

시맨틱 기술을 활용한 OpenAPI 조합 가능 서비스 검색에 관한 연구⁺

최영호, 차승준, 이규철^{**}
충남대학교 컴퓨터공학과

e-mail: {iloot, junii, kclee}@cnu.ac.kr

A Survey of Mashup Capable Services' Retrieval Methods for OpenAPI using Semantic Technology

Young-Ho Choi, Seung-Jun Cha, Kyu-Chul Lee
Dept of Computer Engineering, Chungnam National University

요 약

본 논문은 시맨틱 기술을 활용한 OpenAPI 조합가능 서비스 검색 기법 개발을 위해 관련 연구들을 분석하여 적용방안을 도출하였다. OpenAPI 조합가능 서비스 검색이란 선택된 서비스의 출력과 매칭이 되는 입력을 가진 서비스나 선택된 서비스의 입력과 매칭되는 출력을 가진 서비스를 찾는 것이다. 본 논문에서는 기존의 키워드 기반의 조합 가능 서비스 검색 기법의 한계를 시맨틱 기술을 활용하여 해결하기 위한 관련연구로 SAWSDL-MX2와 LOG4SWS.KOM에 관한 논문들을 분석했다. SAWSDL-MX2에서는 세가지 매칭 기법과 이에 따른 유사도 분석 기법을 제시하였고, LOG4SWS.KOM에서는 두가지 매칭 기법과 이에 따른 유사도 분석 방법을 제시하였다. 관련 연구들에서 분석된 내용을 바탕으로 OpenAPI 조합 가능 서비스에 대한 매칭 기법의 정의, 유사도 분석 기법의 정의가 추후 진행되어야 한다.

1. 서론

매쉬업 기술이란 서로 다른 두 OpenAPI 서비스들을 조합하여 새로운 서비스를 창출해내는 기술[1]로 웹 2.0 환경에서 다양하게 제공되는 OpenAPI 서비스들을 기반으로 활성화되었다. 서비스들간의 매쉬업 가능 여부는 서비스들이 제공하는 오퍼레이션의 입력/출력 값을 기반으로 비교하여 확인할 수 있다.

조합 가능 서비스 검색은 선택된 서비스의 출력과 매칭되는 입력을 가진 서비스나, 선택된 서비스의 입력과 매칭되는 출력을 가진 서비스를 찾는 것이다. 이러한 검색 기법은 매쉬업을 위해 반드시 필요하지만, 일반적인 검색 시스템(예: Programmableweb[2])은 이와 같은 검색을 제공하지 않는다.

이런 문제점을 해결하기 위해 [3][4]에서는 OpenAPI에 대한 시맨틱을 활용한 통합 검색 및 키워드 기반의 조합 가능 서비스 검색 기법을 제안하였다. 그리고 조합가능 서비스 검색을 위해 선택된 서비스의 오퍼레이션 입력/출력 값과 매칭되는 서비스의 출력/입력값을 가진 서비스들을 키워드 기반으로 검색하여 각각 이전 조합 가능 서비스(front mashup capable service), 이후 조합 가능 서비스

(rear mashup capable service)로 정의하였다. 또한 추가 검색어를 통해 검색된 서비스들에 대해 순위화 된 검색 결과를 제공했다.

키워드 매칭은 같은 의미로 활용되지만 키워드가 정확히 일치하지 않는 경우 검색되지 않은 한계점을 가진다. 이러한 키워드 기반의 조합 가능 서비스 검색의 한계를 극복하기 위해 시맨틱 기술의 활용이 필요하다.

시맨틱 기술이란 의미정보를 기계가 이해할 수 있도록 표현하여 이를 기반으로 검색을 수행할 수 있도록 해주는 기술[4]이다. 시맨틱 기술을 활용하게 될 경우 정의된 의미정보를 바탕으로 키워드 매칭만을 통한 검색 기법을 보완하여 사용자에게 정제된 검색 결과를 제공한다.

