

시맨틱 기술을 활용한 RESTful 웹서비스의 검색 기법 개발

(Development of Search Method
using Semantic technologies
about RESTful Web Services)

차 승 준* 최 윤 정** 이 규 철***
(Seung Jun Cha) (Yun Jeong Choi) (Kyu Chul Lee)

요 약 최근 웹 2.0의 등장과 함께 플랫폼으로의 웹이 강조되면서, SOAP 기반의 웹서비스에 비해 RESTful 웹서비스가 크게 증가하고 있다. 하지만 서비스들은 이미 많이 존재하며 빠르게 증가하기 때문에 키워드를 기반으로 사용자가 원하는 서비스를 정확하게 찾는 것은 어렵다. 본 논문에서는 이러한 문제를 해결하기 위해서 시맨틱을 활용한 RESTful 웹서비스 검색 기법을 개발하였다. 이를 위해 우선 OpenAPI 통합 검색 시스템을 바탕으로 시맨틱을 활용하기 위한 시스템 구조를 구성하고, 시맨틱 검색을 위한 기술 형식을 모델링하였다. 이를 바탕으로 의미 마크업(태깅, 시맨틱 어노테이션)을 수행하여, 추출된 결과인 RDF 문서를 서비스 저장소에 저장하여 이를 바탕으로 검색을 수행한다. 온톨로지를 활용하여 입력받은 키워드를 확장하고, 이를 바탕으로 검색을 수행하여 사용자에게 유사도 기반의 키워드 검색 기법에서의 검색 결과보다 확장 / 정제된 검색 결과를 제공한다.

키워드 : OpenAPI, RESTful 웹서비스, 시맨틱, 검색 기법

Abstract Recently with advent of Web 2.0, RESTful Web Services are becoming increasing trend to emphasize Web as platform. There are already many services and the number of service increases in very fast pace. So it is difficult to find the service what we want by keyword based search. To solve this problem, we developed the search method using semantic technologies about RESTful Web Services. For that, first we define the system structure and model the description format based on the integrated search system for OpenAPIs, and then we add Semantic Markup (tagging, semantic annotation) on

the HTML description pages. Next we extract RDF document from them and store it in service repository. Based on the keywords that are extended by means of ontology, the developed system provides more purified and extended results than similarity-based keyword searching system..

Keywords : OpenAPI, RESTful Web Services, Semantic, Search method

1. 서론

서비스 지향 아키텍처(SOA, Service-Oriented Architecture)는 개방형 표준을 따라 자치적인 서비스가 메시지 기반으로 서로 커뮤니케이션하도록 소프트웨어 시스템을 설계하는 방식을 의미한다[1]. 이러한 SOA의 개념을 가장 잘 구현해낸 웹서비스(Web Services)는 서로 다른 그룹의 사람들이, 서로 다른 장소에서, 서로 다른 시간에, 서로 다른 플랫폼 상에서의 애플리케이션 통합을 가능하게 한다.

2003년에 웹 2.0이라는 웹의 새로운 트렌드가 정의되고 사용되면서, 기존에 서비스 제공자가 제공하는 서비스만 사용하던 시대에서 사용자가 화면에 어떠한 데이터가 어떠한 형식으로 출력될 것인지 정하여 변경할 수 있고 동시에 서비스도 추가할 수 있는 시대로 변화였다. 따라서 사용자는 서비스의 작성과 배포가 용이한 방식을 찾게 되었고 이에 따라 OpenAPI가 발전하게 되었다[2][3].

OpenAPI(Application Programming Interface)는 웹 2.0의 근본 개념으로 서비스, 정보, 데이터 등을 언제, 어디서, 누구나 쉽게 이용할 수 있도록 개방시켜 놓은 API를 의미한다. OpenAPI는 웹서비스에서 사용되는 SOAP의 통신방식 뿐만 아니라 REST(Representational State Transfer) 통신방식을 사용하여 사용자들에게 편리함을 제공한다. 이를 RESTful 웹서비스라고 한다[4].

