

# 적응적인 시맨틱 정보 추출을 위한 대화형 사용자 질의 인터페이스 개발\*

조영현, 손진현  
한양대학교 컴퓨터 공학과  
{mhjo, jhson}@cse.hanyang.ac.kr

## Developing an Interactive User Interface System for Adaptive Semantic Information Retrieval

Myung Hyun Jo, Jin Hyun Son  
Department of Computer Science and Engineering, Hanyang University

### 요 약

기존의 웹 서비스는 많은 애플리케이션으로 개발되었고, 다양한 분야에서 잘 활용되어 왔다. 하지만, HTML을 기반 한 기존의 웹은 더 이상의 기능적 성장은 어렵게 되어 기존 웹 서비스의 성장에도 제약을 가져왔다. 이런 문제점을 극복하기 위해 시맨틱 웹이 대두되었다. 시맨틱 웹은 자원의 공유, 통합, 자동 생성의 특징을 가진 기존 웹의 확장이다. 최근 기존의 웹 서비스는 시맨틱 웹의 이런 특성들을 받아들여려는 시도가 있었다. DAML-S가 대표적인 예이다. 본 논문은 기존 웹 서비스 검색처럼 일반 사용자도 편리하게 시맨틱 웹 서비스를 검색할 수 있도록 대화형 사용자 질의 인터페이스를 제시한다. 또, 사용자가 선택한 질의가 시맨틱 질의어(RDQL)로 자동 변환할 수 있는 시맨틱 질의어 자동 생성기 모듈을 제시한다.

키워드: 시맨틱 웹, 시맨틱 언어, 시맨틱 질의어, 웹 서비스, 대화형 사용자 질의 인터페이스

### 1. 서 론

HTML(Hyper Text Markup Language)을 이용한 웹은 네트워크의 발전에 따라 상당한 성장을 이루었다. 이것은 HTML 언어가 가공하거나 관리하기 편리해 일반 사용자 누구나 쉽게 이용할 수 있기 때문이다. 하지만 HTML은 데이터들의 연관성보다는 특정 데이터의 디스플레이 목적을 위해 만들어졌다. 그래서 지금의 웹은 더 이상의 기능적 성장을 기대하기 어려운 상황이다. 이러한 웹의 한계점을 극복하기 위해서, 월드 와이드 웹 컨소시엄은 시맨틱 웹(Semantic Web)을 주장하였다.

시맨틱 웹은 데이터의 디스플레이보다는 다양한 응용 프로그램에서 데이터의 자동 생성(automation), 통합(integration), 재사용(reuse)을 가능하도록 하기 위한 기능들을 부각시켰다. 시맨틱 웹은 웹을 사용하는 주체가 사람뿐만 아니라 기계도 자동적으로 데이터를 이해하고, 해석할 수 있도록 만들어졌다.

기존의 웹이 존재하기 위해서 HTML, XML등의 기술언어가 존재하듯이 시맨틱 웹도 이를 가능하게 하기 위한 기술 언어가 필요하다. 시맨틱 웹 기술 언어는 온톨로지 언어라 말하기도 하는데, 이것은 시맨틱 웹의 개념이 발전하기 이전부터 존재해 시맨틱 웹을 구성하기 위한 중요한 도구로써 활용되었다.

온톨로지 언어에는 DAML+OIL[1], OWL, Ontolingua 등이 있다. 현재 월드 와이드 웹 컨소시엄은 OWL을 온톨로지 언어로 추천하고 있는 상황인데, 본 논문은 OWL의 모태가 되는 DAML+OIL[1]을 이 논문의 기본 온톨로지 언어로 선택하였다.

웹의 기능적 성장이 어렵게 되면서, 가\*존 웹에 기반을 두고 있는 웹 서비스도 문제점들이 대두되고 있다. 이러한 문제점을 해결하기 위해 기존 웹 서비스도 시맨틱 웹이 갖고 있는 특성

을 받아들였는데, 이것을 시맨틱 웹 서비스라 한다. 기존 웹 서비스가 SOAP, UDDI, WSDL로 구성되는 것처럼 시맨틱 웹 서비스도 Service, Profile, Process, Grounding으로 구성된다. 시맨틱 웹 서비스는 DAML+OIL을 기반으로 했기 때문에, DAML-S[2] (DARPA Agent Markup Language - Service) 라 한다.

