

Semantic Web 환경에서의 자원발견

박재홍*, 임유정*, 김도완*, 박찬규**, 조현규**
*배재대학교 정보통신공학과
**ETRI 전자상거래기술연구팀
e-mail : middle75@lycos.co.kr

Resource Discovery on the Semantic Web

Jae-Hong Park*, You-Jeong Lim*, Do-Wan Kim*,
Chan-Kyu Park**, Hyun-Kyu Cho **

*Dept. of Information and Communications, Pai-Chai University
**Team of Electronic Commerce Technology Research, ETRI

요 약

Semantic Web 환경의 등장 배경과 Semantic Web 을 구성하는 구성요소 및 기술분석에 대해 살펴볼 것이며 이를 바탕으로 제주도의 숙박, 항공, 렌트카 정보 온톨로지 구축과 Jena 에서 지원하는 RDQL 을 이용하여 사용자가 원하는 서비스 자원 발견을 수행할 수 있는 프로토타입 테스트 베드 구축에 대해 살펴볼 것이다.

1. 서론

웹의 창시자 팀버너스리가 제안한 Semantic Web[1]은 컴퓨터(Intelligent Software Agents)가 웹상의 정보를 이해하고, 정보를 창출할 수 있는 웹 환경으로, 정보의 탐색과 의사결정이 인간이 아닌 컴퓨터가 할 수 있도록 만들어진 Web 환경이다. 즉 Semantic Web 은 기계를 이용한 정보의 탐색과 의사결정을 지원하는 진정한 의미의 인터넷 환경이라 할 수 있다.

Semantic Web 환경에서는 사용자가 필요로 하는 "What(무엇)"에 대한 내용을 컴퓨터에 알려주고, 컴퓨터는 입력된 조건에 맞는 작업을 수행한 후 사용자에게 제공하게 되는 것이다. 이때 사용자는 "How(어떻게)" 작업해야 하는지 알 필요가 없으며, "How" 문제는 컴퓨터가 담당하게 된다.

차세대 웹인 Semantic Web 의 등장으로 사용자가 원하는 자원의 발견, 수행, 구성과 상호운영 등과 같은 어려운 작업들을 소프트웨어 에이전트들이 담당할 수 있게 되었다. 그리고 Semantic Web 은 모든 분야에 적용될 것이라 생각된다. 그 중에서 크게 지식베이스, 전자상거래, 검색엔진에 많은 영향을 미칠 것이다.

2 장에서는 Semantic Web 환경의 구성요소 및 특징에 대해 살펴볼 것이며 3 장에서는 지능형 웹 서비스 프로토타입 테스트 베드 구축과 검색에 의한 자원 발견에 관하여 마지막으로 4 장에서는 결론 및 향후연

구 방향에 대해 살펴볼 것이다.

2. Semantic Web 환경의 구성요소 및 분석

[그림 1]은 lower level 의 문법과 semantics 를 사용하는 higher-level 언어를 가진 Semantic Web 의 계층구조를 보여준다. Semantic Web 의 계층구조에서 나오는 언어들에 대한 간략한 특징과 개요에 대해 살펴볼 것이다. 그리고 언어에 관련된 문법은 언급하지 않을 것이다.

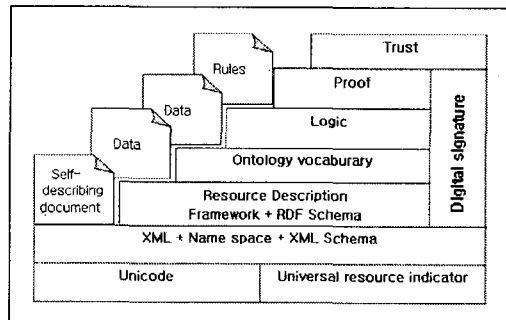


그림 1 The Semantic Web "layer cake" presented by Tim Bernes-Lee at the XML 2000 conference

2.1 XML, Name space, XML Schema

XML[2](eXtensible Markup Language)은 사용자의 필요에 따라 확장 가능한 마크업 언어이다. 그러므로 문서 내용에 관련된 태그를 사용자가 직접 정의할 수 있으며, 정의한 태그를 다른 사람들이 사용할 수 있도록 할 수 있다.

