

OWL-S 기반 구조적인 온톨로지 매칭 방법에 관한 연구

Towards Structural Ontology Matching Approach based on OWL-S

고혜경^{1*}

Hye-Kyeong Ko^{1*}

요약

OWL-S는 서비스를 위한 웹 온톨로지 언어로서 기존의 웹 서비스와 시맨틱 웹 기술의 결합을 통한 웹서비스 지원 모델을 의미한다. 본 논문은 OWL-S 기반의 시맨틱 웹 서비스에서 검색 효율을 높여주기 위한 서비스 검색 방법 및 프레임워크를 제안한다. 제안된 프레임워크에서는 온톨로지 문서를 트리 구조 형태의 온톨로지 패턴으로 구조화하고 온톨로지 패턴으로부터 질의 패턴을 구조화한 뒤 캐쉬 데이터베이스에 각각 저장한다. 또한 사용자의 질의 요청 시 캐쉬 데이터베이스에 저장된 질의 패턴을 직접 검색하여 질의 요청을 빠르게 처리할 수 있는 검색 방법을 제안한다. 본 논문에서 제안한 방법은 전체 온톨로지 서비스를 검색하지 않고 미리 질의 결과로 저장된 질의 패턴을 검색하기 때문에 불필요한 검색 없이 사용자 질의에 대한 처리 속도를 현저하게 감소시킬 수 있다.

핵심어 : OWL-S, 온톨로지 매칭, 서비스 검색, 온톨로지 패턴, 캐쉬 데이터베이스

Abstract

OWL-S is a web ontology language for service and refers to the web service support model by combination of an existing web service and a semantic web technology. This paper proposes an OWL-S based service discovery approach and a framework in order to increase the search efficiency at an OWL-S based semantic web service. Within the proposed framework, the ontology documents are patterned as a tree structure whereas the query patterns are structured from the ontology patterns and individually saved at cache database. When there is a user's query request, the query patterns saved at the cache database are directly searched, and a search method to quickly process the request is proposed. The proposed approach does not search the entire ontology service but only the query patterns that have been saved from the query results. Therefore, the speed of processing user's requests can significantly decrease by reducing unnecessary searches.

Keyword : OWL-S, Ontology Matching, Service Discovery, Ontology Pattern, Cache Database

1 Department of Computer Engineering, Sungkyul University, Gyeonggi-do, 14097, Korea
e-mail : hkko@sungkyul.ac.kr (Corresponding author)

* 이 논문은 성결대학교 미래발전연구원의 지원으로 작성되었습니다.

Received(October 15, 2016), Review(October 28, 2016), Accepted(December 05, 2016), Published(December 31, 2016)

1. 서론

OWL-S (Web Ontology Language for Services)는 서비스를 위한 웹 온톨로지 언어로서 기존의 웹 서비스와 시맨틱 웹 기술의 결합을 통한 웹서비스 지원 모델을 의미한다[1]. 이러한 OWL-S는 웹 서비스에 대해서 추론과 서비스 조합을 위한 계획과 에이전트들에 의한 서비스 이용의 자동화를 가능하게 하는 목적을 가지고 있었다. 이로써 서비스 제공자들은 위와 같은 OWL-S를 이용하여 잠재적인 사용자들에게 자신들의 서비스의 기능을 알릴 수 있게 되었다[1][2].

시맨틱 웹 서비스는 온톨로지를 활용하여 서비스를 기술하고, 온톨로지를 이용해서 서비스 관련 처리를 자동화 한다[1]. 그 중 서비스 검색은 사용자 요청으로부터 만족하는 서비스를 찾는 과정을 의미한다. 이를 위한 프레임워크들은 매치메이킹(matchmaking), 입출력 타입 매칭 등으로 검색 서비스를 지원한다[2].

