Journal of The Korea Society of Computer and Information
Vol. 17 No. 8 , August 2012

2012-17-8-1-3

온톨로지 기반 공공정보 번역 매쉬업 시스템

오 경 진*, 권 기 영*, 홍 명 덕*, 조 근 식*

Ontology-based Machine Translation Mashup System for Public Information

Kyeong-Jin Oh*, Kee-Young Kwon*, Myung-Duk Hong*, Geun-Sik Jo*

요 약

본 논문에서는 외국인이 언어적인 문제점 없이 국내 문화정보 서비스를 이용하기 위한 온톨로지 기반의 번역 매쉬업 시스템을 제안한다. 문화정보 서비스를 위해 서울시 모바일 공공정보 Open API가 제공하는 문화정보를 활용하였으며, 번역을 위해서 구글 AJAX 언어 API를 이용하였다. 또한, 발생하는 번역오류를 최소화하기 위해 번역에 온톨로지를 적용하였다. 온톨로지 모델링을 위해 문화정보 도메인을 분석하고, 문화정보 어휘 온톨로지의 클래스 및 관계, 속성을 정의하였다. 문화정보를 번역하여 활용할 경우 가장 많은 번역 오류가 발생하는 제목, 장소, 주최자 등의 내용을 중심으로 온톨로지 인스턴스를 생성하였다. 온톨로지 기반 번역 매쉬업 시스템을 구현하고, 문화정보의 번역의 정확성을 실험을 통해 번역 API와의 정보 제공 정확도를 비교하였다. 실험 결과를 통해 제안하는 온톨로지 기반 번역 매쉬업 시스템의 유효성을 검증하였다.

▸ Keywords : 온톨로지, 공공정보, 매쉬업 시스템

Abstract

We have proposed an ontology-based translation mashup system for foreigner to enjoy Korean cultural information without any language barrier(linguistic problem). In order to utilize public information, we use a mobile public information open API of Seoul metropolitan city. Google AJAX language API is used for translations of public information. We apply an ontology to minimize errors caused by the translations. For ontology modeling, we analyze the public information

• 제1저자 : 오경진 • 교신저자 : 오경진
• 투고일 : 2012. 4. 29, 심사일 : 2012. 5. 27, 게재확정일 : 2012. 7. 12.
* 인하대학교 컴퓨터정보공학과(Dept. of Computer and Information Engineering, Inha University)

domain and define classes, relations and properties of cultural vocabulary ontology. We generate ontology instances for titles, places and sponsors which are the most frequently occurring translation errors. We compare the accuracy of translations through our experiment. Through the experimental results using the proposed ontology-based translation mashup system, we verify the validity of the system.

▸ Keywords : Ontology, Public Information, Mashup system

I. 서 론

문화체육관광부 통계에 의하면 우리나라를 방문하는 외국인 관광객 수는 800만 명을 넘어섰고, 국내 거주하는 외국인 수도 120만 명 이상임을 알 수 있다. 스마트 기기가 보급되면서 다양한 문화콘텐츠가 앱을 통해 제공되고 있지만, 외국인을 위한 서비스의 부족으로 많은 외국인들이 문화콘텐츠를 이용하는데 어려움을 느끼고 있다. 본 논문에서는 공공정보의 이용을 활성화 하고, 외국인들이 스마트 기기의 앱을 통해 문화관련 정보를 이용함에 있어 언어적 소통의 한계로 인한 불편함을 최소화하기 위해 온톨로지 기반 공공정보 번역서비스를 제공하는 시스템을 제안한다. 공공정보를 외국인에게 제공하기 위해 번역관련 Open API를 사용하면 되지만, 키워드 기반의 기계 번역이 수행되기 때문에 표현된 정보를 정확하게 번역하는데 어려움이 있다[1]. 예를 들어, 구글의 번역 서비스를 이용할 경우 '산울림극단'은 'Sanulrim Extreme', 서울역은 "Station"으로 번역되어 의도하지 않은 결과가 나오는 일이 발생한다. 이러한 번역의 오류로 인해 Open API를 통해 제공되는 공공정보가 다른 의미로 전달될 수 있다. 최근 패턴기반의 영한 자동번역[2]의 경우 제한된 도메인에 자동번역 시스템을 도입함으로써 높은 번역률을 보였다. 하지만 패턴기반의 자동번역은 도메인 특화된 패턴의 구축이 요구되고, 이 과정이 수동으로 진행되는 단점이 있다. 다양한 도메인 정보를 포함하는 공공정보의 경우 각 도메인별 패턴을 구축하는 것이 어렵기 때문에 패턴기반의 자동번역을 공공정보에 적용하는 것은 적합하지 않다.

