

Restful 기반 시맨틱웹 온톨로지 구축 방안

최로환*, 박석천**, 이철식***

*가천대학교 일반대학원 모바일소프트웨어학과

**가천대학교 컴퓨터공학과 정교수(교신저자)

***데이터스트림즈 기술연구소 선임연구원

romance213@gc.gachon.ac.kr

Implementation Method of Semantic Web Ontologies Based on Restful

Ro-Hwan Choi*, Seok-Cheon Park**, Cheol-Sik Lee***

*Dept of Mobile Software, Gachon University

**Dept of Computer Engineering, Gachon University(Corresponding Author)

***Research Engineer, DATASTREAMS co., ltd

요 약

웹 2.0의 등장과 함께 Restful 웹 서비스의 활용이 전통적인 SOAP 기반 웹 서비스에 비해 크게 증가되었다. 최근 웹상에 이용 가능한 Restful 웹 서비스들의 수가 급격하게 증가됨에 따라 사용자들이 적합한 웹 서비스를 찾는 것은 매우 중요한 이슈로 대두되었다. 그러나 기존의 키워드 기반 검색 방법은 낮은 검색의 정확도 때문에 문제가 많다. 따라서 본 논문에서는 Restful 기반 온톨로지 구축 방법을 제안하고자 한다.

I. 서론

최근 웹 2.0의 새로운 패러다임에 맞춰 그동안 폐쇄적으로 관리하던 데이터베이스 정보를 OpenAPI를 활용하여 불특정 다수의 이용자들에게 개방하고 있는 추세에 있다.

구글의 Google Map, Google Trend와 네이버의 지식 서비스 등은 데이터들을 웹서비스로서 공개하고 있다.

그리고 OpenAPI의 구현 형태가 RPC(Remote Procedure Call) 기반의 SOAP(Simple Object Access Protocol) 방식에서 최근에는 REST(REpresentational State Transfer) 방식으로 바뀌고 있다[1].

최근 동향을 살펴보면 아마존의 웹 서비스 중 90%가 REST 방식을 따르고 있으며, 구글은 더 이상 SOAP 방식으로는 웹 서비스를 제공하지 않겠다고 선언하였다.

이렇듯 REST 방식의 웹 서비스 개발 방법론이 최근 많은 주목을 받고 있다.

이 아키텍처는 웹상의 모든 자원들을 URI로 표현하며 이러한 URI 표현 방식을 통해 웹 애플리케이션을 구동하고 그 결과를 전달 받아 처리하는 것에 중점을 두고 있다. 따라서 REST 방식은 웹상에 존재하는 자원의 CRUD(Create, Retrieve, Update, Delete - 만들고, 읽고, 고치고, 지우는 처리)를 HTTP의 메소드인 POST, GET, PUT, DELETE와 대응한 URI 표현 방식으로 요구 자원을 요청한다[2].

이는 매쉬업과 비슷한 개념으로, 한편으론 매쉬업을 RESTful 웹 서비스의 조합으로 생각 할 수 있다.

하지만, 매쉬업을 만들기 위해서는 프로그래밍 기술뿐만 아니라 매쉬업에 포함되는 모든 서비스들에 대한 API를

파악해야 하기 때문에 현실적으로 어려운 일이다.

따라서 최근엔 매쉬업을 자동적으로 생성해 줄 수 있는 여러 가지 툴들(예를 들어, 야후의 Pipes[3], 마이크로소프트의 Popfly[4] 등)이 개발되었지만, 이는 키워드 기반 웹 검색에서만 지원이 된다는 단점이 있다.

왜냐하면, 키워드 기반 검색에서는 키워드가 정확히 일치하는 웹 서비스일 경우에만 발견이 되고 원하는 정보가 아니지만 키워드가 포함이 되었다는 이유로 그다지 관련 없는 정보가 다수 포함될 수가 있기 때문이다.

따라서 이러한 키워드 기반 검색 방법의 한계를 극복하기 위한 기법으로서 시맨틱 정보를 이용한 온톨로지(ontology) 활용 방법이 있을 수 있다[5].

본 논문의 구성은 다음과 같다. 1장 서론에 이어 2장에서는 본 논문과 연관된 관련 연구들을 살펴보고 3장에서 RESTful 시맨틱웹 온톨로지 구축 방안을 제시하고 4장에서 결론을 맺는다.