XML 은 기본적인 구문을 제공하지만 실제 태그들을 정의하고 있지는 않으며, 실제 태그들의 집합은 각각의 목적에 따라 누구든지 확장할 수 있다.

XML 이름공간(Name space)은 XML 의 주요한 구성요소 중 하나이다. xmlns 라는 키워드를 사용한다. XML 은 문서 작성자가 자기 나름대로의 태그를 생성하는 것을 허용하므로, 이름충돌이 생길 수 있다. 이름공간은 문서 작성자가 이런 충돌 현상을 피할 수 있는 수단을 제공한다.

XML Schema[3]는 DTD(Document Type Definition)의 개선안이라고 볼 수 있다. XML 문서에서 스키마란 문서가 마크업되는 방식을 기술하는 것이다. 즉, 문서의 문법, 어휘, 구조, 데이터형 등을 규정하는 모든 규칙들을 통틀어서 스키마라고 칭한다.

2.2 RDF, RDFS

웹 콘텐츠의 수가 기하급수적으로 증가함에 따라, 찾고자 하는 문서를 단순한 키워드 매칭에 따른 기존의 검색방법은 한계를 드러내게 되었다.

RDF(Resource Description FrameWork)[4]는 웹 자원(Resource)을 기술하기 위한 기술을 제공한다. 즉 RDF 는 웹 문서에 대한 정보를 제공하여, 위 문제를 해결할 수 있는 가능성을 제공한다. RDF 는 웹에서 전송될 수 있고, 어떤 문서의 의미 정보를 보유할 수 있다. 또한 정확성을 위하여 파싱(Syntactic Parsing) 될 수 있기 때문에 XML 로 표현된 RDF 는 메타데이터를 인터넷 상에서 전송하는데 유리하다. RDF 는 다양한 메타데이터 사이의 연결을 위하여 Semantics, Structure, Syntax 에 대한 공통적인 규칙을 지원한다. 즉 RDF 는 기계(Software Agents)에게 문서의 의미를 이해하도록 할 수 있는 메커니즘을 제공한다.

RDF 는 트리플 테이블 형식으로 되어있다. 트리플 테이블은 {Subject, Predicate, Object} 쌍으로 이루어진다. 즉 Resource(Subject), Property type(Predicate), Value(Object)로 구성되는 트리플은 XML 문서에 대한 메타데이터를 표현하고, RDF Schema[5]는 이러한 메타데이터의 무결성을 보충하기 위해 메타데이터 자원에 대한 유효성을 검증한다. RDF Schema 는 자원과 자원 사이의 관계와 제약 조건을 기술하기 위한 특성을 기술하고 정의한다. 즉 RDF Schema 는 자원을 소프트웨어 에이전트가 읽고 이해할 수 있는 형태로 정보의 의미와 제약조건을 구체화 한 것이다.

2.3 Ontology

온톨로지(Ontology)는 Semantic Web 의 핵심기술 중 하나이다. 그리고 온톨로지는 지식표현(Knowledge Representation)의 수단 중 하나이다. 따라서 온톨로지

에 관한 연구는 인공지능 구현을 위한 지식표상 방법 중 중요한 기술로서, 최근 Semantic Web 을 구현하기 위한 중요한 요소가 되었다.

RDF Schema 에서 정의된 어휘들은 문서에 대한 비공식적인 지식 표현이다. 따라서 웹이라는 전체적인 측면에서 살펴보면, Web Service 를 제공하는데 분명한 한계를 드러나게 된다. 즉 [그림 1] Semantic Web 구성도가 보여주는 것처럼 웹 환경에 대한 공식적인 온톨로지 계층을 필요로 하게 되는 것이다.

DAML+OIL(DARPA Agent Markup Language + Ontology Inference Layer)[6,7]은 웹 자원을 위한 의미 마크업 언어이다. DAML+OIL 은 DAML+OIL Vocabulary 를 사용하는 트리플들을 위한 명확한 의미를 규정한다. OIL[7]은 Semantic Web 에서 사용할 목적으로 온톨로지를 표현하기 위한 Logic 이며 RDF 및 RDF Schema 와 비교적 호환성을 지닌다.

W3C 는 DAML+OIL 의 개정판인 OWL Web Ontology Language [8]을 발표하였다. OWL 은 World Wide Web 상에 온톨로지 공유와 출판을 위한 Semantic 마크업 언어이다. OWL 은 DAML[6]의 장점과 OIL[7]의 장점을 수용하였으며, 애매모호한 표현을 제거하여 훨씬 간결하게 표현하였다.