현재 RESTful 웹서비스는 키워드를 기반으로 하는 검색만 제공한다. 대표적인 예는 'Programmableweb'(http://www.programmableweb.com/)이다. 키워드를 기반한 검색은 사용자가 입력한 키워드를 포함하고 있는 정보에 대한 검색결과를 제공한다. 이는 양분화된 검색 결과만을 제공해주는 한계점이 있다.

이러한 한계점을 개선하기 위해 유사도를 기반한 순위화된 검색결과를 제공하는 방식이 연구되었다. 이를 통해 사용자는 자신이 찾고자 하는 서비스에 대해 유사도를 기반으로 하여 순위화된 결과를 제공받을 수 있게 되었지만 여전히 키워드 기반 검색의 한계는 극복하지 못하였다.

키워드 기반 검색의 한계를 극복하기 위해 본 논문에서는 유사도를 기반한 키워드 검색 기법을 기반으로 시맨틱 기술을 활용한 RESTful 웹서비스의 검색 기법을 개발하였다. 이를 통해 사용자는 기존의 검색 기법보다 정제/확장된 서비스 검색결과를 제공받을 수 있다.

이에 따른 본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 OpenAPI 통합 검색 시스템 및 관련 시스템

* 본 연구는 건설교통부 첨단도시기술개발사업 - 지능형국토정보기술혁신 사업과제의 연구비지원(07국토정보C05)에 의해 수행되었습니다.

* 충남대학교 컴퓨터공학과 박사과정, junii@cnu.ac.kr

** 충남대학교 컴퓨터공학과 석사과정, yunjeong@cnu.ac.kr

*** 충남대학교 컴퓨터공학과 교수, kclee@cnu.ac.kr(교신저자)

논문접수 : 2010.02.23

수정일 : 1차 2010.03.11 / 2차 2010.03.22

심사완료 : 2010.03.24

에 대해 분석하고, 이를 바탕으로 3장에서는 시스템 구조를 정의하고 검색을 위한 기술 형식 모델링에 대해 설명한다. 4장에서는 HTML 의미 마크업에 대해 기술하며, 5장에서는 시맨틱 검색 기법에 대해 기술하고, 마지막으로 6장에서는 결론과 향후 연구에 대해 설명한다.

2. 관련 연구

2.1 키워드 기반 OpenAPI 통합 검색 시스템

키워드 질의를 이용한 OpenAPI 통합 검색 시스템은 (이하 OpenAPI 통합 검색 시스템) 키워드 질의를 이용하여 웹서비스 및 OpenAPI에 대하여 유사도 연산을 통해 사용자에게 순위화된 결과를 보여주는 시스템이다[5].

OpenAPI 통합 검색 시스템은 XML 형태의 통합 서비스 정보모델(ISIM, Integrated Service Information Model)[그림 1]을 정의하여 다양한 프로토콜의 OpenAPI를 수용하였으며, 관계형 데이터베이스를 이용하여 구현되었다.

서비스 저장소에 ISIM을 파싱하여 테이블과 이들에 걸쳐지는 인덱스를 구성하여 저장하고 SQL 연산을 통해 검색을 수행한다.

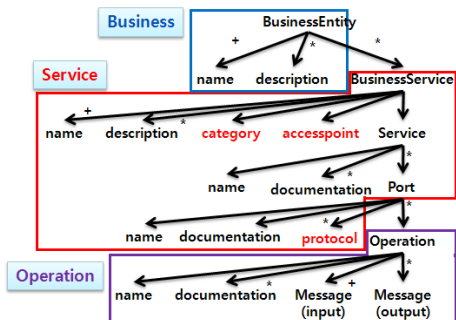


그림 1. 통합 서비스 정보 모델

OpenAPI 통합 검색 시스템의 사용자는 키워드 질의를 통해 순위화 된 서비스 검색 결과를 제공받는다[그림 2]. 이는 OpenAPI에 대한 통합검색을 제공하며 또한 시스템은 기존 관계형 데이터베이스의 기능을 이용하여 구현됨으로써 의미적인 서비스 발견에 비해 구현이 용이하고, 데이터베이스의 여러 장점을 취할 수 있다.