II. 관련연구

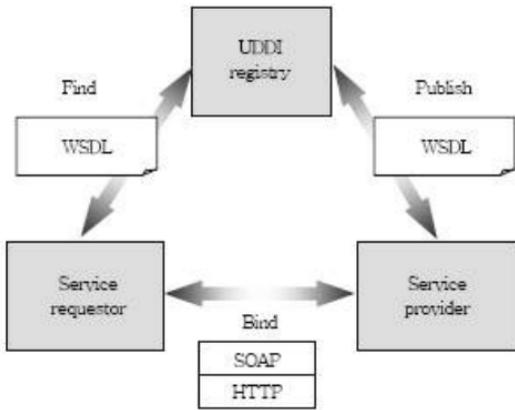
2.1 SOAP 기반 웹서비스

SOAP은 특정 분산 기술 또는 플랫폼에 의존하지 않고 분산 객체를 액세스 할 수 있는 프로토콜로서 HTTP 상에서 전송된다.

모든 SOAP 메시지는 SOAP 봉투(envelope), SOAP 헤더(header), SOAP 바디(body)로 구성된 하나의 XML 문서로 표현되는데 이러한 복잡한 구성으로 인해 HTTP 상에서 전달되기에 무겁고, 메시지 인코딩/디코딩 과정 등

웹서비스개발의 난이도가 높아 개발 환경의 지원이 필요하다.

그림 2.1은 SOAP 기반 웹서비스의 구조를 보여주고 있다. SOAP 기반 웹서비스는 SOA 구조에 따라 UDDI 레지스트리를 통해 웹서비스를 등록하고(publish), 탐색하고(find), 바인딩하여(bind) 이용한다.



(그림 2.1) SOAP 기반 웹서비스 구조

서비스 실행 관점에서, SOAP 기반 웹서비스에서는 서비스 제공자와 요청자 간에 SOAP 프로토콜로 메시지를 주고 받는 방식으로 서비스를 이용한다. 즉, 서비스 요청자가 웹서비스 요청을 SOAP으로 인코딩하여 서비스 제공자에게 전달하면, 서비스 제공자는 이를 디코딩하여 적절한 서비스 로직을 통하여 결과를 얻고, 그 결과를 다시 SOAP 인코딩하여 서비스 요청자에게 반환한다.

2.2 RESTful 웹서비스

REST 방식은 대규모 네트워크 시스템을 위한 아키텍처로서 2000년 Roy Fielding에 의해서 최초로 제안되었다 [6]. 그의 박사학위논문에서, 현재의 웹 아키텍처가 웹의 본래 설계의 우수성을 활용하지 못하므로 웹의 장점을 최대한 활용할 수 있는 네트워크 기반의 아키텍처를 제안했는데 이것이 REST이다. REST는 부수적인 레이어나 세션 관리를 추가하지 않고도 HTTP 프로토콜로 데이터를 전달하는 프레임워크이다. 또한 클라이언트/서버 간의 구성요소를 엄격히 분리하여 구현은 단순화시키고 확장성과 성능은 높일 수 있는 아키텍처다.

최근 들어 REST는 웹에 개방된 리소스들을 원격에서 또는 지역적으로 쉽게 이용하려는 웹 응용에 정착하게 되었고, REST 아키텍처 스타일에 따라 정의되고 이용되는 서비스나 응용을 RESTful 웹서비스라 한다[7].

여기서 리소스(resource)란 REST 아키텍처의 핵심 요소로서 웹사이트, 블로그, 이미지, 음악, 이용자, 지도, 검색 결과 등 웹에서 다른 이들과 공유하고자 개방된 모든 자원을 의미한다. REST 구조에서의 리소스는 그들의 고유한 URI를 가지며, HTTP의 기본 메소드인 GET/PUT/POST/DELETE만으로 접근을 할 수 있다.

리소스의 다양한 표현 방식은 HTTP의 accept 헤더 값 또는 URI 파라미터로 지정하면 되는데, 리소스와 그를 지칭하는 URI, HTTP로 단일화된 인터페이스, 다양한 리소스 표현 등과 더불어 빼놓을 수 없는 RESTful 웹서비스의 특성이 스테이트리스(statelessness)이다.

HTTP의 특성을 상속하여 REST 역시 스테이트리스 특성을 가지게 되는데 스테이트리스란 웹서비스 제공 서버측에서 클라이언트의 상태 정보를 저장, 관리하지 않는 것을 의미한다.

