

자연어 입력으로부터 복합 웹 서비스의 자동 생성

임종현^o 이경호

연세대학교 컴퓨터과학과

jhlm@icl.yonsei.ac.kr, khlee@cs.yonsei.ac.kr

Building Composite Web Services from Natural Language Requests

JongHyun Lim^o Kyong-Ho Lee

Dept. of Computer Science, Yonsei University

요 약

복합 웹 서비스에 대한 요청이 증가하고 있지만 유비쿼터스 환경에서 일반 사용자들이 쉽게 복합 웹 서비스를 사용할 수 있는 환경은 마련되지 않고 있다. 본 논문에서는 자연어 기반의 사용자 친화적인 인터페이스를 개발하여 이를 통해 웹 서비스나 해당 서비스 도메인 영역에 대해 지식이 전혀 없는 일반 사용자가 쉽게 자신의 요구사항을 표현하고, 사용자의 표현을 효율적으로 분석하여 복합 웹 서비스를 자동 구성하는 시스템을 제안한다. 특히 복수 문장으로 구성된 복잡한 자연어 요청으로부터 정교한 수준의 추상 워크플로우를 추출하는 방법을 제안한다.

1. 서 론

웹(World Wide Web) 기술의 비약적인 발전으로 인해 웹을 통해서 정교한 수준의 다양한 서비스를 제공할 수 있는 환경이 조성되고 있다. 웹 서비스(Web services)는 SOAP, WSDL 등의 표준 프로토콜에 기반한 일종의 소프트웨어 컴포넌트로서 분산 환경에서 이중 플랫폼 간의 상호운용성(interoperability)을 지원한다. 특히 유비쿼터스 기술의 발전과 모바일 디바이스의 보급으로 인해 사용자가 언제 어디서든 원하는 서비스를 이용할 수 있는 환경이 조성되고 있다. 이에 유비쿼터스 환경에서 모바일 디바이스를 이용하는 사용자가 서비스를 보다 편리하게 호출할 수 있는 방법이 요구된다.

사용자의 요구수준이 높아짐에 따라 여러 개의 단위 서비스로 구성된 보다 정교한 수준의 복합 웹 서비스(composite Web services)에 대한 수요가 증가할 것으로 예상된다. 현재 웹 서비스를 조합하여 복합 웹 서비스를 자동으로 생성하는 방법에 대한 연구가 활발히 진행 중이다[1]. 기존에 복합 웹 서비스의 자동 조합에 관한 연구들은 크게 두 가지 종류로 구분할 수 있다. 첫 번째는 비교적 형식적인 사용자 요구사항인 추상 워크플로우(abstract workflow)를 입력으로 추상 워크플로우를 구성하는 각각의 태스크에 가장 적합한 단위 웹 서비스

를 바인딩(binding)하여 실행 가능한 복합 웹 서비스를 생성하는 방법이다. 두 번째는 전통적인 AI planning 기법에 기반한 방법으로 사용자가 온톨로지(ontology)를 사용하여 IOPE (입력, 출력, 전제조건 및 효과)로 구성된 요구사항을 기술하면 이로부터 복합 웹 서비스를 자동으로 조합하는 방법이다.

이와 같이 복합 웹 서비스의 자동 조합과 관련된 기존 연구들은 사용자로부터 정보를 받아들이는 방법 보다는 받아들이는 정보를 기반으로 단위 서비스를 조합하는 과정에 초점을 맞추고 있다. 이를 위해 기존 연구들은 복합 웹 서비스 조합을 위한 입력을 도메인 온톨로지를 기반으로 기술되는 추상 워크플로우나 IOPE와 같이 정형화된 형태라고 가정한다. 그러나 이러한 형태의 입력을 구성하기 위해서는 워크플로우 디자인이나 도메인 온톨로지에 대한 학습이 필요하고 워크플로우와 온톨로지를 표현하기에 충분한 화면 사이즈가 필요하다. 따라서 사전 지식이 없는 일반 사용자들이 제한된 리소스를 갖는 모바일 디바이스를 통해 웹 서비스를 호출하는 유비쿼터스 환경에서는 기존의 정형화된 입력을 통해 복합 웹 서비스를 구성하는 것이 매우 어렵다.