시맨틱 웹의 주요 연구는 시스템 간에 주고받는 정보를 파악하여 이를 자동으로 처리하고자 하는 것을 감안할 때, 앞에서 언급된 온톨로지 매칭은 시맨틱 웹에서 매우 중요한 기술이라 할 수 있다[2]. 한편, SOA에서 검색은 이미 UDDI라는 웹서비스 저장소의 표준이 제정되어 안정적인 서비스를 제공한다. 그러나 현재 사용하고 있는 UDDI 표준은 의미 기반의 단순 키워드 매칭을 지원하는 표준으로 다양한 정보의 저장과 자세한 서비스의 검색과 사용자 기반 검색을 지원하는데 많은 한계가 있다[3]. 기존의 시맨틱 웹 기술에 기반을 둔 매치메이킹 방법들은 직접적인 매칭을 제공하지 못하고 사용자 기반 검색을 제공하지 못한다[2][3].

본 논문은 OWL-S기반 웹 서비스의 효율적인 검색을 위한 프레임워크를 제공하고 사용자 질의를 색인하여 질의를 재빠르게 결정하기 위한 방법을 제공한다. 본 논문의 구성은 다음과 같다. 제 2장에서는 온톨로지 매칭에 관한 관련 연구를 기술하고 3장에서는 본 논문에서 제안하는 OWL-S 기반 효율적인 검색 프레임워크에 대해서 자세하게 기술한다. 4장에서는 실험 평가를 통하여 제안된 방법을 분석하고, 마지막으로 5장에서 결론과 향후 연구에 대하여 다루고자 한다.

2. 관련 연구

온톨로지의 매칭과 관련된 연구 분야는 크게 온톨로지 매칭 검색(Mapping discovery)과 매칭을 위한 추론(Reasoning with mappings) 연구로 나눌 수 있다[3]. 매칭 검색은 두 온톨로지에 대해서 서로 유사성을 보이는 속성을 찾는 방법을 연구하고 매칭을 위한 추론은 매칭이 이루어진 후에 그것

을 이용하여 온톨로지를 추론하는 것이다. 본 연구에서 주로 다룬 내용은 온톨로지 간의 유사성을 알아보는 매칭 검색에 대해서 연구한다고 할 수 있다.

온톨로지 매칭과 관련한 연구는 지금까지 다양한 방법론들이 연구되고 개발되어 왔다[4][5]. 온톨로지 매칭은 대상이 되는 두 온톨로지에 존재하는 여러 클래스들에 대해서 매칭이 이루어지는 범위에 따라 온톨로지의 클래스와 서브클래스들만 매칭을 수행하는 것과 해당 클래스들의 인스턴스들까지 매칭의 대상으로 삼는 방법이 존재한다. 온톨로지 매칭의 정확성을 위해서는 인스턴스까지 고려하는 것이 더 정확하나, 이 방법의 경우 매칭 시간이 온톨로지에 포함되어 있는 인스턴스의 양에 따라 기하급수적으로 증가할 수 있는 단점이 존재한다[6].

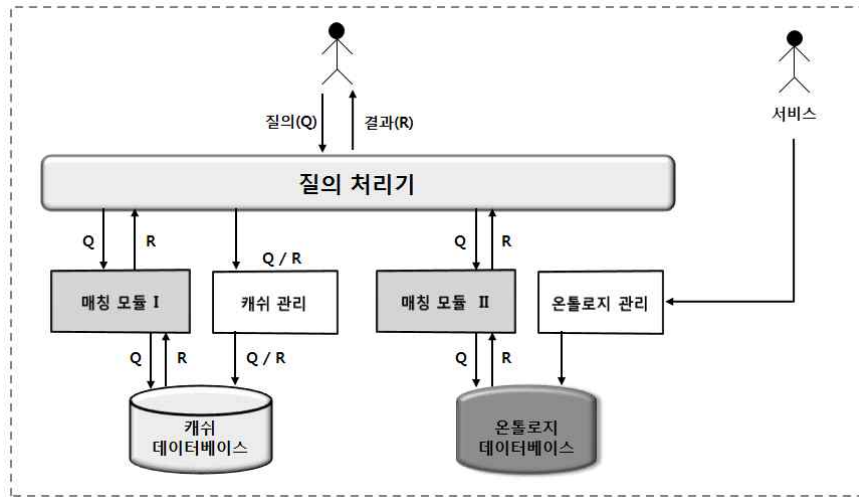
기존에 제안된[7]방법은 시맨틱 웹 기술에 기반을 두어 매치메이킹(matchmaking)을 위한 소프트웨어 프레임워크를 제안하였고 DL(description logic) 표현을 이용하여 서비스 제약조건을 표현하였다. 매칭 방법으로 온톨로지 트리를 생성하여 엘리먼트를 매칭하는 방법을 이용하였지만, 직접적인 매칭(indirect matching)을 제공하지 못한 문제점이 있었다.