3. 지능형 웹 서비스 프로토타입 테스트 베드 구축

Semantic Web[1] 환경 프로토타입의 도메인은 “제주도 여행”으로 설정하였다. 크게 항공정보, 렌트카 정보, 숙박 정보로 나누어져 있다.

항공정보에는 스케줄 정보, 요금, 할인율로 구성되어 있으며, 렌트카 정보에는 렌트카 상점정보와 요금으로 구성되어 있으며, 숙박정보에는 호텔, 여관/모텔, 콘도 정보 등으로 구성되어 있다. [그림 2]는 숙박정보에 대한 홈페이지로 XML 형식으로 되어있으며 XSLT(eXtensible Stylesheet Language Transformations)를 적용하였다.

[그림 2]의 홈페이지 자원을 기초로 하여 각각의 도메인은 [그림 3]와 같은 형식으로 쓰여져 있다. [그림 3]에서 보는 것과 같이 위에서 언급한 언어를 이용하기 위해 네임스페이스를 사용하였고 온톨로지에 관한

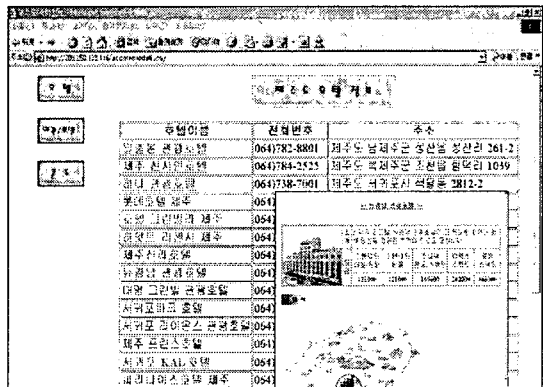


그림 2 숙박정보 홈페이지

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:owl2="http://www.w3.org/2002/07/owl2#"
  xmlns="http://203.250.133.116/accommodations/accom-owl.rdf#">
  <owl:Ontology rdf:about="http://203.250.133.116/accommodations/">
    <owl:versionInfo id="accom-owl.rdf, v 0.1 2002/09/03 15:02:33" content:Exp />
    <rdf:comment id="Information ontology of Jeju-do Accommodations, with data types taken from XML Schema" />
    </owl:Ontology>
    <owl:Class rdf:ID="Region">
      <rdf:type rdfs:label />
    </owl:Class>
    <owl:Class rdf:ID="Province">
      <rdf:type owl:multiset rdfs:resource="#Region" />
    </owl:Class>
    <owl:Class rdf:ID="Person">
      <rdf:type rdfs:label />
    </owl:Class>
    <owl:Class rdf:ID="Jeju-do">
      <rdf:type owl:multiset rdfs:resource="#Province" />
    </owl:Class>
    <owl:Class rdf:ID="Travel">
      <rdf:type rdfs:label />
    </owl:Class>
    <owl:Class rdf:ID="Accommodation">
      <rdf:type owl:multiset rdfs:resource="#Travel" />
      <rdf:type owl:multiset rdfs:resource="#Jeju-do" />
    </owl:Class>
  </rdf:RDF>
  
```

그림 3 OWL 형식의 제주도 숙박 정보

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:accom="http://203.250.133.116/accommodations/accom-owl.rdf#">
  <accom:Accommodations rdf:about="http://203.250.133.116/accommodations/">
    <accom:title>제주도 숙박정보</accom:title>
    <accom:creator>
      <accom:Person>
        <accom:name>박재훈</accom:name>
        <accom:email>middle73@hanmail.net</accom:email>
        <accom:jobtitle>매제대 정보통신공학과 대학원생</accom:jobtitle>
      </accom:Person>
    </accom:creator>
  </accom:Accommodations>
  <accom:Hotel rdf:ID="H_1">
    <accom:name>임출물 관광호텔</accom:name>
    <accom:telephone>064)782-8801</accom:telephone>
    <accom:address>제주도 남제주군 성산읍 성산리 261-2</accom:address>
  </accom:Hotel>
  <accom:Inn rdf:ID="I_1">
    <accom:name>정비스타어관</accom:name>
    <accom:telephone>064)794-6400</accom:telephone>
    <accom:address>제주도 남제주군 대정읍 삼모리 3967</accom:address>
  </accom:Inn>
  <accom:Condominium rdf:ID="C_1">
    <accom:name>포비스콘도미니엄 제주</accom:name>
    <accom:telephone>064)799-9901</accom:telephone>
    <accom:address>제주도 북제주군 어월읍 어월리 2505</accom:address>
  </accom:Condominium>
  </rdf:RDF>
  