이에 반해 자연어(natural language) 입력을 통해서 서비스를 호출하는 방법은 대부분의 사용자들이 학습 없이 사용할 수 있는 친숙한 방법이다. 특히 자연어 입력은 대부분의 휴대용 모바일 디바이스들이 기본적으로 지원하는 인터페이스이다. 이에 사용자의 자연어 질의로부터

본 연구는 한국과학재단 특장기초연구 (R01-2006-000-10864-0(2007))지원으로 수행 되었음.

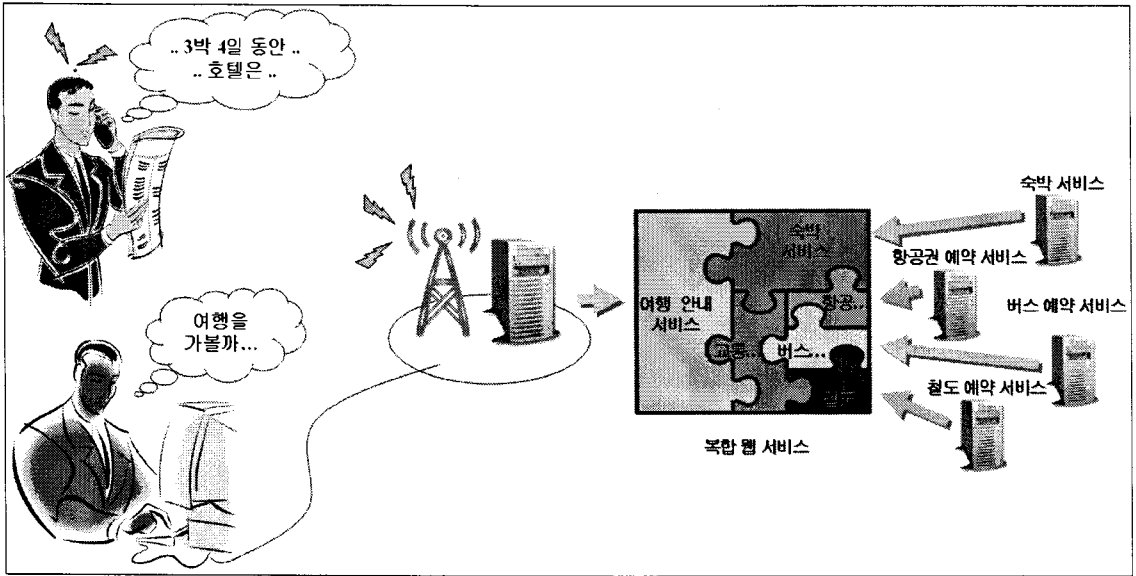


그림 1. 자연어를 이용한 복합 웹 서비스 이용 흐름.

웹 서비스 조합에 필요한 정보들을 추출할 수 있다면 보다 편리한 웹 서비스 환경을 제공할 수 있을 것이다 <그림 1>은 자연어 입력을 통해 복합 웹 서비스를 이용하는 예를 보여준다.

본 논문에서는 보다 편리한 웹 서비스 사용 환경을 구축하기 위해서 사용자가 자연어 형태로 입력한 요구사항으로부터 복합 웹 서비스를 조합 및 생성하는 방법을 제안한다. 특히 본 연구에서는 사용자의 비교적 복잡한 자연어 요청으로부터 정교한 수준의 추상 워크플로우를 추출하는데 초점을 맞추고 있다.

2. 관련연구

사용자의 자연어 질의로부터 문법 구조를 추출하거나 의미 정보를 분석하는 방법에 대한 연구는 오랜 기간에 걸쳐 진행되어왔다. 특히 단순히 자연어 문장을 분석하는 것은 물론이고 다른 분야의 연구와 연계하여 지속적으로 응용되어 왔다. 예를 들어, 자연어를 기반으로 데이터베이스의 인터페이스를 생성하는 방법 [2][3] 이나 자연어를 이용하여 프로그래밍 코드를 구현하는 방법 [4] 과 같은 연구들이 진행되었다.