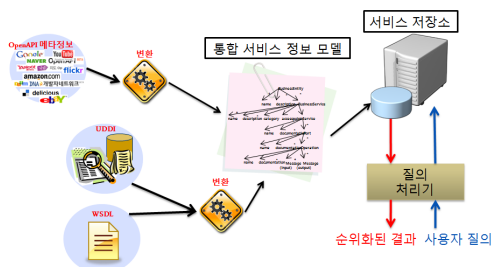


그림 2. OpenAPI 검색 시스템의 시스템 구조

2.2 웹서비스 및 OpenAPI 검색에 관련된 시맨틱 기술

SAWSDL(Semantic Annotations for WSDL and XML Schema)[6]에서는 WSDL의 문서에 시맨틱 어노테이션을 추가하는 방법을 정의한 문서이다. SAWSDL은 온톨로지로부터 XML 스키마로 매핑하는 정보를 구체화하는 어노테이션 매커니즘을 사용한다. 이러한 매핑은 서비스의 호출, 특히 미디어이션을 수행할 경우 사용된다.

SA-REST[7]는 웹 2.0 환경에서 일반 사용자들이 경량화된 서비스를 작성하기 위해 RESTful 웹서비스를 개발하기 위한 것이다. 하지만 매쉬업을 위한 도구들이 표준화된 형식을 사용하지 않기 때문에 오퍼레이션의 미디어이션이 어렵게 되었다.

이러한 문제점을 해결하기 위해 RESTful 웹서비스에 시맨틱 정보를 활용해야 한다. 이를 Semantic annotations to a RESTful Web Services(SA-REST)라고 명명하였다. SA-REST는 WSDL에 시맨틱 어노테이션을 추가한 SAWSDL을 기반으로 구성되었다.

2.3 기존 연구와의 차이점

현재 OpenAPI를 대상으로 시맨틱 검색은 제공되지 않는다. OpenAPI는 ProgrammableWeb에서 키워드 기반의 검색을 제공하고 있으며, 키워드 기반 검색의 한계점을 해결하기 위해 정보검색의 TF*IEF 기법을 활용한 검색을 OpenAPI 통합 검색 시스템[5]에서 제공하고 있다. 하지만 여전히 키워드 기반의 검색으로는 사용자가 찾고자 하는 유사 서비스 검색 더 나아가 시맨틱을 위해 공통된 오퍼레이션을 찾는게 불가능하다.

웹서비스의 시맨틱 어노테이션을 위한 SAWSDL과 OpenAPI에 시맨틱 기술을 활용한 SA-REST를 통해 웹서비스와 OpenAPI에 시맨틱 정보를 활용하는 방안을 살펴보았지만 아직 구체적인 시스템이 존재하지 않는다.

따라서 본 논문의 목표는 OpenAPI 통합 검색 시스템을 확장하여 시맨틱 기술을 활용한 검색 기법의 개발하는 것이다. 이를 통해 사용자는 정제/확장된 검색 결과를 제공받을 수 있으며 또한 시맨틱 매쉬업을 위한 검색에도 활용될 수 있다.

3. 시스템 구조 및 기술 형식 모델링

본 장에서는 시맨틱 기술을 활용하기 위해 설계한 시스템 아키텍처에 대해 설명하고, 이를 바탕으로 정의한 기술 형식 모델링에 대해 기술한다.

3.1 시스템 구조

시맨틱 기술을 활용한 RESTful 웹서비스의 검색을 위한 시스템 구조는 [그림 3]과 같다. OpenAPI 통합 검색 시스템과는 달리 서비스에 대한 설명정보는 HTML 페이지에 작성되며, 사용자는 이를 바탕으로 설명정보 페이지의 변화 없이 필요한 정보를 얻기 위한 태그 및 시맨틱 어노테이션을 추가한다. 이를 의미 마크업이라고 한다.

의미 마크업이 끝난 문서는 추출기를 이용하여 시맨틱 정보가 정의된 RDF 문서를 얻을 수 있으며, 이는 온톨로지에 반영될 뿐만 아니라 서비스 저장소에 저장되어 검색에 활용된다.