본 논문에서는 문화정보 등의 공공정보가 번역되어 제공될 경우, 발생되는 오류를 해소하기 위하여 공공정보 Open API를 이용한 온톨로지 기반 번역 매쉬업 시스템을 제안한다. 제안하는 시스템은 서울시에서 제공하는 서울시 모바일 공공정보 Open API를 통해 공연정보, 문화시설정보, 문화재정보 등의 공공정보를 이용한다. 공공정보 번역을 위해 구글 AJAX 언어 API를 활용한다. 번역 과정에서 발생하는 오역 문제를 해결하기 위해 문화정보 어휘 온톨로지를 모델링하고, 이를 바탕으로 공공정보 번역에 활용하였다. 지도 및 환율 API를 이용하여 매쉬업 시스템을 구축하고, 제안하는 번역 시스템의 성능을 평가하기 위해 각 부분별 공공정보를 이용하여 검증한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 기술하며, 3장에서는 제안하는 공공정보 Open API를 활용한 온톨로지 기반의 번역 매쉬업 시스템에 대하여 설명한다. 4장에서는 제안 시스템의 타당성 검증을 위한 실험을 수행하고 결과를 분석한다. 마지막으로 5장은 제안한 시스템에 대한 결론을 맺고 향후 연구의 방향을 제시한다.

II. 관련 연구

2.1 Open API 및 매쉬업 서비스

Open API는 웹2.0의 근본 개념인 '데이터의 개방 및 공유'를 구현하는 핵심 기술로 서비스, 정보, 데이터 등을 언제, 어디서나, 누구나 쉽게 이용할 수 있도록 개방된 웹서비스 API를 의미한다[3]. 공공기관도 투명하고, 참여적이며 협력적인 공공정보의 사용을 위해 Open API를 통해 보유한 공공정보를 제공하고 있다. 스마트 기기의 확산으로 다양한 스마트 앱이 등장하면서 공공정보의 활용이 증가하고 있으며, 공공기관의 정보를 통해 새로운 사회적·경제적 가치 창출이 가능하게 되었다[4]. 여러 Open API로부터 다양한 분야의 정보가 결합되면 편리하고 유용한 정보 서비스가 가능해지고, 정보의 활용도가 커져 정보 공개의 가치가 높아지게 된다. 본 논문에서는 외국인을 위한 문화정보 앱 구현을 위해 서울시가 제공하는 모바일 공공정보 Open API를 활용한다. 구글, 야후 바벨피쉬, 마이크로소프트 Bing 번역기를 문화정보 번역 성능을 비교하고, 실험결과 오류가 가장 적은 구글 번역기를 제안하는 시스템에서 활용하였다.

매쉬업 서비스는 Open API 서비스를 조합하여 완전히 새롭고 창의적인 서비스를 만들어 내는 것을 의미한다. 2011년 5월을 기준으로 3,205개의 Open API와 5,787개의 매쉬업 서비스제공 사이트가 등록되어 있고, 지도, 검색, 사진공유, 쇼핑정보, 소셜 네트워크 서비스 등이 매쉬업 서비스의 다수를 차지하고 있다[5]. 본 논문에서는 공공정보, 번역, 지도 및 환율 API를 매쉬업 시스템에 활용한다.

2.2 선행 연구 고찰

[6, 7]은 Open API를 이용한 다국어 검색 서비스에 대한 모델링을 다루었다. 두 논문은 다국어 번역을 위한 교차 번역 시스템에 번역 Open API를 활용하였고, 동의어 처리를 위해 온라인 사전 및 AJAX를 이용한 자동 완성 기능을 적용하였다. 하지만 번역 오류를 해결하기 위한 방법을 제시하지 않았고, 사용자가 일일이 자동완성기능을 이용해야 한다.

[8]에서는 SMS의 문장을 수신자가 원하는 언어로 번역하기 위한 시스템을 제안하였다. 단어의 축약형 및 완곡 어구에 대한 자동 번역이 어렵고, 구글 번역 API 사용으로 인한 오역 문제가 여전히 남아있다.