특히 복합 웹 서비스와 관련한 연구는 웹 서비스 기술의 등장과 함께 최근에 대두된 분야로서 현재 활발히 연구가 진행중이다. 그러나 아직까지 복합 웹 서비스를 위한 자연어 인터페이스에 대한 연구 [5][6][7][8] 는 기초적인 단계에 머물러 있다. 기존 연구들의 특징 및 문

제점에 대한 간략한 기술은 다음과 같다.

Bosca 등 [5][6][7] 은 사용자로부터 자연어 기반의 질의를 받아들여서 이에 적합한 복합 웹 서비스의 워크플로우를 생성하는 방법을 제안한다. 이 방법은 사용자의 문장에서 직접적으로 워크플로우의 형태를 이끌어 내고 사용자의 의도에 맞는 서비스를 찾는다. 그러나 제안된 방법은 단순한 형태의 문장을 대상으로 하기 때문에 둘 이상의 문장으로 이루어진 질의를 처리할 수 없으며, 따라서 실제 생성되는 워크플로우의 형태가 매우 단순하다.

Englmeier 등 [8] 은 비즈니스 프로세스를 자연어 기반의 스토리북 형태로 입력받아 스토리북 내의 각 문장에 대해 적합한 서비스를 대응시키는 WS-Talk 시스템을 제안한다. 제안된 시스템은 사용자가 기술한 각각의 문장에 대응하는 명령어를 찾는 방식으로 전체적인 프로세스를 구성하지는 않는다.

Al-Muhamed 등 [9][10][11] 은 서비스의 도메인 온톨로지(domain ontology)를 이용하여 사용자의 자연어 요청을 술어계산의 형식으로 변환하여 보다 논리적이면서 형식적으로 사용자의 의도를 표현하는 방법을 제안한다. 이 방법은 일반적으로 자연어 분석기를 사용하는 다른 연구에 비해 질의가 문법적으로 올바르지 않을 경우에도 크게 영향을 받지 않는 특징이 있다. 그러나 둘 이상의 문장에서 교차 참조가 불가능하고, 결합(conjunction)관계만을 지원한다.

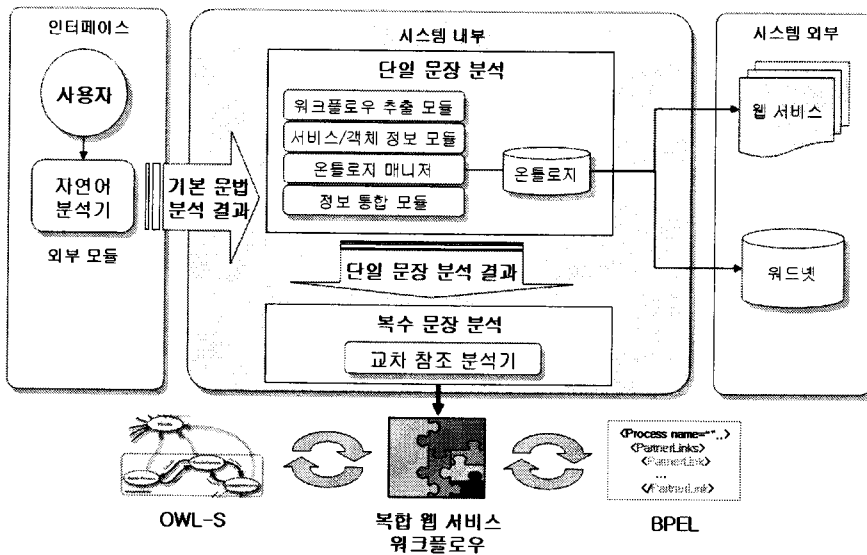


그림 2. 자연어 기반 요청으로 부터 복합 웹 서비스 구성을 위한 시스템.

