을 갖는다. 즉 어떤 사람이나 조직은 누가 URI를 만들고 어떻게 사용되는지에 통제 권한을 가질 수 있다. 우리가 링크를 위해 어떤 URI를 만들든지 할 때 어떤 허가나 권한도 필요없이 자유롭게 언제든지 만들 수 있다. 심지어는 소유하고 있지 않은 것, 실존하지도 않는 추상적인 것을 위한 URI도 만들 수 있다. 이렇게 URI를 자유롭게 만들 수 있는 특성은 초기 웹이 성장하는데 중요한 요인 중의 하나였다. SW에서도 이러한 URI의 특징을 이어받아 사용한다.

### 3. XML(Extensible Markup Language)+NS+xml schema 계층

XML은 HTML 후속으로 웹상에서 자료를 표현하고 교환하는데 표준으로 자리잡아가고 있다. SW에서도

XML을 문서작성의 기반 기술로 사용한다. SW이 전혀 새로운 웹이 아니라 현재 웹을 기반으로 의미를 처리할 수 있는 차세대 지능형 웹을 만들고자 하기 때문에 XML을 기반 기술로 선택한 것은 당연하다.

Namespace은 XML 문서에서 엘리먼트나 어트리뷰트의 이름으로 쓰일 수 있는 이름들의 집합이다.

Namespace은 두가지 이상의 소스로부터 문서들을 조합해 낼 수 있어 여러 소스들이 섞인다고 해도 어떤 용어가 어떤 소스로부터 왔는지를 식별해낼 수 있게 해준다. 즉 XML을 사용할 때 상이한 의미를 가지는 동

일한 용어간의 유효성 확성을 해결하는데 사용된다.

XMLSchema는 문서가 마크업되는 방식에 대한 정의를 뜻한다. DTD도 하나의 스키마라고 할

수 있다. XMLSchema를 만들면 서로 다른 시스템에서 여러 개발자가 XML 문서를 생성할 때나 개발할 때, 이것을 참고해 문서가 올바른지 체크할 수 있다.

### 4. RDF(Resource Description Framework)/rdf schema 계층

XML은 누구나 자신만의 고유한 태그를 만들어 사용할 수 있기 때문에 사람에게서는 어느 정도 의미를 제공할 수 있다. 그러나 XML은 임의의 구조를 갖는 문서를 만들 수 있기 때문에 컴퓨터가 그 구조를 통하여 의미를 파악하기는 쉽지 않다. 예를 들어 (Author)라는

XML(Extensible Markup Language)  
+NS+xml schema 계층

태그를 사용한 문서에서 사람들은 (Author)의 의미를 쉽게 유추할 수 있지만 컴퓨터 프로그램은 Author가 무슨 의미인지 파악하기 쉽지 않고 의미를 파악한다 해도 어떻게 Person과 연결되는지 등에 대한 정보를 정의하기 쉽지 않다. 이처럼 동일한 내용에 대해서 태그를 만드는 사람마다 각각각색으로 만들 수 있고 문서 구조도 다양하게 만들 수 있기 때문에 이런 XML 문서를 이해하는데는 많은 어려움이 있다. 이러한 한계 때문에 SW에서는 의미표현을 위한 수단으로는 RDF를 사용한다.

RDF는 정보자원이나 자원의 타입을 기술하는 언어이다. 즉 데이터에 대한 데이터인 메타데이터를 기술하기 위한 사양으로 컴퓨터가 이해할 수 있는 형태의 정보를 응용 프로그램 사이에서 교환하기 위한 수단을 제공한다.

지금까지 비교적 많은 연구가 진행되어 class, property, subproperty, subclass, domain, range, comment, label 등을 지원한다. RDF 문장은 대문분의 단어가 URI라는 사실만 제외하면 단순한 영어 문장과 매우 유사하다. subject, predicate, object로 구성되는데 예를 들어 다음과 같은 문장이 있다고 하자.

Ora Lassila is the creator of the resource  
<http://www.w3.org/Home/Lassila>

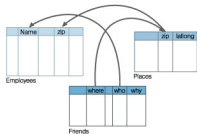
이 문장을 RDF로 나타내면 다음과 같다.

```
<?xml version="1.0" ?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/"
  <rdf:Description about="http://www.w3.org/Home/Lassila">
    <s:CreatorXtra Lassila</s:Creator>
  </rdf:Description>
</rdf:RDF>
```

향후 더 많은 RDF 문서가 웹상에 구축될 때 SW의 성공이 앞당겨질 것으로 예상된다. 다행스러운 것은 현

재 데이터베이스를 통해 구축된 많은 정보들은 RDF로 변환하는데 적합한 구조를 가지고 있다는 것이다. 데이터베이스 정보를 RDF로 변환하여 지능형 프로그램들이 그 정보를 이용할 수 있게 되면 현재 데이터베이스 응용 프로그램처럼 다양하게 정보를 조회할 수 있게 될 것이다.