[9]은 약재 온톨로지를 활용한 약재검색 매쉬업 시스템을 제안하였다. 온톨로지 기반 매쉬업시스템이라는 점에서 본 연구와 유사하나 약재라는 전문 도메인에서 온톨로지 기술을 활용하였고, 전문용어의 유의어 및 동의어 문제 해결을 위해 온톨로지를 사용하였다.

지능형 번역 에이전트를 이용한 실시간 번역 연구[10]는 번역에 온톨로지를 적용한 연구이다. 세종사전, 가도카와 시소러스, 한일 기계번역사전, 전문용어사전을 이용하여 한국어와 일본어 지식베이스를 위한 사전 온톨로지를 정의하고, 실시간 번역에 사용하였다.

영어 및 다른 언어로 많은 웹 문서가 표현되면서 웹 문서로부터 필요한 정보를 검색하는 것이 어려워졌다. 다국어로 구성된 웹 문서에 대한 색인 및 검색을 수행하기 위한 다양한 연구들이 수행되었고, 온톨로지 기반 번역을 통한 정보 검색 기법이 연구되었다[11][12][13].

III. 온톨로지 기반 공공정보 번역 매쉬업 시스템

3.1 시스템 구조

[그림 1]은 공공정보 Open API를 활용한 온톨로지 기반 번역 매쉬업 서비스를 위한 시스템 구조를 나타낸다.

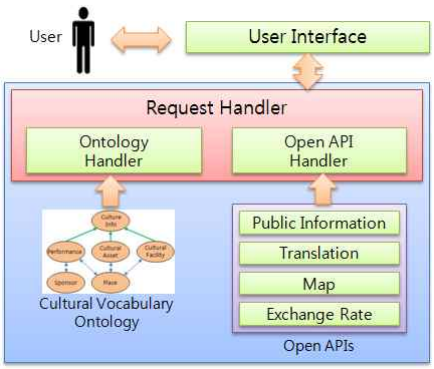

그림 1. 시스템 구조
Fig. 1. System Architecture

시스템은 크게 사용자 인터페이스(User Interface)와 요청핸들러(Request Handler)로 이루어진다. 사용자 인터페이스는 [그림 3]과 같이 세 가지 요소로 구성되어 있다. 사용자는 사용자 인터페이스를 통하여 검색하고자 하는 문화의 종류 및 언어를 선택하게 된다. 사용자가 선택한 사항은 요청 핸들러에 전달되어 처리된 후, 사용자 인터페이스에 나타난다. 요청 핸들러에 의해 생성된 결과 중 하나를 선택하면 요청한 정보의 번역된 내용이 지도 위에 나타나게 된다. 요청 핸들러는 정확한 번역을 위해 모델링된 온톨로지 정보를 활용하는 온톨로지 핸들러(Ontology Handler)와 공공정보 추출 및 부가적인 정보 활용을 위한 Open API 핸들러(Open API Handler)로 구성된다. 사용자 인터페이스에서 생성된 요청은 두 핸들러를 통해 처리되고 통합된다. 사용자 요청에 대한 처리가 완료되면 번역된 공공정보는 사용자 인터페이스에 표시된다.

온톨로지 핸들러는 공공정보 번역을 하는 과정에서 발생할 수 있는 오류를 제거하기 위해 모델링된 온톨로지를 활용한다. Open API 핸들러를 통해 추출된 공공정보의 내용 중에서 기계번역에 오류가 발생할 수 있는 부분을 추출해 온톨로

지 핸들러를 통해 처리하게 된다. 온톨로지는 서울시 공공정보 및 Open API를 분석하여 모델링되었다. 공공정보의 분류체계 및 연동 규격서에 정의된 각 변수명은 바뀌지 않기 때문에 이 부분에 대한 온톨로지 모델링은 불필요하다. 본 논문에서는 공공정보 중 문화정보를 중심으로 번역 오류가 발생하기 쉬운 제목, 장소, 주최자 등의 정보를 중점으로 온톨로지를 모델링하고, 인스턴스를 생성하였다. 모델링한 온톨로지의 세부사항은 3.2절에서 기술한다.