3. 자연어 질의 기반 복합 웹 서비스 구성 시스템

본 논문에서 제안하는 시스템은 크게 5개의 내부 모듈과 1개의 외부 모듈로 구성되어 있다. 5개의 내부 모듈은 워크플로우 추출 모듈, 서비스/객체 정보 모듈, 온톨로지 매니저, 정보 통합 모듈, 교차 참조 분석기이며, 1개의 외부 모듈은 기존에 개발된 자연어 분석기를 연동하여 사용하기 위한 모듈이다.

전체적인 작동 흐름은 <그림 2>에서 보이는 바와 같이 먼저 사용자가 주어진 인터페이스를 통해 원하는 서비스에 대해 질의를 던지게 되면 이를 자연어 분석기가 받아서 문법적인 분석을 하게 된다. 자연어 분석기를 통해 구한 기본적인 분석 결과를 바탕으로 워크플로우 추출 모듈에서는 워크플로우를, 서비스/객체 정보 모듈에서는 서비스와 객체 각각에 대한 정보를, 온톨로지 매니저에서는 문장이 속하는 도메인 온톨로지에 대한 정보를 구하게 된다. 그 과정에서 서비스/객체 정보 모듈은 자신이 추출한 서비스와 객체에 대한 정보를 온톨로지 매니저를 통해 온톨로지와 매핑하는 과정을 거치게 되며 온톨로지 매니저는 이 과정에서 정확하게 매핑되지 않는 서비스/객체 정보에 대해 적절한 추론을 하게 된다. 이러한 추론 과정에서 워드넷을 이용한 유사의미 탐색과정 또한 포함이 되게 된다. 3개의 모듈들의 처리 결과를 받게 되는 정보 통합 모듈에서는 이들 정보들을 취합하여 하나의 복합 웹 서비스 워크플로우를 구성하게 된다. 이

렇게 얻은 결과는 단일 문장에 대한 결과로 만약 여러 문장으로 이루어진 질의일 경우 단일 문장에 대한 결과들을 수렴하여 교차 참조 분석기로 넘기게 된다. 교차 참조 분석기는 내부에서 문장 간의 도메인 비교 유사성 비교, 공통요소 추출 등의 과정을 거쳐 최종적인 복합 웹 서비스 구성을 위한 추상 워크플로우를 생성하게 된다. 이렇게 생성한 추상 워크플로우는 OWL-S나 BPEL과 같은 특정 워크플로우 언어에 독립적인 형태로 저장되며 OWL-S나 BPEL로의 변환이 가능하다.

3.1 워크플로우 추출 모듈

사용자의 요청이 단일 서비스에 대한 요청이 아니고 복수의 서비스에 대한 복합적인 요청일 경우 사용자의 질의에서 서비스가 요청되는 순서나 요청되는 서비스간의 관계를 이끌어내야 한다. 워크플로우 추출 모듈에서 그러한 작업이 이루어지는데 워크플로우 추출을 위해선 먼저 자연어 분석기에서 질의를 분석한 결과를 이해해야 한다. 자연어 분석기를 통과한 질의는 <그림 3> 과 같이 문장의 구조적인 분석 결과를 가지게 된다. 이러한 결과들을 분석하여 워크플로우를 추출한다. 분석 과정은 다음과 같다.

먼저 분석을 위한 준비 단계로 템플릿을 등록하는 단계가 필요하다. 워크플로우가 필요한 적절한 문장을 분석하고, 해당 결과를 추출 모듈에 등록한다. 그리고 등록

한 문장 분석 결과에 사용될 워크플로우를 함께 등록한다. 이러한 과정을 다양한 문장에 대해 수행하게 되면 사용자의 다양한 질의 형태에 대해서도 적절한 워크플로우를 추출하는 것이 보다 용이해진다.