본 논문은 두 가지를 제안 한다. 먼저, 시맨틱 언어/시맨틱 질의어를 모르는 일반 사용자에게 대화형 사용자 질의 인터페이스를 제공한다. 제공된 인터페이스는 시맨틱 웹 서비스의 메타정보를 담고 있는 DAML-S Profile[2]에 대한 정보를 보여준다. 또, 대화형 사용자 질의 인터페이스를 시맨틱 질의어로 자동 변환하는 SQA (Semantic Query Adapter) 시스템을 제안 한다. SQA는 서비스 속성을 표현한 DAML-S Profile[2]에 대한 질의 패턴을 분석해서 시맨틱 질의어 RDQL[3]로 변환하는 시스템이다. RDQL을 선택한 이유는 RQL, DQL등의 다른 시맨틱 질의어보다 표현력과 완전성이 뛰어나기 때문이다.

논문의 구조는 다음과 같다. 제 2장에서는 관련 연구로써 시맨틱 언어와 시맨틱 질의어를 분석한다. 제 3장에서는 DAML-S Profile 구조를 통한 대화형 사용자 질의 인터페이스 설계와 대화형 사용자 질의 인터페이스에서 시맨틱 질의어로 자동 변환하는 SQA 모듈을 제안한다. 제 4장에서는 SQA의 구현 및 결과를 소개하고, 제 5장에서는 결론을 맺는다.

### 2. 관련 연구

#### 2.1 DAML+OIL

시맨틱 웹을 구축하기 위해 기존 웹에서 사용하던 HTML, XML을 이용한다는 것은 상당한 문제점을 가지고 있다. HTML은 디스플레이의 목적을 위해 생성된 언어이기 때문에 에이전트가 자동으로 데이터를 생성, 공유하기에 적합하지 않다. XML은 HTML 보다 구조화된 언어지만, 다양한 형태 구조를 가

\* 본 연구는 대학 IT연구 센터 육성·지원 사업의 연구 결과로 수행되었음

\* 본 연구는 한국 과학재단 목적 기초 연구 (R08-2003-000-10464-0) 지원으로 수행되었음

지기 때문에 에이전트가 데이터의 의미를 파악하기 힘들다. 즉, HTML과 XML은 시맨틱 웹을 구성하기 위해서 적합하지 못하기 때문에 새로운 시맨틱 웹 기술 언어의 필요성이 대두되었다.

DAML+OIL[1]은 XML 기반으로 데이터의 의미적 상호 운용성을 가능하게 하기 위해 DARPA 프로젝트로 개발된 온톨로지 언어이다. DAML+OIL은 RDF(S)[4]와 OIL을 기반 하여 개발되었다. DAML+OIL은 기존의 RDFS보다 데이터에 대한 정의가 일반적이고, 더 많은 기능들을 추가시켰다. 크게 두드러진 요소는 First-Order Logic을 기반한 공리개념을 포함시킨 것이다. 기본적으로 DAML +OIL로 표현된 온톨로지는 크게 클래스 요소(class element)와 특성 요소(property element)로 구성된다.

2.2 RDQL

시맨틱 웹은 시맨틱 웹 기술 문서 정보 추출을 위한 질의어가 필요하다. XML 문서들을 질의 하기위해 XPath, XQuery 등이 필요한 것과 같은 맥락이다.

RDQL(RDF Data Query Language)[3]은 하나의 RDF[4] 문서에 대한 질의를 목적으로 하며, SQL과 유사한 형태를 가진다. DAML+OIL도 RDF와 같은 구조로 구성되어 있기 때문에 RDQL로 질의가 가능하다. 질의는 하나의 RDF 모델에 대한 제약과 트리플 패턴(tripple pattern)으로 구성된다. RDQL은 다음과 같은 형태를 갖는다.

- SELECT : 결과로 반환될 변수를 설정한다.  
예) *SELECT ?x, ?y*
- FROM : 질의 대상 문서를 설정한다.  
예) *FROM <http://www.daml.org/example.rdf>*
- WHERE : 반환될 변수를 이용한 트리플의 결합 (conjunctive) 질의를 형성한다.  
예) *WHERE (<?x, <foo:hasFather>, ?y>), ...*
- AND : 변수에 대한 조건을 설정한다.  
예) *AND ?y >= 20*
- USING : 질의 중에 사용한 이름 공간을 설정한다.  
예) *USING foo for <http://www.daml.org/a.rdf>*

3. 시맨틱 질의어 자동생성기

3.1 DAML-S Profile 구조를 통한 대화형 사용자 질의 인터페이스 설계

시맨틱 웹 서비스는 에이전트가 데이터를 이해하고 분석할 수 있도록 다음의 자동화 서비스를 가능하게 한다. 웹 서비스 검색(discovery), 웹 서비스 삽입(invocation), 웹 서비스 선택 및 통합(selection & composition), 웹 서비스 감시(execution monitoring).[2] 이러한 특성들 중에 가장 우선시 되는 것은 웹 서비스 검색이다. 사용자 또는 에이전트가 웹 서비스를 사용하려면, 웹 서비스 검색이 먼저 이루어져야 하기 때문이다.