시퀀스 기반 온톨로지 매칭 연구[1]은 온톨로지를 트리 형태로 변환하여 프루퍼 시퀀스(prufer sequence)를 이용하여 온톨로지 레이블 정보에 대한 시퀀스 정보를 이용한다. 본 연구는 [1]방법과 다르게 캐쉬 데이터베이스를 이용하여 온톨로지 트리과 질의 트리를 트리 패턴으로 캐쉬에 저장하는 방법을 이용한다.

반면에 [8]에서 제안한 방법은 서비스의 매치메이킹을 위한 방법으로 그래프 기반 접근 방법을 제안하였다. 기존의 그래프 매칭 알고리즘을 적용하여 동일한 구조를 가지면 서비스가 매칭되는 방법으로 제안하였다. 그러나 이 방법도 직접적인 매칭을 제공하지 못한 문제점이 있었다. 본 연구에서는 온톨로지 매칭에 인스턴스를 사용하지 않고 스키마를 이용한 매칭 방법에 초점을 맞춘다.

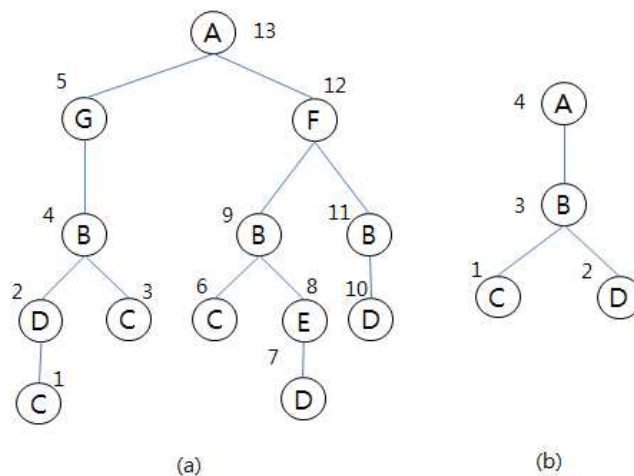
3. 제안된 OWL-S 기반 서비스 검색 프레임워크

본 논문에서는 OWL-S기반 웹 서비스의 효율적인 검색을 위한 프레임워크를 제안한다. 이를 위해 질의를 색인하여 입력된 서비스에 매치되도록 하여 질의가 어떤 것인지 재빠르게 결정하기 위한 시스템을 제안한다. 제안된 프레임워크는 온톨로지 문서를 트리 구조 형태인 온톨로지 패턴으로 구조화하고 구조화된 온톨로지 패턴으로부터 질의 패턴을 구조화하여 캐쉬 데이터베이스에 각각 저장한다.



[그림 1] 제안된 OWL-S 기반 서비스 검색 프레임워크
 [Fig. 1] Proposed framework for service discovery based on OWL-S

[그림 1]은 본 논문에서 제안한 OWL-S 기반 서비스 검색 프레임워크를 나타낸다. 제안된 프레임워크에서는 사용자의 질의가 요청되면 캐시 데이터베이스에 저장된 질의 패턴을 검색하여 질의 결과를 돌려준다. [그림 1]과 같이 제안된 프레임워크는 매칭 모듈 I, 캐시 데이터베이스, 온톨로지 데이터베이스 및 매칭 모듈 II로 구성된다. 이 때 매칭 모듈 I과 매칭 모듈 II는 질의 처리기에 의해 제어된다.