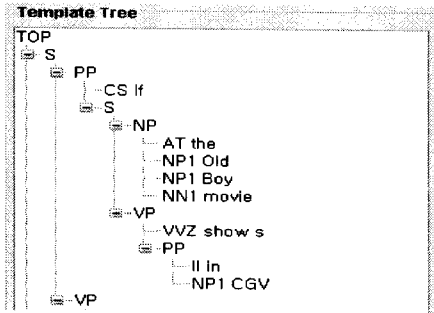


그림 3. 기초적인 문장 분석 결과

다음 과정은 실제 사용자 질의의 문장 분석 결과를 워크플로우 추출 모듈에 등록되어있는 템플릿들과 비교하는 과정이다. 이 과정에서 템플릿이 중첩적으로 적용이 되게 된다. 왜냐하면 하나의 문장이더라도 문장 안에 여러 개의 구절로 이루어져 있을 수 있기 때문에 각각의 모든 구절을 잇는 전체적인 템플릿도 적용 가능하며 각각의 구절 마다 혹은 두세 개의 구절을 잇는 부분적인 템플릿도 적용이 가능하기 때문이다. 그래서 템플릿이 중첩적으로 적용이 되며 <그림 3>에서 그 간단한 예를 살펴볼 수 있다.

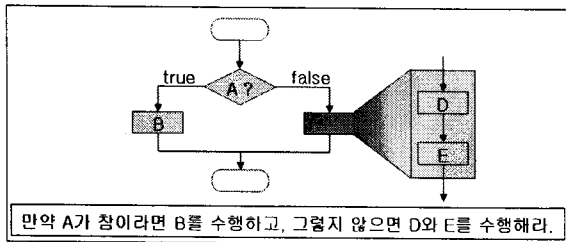


그림 4. 워크플로우 템플릿의 중첩 적용.

템플릿이 중첩적으로 적용 가능하기 때문에 문장 구조를 보다 효율적으로 템플릿으로 저장하기 위해 다음과 같은 원칙을 세울 필요가 있다. 이미 등록된 다른 템플릿의 조합으로 나타날 수 있는 문장 구조는 새로 등록하지 않는다. 만약 이러한 등록이 이루어 질 경우 동일한 문장 구조에 대해 여러 가지 결과가 나타날 수도 있고, 간단한 문장 구조들의 조합으로 이미 표현할 수 있는 문장 구조를 복잡한 하나의 문장 구조로 등록할 필요성이 없기 때문이다.

3.2 서비스/객체 정보 모듈

사용자가 요청하는 서비스를 찾기 위해선 사용자의 자연어 질의에서 충분한 정보를 추출해야 한다. 일반적으로 자연어 질의에서 사용된 단어들의 품사를 조사하게 되면 동사는 서비스의 내용과 관련이 깊고, 명사와 같은 경우 서비스와 관련된 객체일 경우가 많다. 그리고 형용사나 부사의 경우 객체나 서비스에 대한 추가적인 정보를 줄 가능성이 크다. 이러한 기본적인 직관에 근거하여 서비스/객체 정보 모듈에서는 사용자의 요청에서 사용자가 원하는 서비스의 종류와 그 서비스의 입/출력 객체, 기타 그와 관련된 사항에 대한 정보를 추출한다.

정보 추출 단계의 첫 번째는 가장 중요한 정보 중 하나라 할 수 있는 서비스 정보 추출이다. 서비스 정보는 문장 내에 포함된 동사를 추출하여 사용한다. 그러나 경우에 따라 문장 내에 여러 개의 동사가 포함될 경우도 있고, 실제적으로 동사가 큰 의미를 가지지 못하는 경우도 있다. 이럴 경우에 중심적인 동사를 찾거나 온톨로지를 탐색해 서비스를 추론한다. 온톨로지를 탐색하여 추론하는 과정은 온톨로지 매니저에서 담당하게 된다. 중심 동사를 찾는 방법은 먼저 조동사 위주로 큰 의미가 없는 동사들을 먼저 제거하고, 남는 동사를 중심 동사로 삼거나 의미적으로 타 동사를 수식하는 동사들을 먼저 제거하는 방법을 사용한다.

다음 단계는 명사로부터 객체에 대한 정보들을 추출하는 과정이다. 동사를 제외하고, 나타난 모든 명사를 일단 객체로 등록한다. 그리고 등록된 명사들을 수식하는 형용사들을 해당 명사가 속하는 객체에 추가하여 정보로 사용한다. 추출한 객체 정보들을 온톨로지 매니저를 통해 온톨로지와 매핑 한다.