이와 같이 RESTful 웹서비스는 리소스 중심의 표현, 전달, 접근방식의 특성으로 인해 리소스 기반 아키텍처(ROA)라고 한다. ROA는 서비스 중심의 SOA에 대응되는 개념으로 일컬어지고 있다. 즉, RESTful 웹서비스는 리소스 URI를 알면 웹서버와 웹클라이언트의 종류에 상관없이 HTTP 프로토콜만으로 접근 가능한 서비스라 할 수 있다. 이러한 단순명료한 접근 방식 때문에 구글, 야후, 트위터 등 대부분의 웹 2.0 API가 RESTful 웹서비스로 제공되고 있으며, 위젯을 이용한 서비스 매시업(mashup)을 활성화시킨 원동력이기도 하다.

표 2.1은 SOAP 기반 웹서비스 기술과 RESTful 웹서비스 기술을 비교하여 요약한 표이다[8].

<표 2.1> 웹서비스 기술 비교

	SOAP기반 웹서비스	RESTful 웹서비스
특징	Machine-Readable Web	Human-Readable Web
	Stateful	Stateless
	엄격한 문법 검사	기본 XML만으로 서비스개발 가능
	웹서비스 개발환경이 지원되어야 함	별도 개발환경을 지원할 필요 없음
전달 메커니즘	Remote Procedure Call	Syndicate Pattern
전달 프로토콜	SOAP / HTTP, SMTP	HTTP
서비스명세	WSDL	WADL, XML, JSON, hREST 등
문제점	무거운 프로토콜	관리가 어렵고 표준의 부재

III. RESTful 시맨틱 온톨로지 구축 기법 제안

온톨로지 구축 작업은 두 가지 이유 때문에 쉬운 일이 아니다. 첫째, RESTful 웹 서비스 매개변수들은 복합단어, 약어, 개발자의 명명(naming) 습관 등으로 인해 매우 다양해 질 수 있다. 따라서 WordNet과 같은 전자 사전을 바로 적용하기 어렵다. 둘째, RESTful 웹 서비스 리소스에서는 일반적으로 매개변수들이 몇 개 존재하지 않으며, 이에 대한 충분한 설명도 거의 제공하고 있지 않다. 따라

서 단어 빈도수를 기반으로 하는 TF/IDF(Term Frequency/Inverse Document Frequency)[9]와 같은 전통적인 정보 검색(information retrieval) 기법들은 잘 적용될 수 없다.

3.1 연관 규칙 기반 분석 기법

RESTful 웹 서비스의 매개변수들을 토큰화하여 용어들로 분리한 후, 관련성이 많은 용어들에 대해 클러스터(cluster)를 형성하면 이 클러스터는 각각의 단어가 아닌 하나의 의미 있는 개념(concept)을 나타낸다. 이러한 클러스터는 “매개변수들이 동시에 자주 나타난다면, 그것들은 같은 개념을 나타내는 경향이 있다”는 가정 하에 하나의 특별한 연관규칙(association rules)[10]에 따라 만들어진다.

	t1	t2	t3	t4	t5	t6	freq
q1	0	1	1	1	0	0	90
q2	1	1	0	0	0	0	80
q3	0	0	1	1	1	0	60
q4	0	0	0	0	1	1	60

(a) 질의 빈도 테이블

	t1	t2	t3	t4	
t1	Affinity (t1,t1)	Affinity (t1,t2)	Affinity (t1,t3)	Affinity (t1,t4)	...
t2	Affinity (t2,t1)	Affinity (t2,t2)	Affinity (t2,t3)	Affinity (t2,t4)	...
t3	Affinity (t3,t1)	Affinity (t3,t2)	Affinity (t3,t3)	Affinity (t3,t4)	...
t4	Affinity (t4,t1)	Affinity (t4,t2)	Affinity (t4,t3)	Affinity (t4,t4)	...
	⋮	⋮	⋮	⋮	⋮

(b) 트리플 간 연관성 테이블

(그림 3.1) 연관성 추출 과정에서 생기는 테이블들

스키마 레벨 트리플 간 연관성을 추출하는 작업은 도메인 온톨로지의 질의 로그로부터 그림 3(a)와 같은 질의 빈도 테이블을 만드는 작업과, 이를 활용하여 그림 3(b)와 같은 트리플 간 연관성 테이블을 생성하는 작업으로 구성된다.

그림 3(a)의 질의 빈도 테이블은 각 열이 도메인 온톨로지 요약 그래프에 존재하는 스키마 레벨 트리플에 대응되고, 각 행이 하나의 질의 형태에 대응된다. 질의 형태란 동일한 스키마 레벨 트리플 조합을 사용해서 수행되는 구조 질의들을 의미 한다.