따라서 본 논문에서는 [3][4]을 확장하여 조합 가능 서비스 검색에 시맨틱 기술을 적용하기 위해 관련된 연구들에 대한 분석 및 적용 가능성에 대해 알아보았다.

[4]의 논문에서 서비스 저장 및 검색을 위한 모델로 SA-OpenAPI를 제안하였고, 서비스는 태깅 및 시맨틱 어노테이션을 통해 SA-OpenAPI로 저장된다. 따라서 본 논문에서는 SAWSDL로 활용되는 조합과 파라미터 매칭에 대한 논문을 대상으로 분석하였다.

파라미터 매칭과 관련된 연구의 분석을 통해 사용자가 선택한 서비스의 입력값과 검색할 서비스의 출력값 사이의 매칭에 대한 기준을 정의할 수 있고, SAWSDL을 이용한 서비스들 사이의 매칭방식을 참조할 수 있다.

⁺ 본 연구는 국토해양부 첨단 도시기술개발사업 - 지능형국토정보기술혁신 사업과제의 연구비지원(06국토정보 B01)에 의해 수행되었습니다.

^{**} 교신저자

이에 따른 본 논문은 다음과 같은 구성을 가지고 있다. 2, 3장에서는 관련연구로 분석한 SAWSDL-MX2와 LOG4SWS.KOM에 대해 정리했다. 4장에서는 관련 논문 분석 결과와 이로부터 얻어진 연구 방향에 대해 설명하며, 5장에서는 결론 및 향후 연구내용에 대해 설명한다.

2. SAWSDL-MX2

SAWSDL-MX2 [5][6][7]는 SAWSDL로 제공되는 서비스들을 위한 시맨틱 웹서비스 매치메이커이다. 기존의 온톨로지를 활용한 시맨틱 매치메이킹에 관련된 연구인 OWLS-MX와 WSMO-MX의 방식을 혼합시켜 새로운 매칭방식을 정의한 SAWSDL-MX는 2008년에 개발된 이후로 계속적인 개선과 테스트를 진행했다. 그 후 2009년에 SAWSDL-MX2를 발표했고, 2010년까지 4개의 논문이 발표되어 현재도 연구가 진행되고 있다.

해당 연구에서는 파라미터 매칭을 위해 우선 세 가지의 매칭 기준을 정의하여 유사성을 체크하고, 이를 바탕으로 유사도를 분석하여 분석 결과를 제공하였다.

2.1 매칭기준 정의

서비스 사이의 매칭에 대한 기준을 정의하기 위해 분류한 세 가지 매칭 기준은 다음과 같다. 첫 번째 기준은 로직 기반 오퍼레이션 매칭으로 요청 서비스와 제공 서비스의 오퍼레이션이 가진 입력 값들과 출력 값들 사이의 1:1 매칭을 기준으로 참조하고 있는 온톨로지 사이의 매칭을 통해 두 서비스의 유사성을 판별한다. 매칭에 사용되는 알고리즘은 선행연구였던 OWLS-MX[8]에서 사용된 'Exact', 'Plugin', 'Subsumes', 'Subsumed-by', 'Fail'의 다섯 가지 매칭 필터를 기준으로 각 서비스가 가지고 있는 파라미터가 모델 레퍼런스를 이용하여 참조하는 온톨로지 사이의 관계를 정의했다([수식 1]).

두 번째 기준은 텍스트 기반 오퍼레이션 매칭이다. 이는 첫 번째 기준인 로직 기반의 매칭 후에 활용되는 것으로, 요청 서비스에서 선택된 토큰과 제공 서비스 사이의 유사성을 측정하는 방식으로 이루어진다. 우선 서비스 문서가 가진 오퍼레이션의 논리적 설명정보에서 선택된 파라미터의 용어에 따라 파라미터를 분배한다. 나뉜 파라미터들은 입력값 별로 TF-IDF 가중치 벡터 문법에 따라 참조된 온톨로지 개념에 대한 키워드벡터 계산을 수행하고, 출력값에 대해서도 동일한 과정을 수행한다.