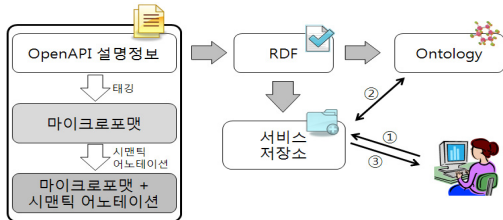


그림 3. 시스템 구조

사용자는 서비스 저장소에 질의를 하게 되고(①), 서비스 저장소는 온톨로지를 활용하여 사용자가 질의한 검색어를 확장하고(②) 이를 바탕으로 검색을 수행하여 결과를 사용자에게 반환한다(③).

3.2 기술 형식 모델링

본 논문에서 검색을 위해 사용되는 서비스 저장소에는 OpenAPI 통합검색 기법에서 개발된 통합 서비스 정보 모델(ISIM)을 기반으로 저장된다. 이는 다양한 프로토콜로 정의되는 모든 OpenAPI에 적용될 수 있는 정보모델로 연구된 것이다.

4. 의미 마크업

의미 마크업은 RESTful 웹서비스의 설명정보에 시맨틱 기술을 적용하는 것으로, 태깅과 시맨틱 어노테이션으로 구성된다.

4.1 태깅

HTML 페이지에서 RDF 문서의 추출을 위해 OpenAPI 설명정보에 RDFa 정보가 저장되는 것을 태깅이라고 한다. 이를 위해 우선 추출될 RDF 스키마의 구성이 선행되어야 한다. ISIM 문서의 추출을 위한 RDF 스키마 문서의 일부는 표 1과 같다. 정의된 스키마에 맞게 HTML 페이지에 태깅을 수행한다.

표 1. RDF 스키마

```
# service model classes and properties
us:BusinessEntity
  rdf:type rdfs:Class .
us:hasBusinessEntityName
  rdf:type rdfs:Property ;
  rdfs:domainus:BusinessEntity ;
  rdfs:range xsd:string .
us:hasBusinessEntityDescription
  rdf:type rdfs:Property ;
  rdfs:domainus:BusinessEntity ;
  rdfs:range xsd:string .
us:BusinessService
  rdf:type rdfs:Class .
us:hasBusinessServiceName
  rdf:type rdfs:Property ;
  rdfs:domainus:BusinessService ;
  rdfs:range xsd:string .
....
```

4.2 시맨틱 어노테이션

시맨틱 어노테이션은 온톨로지를 통한 키워드 확장을 위해 오퍼레이션에 관련된 시맨틱 정보를 기술하는 것이다. SAWSDL에서 정의한 시맨틱 어노테이션을 위해 오퍼레이션의 입출력 정보에 모델 참조(Model Reference)를 통한 어노테이션 방법을 이용하여 설명페이지에 어노테이션을 추가한다. 태깅과 시맨틱 어노테이션이 추가된 HTML 설명정보 페이지는 표 2와 같다.

표 2. 시맨틱 어노테이션

```
<tr class="even">
  <td>
    <span property="us:hasOperationInput">appid</span>
  </td>
  <td>string (required)</td>
  <td>The application ID. See <a href="/faq/index.html#appid">Application IDs</a> for more information.</td>
</tr>
<tr class="odd">
  <td>
    <span property="us:hasOperationInput">
      <a rel="sawSDL:modelreference" href="http://dbslab.cnu.ac.kr/onto/geo.owl#address">①</a>
      <a rel="sawSDL:liftingSchemamapping" href="http://dbslab.cnu.ac.kr/lifting.xslt">②</a>
      <a rel="sawSDL:loweringSchemamapping" href="http://dbslab.cnu.ac.kr/lowering.xslt">③</a>
    </span></td>
  <td>string</td>
  <td>Street name. The number is optional.</td>
</tr>
```

HTML 설명정보에는 오퍼레이션의 입/출력 값에 대해 참조하는 온톨로지에 대한 설명(①), 리프팅 스키마(②), 로어링 스키마(③)를 정의하였다.

리프팅 스키마는 서비스 요청에 필요하여 입력받은 인자에 대해, RDF로 정의된 그라운드링 스키마(Grounding Schema)로 변환하는 스키마이다.

로어링 스키마는 그라운드링 스키마 형식으로 기술된 문서를 특정 서비스를 요청하기 위해 적합한 형식으로 변환하는 역할을 한다.