Open API 핸들러는 공공정보, 번역, 지도, 환율에 관련된 Open API를 다루기 위한 모듈로 구성된다. 모든 사용자의 요청은 공공정보를 기반으로 이루어지기 때문에 요청에 해당하는 주요 정보는 서울시 공공정보 Open API를 통해 추출한다. 추출된 공공정보는 우리말로 구성되어 있기 때문에 번역 API를 활용하여 번역되고, 사용자 인터페이스를 통해 사용자에게 전달된다. 요청된 정보의 결과를 나타낼 경우 사용자가 정보 주체의 위치를 직관적으로 파악할 수 있도록 지도 Open API를 사용하였다. 사용자의 요청에 해당하는 공공정보로부터 장소가 추출되고, 지도 Open API를 통해 정확한 위치에 번역된 정보가 표현된다. 공공정보 중 문화시설 및 공연에 관련된 정보는 입장료 및 관람료와 같은 가격 정보를 포함한다. 이러한 가격 정보는 사용자가 쉽게 이해할 수 있도록 사용자가 선택한 언어권에서 사용되는 화폐단위로 변환하여 제공되어야 한다. 이를 위해 환율 Open API를 활용하여 적합한 화폐단위로 계산된 가격 정보를 사용자에게 전달한다.

본 논문에서 매쉬업을 위해 사용되는 Open API는 [표 1]과 같다. 공공정보는 서울시 모바일 공공정보 Open API, 지도정보는 다음지도 API, 번역은 구글 AJAX 언어 API 그리고 환율이 적용된 가격 정보 제공을 위해 네이버 환율 API를 이용하였다.

표 1. 사용 Open API
Table 1. Open API for a proposed system

Open API	내 용
공공정보 API	서울시 모바일 공공정보(문화정보) API
지도 API	다음지도 API, 야후지도API(POI)
번역 API	구글 AJAX 언어 API
환율 API	네이버 환율 API

사용자가 요청한 정보를 번역하기 위해 요청 핸들어가 수행되는 과정은 다음과 같다. 우선 사용자가 요청한 정보를 Open API 핸들러를 통해 추출한다. Open API 핸들러는 공공정보 Open API를 통해 사용자가 요청한 공공정보를 추출한다. 추출된 정보의 번역에 앞서, 오역이 일어날 수 있는 부분을 선별하는 과정이 필요하다. 추출된 공공정보로부터 각 단어를 추출하여 온톨로지 핸들러에 전달한다. 온톨로지 핸들러는 각 단어를 모델링된 인스턴스와 비교하여 번역될 단어를 결정하고, 단어와 번역된 의미의 쌍을 반환한다. 요청 핸들러는 번역 API를 통하여 추출된 공공정보의 번역을 수행한다. 번역된 정보는 번역 오류를 포함할 수 있는데, 요청 핸들러는 이를 반환된 단어와 번역 쌍을 이용하여 해결한다. 반환된 쌍의 단어가 번역된 부분을 추출하고, 번역 API를 이용하여 변역된 정보와 단어와 번역 쌍에 포함된 정보가 서로 다를 경우, 오역이 발생한 것으로 간주한다. 요청 핸들러는 차이가 있는 번역 정보는 수정되어 통합하고 나머지 작업을 수행한다. 가격 정보는 실시간 환율에 따라 결정이 되기 때문에 번역 API를 통해 해결할 수 없다. 따라서 본 논문에서는 환율 Open API를 이용하여 가격 정보를 변환하게 된다. 사용자 인터페이스에서 선택한 언어권에서 사용하는 화폐 단위를 이용하여 가격 정보를 변환하고, 변환된 가격 정보를 이용하여 번역된 공공정보를 업데이트한다. 번역된 텍스트 정보를 사용자에게 직접 제공할 경우, 위치를 파악하는데 어려움이 발생할 수 있다. 이를 해결하기 위하여 지도 API를 통해 사용자가 요청한 대상의 정확한 위치를 검색하고 번역된 정보를 지도위에 나타낸다. 이러한 과정을 통해 사용자가 요청한 정보는 정확하게 번역된 공공정보와 위치를 이용하여 표현되고, 사용자는 쉽게 정보를 파악할 수 있게 된다.