[그림 2] 온톨로지 트리 T (a) 와 질의 트리 Q (b)
 [Fig. 2] Ontology tree T (a) and Query tree Q (b)

3.1 프루퍼 시퀀스를 이용한 온톨로지 트리 구성

제안된 프레임워크에서는 먼저 온톨로지 문서를 온톨로지 패턴으로 구조화하기 위하여 프루퍼 시퀀스[9]를 이용하여 트리 형태로 변환한다. [그림 2]는 온톨로지 트리와 질의 트리를 나타낸다. [그림 2]에서 ‘트리 T’는 온톨로지 트리이고 복수개의 노드로 구성될 수 있다. 질의 트리는 [그림 2]의 (b)와 같이 트리 ‘Q’로 나타낼 수 있으며 복수개의 노드로 구성될 수 있다.

[그림 2]의 트리 구조에서 노드의 방문 순서는 전위 운행법, 중위 운행법, 후위 운행법 등과 같이 여러 방법이 적용될 수 있으나, 본 논문에서는 후위 운행법을 적용하여 트리의 방문 순서를 정한다. 후위 운행법을 이용하여 변환된 온톨로지 트리와 질의 트리로부터 각 노드에 대한 방문 순서를 결정한 후 방문 순서를 레이블 형태로 표현할 수 있다. 노드의 방문 순서를 이용하여 질의 순서와 순서 인덱스를 생성한다.

[표 1] 온톨로지 트리의 LPS와 NPS 구성

[Table 1] LPS and NPS of Ontology tree

온톨로지 트리 T	
LPS	NPS
C	1
D	2
B	4
C	3
G	5
A	13
C	6
B	9
D	7
E	8
B	9
F	12
D	10
B	11
F	12
A	13

질의 트리 Q	
LPS	NPS
C	1
B	3
D	2
B	3
A	4

질의 순서는 LPS(Label Prufer Sequence) 값을 나타내고 LPS에 대한 순서 인덱스는 NPS(Number Prufer Sequence) 값으로 나타낸다. [표 1]은 온톨로지 트리 T와 질의 트리 Q의 LPS값과 NPS 값을 보여준다.

예제 1. [그림 2]에서 LPS 값은 후위 운행법을 적용하게 되면 ‘ $C D B C G A C B D E B F D B F A$ ’와 ‘ $C B D B A$ ’로 나타낼 수 있다. NPS에 후위운행법을 적용하면 레이블 값을 나타낼 수 있는데, ‘ $1 2 4 3 5 13 6 9 7 8 9 12 10 11 12 13$ ’와 ‘ $1 3 2 3 4$ ’의 레이블 값으로 나타낼 수 있다.

전체 온톨로지 문서를 캐쉬 데이터베이스에 저장되는 정보는 앞에서 획득한 질의 순서와 순서 인덱스의 매칭된 정보를 나타낸다. 사용자의 질의 요청 시 캐쉬 데이터베이스에 저장된 순서 인덱스에 해당하는 NPS 값을 찾아 질의 요청에 응답하게 된다.

예를 들면, 예제 1에서 질의 패턴 Q에 대한 프루퍼 시퀀스 ‘ $C B D B A$ ’를 인덱스로 저장하며 동일한 질의가 다시 요청이 되었을 때 ‘ $C B D B A$ ’라는 질의에 대한 결과 순서 인덱스 ‘ $1 3 2 3 4$ ’를 캐쉬 데이터베이스에서 검색하여 응답하게 된다.

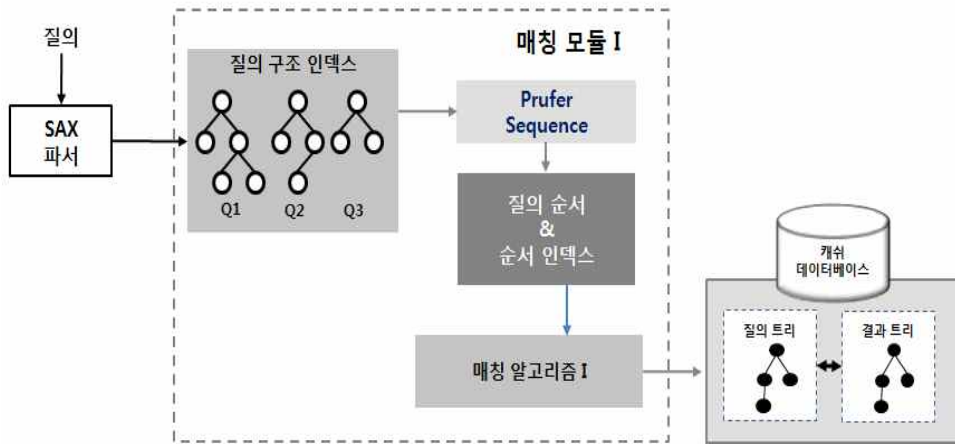
이와 같이 본 논문에서는 온톨로지 문서에 프루퍼 시퀀스를 주어 질의 패턴 Q에 대한 프루퍼 시퀀스를 저장한 후, 동일한 질의 수행 시 질의 패턴에 대한 프루퍼 시퀀스를 매치하여 원본 온톨로지의 전체 검색 없이 결과를 빠르게 돌려줄 수 있는 장점이 있다.