그림 1은 본 논문의 시맨틱 웹 서비스 DAML-S 구조이다. 하나의 시맨틱 웹 서비스를 구축하기 위해 총 4개의 파일 (Service.daml, Profile.daml, Process.daml,

Grounding.daml)[2] 이 필요한데, 시맨틱 웹 서비스 검색을 위해 ServiceProfile만이 필요하다. ServiceProfile은 서비스의 기능에 대한 설명을 하는 부분으로서, 찾고자 하는 서비스를 결정하기 위해 서비스 에이전트가 필요로 하는 정보의 타입을 기술한다.

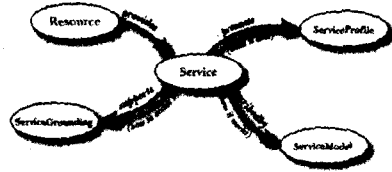


그림 1. DAML-S 구조

서비스 프로파일은 표 1에서와 같이 크게 세가지 정보를 구분해서 기술한다.

- 서비스를 제공하는 업체에 관련된 정보(Organization)
- 서비스를 연산하기 위한 정보 (Function)
- 서비스의 특징을 나타내는 정보(Profile Attribute)

표 1. Service Profile Elements & Attributes

	element	attribute	
Organization	serviceName		
	textDescription		
	contactInformation		
	Actor	name	
		title	
		phone	
fax			
email			
Function	physicalAddress		
	webURL		
	input		
	output		
	precondition		
Profile Attribute	effect		
	serviceParameter	serviceParameterName	
		sParameter	
	QualityRating	ratingName	
		rating	
	ServiceCategory	categoryName	
taxonomy			
value			
		code	

시맨틱 웹 서비스는 인간보다 에이전트가 쉽게 이해할 수 있도록 구성되었다. 하지만 서비스 프로파일 모든 정보가 에이전트를 위한 것은 아니다. 업체에 관련된 정보는 서비스를 실제로 사용하는 일반 사용자와 업체의 연결을 위한 정보로서 인간을 위한 정보이다. 그래서 Organization 은 웹 서비스 검색의 입력 파라미터보다는 출력 데이터 또는 출력 후 업체와 연결을 위한 데이터로 사용된다. 이것은 이 시스템의 입출력 정보를 결정짓기 위한 정책에 중요한 요소로 작용되었다.

이 시스템은 일반 검색 엔진이 갖고 있는 표현력 및 기능성을 제공하기 위해, 사용자 질의 인터페이스에 논리연산자(AND, OR) 및 비교 연산자(<, >) 특성을 포함시켰다.

3.2 사용자질의 패턴으로부터 RDQL 자동생성(SQA)

이 시스템의 아키텍처는 그림 2와 같다. 모든 프로시저의 컨트롤은 컨트롤러에서 담당하며, 대화형 사용자 질의 인터페이스에서 선택된 질의를 질의어 생성기가 시맨틱 질의어로 생성한다. 시맨틱 질의어(RDQL)를 생성할 때, 논리 연산자 및 비교 연산자의 특성을 제공하기 위한 컴포넌트는 다중 생성 프로시저이다. 이것은 논리 연산자 및 비교 연산자에서 사용하는 포괄적인 의미 관계(OR or <, >)를 여러 개의 질의어로 쪼개어 가능하게 하는 프로시저이다.

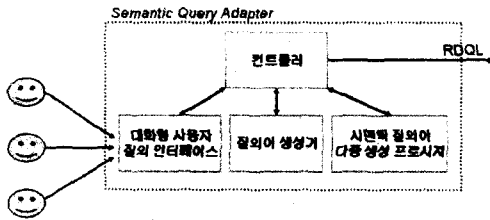


그림 2. SQA 아키텍처

은 오른쪽 패널에 질의 창으로 추가/삽입되어 사용자에게 질의를 가능하게 한다.

오른쪽 패널은 각 엔티티마다 크게 두 가지로 구분된다. 먼저, 서비스 프로파일의 정보를 트리형태로 보여준다. 일반 사용자가 시맨틱 웹 서비스를 알지 못하기 때문에, 트리를 통해 자신이 원하는 정보를 선택할 수 있도록 한다. 이것은 시맨틱 웹서비스가 트리를 구조를 가진 DAML+OIL을 기반으로 했기 때문에 가능한 것이다. 둘째, 트리로부터 선택된 사용자 질의는 오른쪽 박스에 추가 포함된다. 오른쪽 박스는 서비스 프로파일의 특성에 맞게 논리 연산자 및 비교 연산자 특성을 내포하고 있다. 또, 박스 하나에 포함시킬 수 없는 연산자 의미는 선택 라디오 버튼, 체크 박스등을 이용했다.

