

웹 서비스를 이용한 계층적 분산 온톨로지 모델

남호영^o, 양정진
가톨릭 대학교 컴퓨터 공학과
furystyle@naver.com^o, jungjin@catholic.ac.kr

Hierarchical Distributed Ontology Model Using Web Service

Hoyoung Nam^o, Jungjin Yang
School of Computer Science and Information Engineering
The Catholic University of Korea

요 약

인터넷 환경은 놀랄만한 속도로 발전하고 있다. 사용자의 수가 늘어나고 그와 동시에 자료의 양도 폭발적으로 늘어나고 있음에 따라, 정확한 정보를 찾고 불필요한 정보를 필터링 하는 기술이 요구되기 시작하였다. 그 대표적인 기술이 시맨틱 웹(Semantic Web)[1]으로 시맨틱 웹은 웹상의 정보에 메타데이터를 추가로 정의하여 사람 뿐 아니라 컴퓨터가 그 정보의 의미를 파악 할 수 있도록 하는 것이다.

이러한 시맨틱 웹을 위해서는 기계가 의미를 이해할 수 있도록 온톨로지(Ontology)를 먼저 구축해야만 한다. 온톨로지는 자원과 개념의 관계를 정의해 놓은 일종의 사전으로 이를 기술하기 위한 언어로는 RDF, OWL등이 있다. 온톨로지 데이터가 증가함에 따라 온톨로지 저장소의 크기가 증가하게 되면 성능을 위해 지역적으로 온톨로지 저장소를 분산해야 한다. 이에 따라 본고는 분산 환경에서의 통합된 질의에 대한 연구를 바탕으로 확장 가능하고 유연한 구조의 분산 온톨로지 모델을 제시한다.

Keyword : 웹 서비스, 온톨로지

1 서 론

현재 월드 와이드 웹(World Wide Web)을 필두로 한 인터넷 환경은 놀랄만한 속도로 발전하고 있다. 사용자의 수가 늘어나고 그와 동시에 자료의 양도 폭발적으로 늘어나고 있음에 따라, 정확한 정보를 찾고 불필요한 정보를 필터링 하는 기술이 요구되기 시작하였다.

그 대표적인 기술이 시맨틱 웹(Semantic Web)으로 시맨틱 웹은 웹상의 정보에 메타데이터를 추가로 정의하여 사람 뿐 아니라 컴퓨터가 그 정보의 의미를 파악 할 수 있도록 하는 것이다. 이러한 시맨틱 웹을 위해서는 기계가 의미를 이해할 수 있도록 온톨로지(Ontology)를 먼저 구축해야만 한다.

온톨로지는 자원과 개념의 관계를 정의해 놓은 일종의 사전으로 이를 기술하기 위한 언어로는 RDF, OWL등이 있다.

OWL은 W3C의 권고안으로 채택된 대표적인 온톨로지 언어로 풍부한 어휘(vocabulary)와 형식적 의미론(formal semantics)을 포함하고 있기 때문에 기계 해석이 가능한 웹 콘텐츠를 제작하는 데 있어서 XML, RDF, 및 RDF 스키마보다 뛰어나다.[2]

온톨로지 언어로 기술 된 온톨로지 데이터를 얻기 위해서는 SPARQL[3]과 같은 질의 언어를 통해 결과를 얻어낸다.

온톨로지 데이터가 증가함에 따라 온톨로지 저장소의 크기가 증가하게 되면 성능을 위해

지역적으로 온톨로지 저장소를 분산해야 한다.

이에 따라 본고는 분산 환경에서의 통합된 질의에 대한 연구를 바탕으로 확장 가능하고 유연한 구조의 분산 온톨로지 모델을 제시한다.

II 관련연구

1) 웹 서비스

웹 서비스(Web Service)는 특정 소프트웨어 패키지에 종속적이지 않은 개방형 구조를 가지고 인터넷 공간을 기본 채널로 이용하여 표준 기술을 바탕으로 한 인터페이스를 갖는 모듈화된 일련의 소프트웨어 구성요소들이라 할 수 있다.[4]

웹 서비스는 내부 또는 외부의 사용자들에게 공개되며 검색이 가능하고 다른 웹 서비스들을 조합하여 새로운 서비스를 구성할 수 있다.