3.2 제안된 온톨로지 매칭 모듈

본 논문에서 제안한 프레임워크에서는 온톨로지 매칭을 위해 두 개의 매칭 모듈을 이용한다. 매칭 모듈 I은 원본 온톨로지 문서는 온톨로지 패턴으로 구조화하고 사용자 질의는 질의 패턴으로 구조화하여 각각 캐쉬 데이터베이스에 저장한다. 매칭 모듈 II는 사용자의 질의 요청 시 캐쉬 데이터베이스에 저장된 질의 패턴을 검색하여 질의 요청에 대응하는 응답 단계로 구성된다.

3.2.1 온톨로지 매칭 모듈 I

매칭 모듈 I은 SAX 파서를 이용하여 사용자의 각각의 질의마다 변환된 온톨로지 트리과 질의 트리를 생성하고 이로부터 각 노드의 인덱스를 결정함으로써 질의 패턴으로 구조화할 수 있다. 각각의 질의마다 복수개의 트리 형태의 질의 인덱스를 만들고 프루퍼 시퀀스를 이용하여 질의 순서와 순서 인덱스 형태로 변화시킨다. 이때, 변환된 질의 순서는 LPS값을 의미하며 LPS에 대한 순서 인덱스는 NPS값을 의미한다. 캐쉬 데이터베이스는 또한 매칭 모듈 II에 의해 변환된 질의 순서와 순서 인덱스를 저장하는 역할을 한다.



[그림 3] 온톨로지 매칭 모듈 I의 데이터 뷰
 [Fig. 3] Data view of ontology matching module I

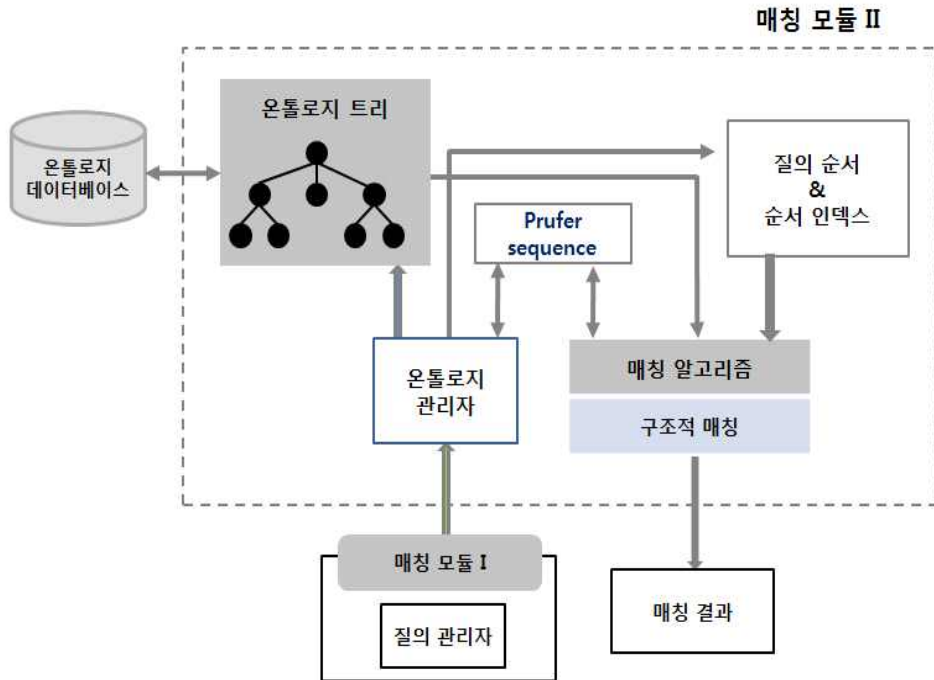
[그림 3]은 매칭 모듈 I의 데이터 뷰를 나타낸다. 예를 들면, [그림 3]에서 주어진 질의는 SAX 파서를 이용하여 각각의 질의마다 Q1, Q2, Q3의 트리 형태를 갖는 질의 구조 인덱스로 표현된다. 질의 구조 인덱스는 매칭 모듈 I에서 프루퍼 시퀀스를 적용하면 질의 순서와 순서 인덱스 형태로 캐시 데이터베이스에 저장된다.