세 번째 기준은 WSDL 문서 구조에 따른 매칭으로 WSDL 분석기를 이용하여 SAWSDL 설명정보를 WSDL 파일로 인식하여 두 서비스를 비교하는 방식이다. 인식된 SAWSDL 파일들은 분류된 트리 형태로 변환하여 최상위 노드부터 차례로 비교한다. 비교 대상은 서비스에서 제공되는 요소 이름, 데이터 타입, 구조적 속성, 데이터 인스턴스의 성격, 사전 및 용어집(dictionaries and thesauri)의 배경지식 형태를 유사성 계산의 대상으로 이용하여 WSDL 분석기를 통한 계산 결과를 유사성 판별에 이용한

다.

Exact

$$(\exists \text{injective assignment } M_{in} : \forall m \in M_{in} : m1 \in in(O_1) \wedge m2 \in in(O_2) \wedge m1 \equiv m2) \wedge (\exists \text{injective assignment } M_{out} : \forall m \in M_{out} : m1 \in out(O_1) \wedge m2 \in out(O_2) \wedge m1 \equiv m2)$$

Plug-in

$$(\exists \text{injective assignment } M_{in} : \forall m \in M_{in} : m1 \in in(O_1) \wedge m2 \in in(O_2) \wedge m1 \sqsupseteq m2) \wedge (\exists \text{injective assignment } M_{out} : \forall m \in M_{out} : m1 \in out(O_1) \wedge m2 \in out(O_2) \wedge m2 \in lsc(m1))$$

Subsumes

$$(\exists \text{injective assignment } M_{in} : \forall m \in M_{in} : m1 \in in(O_1) \wedge m2 \in in(O_2) \wedge m1 \sqsupseteq m2) \wedge (\exists \text{injective assignment } M_{out} : \forall m \in M_{out} : m1 \in out(O_1) \wedge m2 \in out(O_2) \wedge m1 \sqsupseteq m2)$$

Subsumed-by

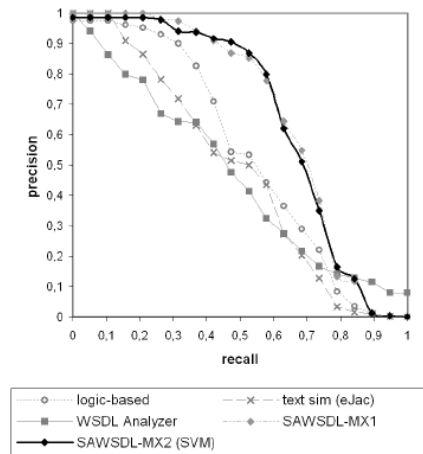
$$(\exists \text{injective assignment } M_{in} : \forall m \in M_{in} : m1 \in in(O_1) \wedge m2 \in in(O_2) \wedge m1 \sqsupseteq m2) \wedge (\exists \text{injective assignment } M_{out} : \forall m \in M_{out} : m1 \in out(O_1) \wedge m2 \in out(O_2) \wedge m2 \in lgc(m1))$$

[수식 1] 각 필터별 매칭 기준

2.2 유사도 분석

이렇게 세 가지의 기준을 통해 계산된 유사도 계산 결과를 응집하여 랭킹화하기 위해 SVM[9]을 이용했다. SVM에서 이용되는 기능(feature) 벡터로 로직 기반 매칭의 5개의 정도(degree)와 텍스트 기반 매칭의 유사성 계산 결과, WSDL구조에 따른 매칭결과, 이렇게 7개의 벡터를 기준삼아 두 서비스의 유사성 결과를 응집했다.

그리고 성능 평가를 위해 SAWSDL 테스트 데이터 중 상용되고 있는 SAWSDL-TC1을 이용하여 정확률과 재현률 계산을 통해 (그림 1)와 같은 테스트 결과를 얻어낼 수 있었다.