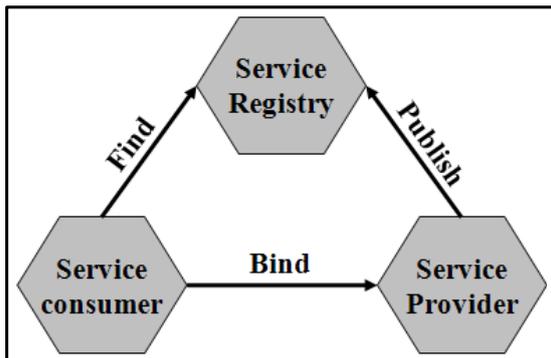


그림 1 웹 서비스

위의 그림 1과 같이 웹 서비스 제공자는 Service Registry(UDDI)에 서비스를 등록하고 웹 서비스 사용자는 원하는 웹 서비스를 찾기 위해 Service Registry를 검색한다.

웹 서비스의 기본적 구성 요소는 다음과 같다.

■ SOAP(Simple Object Access Protocol)

SOAP는 웹서비스가 인터넷 상에서 어떤 방식으로 통신해야 하는지를 규정하고 있다. SOAP는 분산된 환경에서 정보를 교환하도록 해주는 XML기반의 통신 프로토콜이다. 따라서 다양한 웹 요청자들이 인터넷을 통해 정보를 교환하는 과정에서 필수적인 통신매개의 역할

을 담당한다.

■ WSDL(Web Service Description Language)

WSDL은 웹 서비스를 정의해 주는 언어이다. 즉, 웹 서비스의 IDL(Interface Definition Language) 버전이며 특정 웹 서비스의 방법과 프로토콜, 데이터 포맷들을 더욱 상세하게 정의하는 일종의 스크립트 언어이다. WSDL은 제공되는 웹 서비스가 어떠한 서비스이며, 어떠한 방식으로 접속하는지를 정의해 주는 역할을 한다. 따라서 개발자가 웹 서비스 프로그램이나 인터페이스 정의 등 웹 서비스를 기술 할 때 WSDL을 이용한다.

■ UDDI(Universal Discovery Description Interface)

UDDI는 웹 서비스의 디렉토리 서비스를 수행한다. 일종의 온라인 전화번호부와 유사하게 웹 서비스를 찾아주고 홍보해 주는 역할을 수행한다. 그러므로 웹 서비스를 제공하는 서비스 제공자는 웹 서비스 내용과 방법을 UDDI 디렉토리에 등록하고 웹 서비스 요청자들은 외부에서 웹 서비스를 검색하는 데에 UDDI를 이용한다.[5]

웹 서비스는 WS-* 확장을 통해 기본적인 프레임워크를 확장시키고 있다.

WS-* 확장은 다음과 같은 예를 들 수 있다.

- WS-Addressing
- WS-Coordination
- WS-BPEL(BPEL4WS)
- WS-ReliableMessaging
- WS-Policy
- WS-Security

이중에서 WS-Addressing 확장은 WS-* 확장의 기본으로 다른 확장에서도 내부적으로 사용한다. WS-Addressing은 다음과 같은 정보를 명시한다.

destination	메시지가 보내질 주소
source endpoint	메시지를 만든 공급자의 엔드포인트
reply endpoint	메시지에 대한 응답을 보낼 엔드포인트
fault endpoint	문제 발생 시에 fault 메시지를 보낼 엔드포인트
message id	메시지의 고유 값
relationship	메시지가 응답해야 하는 관련 있는 메시지의 id

표 1 WS-Addressing

2) 소수 레이블링

소수 레이블링 기법은 소수의 특성인 1과 자기 자신으로만 나누었을 때만 나누어 떨어지는 성질을 이용하여 XML 트리의 각 노드에 레이블을 부여하는 방법이다.