4.3 RDF 추출

RDFa로 태깅된 HTML 문서는 RDFa 추출기(RDFa Extractor, Distiller)를 이용하여 RDF 문서를 얻을 수 있다. RDFa 추출기를 이용하여 본 논문에서는 OpenAPI 설명정보에 태깅 및 시맨틱 어노테이션을 저장한 페이지를 대상으로 RDF를 추출할 수 있다. 표 3은 추출된 RDF 문서이다.

5. 시맨틱 기술을 활용한 검색 기법 개발

시맨틱 기술을 활용함으로써 입력받은 키워드를 바탕으로 도메인/일반 온톨로지를 통해 키워드를 확장함으로써 정제/확장된 검색 결과를 제공할 수 있다.

5.1 도메인 온톨로지의 활용

도메인 온톨로지는 특정한 분야에서 이용되는 온톨로

표 3. RDF 문서

```
<us:BusinessEntity>
  <us:hasBusinessEntityName>
    yahoo</us:hasBusinessEntityName>
  <us:hasBusinessService>
    <us:BusinessService>
      <us:hasCategory>mapping</us:hasCategory>
      <us:hasBusinessServiceName>Yahoo! Maps Web
        Services - Geocoding
      API</us:hasBusinessServiceName>
      <us:hasAccessPoint>
        http://developer.yahoo.com/
      </us:hasAccessPoint>
      <us:hasPort>
        <us:Port> ...
      </us:Port>
      </us:hasPort>
    </us:BusinessService>
  </us:hasBusinessService>
</us:BusinessEntity>
</rdf:RDF>
```

지로 주로 해당분야의 전문가들이 사용되는 용어와 지식을 바탕으로 작성되어야 한다. 본 연구에서는 지리정보시스템(GIS)와 관련된 온톨로지를 대상으로 테스트를 수행하였다. 도메인 온톨로지는 오퍼레이션 입출력 인자에 추가된 시맨틱 어노테이션을 대상으로 활용된다.

5.2 일반 온톨로지의 활용

일반 온톨로지는 주위의 일반적 개념들을 대상으로 구축한 온톨로지이다. 일반 온톨로지의 대표적인 것으로 워드넷이 있다. 일반 온톨로지를 활용한 시맨틱 검색 기법은 시맨틱 어노테이션이 추가되지 않은 이름, 설명정보를 대상으로 활용된다.

5.3 유사도 기반 키워드 검색 기법과의 비교

제안된 검색 기법이 기존 키워드를 대상으로 유사도 기반으로 정보를 제공해 주던 기법과의 차이점은 다음과 같다. 두 시스템은 프로토타입 시스템을 통해 유사도를 측정하였다.

우선 확장된 서비스의 결과를 얻을 수 있다. 의미 마크업에서 추가한 시맨틱 어노테이션에 대해서는 도메인 온톨로지를 활용하여 키워드를 확장하였으며, 시맨틱 어노테이션이 추가되지 않은 이름과 설명정보는 일반 온톨로지를 활용하여 관련 키워드를 확장하였다. 그림 4-(a)는 확장된 검색에서의 'zip'을 검색했을 경우, 기존 검색에서는 3건의 검색결과를 보여준 반면 제안된 검색 기법을 통해 'zip, zipcode, citycode'로 키워드를 확장하여 관련

된 2건의 검색 결과가 추가된 것을 확인할 수 있다.

다음으로는 정제된 서비스의 결과를 얻을 수 있다. 기존의 단순히 입력받은 키워드를 통해 검색하는 것을 벗어나서 유사어에 대한 검색을 수행하여 사용자에게 찾고자 하는 서비스와 더욱 근접한 결과를 얻게 할 수 있다. 그림 4-(b)는 'address'를 검색하였을 경우의 검색 결과로 기존의 검색에서의 각 서비스의 유사도가 제안된 검색 기법에서 변화된 것을 확인할 수 있다. 이는 제안된 시스템에서 도메인 온톨로지를 활용하여 키워드를 'address, location, state'로 확장하여 검색 결과가 정제되었기 때문이다.

6. 결론 및 향후 연구

본 논문에서는 시맨틱 기술을 활용하여 RESTful 웹서비스에 대한 검색 기법을 개발하였다. 이를 위해 본 논문에서는 다음과 같은 연구를 수행하였다.