매칭 알고리즘 I은 질의가 사용자에게 의해 요청될 경우 질의에 대한 질의 순서와 순서 인덱스를 캐시 데이터베이스에서 검색한 후 해당하는 순서 인덱스를 찾아 질의 요청에 응답한다. 만일 만족하는 질의 결과가 없을 경우에는 다시 매칭 알고리즘 I을 이용하여 매칭 모듈 II에 재검색을 요청한다.

3.2.2 온톨로지 매칭 모듈 II

매칭 알고리즘 I을 이용하여 질의에 대한 결과를 찾을 수 없는 경우 매칭 모듈 II는 매칭 모듈 I으로부터 재검색에 대한 질의 요청을 받는다. 요청을 받은 후 온톨로지 데이터베이스에서 재검색을 진행하여 질의 요청에 대응하는 순서 인덱스를 찾아 돌려주게 된다.

[그림 4]는 매칭 모듈 II의 데이터 뷰를 보여준다. [그림 4]에서 매칭 결과는 온톨로지 문서를 트리 형태로 표현하여 온톨로지 데이터베이스에 저장한다. 온톨로지 문서가 온톨로지 트리 형태로 표현되면 프루퍼 시퀀스를 이용하여 질의 순서와 순서 인덱스 형태로 온톨로지 데이터베이스에 저장할 수 있게 된다.



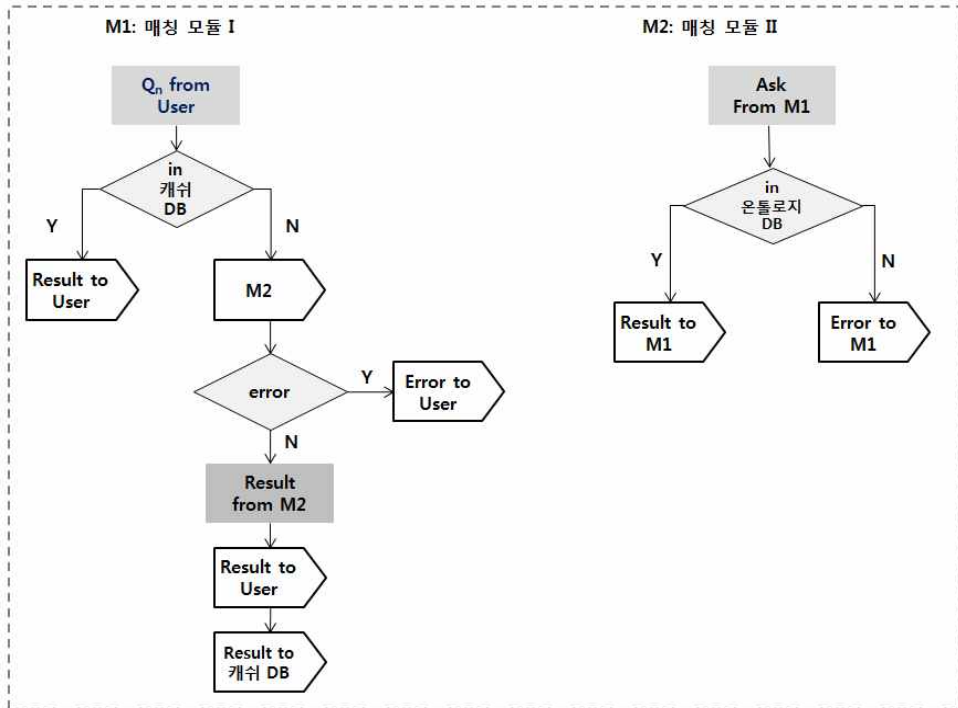
[그림 4] 온톨로지 매칭 모듈 II의 데이터 뷰
 [Fig. 4] Data view of ontology matching module II