```

그림 5 숙박정보에 대한 인스턴스

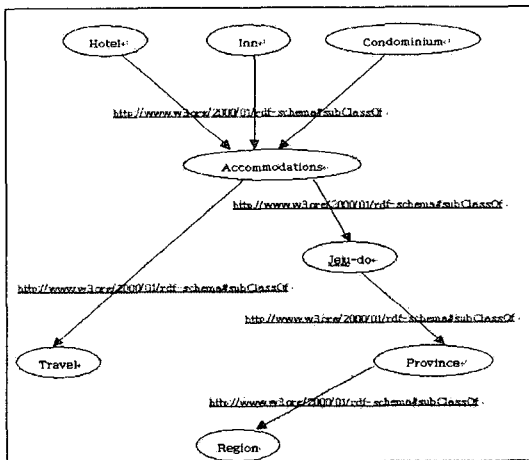


그림 4 제주도 숙박에 관련된 클래스 계층구조

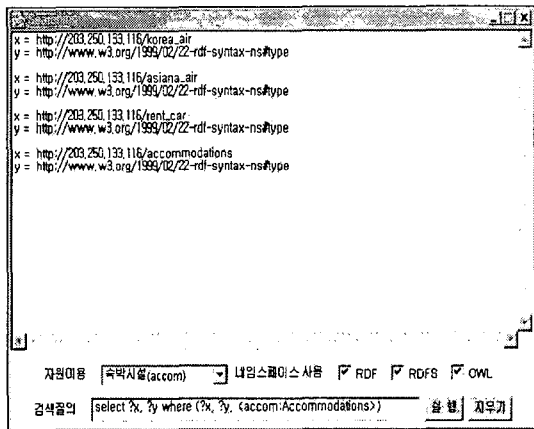


그림 6 RDQL 을 이용한 자원발견

설명과 도메인에 관련된 클래스와 속성을 정의하고 있다.

유효성을 검증하기 위해 W3C 에서 제공하는 RDF Validation Service[9]을 통해서 [그림 3]의 숙박정보 온톨로지의 유효성을 검증하면 클래스와 속성들의 관계가 표시되어지는 것을 볼 수 있다. 그림의 크기가 큰 관계로 클래스의 계층구조만 보여주기 위해서 [그림 4]와 같이 표현하였다. [그림 4]에서 보여지는 것처럼 숙박은 여행과 제주도의 서브클래스로 표현되었다. 그리고 호텔, 여관/모텔, 콘도는 숙박의 서브클래스로 표현되었고 각각은 이름, 전화번호, 주소의 속성을 가지고 있으며, 속성의 데이터형을 XML Schema로 제한을 하였다.

[그림 3]에서 정의되어진 클래스와 속성들을 토대로 [그림 5]는 숙박정보에 대한 인스턴스로 클래스와 속성들의 관계를 표시하고 있으며 속성에 관련된 문자열이나 자원을 표시하고 있다. 그리고 호텔, 여관/모텔,

콘도의 인스턴스를 볼 수 있다.

자원의 발견은 Semantic Web 툴킷의 하나인 Jena[10] 를 이용할 것이다. 그 중에서 Jena 에서 지원하는 RDQL(a query language for RDF) API 를 이용할 것이다. RDQL API 를 이용하여 Java 로 짠 애플리케이션을 [그림 6]과 같이 볼 수 있을 것이다. 애플리케이션은 네임 스페이스를 사용하여 간결하게 검색질의를 할 수 있도록 하였으며 숙박시설, 렌트카, 항공정보와 같은 자원을 사용할 수 있도록 하였다.