(그림 1) 정확률/재현률 곡선

평균 질의에 대한 응답 시간과 평균 정확률에 대해 다른 도구들과 비교한 결과는 <표 1>과 같다.

<표 1> 매치메이커 사이의 기능 비교표

	SAWSDL-MX1	SAWSDL-MX2	COM4SWS	URBE	SAWSDL-iMatcher3
AP	0.7	0.68	0.68	0.73	0.63
AQRT	3.1s	7.9s	6.14s	20s	0.75s

3. LOG4SWS.KOM

LOG4SWS.KOM[10]은 “Logic-based Matchmaking for Semantic Web Services”의 약자로 WSDL 2.0과 SAWSDL 기반의 설명정보를 가진 서비스들에 대한 매치 메이커이다. LOG4SWS.KOM에서는 두 가지 매칭 기준을 토대로 서비스 사이의 매칭을 판별한다.

3.1 매칭기준 정의

첫 번째 기준은 오퍼레이션 중심의 매칭으로 두 서비스가 가진 인터페이스, 오퍼레이션, 입력값, 출력값 이렇게 4개의 컴포넌트 별로 각각 유사성을 계산하여 [수식 2]와 같은 유사성 응집수식을 통해 가중치 별로 유사성 계산 결과를 응집한다.

$$w_{iface} + w_{op} + w_{in} + w_{out} = 1$$

$$sim_{agg}(a, b) = sim_{iface}(a, b) * w_{iface} + sim_{op}(a, b) * w_{op} + sim_{in}(a, b) * w_{in} + sim_{out}(a, b) * w_{out}$$

[수식 2] 유사성 응집 수식

응집된 결과는 요청 서비스의 오퍼레이션 집합 I와 제공 서비스의 오퍼레이션 집합 J를 이용하여 [수식 3]과 같은 수식을 이용하여 서비스 사이의 유사성 계산결과로 변환된다.

$$sim_{serv}(R, O) = \frac{1}{|I|} * \sum_{i \in I, j \in J} x_{ij} * sim_{agg}(i, j)$$

[수식 3] 서비스 유사성 계산 방식

각 컴포넌트 별 유사성 계산은 [11]에서 제안한 이분 그래프(bipartite graph) 계산 방식을 이용했다. 요청과 제공의 컴포넌트 집합을 그래프의 두 부분으로 분할된 노드들의 집합으로 인식하고 두 부분의 노드들 별로 가중치를 두고 1:1로 연결하여 이분 그래프 매칭 알고리즘을 통해 유사성을 계산했다.

두 번째 기준은 유사성 배정방식으로 두 서비스가 참조하고 있는 온톨로지 개념 사이의 포함관계를 이용한 유사성 계산 방식이다. 기본적인 문법은 [12]의 매칭 알고리즘을 따르면서 시스템의 필요에 따라 변형시킨 방식으로 요청 서비스가 참조하는 온톨로지 A와 제공 서비스가 참조하는 온톨로지 B 사이의 매칭 정도(Degree of Match)을 [수식 4]과 같이 정의했다.

$$DoM(A, B) = \begin{cases} exact & if A \equiv B \\ super & if A \sqsubseteq B \\ sub & if A \sqsupseteq B \\ fail & else \end{cases}$$

[수식 4] 매칭 정도

매칭 순위는 exact > super > sub > fail로 0~1 사이의 수치를 이용하여 각 유사성 별 수치를 적용했다. 또한 컴포넌트 별로 매칭 정도에 따른 유사성의 측정을 위해 [수식 5]과 같이 각 컴포넌트 별로 유사성에 따른 수치를 0~1 사이의 숫자로 정의했다.

$$d_{L,D} \in [0; 1] \quad \forall L \in \{iface, op, in, out\}, D \in \{exact, super, sub, fail\}$$