3.2.3 구조적 온톨로지 매칭 알고리즘

사용자로부터 질의가 들어오면 매칭 알고리즘 I은 캐시 데이터베이스에 결과가 있는지 먼저 확인하고 결과가 있으면 결과를 사용자에게 돌려준다.

[그림 5]는 매칭 모듈들과 같이 동작하는 매칭 알고리즘 I의 동작 순서도를 나타낸다. 만일 만족하는 결과가 없으면 매칭 알고리즘 I은 매칭 모듈 II를 호출한다. 결과를 찾지 못하고 에러가 발생하게 되면 사용자에게 에러 메시지를 보내주게 되고 에러가 없으면 매칭 모듈 II가 온톨로지 데이터베이스에서 맞는 결과를 검색하여 만족하는 결과가 있으면 매칭 모듈 I에게 결과를 보낸다.

또한 결과에 맞는 검색이 없으면 매칭 모듈 I에게 에러 메시지를 보내준다. 이와 같이 동작하면 동일한 질의 요청 시 전체 온톨로지 서비스에 대한 검색을 차단시켜 질의에 대한 처리 속도를 향상시킬 수 있다.



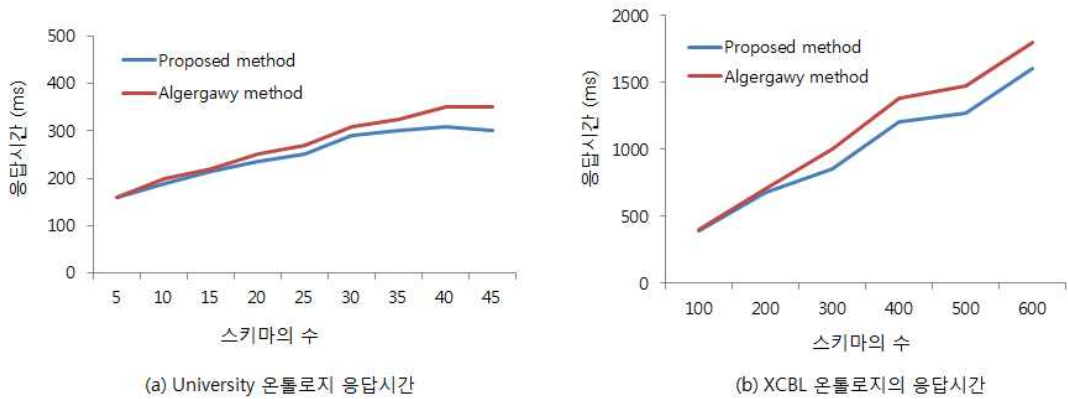
[그림 5] 구조적 온톨로지 매칭 알고리즘 I의 동작 순서도
 [Fig. 5] Operation flowchart of structural ontology matching algorithm I

4. 실험 평가

이 장에서는 실험을 통해 제안한 알고리즘의 성능 평가를 알아보고자 한다. 본 연구에서는 성능 평가를 위해 제안된 방법과 Algergawy 방법[1]을 자바로 구현하여 실험하였다. 실험은 Windows 7에서 실행되었으며 3GB의 RAM과 3.00GHZ 펜티엄 프로세서에서 수행되었다. 또한 실험에서 사용되는 온톨로지는 웹상의 온톨로지를 이용하였다.