3.2 패턴 기반 시맨틱 분석 기법

패턴 기반 시맨틱 분석 기법의 주된 목표는 매개변수 내의 각 단어들 사이의 상관관계를 취득하고, 비교되는 단어들이 서로 유사하고 상관관계가 조건에 일치한다면 그 비교를 매치하는 것이다.

매개변수 내의 각 단어들 사이의 상관관계를 취득한 후 그들을 온톨로지에 저장한다.

시맨틱 웹의 논리적 기반이 되는 온톨로지는 “특정분야에 대한 공유할 수 있는 개념들의 정형화된 기술”로 정의되며, 개념은 그것의 특징을 설명하기 위한 속성(property) 과 개념 사이의 상-하위(subclass) 관계로 표현된다.

따라서 본 논문에서는 각 단어들 간의 상관관계로써 속성과 상-하위 관계를 중점적으로 취급한다. 위에 서술된 패턴 조사 결과로부터 RESTful 웹 서비스 매개변수에 대한 온톨로지 변환 규칙(Transformation Rules)은 아래 (표 3.1)와 같다.

● 명사₁ + ... + 명사_n 형태 : 규칙-1, 규칙-4, 규칙-5 매개변수가 중심어(명사₁)의 속성(property)이 된다.

즉 ‘매개변수 **propertyOf** 명사₁’의 규칙이 생성된다. 예를 들면, `companyId`는 규칙-1을 따르며 다음과 같은 규칙이 온톨로지에 첨가된다.

- `companyId` → `companyId propertyOf Company`
 하지만 규칙-5의 경우에는 전치사의 역할로부터 중심어가 명사₂로 바뀌어 ‘매개변수 **propertyOf** 명사₂’의 규칙이 생성된다. 예를 들면, `dateOfBirth`는 규칙-5를 따르며 다음과 같은 규칙이 온톨로지에 첨가된다.

- `dateOfBirth` → `dateOfBirth propertyOf Birth`

● 형용사/동사+명사 형태: 규칙-2, 규칙-3 매개변수가 중심어(명사)의 자식관계(subclass)가 된다.

즉 ‘매개변수 **subClassOf** 명사’의 규칙이 생성된다. 예를 들면, `availableCredit`는 규칙-2를 따르며 다음과 같은 규칙이 온톨로지에 첨가된다.

- `availableCredit` → `availableCredit subClassOf Credit`

표 3.1과 같은 변환 규칙의 생성은 다음과 같은 관찰로부터 유도되었다[11].

<표 3.1> 온톨로지 변환 규칙

규칙	패턴	상관관계
1	명사 ₁ + 명사 ₂	매개변수 propertyOf 명사 ₁
2	형용사+명사	매개변수 subClassOf 명사
3	동사+명사	매개변수 subClassOf 명사
4	명사 ₁ + 명사 ₂ + 명사 ₃	매개변수 propertyOf 명사
5	명사 ₁ + 전치사 + 명사 ₂	매개변수 propertyOf 명사 ₂

표 3.1과 같은 규칙들을 사용하여 온톨로지가 구축되고 나면, 두 번째 단계는 두 개념 간 매칭을 시키는 것이다. 하나의 온톨로지는 원천(source) 웹 서비스 집합으로부터

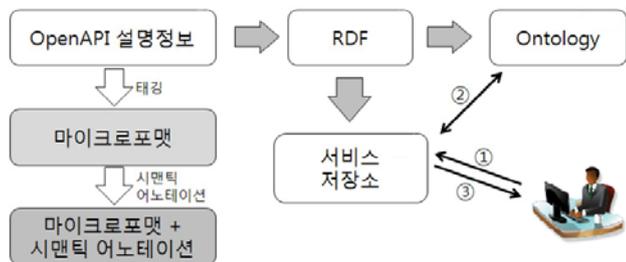
취득하고, 다른 온톨로지는 목표(target) 웹 서비스 집합으로부터 취득한다.

3.3 제안하는 시스템 구조

제안하는 시맨틱 온톨로지 구축 방법의 핵심 내용은 RESTful 웹 서비스의 매개변수들에 대해 의미적으로 (semantically) 같은 개념들을 묶고(clustering), 매개변수 내에 있는 단어들 간의 계층관계(hierarchical relationship)를 구축하여 단어들 사이에 숨겨져 있는 시맨틱 개념을 활용한다.

3.3.1 시맨틱 어노테이션

시맨틱 기술을 활용한 RESTful 웹서비스 검색에 대한 시스템 구조는 아래 그림 3.2와 같다.