[수식 5] 개별 매칭 레벨과 DoM에 대한 집합

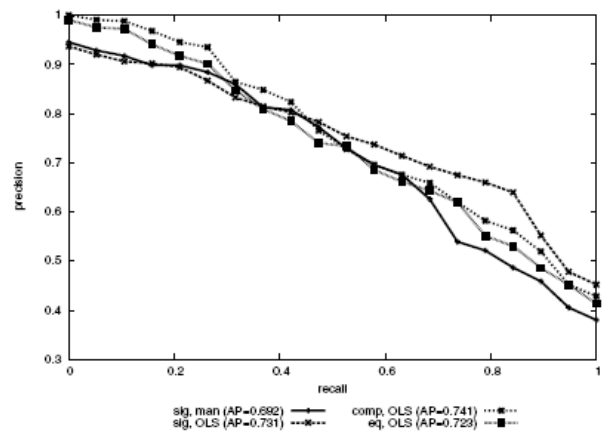
또한 요청 서비스가 참조하는 온톨로지 개념을 A, 제공 서비스가 참조하는 온톨로지 개념을 B라 했을 때 둘 사이의 최단 거리를 $PL(A, B)$ 라 정의하고 이를 이용하여 유사도인 $cs(A, B)$ 를 정의했다. 이렇게 결정된 요소들을 이용하여 [수식 6]과 같은 컴포넌트 유사성 규칙을 정의했다.

$$cs(A, B) = \begin{cases} d_{L,exact} & if A \equiv B \\ d_{L,super}/PL(A, B) & if A \sqsubseteq B \\ d_{L,sub}/PL(A, B) & if A \sqsupseteq B \\ d_{L,fail} & else \end{cases}$$

[수식 6] 컴포넌트 유사성

3.2 유사도 측정

앞에서 정의한 컴포넌트 별로 다른 매칭 정도를 이용하여 최적의 수치적인 관계성을 계산하기 위해서 본 논문에서는 OLS 측정기(Ordinary Least Squares Estimator)를 이용했다. 성능평가를 위한 테스트 데이터로 SAWSDL-MX2와 동일하게 SAWSDL-TC1을 이용했으며 정확률과 재현률 테스트 결과 (그림 2)와 같은 정확률, 재현률 결과를 얻어낼 수 있었다.



(그림 2) 정확률과 재현률을 통한 성능평가

3. 관련 논문 분석 결과

매칭 기법은 두 서비스 간의 입력-입력 또는 출력-출력에 해당하는 기법에 대한 연구이고 조합가능 서비스 검색 기법에서는 두 서비스의 입력-출력 또는 출력-입력의 비

교이기 때문에, 조합 가능 서비스의 매칭 기법 및 유사도 측정 방식을 매칭 기법의 것을 활용하여 적용할 수 있다.

SAWSDL-MX2와 LOG4SWS.KOM의 연구를 분석해 본 결과 모두 시맨틱을 이용한 파라미터 사이의 매칭을 위해서 매칭 기준을 정의하고 테스트 데이터를 통해 유사성을 측정하고 검증하는 과정을 거쳐 시스템의 우수성을 판단했다.

SAWSDL-MX2의 경우 OWLS-MX에서 정의한 매칭 정도(Degree of Match)를 참조하여 매칭 기준을 정의함으로써 서비스들이 가진 입/출력값들 사이의 매칭이 가능하도록 했고, LOG4SWS.KOM의 경우 [12]의 매칭 정도를 활용하여 서비스들이 가진 입/출력값 사이의 매칭을 통해 매칭되는 서비스를 판별했다. 그러므로 시맨틱 기술을 이용한 조합가능 서비스 검색기법 개발에서도 위와 같이 매칭 기법에 OpenAPI 서비스의 특징을 추가한 매칭 기법이 정의되어야 한다.

그리고 시맨틱 매칭을 위한 테스트 데이터로 사용된 SAWSDL-TC를 이용하여 정확률과 재현률을 통해 타 논문과의 비교해야 한다.