3.2 문화정보 어휘 온톨로지

요청 핸들러가 공공정보를 번역할 때 기계 번역으로 인해 발행되는 오역을 줄이기 위한 목적으로 문화정보 어휘 온톨로지(Culture Vocabulary Ontology)를 모델링하였다. 문화정보 도메인의 분석을 통해 정의된 클래스는 [표 2]와 같다. 제공될 문화 콘텐츠는 문화정보로 표현되고 CultureInfo 클래스로 정의하였다. 문화 정보는 공연 정보, 문화시설 정보 및 문화재 정보로 구분된다. 각 정보는 Performance, CulturalFacility 그리고 CulturalAsset 클래스로 정의하였다. 문화정보 및 하위 클래스들은 그 인스턴스가 이뤄지는 장소를 포함하고 있다. 장소를 위한 클래스는 Place로 표현된다.

표 1. 문화정보 어휘 온톨로지 클래스
Table 1. Classes of Culture Information Vocabulary

클래스	설 명
Place	문화정보의 위치를 의미하는 클래스
Sponsor	공연주최를 의미하는 클래스
CultureInfo	문화정보의 상위 클래스
Performance	공연을 나타내는 클래스
CultureAsset	문화재를 나타내는 클래스
CultureFacility	문화시설을 나타내는 클래스

CultureInfo 클래스는 문화정보의 이름(제목)을 표현하는 속성을 갖는다. CultureInfo 클래스의 하위 클래스 Performance, CultureAsset 그리고 CultureFacility 클래스는 CultureInfo 클래스로부터 이름 속성을 상속받는다. 세 클래스는 문화정보가 존재하는 위치를 표현하기 위해 Place 클래스와 hasPlace 관계를 갖는다. 도메인(domain)은 CultureInfo의 세 하위 클래스이며 범위(range)는 Place 클래스로 정의된다. Place는 문화정보의 주체인 공연장 및 문화시설 등이 존재하는 장소를 의미하며 장소의 이름을 위해 hasPlaceName 속성을 지닌다. 장소 클래스는 오역을 줄이기 위해 각 언어 표현을 위한 속성을 가지고 있다. [그림 2]는 문화정보 도메인을 분석한 결과를 이용하여 모델링된 문화정보 어휘 온톨로지가 갖는 클래스, 속성과 클래스 사이에 존재하는 관계를 나타낸다.

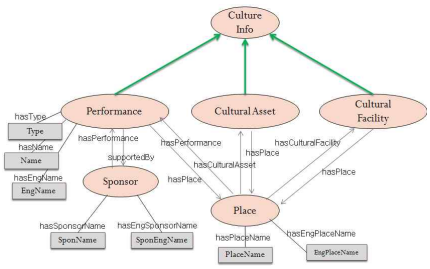

그림 2. 온톨로지 클래스 관계도
Fig. 2. Relations of Ontology Class

문화정보 어휘 온톨로지를 통해 오역을 줄일 수 있는 예는 다음과 같다. 특정 공연이 진행되는 장소 중 "우리극장"이 존재한다. "우리극장"은 공공정보에서 장소의 개념으로 정의되어 있다. 본 논문에서 활용하는 구글 번역 API를 통해 "우리극장"을 번역하면 "Our Theater"라는 결과를 돌려준다. 이는 "우리"라는 의미를 기계번역을 통해 번역함으로써 "our"로 번역하게 되는 것이다. "우리극장"은 고유명사로써 "Woori Theater"와 같이 표기되기 때문에 번역 API를 통한 번역에 오역이 발생하게 된다. 하지만 우리극장의 장소 개념과 정의된 온톨로지를 통해 "우리극장"을 정확하게 번역하여 제공할 수 있게 된다. 문화정보 및 하위 클래스의 인스턴스에 대한 내용 중 오역이 일어날 수 있는 정보에 대해 온톨로지가 정의되어 있는 경우 정확한 번역을 통해 정보 제공이 가능하다.

IV. 실험 및 평가

4.1 시스템 구현

본 논문에서 제안하는 시스템은 윈도우 7 운영체제에서 개발되었다. 사용자 인터페이스는 Apache HTTP Server 2.2, PHP 및 AJAX 기술을 이용하여 구현되었다. 문화정보 어휘 온톨로지를 모델링하기 위해 Protege 4.0을 이용하였고, 시스템에서 온톨로지 인스턴스 생성 및 활용을 위해 Jena 프레임워크[14]를 사용하였다.