[표 2] 데이터 셋
 [Table 2] Data set

도메인	온톨로지/노드 수	온톨로지 크기
University	44/500	<1KB
XCBL	570/3500	<10KB



[그림 6] 온톨로지의 응답 시간 분석
 [Fig. 6] Response time analysis of ontologies

본 논문에서는 매칭 효율성을 평가하기 위해서 온톨로지의 스키마 수에 따른 응답 시간을 평가하였다. [그림 6]은 두 개의 도메인에서 제안된 방법이 Algergawy method[9]보다 스키마의 크기가 커질수록 응답시간이 10% ~15% 정도 줄어든 것을 알 수 있다. [그림 6]에서 제안된 방법은 University 도메인으로부터 500개의 노드를 포함하는 44개의 스키마들을 매칭할 때 0.3분의 시간이 걸리고, XCBL 도메인으로부터 3500개의 노드를 포함하는 570개의 스키마들을 매칭할 때 1.5분의 시간이 걸렸다. 제안된 방법에서는 동일한 스키마를 질의 트리로 캐쉬 데이터베이스에 미리 저장했기 때문에 온톨로지 안에 동일한 스키마의 수가 많을수록 매칭 처리 속도가 향상되었다.

5. 결론

본 논문에서는 OWL-S기반 웹 서비스의 효율적인 검색을 위한 프레임워크를 제공하였다. 또한 사용자 질의를 인덱스하여 입력 서비스에 매치되도록 하여 질의가 어떤 것인지 재빠르게 결정하기 위한 매칭 알고리즘을 제안하였다. 제안된 프레임워크는 SAX 파서를 이용하여 사용자의 각각의 질의마다 복수개의 트리 형태의 구조 인덱스를 프루퍼 시퀀스를 이용하여 질의 순서와 순서 인덱스 형태로 변환시킨다. 그리고 매칭 모듈 I과 매칭 모듈 II에 의해 변환된 질의 순서와 순서 인덱스를 캐쉬 데이터베이스에 저장하였다. 매칭 모듈 II는 질의에 대한 요청 결과를 찾을 수 없을 경우 온톨로지 데이터베이스에서 재검색을 하여 질의 요청에 해당하는 순서 인덱스를 찾아 되돌려줄 수 있다. 실험 결과 온톨로지의 동일한 스키마에 대한 반복 질의 시 전체 온톨로지를 검색할 필요 없이 질의 패턴을 캐쉬 데이터베이스에서 검색하여 돌려줌으로써 응답 속도를 감소시켰다.

향후 연구로는 다양한 도메인의 온톨로지로 실험 환경을 확장하여 연구의 보편성을 확보할 필요가 있다.

References

- [1] A. Algergawy, E. schallehan, and G. Saake, A sequence-based ontology matching approach, Proceedings of 18th European Conference on Artificial Intelligence Workshops, (2008), pp. 26-30.
- [2] Noy, N. F. Semantic integration: a survey of ontology-based approaches , Sigmod Record, Special Issue on Semantic Integration, (2004), Vol.33, No.4, pp. 65-70.
- [3] Giunchiglia, F., Shvaiko. P., and Yatskevich, M., Semantic schema matching, Proceeding of the 13th International Conference on cooperative Information Systems, (2005) October, pp. 347-365.
- [4] Ehrig, M. and S. Staab, QOM: Quick ontology mapping, Lecture Notes in Computer Science, No. 3298, (2004), pp. 683-697.
- [5] Noy, N. F. and M. A. Musen, The PROMPT Suite: interactive tools for ontology merging and mapping, International Journal of Human-Computer Studies, (2003), Vol.33, No.4, pp. 65-70.
- [6] Magnini B., M. Speranza, and C. Girardi, A semantic-based approach to interoperability of classification hierarchies : Evaluation of linguistic techniques, Proceedings of the 20th International Conference on Computational Linguistics, (2004) August pp. 23-27.
- [7] Lei Li and Ian Horrocks, A software framework for matchmaking based on semantic web technology, Proceedings of the 12th International Conference on World Wide Web, (2003) May 20-24, pp. 331-339.
- [8] David Trastour, C. Bartolini, and J. Gonzalez-Castillo, A Semantic web approach to service description for matchmaking of services, Proceedings of the first International Conference on Semantic Web Working, (2001) July 30-August 01, pp. 447-461.
- [9] Tam'as Fleiner, On pruffer codes, Technical reports, (2005).