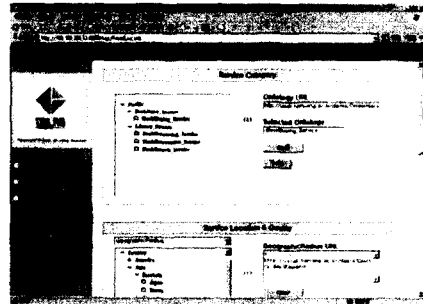


그림 3. 대화형 사용자 질의 인터페이스

각 연산자는 표 1에 나타난 서비스 프로파일 엔티티들의 의미에 따라 적용 여부를 결정 받는다. 입력 질의로써 선택된 엔티티들은 Function 과 Profile Attribute에 포함된 엔티티들이다. 각 엔티티들은 DAML+OIL 제약 조건에 따라 의미 해석을 달리 하는데, 이것은 연산자 적용 여부의 판단 근거로 적용되었다. 이것을 정리한 것은 표2와 같다.

표 2의 설명을 예로 들면 다음과 같다. ServiceCategory는 업체에서 제공하는 서비스의 유형을 구분짓기 위한 엔티티이다. 대표적으로, NAICS 나 UNSPSC 와 같은 분류 단계에서 제공하고 있는 분류 기준을 사용해서 서비스의 유형을 기술하고 있다. 서비스를 제공하는 업체는 NAICS와 UNSPSC 중 하나만 또는 모두에 가입하여 서비스를 분류 받을 수 있다. 즉, ServiceCategory는 사용자 질의에서 AND, OR 연산자를 사용할 수 있도록 개발되어야 한다. ServiceCategory는 각 인스턴스들(NAICS, UNSPSC)이 의미적으로 비교 대상이 아니기 때문에, 비교 연산자를 사용할 수 없다.

표 2. 연산자와 DAML-S Profile의 의미적 결함표

	AND	OR	<, >
input	Yes	Yes	No
output	Yes	Yes	No
effect	Yes	Yes	No
precondition	Yes	Yes	No
serviceParameter	Yes	Yes	No
QualityRating	Yes	No	Yes
ServiceCategory	Yes	Yes	No

4. 대화형 사용자 질의 인터페이스 및 SQA 구현

그림 3은 이 시스템의 사용자 질의 인터페이스 일부로써, 왼쪽 패널은 표1에서 정의한 Function과 Profile Attribute의 엔티티들을 선택하기 위한 것이다. 왼쪽 패널에서 선택된 엔티티들

5. 결론

기존 검색 시스템에서 제공한 기능성을 제공하면서, 시맨틱 웹 서비스를 질의 할 수 있는 이 시스템은 시맨틱 웹을 모르는 일반 사용자 특의 확장성을 가져올 것이다. 또, 이런 확장성은 시맨틱 웹이 기존의 웹과 잘 결합할 수 있는 계기를 마련할 것이다.

월드 와이드 웹 컨소시엄에서 OWL을 온톨로지 언어로 추천함으로써, 이 언어를 통한 향후 발전이 기대된다. 본 논문의 향후 연구는 OWL-S의 대화형 사용자 질의 인터페이스를 구축하는 것이다. 또, RDQL 이외의 다른 시맨틱 질의 언어(RQL, DQL, ...)로의 자동 생성을 가능하게 하는 시스템을 구축할 것이다. 이것은 에이전트 간의 통신 프로토콜로 작용되어, 확장적 호환 효과를 가져올 것이다.

6. 참고 문헌

- [1] DAML language home page (<http://www.daml.org/languages>)
- [2] The DAML Services Coalition, "DAML-S: WebServiceDescription for the Semantic Web", to appear in The First International SemanticWeb Conference (ISWC), Sardinia, Italia, June 9-12th, 2002
- [3] Andrew Seaborne, RDQL: A Data Oriented Query Language for RDF Models. Technical report. Hewlett-Packard, 2001. Peter Mika, Victor Iosif, York Sure, and Hans Akkermans
- [4] O.Lassila and R. Webick, "Resource Description Framework (RDF) Model and Syntax Specification." W3C Recommendation, Jan. 1999, [www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/](http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/) (current Feb. 2004)