온톨로지 기반 공공정보 매쉬업 시스템의 인터페이스는 [그림 3]과 같다. 사용자는 [그림 3]의 ①에서와 같이 언어 및 얻고자하는 공공정보를 선택할 수 있다. 선택사항에 해당되는 정보들은 ②번과 같이 목록으로 표현된다. 목록에서 사용자가 원하는 정보를 선택하면 ③번과 같이 지도위에 공공정보의 내용이 표현된다.

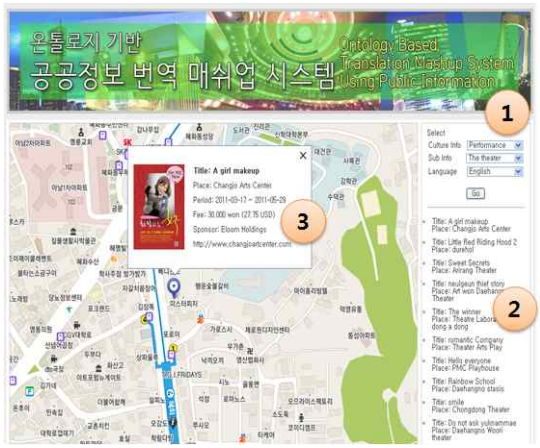

그림 3. 정보 조회 사용자 인터페이스
Fig. 3. User Interface

[그림 3]은 연극 “화장하는 여자”에 대한 정보를 보여준다. ①번에서 문화정보는 공연으로 선택되었고, 하부정보는 연극이 선택되었다. ②번에 나타난 연극의 리스트 중 “화장하는 여자”가 선택되어 인터페이스에 지도를 표시하고 공연이 진행되는 장소에 표시를 하고, 조회된 정보는 제안된 시스템의 번역 기능을 통해 번역되어 ③번과 같이 표시된다.

그림 4. 문화시설 정보 조회 예
Fig. 4. Screen shot for culture facility

그림 5. 일본어 및 중국어 정보 조회 화면
Fig. 5. Screen shots of Japan and China

[그림 4]는 문화시설 정보를 조회한 예를 나타낸다. 문화시설정보에서 국립고궁박물관이 선택되었으며, 지도에 국립고궁박물관이 위치한 위치가 표시되고 [그림 3]과 같은 방법으로 국립고궁박물관에 대한 정보가 번역되어 제공된다. [그림 5]의 왼쪽과 오른쪽은 [그림 4]와 동일한 연극 “화장하는 여자”를 각각 일본어와 중국어로 조회한 결과이다.

4.2. 번역 정확도 평가

번역 정확도 평가를 위해 서울시 공공정보 Open API로부터 문화정보에 대한 제목, 장소, 주최자를 추출하고 이를 기반으로 정확도 평가를 수행하였다. 다른 정보들과 달리 제목, 장소, 주최자는 번역 API를 통한 오역이 가장 많이 나타나는 부분이다. 기간, 날짜, 요금 등의 정보는 번역 API와 환율 API를 통해 직접적으로 변환이 가능하고 번역의 오류가 적어 평가를 위한 부분에서 제외하였다. 실험을 위해 제목, 장소, 주최에서 나타나는 고유명사에 대한 오역을 줄이기 위해 문화정보 어휘 온톨로지의 인스턴스를 각각 74, 50, 53개씩 생성하였고, 생성된 온톨로지 인스턴스는 온톨로지 핸들러에 의해 활용되었다. 제목은 ‘난타’, ‘세종’, ‘국물’, 장소는 ‘우리극장’, ‘네모’, 주최는 ‘숙명’, ‘극단’과 같이 고유명사 중심의 인스턴스 생성이 이루어졌다. 이는 다른 의미로 번역될 가능성이 높거나, 우리말 표현은 한 단어이지만 영어를 이용해서 한 단어로 표현하기 어려운 단어가 고유명사에 많기 때문이다. 공공정보 중 문화정보에서 제목은 300개, 장소 및 주최자는 각 100개씩 총 500개의 데이터를 선정하여 정확도를 평가한다. 제안하는 온톨로지 기반 번역의 정확도를 평가하기 위해 온톨로지 기반의 번역과 온톨로지 사용이 없는 구글 번역 API 기반 번역의 정확도를 비교한다. 논문에서 정의한 온톨로지에 다양한 언어에 대한 정의를 위해 여러 속성을 정의하였지만, 실험에서는 영어 번역을 기준으로 문화정보 번역의 정확도를 평가하였다.