먼저 XML 트리의 최상위 노드에 소수로 레이블을 주고 그 이하 서브 트리를 구성하는 각 노드에도 마찬가지로 소수로 레이블을 부여한다. 이 레이블을 셀프-레이블(self-label)이라고 부르며 XML 문서를 파싱하는 순서대로, 즉 트리에서 전위순회(preorder traversal)로 부여한다.

또한 자식노드의 레이블은 부모노드의 레이블과 자신에게 부여된 셀프-레이블의 곱으로 만들어진다. 그러므로 부모-자식 관계를 판단하기 위해서는 두 노드의 소수 레이블 값을 비교하여 나눈 값의 나머지가 0이라는 것으로 두 노드가 부모-자식 관계로 연결된 노드인지를 판단할 수 있다.[6]

본고에서 제시하는 모델의 가장 주요한 기능은 온톨로지가 분산 되어 있는 환경에서 시스템 사용자의 질의에 대해 분산 되어 있는 모든 온톨로지를 고려한 응답이다.[7]

III 계층적 분산 온톨로지 모델

1) 모델의 기능

이와 같은 응답은 단일 온톨로지와 같이 항상 질의에 대한 정확한 응답을 주어야 하며 시간적인 효율성면에서는 계층적으로 분산 된 깊이에 따르는 병렬성을 나타내야한다.

그 밖에도 최상위 서비스의 병목 현상을 줄이고 트래픽을 분산시키는 구조와 다양한 쿼리언어를 이해하고 변환 시키며 벤더에 특화 되어있는 저장소를 일반화된 방법으로 연결시키는 기능이 부가적으로 필요하다.

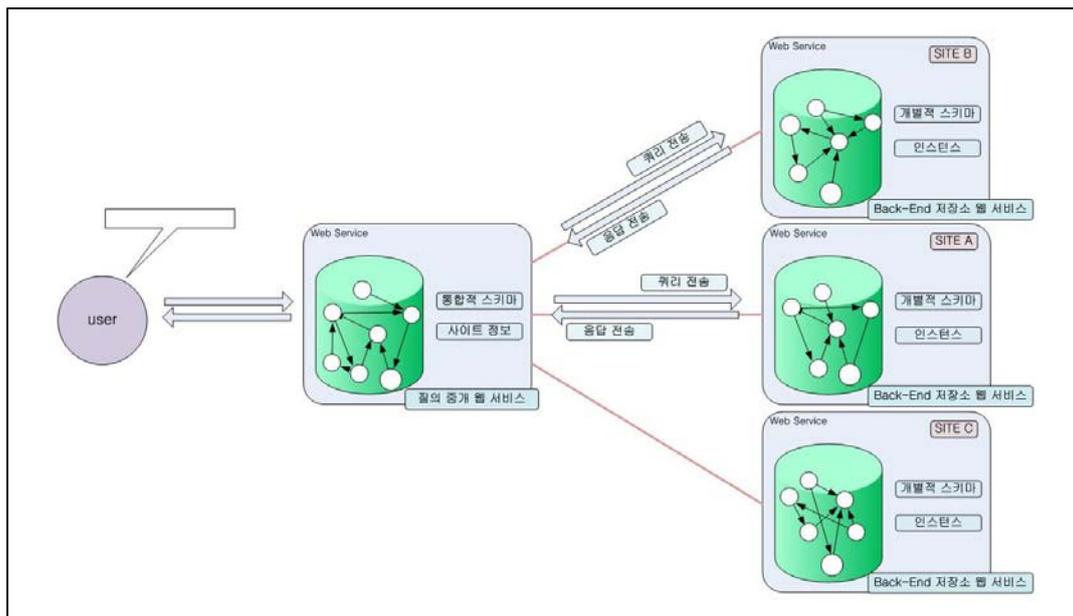


그림 2 질의 중개 웹 서비스와 Back-End 온톨로지 저장소 웹 서비스

2) 모델의 구성

본고에서 제시하는 계층적 분산 온톨로지 모델의 구성요소는 크게 두 가지이다.