다음은 각각의 데이터베이스가 독립적으로 구축되어 웹상에 공개된 후 다른 사람에 의해 의미적 링크 정보가 추가되면 추가된 테이블까지 검색이 확장되는 경우를 보여준다. 즉 어떤 사람이 Friends의 where와 Places의 zip이 동일한 것을 알고 있을 때, 또 다른 사람이 화살표 지시처럼 Places의 zip과 Employees의 zip이 동일하다는 정보를 정의하면 Employees를 통하여 '홍길동'이라는 이름과 '123456'이라는 우편번호를 가진 중얼린 왓는 검색은 Friends까지 포함하도록 확장할 수 있다. 이때 사용된 것은 RDF의 equivalent 특성이다.



(그림 2) Semantic Link를 통한 검색확장

### 5. Ontology 계층

원래 Ontology란 특정 도메인에 대한 공유되는 일반적인 이해, 개념의 표현, 개념과 관계에 대한 공식적인 기술 등의 의미로 사용된다. SW에서는 정보자원에 대한 의미와 이들 간의 연관성을 기술하는 것으로 사용되는데 기본적으로 RDF를 확장하여 transitivity,

unique, unambiguity, cardinality 등의 특성을 지원하기 위해서 OIL, SHOE, DAML, WebOnt 등과 같은 연구가 활발하게 진행되고 있다.

### 6. Logic 계층

SW의 Logic 계층은 아직까지 연구개발 초기에 있지만 그동안 기존 Logic 등의 연구 결과를 SW에 적용하는 노력이 필요할 것이다. SW의 Logic 계층에서는 기존 정보를 바탕으로 새로운 결론을 도출하는 추론 기능을 지원해야 한다. 예를 들어 어떤 어떤 자동차 회사에서 사원이 한달에 자동차를 10대 이상 판매하면 A클래스로 지정한다고 하자. 그러면 SW 프로그램은 다음과 같은 추론을 할 수 있어야 한다.

Proof 계층은 미래에 매우 중요한 부분을 차지할 것이다. 그러나 아직 이 계층에 대해서는 별로 연구된 바가 없다. Logic을 통하여 도출된 결론을 Proof로 증명할 수 있는 방법 등을 연구해야 한다. SW에서 이러한 Proof를 얻기 위해서는 수많은 링크를 방문해야 할 경우도 있을 것이다. 따라서 heuristic 검색 방법 등에 대한 연구도 필요한 것이다.

홍길동은 차를 15대 팔았다. 그러므로 홍길동은 A 클래스의 구성원이다.

### 7. Proof 계층

Proof 계층은 미래에 매우 중요한 부분을 차지할 것

이다. 그러나 아직 이 계층에 대해서는 별로 연구된 바가 없다. Logic을 통하여 도출된 결론을 Proof로 증명할 수 있는 방법 등을 연구해야 한다. SW에서 이러한 Proof를 얻기 위해서는 수많은 링크를 방문해야 할 경우도 있을 것이다. 따라서 heuristic 검색 방법 등에 대한 연구도 필요한 것이다.

### 8. Trust/Digital Signature 계층

SW이 성공하기 위해서는 문서에 신뢰를 줄 수 있는 방법이 있어야 한다. 이러한 목적으로 Digital Signature를 SW에 적용하는 연구가 진행될 것이다. 어떤 사람이 어떤 문서나 문장을 작성했다는 증거로 Digital Signature를 제공하면 이러한 문서를 읽었을 때 어느 정도 신뢰할 수 있는지를 결정할 수 있을 것이다.

## IV. 결 론

인터넷 웹이 우리의 생활방식을 많이 변화시키고 있다. 미래 생활은 웹의 의존도가 한층 강화될 것이다. 지금까지 주로 PC를 위주로 한 웹 이용 환경도 PDA, 랩탑, 휴대전화, 가전제품 등으로 다변화 될 것이다.

우리 생활이 웹에 의존하면 할수록 사람들은 좀더 정확하고 편리한 웹뿐만 아니라 주어진 정보를 이용해서 새로운 결론을 추론할 수 있는 기능을 요구할 것이다.

이러한 요구에 대한 대안으로 SW이 사용될 수 있을 것으로 기대된다.

그러나 국내에서는 그동안 XML, 메타데이터 등에 대한 연구는 활발히 진행해 왔으나 상대적으로 SW에 대한 연구는 부족한 것으로 판단된다.

향후 10년을 내다보고 SW에 대한 연구가 필요한 시점이다.



#### 참고자료

- (1) The Semantic Web: An Introduction - <http://infometa.net/2001/swintro/>
- (2) Semantic Web Activity - <http://www.w3.org/2001/sw/>
- (3) The Semantic Web Community Portal - <http://www.semanticweb.org/>
- (4) Dave Beckett's Resource Description Framework(RDF) Resource Guide - <http://www.kit.net.ac.uk/discovery/html/resources/>