[그림 6]에서 보이는 검색질의는 SQL 문과 비슷하여 쉽게 사용할 수 있으며 RDF[4]의 트리플 형식을 이해하고 있으면 손쉽게 검색을 할 수 있다. [그림 6]의 검색질의는 숙박에 관련된 자원을 찾는 것이고 결과값들은 이와 같은 질의를 통해서 추출되는 각각의 도메인에 대한 자원과 속성을 나타낸 것이다.

[그림 7]은 [그림 5]의 숙박정보에 대한 인스턴스를 이용하여 호텔에 관련된 모든 속성과 문자열(자원)을

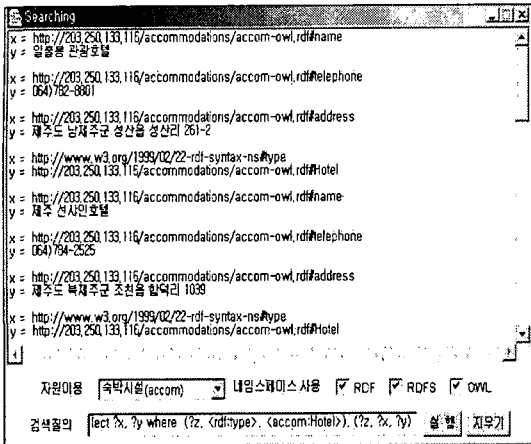


그림 7 RDQL 을 이용한 호텔의 인스턴스 정보추출

찾는 질의이다. 이 질의를 통해서 호텔의 이름, 전화번호, 주소의 속성과 각 속성에 대한 문자열을 볼 수 있다. 항공 스케줄이나 렌트카 정보도 이와 같은 질의를 통해서 원하는 값을 얻을 수 있다.

4. 결론 및 향후 연구방향

사용자가 원하는 서비스 자원발견을 수행할 수 있는 프로토타입 테스트 베드 구축을 살펴보았다. 이것은 컴퓨터가 이해할 수 있는 언어로 표현한 것이다. 즉 Semantic Web 환경을 갖추기 위한 밑바탕이라 볼 수 있다. 진정한 의미의 Semantic Web 은 자동화된 자원 발견과 수행, 구성과 상호운영이라 볼 수 있다. 이와 같은 Semantic Web 을 구축하기 위해 프로토타입을 기반으로 하여 소프트웨어 에이전트가 문서내용을 파악하며 사용자 제약조건을 이용하여 자동적인 자원의 발견, 검색의 실행, 작업의 실행 등을 할 수 있도록 할 것이다. 다시말해서 에이전트가 사용자의 제주도 여행채비준비에 관련된 모든 수행을 자동적으로 수행할 수 있도록 할 것이다.

참고문헌

- [1] S.A. McIlraith, T.C. Son, H.Zeng. Semantic Web Service, IEEE Intelligent Systems, March/April 2001
- [2] Tim Bray, Jean Paoli and Eve Maler. Extensible Markup Language(XML) 1.0 (Second Edition). 2000, W3C: Recommendation(October 2000) [http:// www.w3.org/TR/2000/REC-xml-20001006](http://www.w3.org/TR/2000/REC-xml-20001006)
- [3] David C. and Fallside(IBM). XML Schema Part 0: Primer. 2001, W3C : Recommendation(May 2001) [http:// www.w3.org/TR/xmlschema-0/](http://www.w3.org/TR/xmlschema-0/)
- [4] Lassila, O. and Swick, R. R., Resource Description Framework(RDF) Model and Syntax Specification. 1999, W3C : Recommendation(February 1999) <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- [5] Brickley, D. and Guha, V. RDF Vocabulary Description Language 1.0 : RDF Schema. 2002, W3C : Working Draft(April 2002) [http:// www.w3.org/TR/rdf-schema/](http://www.w3.org/TR/rdf-schema/)

- [6] DAML, Darpa Agent Markup Language, <http://www.daml.org>
- [7] OIL, Ontology Inference Layer, <http://www.ontoknowledge.org/oil/>
- [8] Hendler, J.and Horrocks, L. OWL Web Ontology Language 1.0 Reference. 2002, W3C : Working Draft(July 2002) [http:// www.w3.org/TR/2002/WD-owl-ref-20020729/](http://www.w3.org/TR/2002/WD-owl-ref-20020729/)
- [9] W3C RDF Validation Service, <http://www.w3.org/RDF/Validator/>
- [10] HPL Semantic Web activity, <http://www.hpl.hp.com/semweb>