첫째는 질의 중개 웹 서비스로 실제적인 시맨틱 데이터는 가지고 있지는 않지만 서비스 요청자와 실제 데이터 사이에서 질의를 중개하는 역할을 한다.

둘째는 Back-End 온톨로지 저장소 웹 서비스로 실제적인 시맨틱 데이터를 가지고 있으며 질의를 처리하여 응답을 보내는 역할을 한다.

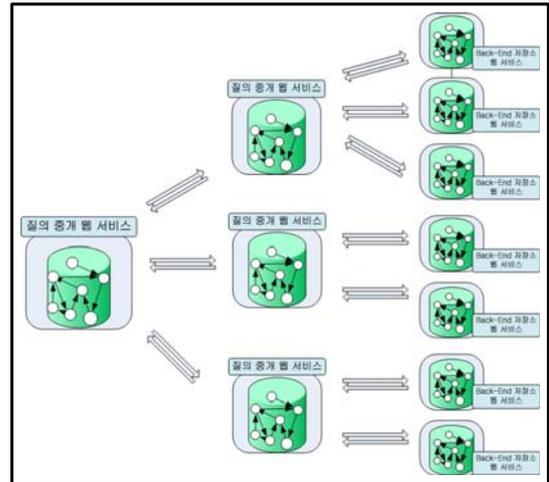


그림 3 계층적 질의 중개 웹 서비스

두 처리하는 집중형 구조이다.

그렇기 때문에 질의가 많아진다면 병목현상이 일어나게 되며 본고에서는 그에 대한 해결책으로 WS-Addressing을 이용한다.

질의의 종류에 따라서는 중개 웹 서비스가 응답을 종합하지 않고 하위 웹 서비스에서 바로 응답을 주어도 상관이 없을 때가 있다.

예를 들면 그림 2에서 SITE A에 있는 데이터에 대한 질의라면 그 응답은 SITE A가 직접 보내도 상관이 없다는 것이다.

이를 통해 요청에 대한 응답의 트래픽을 분산 시키는 효과를 가져온다.

2.1) 질의 중개 웹 서비스

질의 중개 웹 서비스는 그림 2에서 사용자와 Back-End 온톨로지 저장소 웹 서비스 사이에 위치하여 질의를 전달해 주는 웹 서비스이다.

질의 중개 웹 서비스의 온톨로지 저장소에는 실제 데이터인 인스턴스가 존재하는 것이 아니라 하위 저장소의 스키마들이 통합되어 저장되어 있다. 여기에는 또한 하위 저장소의 스키마들의 관계를 정의하는 소수 레이블링 방식의 테이블이 추가로 저장 되어있다.

이 정보를 통해 쿼리문을 변환하여 하위 저장소에 전달하는 기능을 하고 그 응답을 종합하여 최종적으로 질의를 보낸 사용자에게 전달하게 된다.

대형의 시맨틱 데이터를 다루는 경우에도 그림 3과 같이 질의 중개 웹 서비스를 계층적으로 연결해서 성능을 높일 수 있다. 이러한 계층 구조에서 고려해야 할 사항은 온톨로지 저장소를 나눌 때 주제에 따라 계층 구조와 지역적인 구분을 하는 것이다. 최상위 서비스 이용자의 질의가 계층 구조의 하위요소로 전달 될 때, 그 전달을 최소로 할 수 있는 구조가 효율적인 구조가 되겠다.