3.3 온톨로지 매니저

온톨로지 매니저는 사용자의 질의가 속한 도메인 분석과 더불어 서비스/객체 정보 모듈에서 추출한 정보와 온톨로지상의 컨셉간의 매핑을 담당하며 추론 모듈을 포함한다. 다른 모듈의 온톨로지 접근 요청을 직접 처리하게 되며 워드넷 상의 정보 접근도 담당한다.

온톨로지 매니저에서 가장 핵심적인 기능은 서비스나 객체의 정보가 넘어왔을 때 그 해당 서비스와 객체가 어떤 온톨로지상의 개념과 매핑되는지를 탐색하는 기능이다. 이러한 기능을 효율적으로 달성하기 위해 워드넷과 연계를 통해 사용자의 다양한 자연어 어휘 사용으로 인해 유추 해석이 필요한 유의어들에 대한 정보를 사용하는

다. 그리고 정확한 매칭이 발견되지 않더라도 적절한 추론으로 의미적으로 가까운 개념 중 대응 가능한 것을 찾는다.

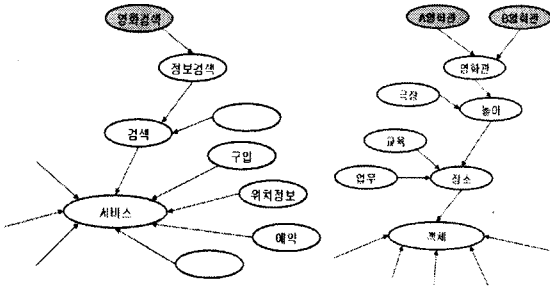


그림 5. 서비스와 객체 온톨로지의 예.

온톨로지 매니저가 기능을 효율적으로 수행하기 위해서는 서비스와 객체 정보를 의미적 온톨로지로 구축하는 것이 필수적이다. 온톨로지를 사용하여 추론을 실시하게 되면 사용자의 의도에 보다 정확한 결과를 반환할 수 있다. <그림 5>는 서비스와 객체 온톨로지를 도식화한 예이며 실제로 이러한 온톨로지는 OWL을 이용하여 기술된다.

3.4 정보 통합 모듈

정보 통합 모듈은 워크플로우 추출 모듈, 서비스/객체 정보 모듈 그리고 온톨로지 매니저에서 추출된 모든 정보를 바탕으로 최종적인 워크플로우를 생성한다. 우선 워크플로우 추출 모듈에서 추출한 워크플로우 템플릿을 바탕으로 템플릿 내에 비어있는 서비스 정보들을 서비스/객체 정보 모듈에서 추출한 정보들로 채워 넣는다. 그리고 온톨로지 매니저에서 주는 정보들은 각각의 서비스에 대한 도메인 정보로 활용한다.

정보 통합 모듈에서 생성된 결과는 단일 문장이나 구절을 분석하여 도출한 워크플로우로서 복수의 문장과 구절을 통합적으로 분석하기 위한 준비 단계의 결과물이다.

3.5 교차 참조 분석기

교차 참조 분석기는 <그림 2>에서 나타난 것과 같이 이전 단계의 모듈들의 수행 결과를 이용하게 된다. 먼저 온톨로지 매니저에서 분석한 각각의 문장이나 구절이 속한 도메인 온톨로지 정보를 이용하게 된다. 공통된 도메인 온톨로지에 속하는 문장들은 의미적으로 좀 더 긴밀히 연결된 관계로 유추 해설할 수 있다. 그리고 서비스/객체 정보 모듈에서 추출한 정보를 바탕으로 공통된 서

비스나 객체 정보를 사용하는 문장들에 대해 해당 서비스나 객체가 동일한 정보를 재사용 하는 것인지 아니면 서로 다른 정보인지 유추하여 의미적 연결 관계를 분석하는데 참조하게 된다. 마지막으로 대명사나 다른 문장을 지칭 가능한 단어를 추출하여 앞서 분석한 결과와 결합하여 정확히 어떤 문장이나 단어를 참조하고 있는지를 분석한다. 이렇게 분석한 결과를 통합하여 최종적으로 복수 문장 질의에 대한 복합 웹 서비스 워크플로우를 구성한다.