먼저 관련 연구를 분석하였다. 관련연구는 OpenAPI 통합 검색 시스템과 OpenAPI 및 웹서비스 검색에 관련된 시맨틱 기술인 SAWSDL과 SA-REST를 분석하였다.

관련 기술들을 활용하여 OpenAPI 통합 검색 시스템을 바탕으로 시스템 구조를 정의하고, 시맨틱을 활용하기 위한 OpenAPI 기술형식을 모델링 하였다. OpenAPI 통합 검색 시스템의 통합 서비스 정보 모델 형식을 유지하고, 시맨틱 정보를 위해 의미 마크업을 정의하였다.

정의된 모델링을 바탕으로 시맨틱 기술을 활용한 검색 기법을 설명하였다. 시맨틱 기술을 적용하기 위해 도메인 온톨로지와 일반 온톨로지를 활용하였다. 입력받은 키워드에 대해 오퍼레이션의 파라미터인 경우 추가된 시맨틱 어노테이션을 바탕으로 도메인 온톨로지를 이용하여 키워드를 확장하였으며, 이름과 설명정보인 경우에는 일반 온톨로지를 활용하여 키워드를 확장하였다.

확장된 키워드를 바탕으로 검색을 수행함으로써 기존 유사도 기반의 키워드 검색 기법에서의 검색 결과보다 더 사용자가 원하는 서비스를 검색 할 수 있게 되었으며, 또한 유사 서비스도 검색 결과로 보여줄 수 있게 되었다.

향후 연구 방향으로는 우선 개발된 시스템의 성능을 평가할 수 있는 프로토타입 시스템과 테스트베드, 테스트 데이터를 정의하여야 한다. 또한 연구된 시맨틱 검색 기법을 다양한 프로토콜과 많은 양의 OpenAPI에 적용할

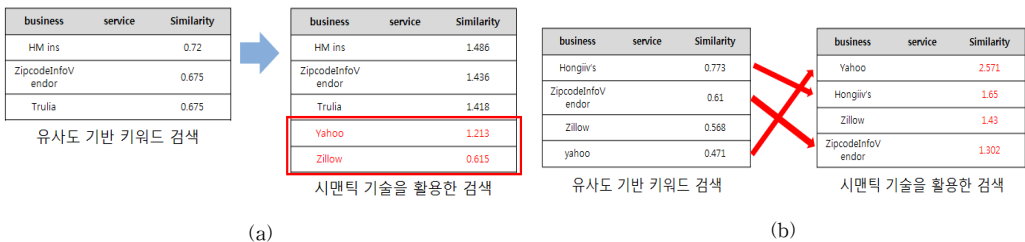


그림 4. 시맨틱 검색을 통한 검색 결과 향상

수 있도록 하는 온톨로지의 개발이 이루어 져야 한다. 또한 태깅, 시맨틱 어노테이션 과정을 사용자가 쉽게 접근할 수 있는 방법론이 제시되어야 한다.

참 고 문 헌

- [1] Lui, X., Hui, Y., and Liang, H., "Towards Service Composition Based on Mashup," In Proc. of IEEE International Conference on Service Computing, 2007, pp. 332-339.
- [2] O'Reilly, T., What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, 2005
- [3] 김재철, 이규철, u-GIS 국토정보 제공 시스템, 한국공간정보시스템학회, 제11권 제1호, 2009, pp.1-8
- [4] Richardson, L., and Ruby, S., RESTful Web Services, O'Reilly, 2008
- [5] 천동석, 차승준, 김경옥, 이규철, "u-GIS 환경에서 OpenAPI와 매쉬업 가능 서비스에 대한 통합 검색 기법 개발" 한국공간정보시스템학회, 제11권 제1호, 2009, pp.25-34
- [6] Farrell, J., and Lausen, H., Semantic Annotations for WSDL and XML Schema, W3C Recommendation
- [7] Lathem, J., Gomadam, K., Sheth, A., "SAREST and (S)mashups: Adding Semantics to RESTful Services," In Proc. IEEE Int'l Conf. Semantic Computing, 2007, pp. 469 - 476