4. 결론

본 논문에서는 시맨틱 기술을 활용한 OpenAPI 조합가능 서비스 검색 기법의 개발을 위해 관련 연구들에 대해 분석한 내용을 정리해 보았다.

첫 번째로 분석한 SAWSDL-MX2에서는 파라미터 사이의 매칭 판별을 위해 각 서비스가 가진 파라미터가 참조하고 있는 온톨로지 사이의 관계성을 조사했다. 그리고 Exact, Plugin, Subsumes, Subsumed by, Fail의 다섯 가지 필터를 정의하여 온톨로지 사이의 관계성을 체크했다. 텍스트 기반 매칭에서는 TF-IDF를 이용하여 특정 서비스 입/출력값과 검색된 서비스의 논리적인 설명정보 사이의 텍스트 유사성을 통해 두 서비스의 유사성을 증명했다.

두 번째로 분석한 LOG4SWS에서는 [12]이 제안한 매칭 알고리즘을 이용하여 매칭정도를 새로 정의하고 인터페이스와 오퍼레이션, 입/출력 값 사이의 유사성을 통해 두 서비스의 유사성을 증명했다.

분석한 두 시스템 모두 테스트 데이터로 SAWSDL-TC를 이용하였고, 성능평가는 정확률/재현률을 통해 연구의 결과를 도출했다. 조합가능 서비스 검색 기법에 시맨틱 기법을 적용하기 위해서 관련 연구와 마찬가지로 매칭 기법의 정의가 선행되어야 하고 이를 바탕으로 유사도를 측정해야 한다.

참고문헌

- [1] 김중태, “새로움을 만드는 혼합 서비스의 물결”, Oracle Korea Magazine, 2007 Winter, pp 14-18
- [2] ProgrammableWeb, <http://www.programmableweb.com/>
- [3] Seung-Jun Cha, Kyu-Chul Lee, “Development of

Integrated Retrieval Methods for OpenAPIs and Mash-up Capable Services”, 12th IEEE International Conference on Computational Science and Engineering, 2009.8, pp374-381

- [4] 최윤정, 차승준, 이규철, “시맨틱 웹 기술을 활용한 OpenAPI 통합 검색 시스템 개발”, 한국정보과학회 2010년 12월 제26권 제3호, pp45-65
- [5] Matthias Klusch, Patrick Kapahnke, Ingo Zinnikus “SAWSDL-MX2: A Machine-learning Approach for Integrating Semantic Web Service Matchmaking Variants”, IEEE2009, ICWS.2009.7, pp335-342
- [6] Matthias Klusch, Patrick Kapahnke, Ingo Zinnikus “Hybrid Adaptive Web Service Selection with SAWSDL-MX and WSDL-Analyzer”, ESWC 2009, pp.550-564
- [7] Matthias Klusch, Patrick Kapahnke, Ingo Zinnikus “Adaptive Hybrid Semantic Selection of SAWSDL Services with SAWSDL-MX2”, IJSWIS, 2010.10, pp 1-26
- [8] M. Klusch, B. Fries, and K. Sycara. “Automated semantic web service discovery with owls-mx”, In Proceedings of 5th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), Hakodate, Japan. ACM Press, 2006.5 pp 915-922
- [9] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [10] Stefan Schulte, Ulrich Lampe, Julian Eckert, Ralf SteinMetz, “LOG4SWS.KOM: Self-Adapting Semantic Web Service Discovery for SAWSDL”, IEEE2010, SERVICES.2010.7, pp 511-518
- [11] U. Bellur and R. Kulkarni. “Improved Matchmaking Algorithm for Semantic Web Services Based on Bipartite Graph Matching”, In 2007 IEEE International Conference on Web Services (ICWS 2007), 2007.7, pp 86 - 93
- [12] M. Paolucci, T. Kawamura, T. R. Payne, and K. P. Sycara. “Semantic Matching of Web Services Capabilities”. In First International Semantic Web Conference (ISWS 2002), volume 2342 of LNCS, 2002.6, pp 333 - 347,