(그림 3.2) 시스템 구조

사용자가 서비스 저장소에 질의 하면(①), 서비스 저장소는 온톨로지를 활용하여 사용자가 질의한 검색어를 확장한다(②). 이를 바탕으로 검색을 수행하여 결과를 사용자에게 반환(③)하게 된다.

사용자가 검색 및 질의를 했을 때 그 질의한 단어에 대한 확장은 온톨로지가 도맡아하게 되는데, 구문의 형태를 인식함에 있어서 사람이 이해하는 수준으로 데이터의 의미를 인식해 처리하고 문장이나 문서에 주석처럼 추가적인 정보를 제공하는 시맨틱 어노테이션 형식을 따른다. 특정 개체를 매핑하여 추가적인 시맨틱 정보를 생성하여 기존의 데이터들로 새로운 의미의 데이터를 확장 생성할 수 있게 한다. 그렇게 되기 위해서는 오류를 최소화 하기위해 주제영역을 확실히 정하여 정해진 목표에 맞춰 타당성 조사가 이루어져야하고 그 영역의 어휘사전 및 전문가가 사용하는 언어개념까지 수집해 개념을 잡아놓는 것이 필요하다.

3.3.2 온톨로지 통합

온톨로지의 목적은 공유 및 재사용이만큼 기존의 것을 활용할 수 있게 하여 통합할 것인지, 아니면 아예 새로운 온톨로지를 구축할 것인지 결정해야하고 통합을 한다면 수집된 어휘와 기존의 온톨로지의 통합 과정을 거쳐 개념과 관계를 재정의해야 한다. 왜냐하면 개발자의 경험이나 인식에 따른 차이로 인한 다른 결과가 나타날 수 있기 때문이다. 또한 유지관리 목적으로 시뮬레이션이나 테스트를

통해 오류사항 및 문제를 개선하여 지속적인 온톨로지 활용이 가능하게 해야 한다.

IV. 결 론

본 논문에서는 연관규칙 기반에 패턴 기반 시맨틱 분석 기법을 추가한 시맨틱 온톨로지 구축 방법을 제안하였다.

본 연구의 핵심 내용은 RESTful 매개변수들에 대해 의미적으로 같은 개념들을 묶고, 매개변수 내에 있는 각 단어들 간의 계층관계를 형성하여 자동적으로 시맨틱 온톨로지를 구축하는 것이다.

향후 연구 과제로는, 본 논문에서 제안되는 패턴 기반 시맨틱 분석 기법은 적용되는 영역이나 응용 분야에 따라 온톨로지 변환 규칙이 약간 달라질 수 있다. 다른 문헌들에 나와 있는 구문 패턴 분석 기법이나 통계 기반 패턴 기법과 같은 다양한 기법들에 대한 분석이 요구된다.

사사의 글

본 논문은 2013년도 지식경제부의 SW 전문인력양성 사업의 재원으로 정보통신산업진흥원의 고용계약형 SW석사 과정 지원 사업(HB301-13-1003)으로부터 지원받아 수행되었습니다.

참고문헌

- [1] L. Richardson and S. Ruby, RESTful Web Services, O'RELLY, 2007.
- [2] S. Vinoski, "Demystifying RESTful Data Coupling," IEEE Internet Computing, Vol. 12, Iss. 2, pp. 87-90, March/April 2008.
- [3] <http://pipes.yahoo.com/pipes>
- [4] T. Loton, Introduction to Microsoft Popfly, No Programming Required, Lotontech Limited, 2008.
- [5] K. Gomadam, A. Ranabahu, and A. Sheth, "SA-REST: Semantic Annotation of Web Resources", 2010.
- [6] R. T. Fielding, Architectural Styles and the Design of Network-based Software Architectures, University of California, Irvine, 2000.
- [7] 박유미 외 4인, "SOAP 기반 웹서비스와 RESTful 웹서비스 기술 비교", 제 25권, 제 25 호, 전자통신동향 분석, 2010년 4월
- [8] Greg Hines, "RESTful Web Services and Drupal," PingVision, 2008.
- [9] G. Salton and M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1983.
- [10] X. Dong, A. Halevy, J. Madhavan, E. Nemes, and J. Zhang, "Similarity Search for Web Services," Proceedings of VLDB, 2004.
- [11] 이용주, "RESTful 웹 서비스에서 시맨틱 온톨로지를 구축하기 위한 클러스터링 및 패턴 분석 기법", 인터넷정보학회, 2011.