2.3) Back-End 온톨로지 저장소 웹 서비스

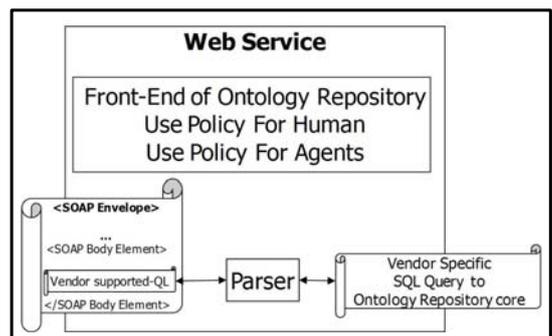


그림 4 Back-End 온톨로지 저장소의 쿼리 형식 변환 기능

2.2) 질의 중개 웹 서비스의 부가적 요소

2.2.1) WS-Addressing 을 통한 직접 응답 처리

그림 2에서와 같이 현재까지의 모델은 Front-End의 웹 서비스가 요청과 응답을 모

실제적인 온톨로지의 인스턴스들과 개별적인 스키마를 가지고 있는 지역적인 온톨로지 저장소를 감싸고 있는 것이 Back-End 온톨로지 저장소 웹 서비스이다.

Back-End 온톨로지 저장소 웹 서비스에서는 SOAP 메시지로 요청된 질의문에 대해서 수행 결과를 응답하는 역할을 하며 내부적인 저장소 구성과 상관없이 일반적인 인터페이스를 갖는다. 즉 벤더에 특화되어있는 쿼리문과 일반적으로 사용하는 질의 언어와의 변환기능을 추가로 가져야 한다.

Back-End 온톨로지 저장소 웹 서비스는 일반적인 웹 서비스의 특성을 갖기 때문에 계층 모델에 속한 하나의 멤버가 될 수도 있지만 직접적으로 요청되는 질의문에 대해서도 동일하게 응답할 수 있다.

IV 결 론

온톨로지 데이터가 증가하면 그에 따른 적재시간과 추론 시간은 데이터양에 대해 지수함수적으로 증가하게 된다.[8]

이와 같은 환경에서는 온톨로지를 계층적으로 분산 시킨 모델이 효과적이다. 온톨로지를 주제별로 계층화 시켜서 분산시킬 수 있다면 본고에서 제시한 웹 서비스 모델로 유연성과 확장성을 확보할 수 있다.

또한 질의 중개 서비스들이 연결되어 있는 계층적 구조와는 상관없이 내부적 상황을 고려하지 않고도 서비스 사용자는 원하는 결과를 얻을 수 있다.

향후 연구에서는 질의 중개 웹 서비스의 실제 구현과 더불어 단일 온톨로지 저장소 모델과 본고에서 제시한 웹 서비스를 이용한 계층적 분산 온톨로지 모델을 비교하여 성능 분석을 할 예정이다. 분산 온톨로지 구성이 필요한 온톨로지 데이터의 크기와 적합한 주제를 찾는 연구도 병행할 예정이다.

V 후 기

본 연구는 21세기 프론티어 연구개발사업의 일환으로 추진되고 있는 지식경제부의 유비쿼터스컴퓨팅 및 네트워크 원천 기반기술 개발사업의 08B3-S1-10M 과제로 지원된 것임.

VI 참고문헌

- [1] T.Berners-Lee, J.Hender, O. Lassila. "The Semantic Web", Scientific American, 2001.
- [2] W3C, OWL, Web Ontology Language (OWL), <http://www.w3.org/2004/OWL/>
- [3] W3C, SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>
- [4] 웹 서비스, 정지훈, 한빛미디어, 2002
- [5] 웹 서비스 기반 지능형 추천 시스템. 김성태, 박수민, 양정진. 한국정보과학회 학술발표논문집 Vol.31, No.2. 2004
- [6] A Prime Number Labeling Scheme for Dynamic Ordered XML Trees, Xiaodong Wu, Mong Li, Lee Wynne Hsu, ICDE, 2004
- [7] 분산된 온톨로지의 통합 질의 처리를 위한 저장 및 인덱싱 기법. 김용욱. 2007.02
- [8] 다양한 OWL-DL 추론 엔진에서 대용량 ABox 추론에 대한 성능평가 (A Performance Analysis of Large ABox Reasoning in OWL-DL Reasoners) 서은석, 박영택. 정보과학회논문지 : 소프트웨어 및 응용 제34권 제7호, 2007. 7