4. 결론 및 향후연구

본 연구에서는 사용자의 자연어 입력으로부터 복합 웹 서비스를 생성하기 위한 추상 워크플로우를 자동으로 구성하는 시스템을 제안하였다. 제안한 시스템을 이용하면 기존의 연구보다 단일 문장에 대해서도 다양한 추상 워크플로우를 생성할 수 있을 뿐만 아니라 기존 연구에서 다루지 않았던 복수 문장 질의에 대한 추상 워크플로우 생성이 가능하고, 서로 다른 문장이나 구절 간에 교차 참조 관계를 밝힐 수 있다. 뿐만 아니라 생성한 추상 워크플로우가 독립적인 언어로 기술이 되고, 다른 워크플로우 기술 형식으로 변환이 가능하기 때문에 추상 워크플로우로부터 상세한 워크플로우를 생성하는 연구에서 쉽게 사용이 가능하다.

제안한 시스템의 효율성을 극대화하기 위해서는 잘 정리된 서비스와 객체 온톨로지가 필수적이다. 그러므로 서비스와 객체 온톨로지를 작성하기 위한 방법에 대한 연구가 필요하다.

현재 제안한 방법을 검증하기 위한 프로토타입을 구현 중이다. 그리고 구현한 시스템을 검증하기 위해 다양한 테스트 데이터를 준비하여 사용할 계획이다.

5. 참고문헌

[1] J. Rao, X. Su, "A Survey of Automated Web Service Composition Methods", Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition, pp. 43-54, 2004.
 [2] Y. Li, H. Yang, H.V. Jagadish, "Constructing a Generic Natural Language Interface for an XML Database", Proceedings of the 10th International Conference on Extending Database Technology, pp. 737-754, 2006.

- [3] A. Popescu, A. Armanasu, O. Etzioni, D. Ko, "Modern Natural Language Interfaces to Databases: Composing Statistical Parsing with Semantic Tractability", Proceedings of the 20th International Conference on Computational Linguistics, pp. 30-39, 2004.
- [4] H. Liu, H. Lieberman, "Programmatic Semantics for Natural Language Interfaces", Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 1597-1600, 2005.
- [5] A. Bosca, G. Valetto, R. Maglione, F. Corno, "Specifying Web Service Compositions on the Basis of Natural Language Requests", Proceedings of the 3rd International Conference on Service Oriented Computing, LNCS 3826, pp.588-593, 2005.
- [6] A. Bosca, A. Ferrato, F. Corno, I. Congju, G. Valetto, "Composing Web Services on the Basis of Natural Language Requests", Proceedings of the IEEE International Conference on Web Services (ICWS'05), PP.817-818, 2005.
- [7] A. Bosca, F. Corno, G. Valetto, R. Maglione, "On-the-fly Construction of Web Services Compositions from Natural Language Requests", Journal of Software Vol. 1, Issue 1, pp. 40-50, Academy Publisher, 2006.
- [8] K. Englmeier, J. Pereira, J. Mothe, "Choreography of Web Services based on Natural Language Storybooks", Proceedings of the 8th international conference on Electronic commerce (Semantic web ontologies, rules, and services track), pp. 132-138, 2006.
- [9] M. J. Al-Muhammed, D. W. Embley, "Conceptual Model Based Semantic Web Services", Proceedings of International Conference on Conceptual Modeling, LNCS 3716, pp. 288-303, 2005.
- [10] M. J. Al-Muhammed, D. W. Embley, "Resolving Underconstrained and Overconstrained Systems of Conjunctive Constraints for Service Requests", Proceedings of the 18th International Conference on Advanced Information Systems Engineering, pp. 223-238, 2006.
- [11] M. J. Al-Muhammed, D. W. Embley, "Ontology-Based Constraint Recognition for Free-Form Service Requests", Proceedings of IEEE 23rd International Conference on Data Engineering, pp. 366-375, 2007.