표 3. 번역 결과
Table 3. Translation results

	한글명	구글 번역 API	온톨로지 적용
제목	난타	Nanta	Nanta
	세종이야기	Three kinds of stories	The story of Sejong
	국물있사옵니다.	Saopnida broth is	Gungmul It Sa Ob ni Da
장소	대학로 우리극장	Daehangno our theater	Daehangno Woori theater
	네모극장	Square Theatre	Nemo Theatre
	나눔문화	Sharing Culture	Sharing Culture
주최	숙명가야금 연주단	Fate Gayageum Orchestra	Sookmyung Gayageum Orchestra
	극단 동	Extreme Winter	Theatrical Company Dong
	파스텔뮤직	A music	A music

[표 3]은 제목, 장소, 주최자로 나누어 구글 번역 API를 이용해서 번역된 결과와 온톨로지를 적용하여 번역된 결과의 일부를 나타낸다. 한글명에 따른 번역 결과를 구글 번역 API만을 이용해 번역된 결과와 온톨로지 기반 번역결과로 나타냈다. 제목 부분에서 뮤지컬 공연 “난타”의 경우 두 번역이 정확했지만, “세종이야기”와 “국물있사옵니다”의 경우 구글 번역 API는 모두 잘못된 번역 결과를 나타냈다. 이는 구글 번역

API가 세종을 왕으로 인식하지 못하고 종류로 인식한 결과이다. “국물있사옵니다”의 경우에도 두 번역은 다른 결과를 도출하였다. “국물도 없다”는 말을 반어적으로 활용하였기 때문에 번역 API가 정확히 해석을 하지 못하였다. 이 경우 [15]에서 사용한 영어 표현을 이용하여 온톨로지를 모델링하였다.

장소 부분에서 “대학로 우리극장”의 경우 ‘우리’가 ‘our’로 번역이 되는 오류가 발생하였다. “네모극장”은 ‘Nemo’가 아닌 ‘Square’로 번역이 되면서 두 번역 결과의 차이를 나타냈다. 장소 표현에 고유명사의 사용으로 인해 많은 오류가 남을 알 수 있었다. “나눔문화”의 경우 현재 ‘Nanum’이라는 표현을 사용하고 있지만 논문에서는 온톨로지로 정의되지 않아 번역 API의 결과로 번역되었다.

주최자 부분은 “극단”을 “Extreme″으로 재단법인의 ”(재)“을 ”Re“로, 사단법인의 “(사)”를 “Four″로 번역하는 등의 오류로 인해 제목, 장소, 주최자 세 항목 중 가장 많은 번역오류가 나타났다. 주최자의 표현이 장소와 마찬가지로 고유한 이름을 이용하여 표현된 경우가 많고, 번역 API가 축약된 표현을 의미적으로 해석하지 못하기 때문에 발생한 것으로 분석된다.

표 4. 정확도 비교
Table 4. Accuracy comparison

항 목	번역 정확도	
	온톨로지 미적용	온톨로지 적용
계	60.8%	78.6%
제 목	68.3%	76.7%
장 소	56%	82%
주최자	43%	81%

[표 4]는 온톨로지의 적용 여부에 따른 실험 결과를 항목별로 나타낸 것이다. 온톨로지 기반 번역 매쉬업시스템을 이용한 실험 결과 온톨로지를 적용한 번역의 결과는 적용하지 않은 결과에 비해 전체적으로 17.8%의 번역 정확성 향상을 나타냈다. 항목별로는 제목부분은 68.3%에서 76.7%로 소폭 향상 되었으며, 장소부분은 56%에서 82%로, 주최자 부분은 43%에서 81%로 가장 좋은 결과 향상을 나타냈다. 장소와 주최자의 표현에 있어 고유한 이름사용과 축약된 표현의 사용으로 인해 많은 오역이 발생하는 것을 알 수 있었다. 정확하게 번역된 내용을 사용자에게 전달하기 위해서는 장소 및 주최자에 대한 온톨로지 모델링이 필요하고, 이를 통해 정보를 전달하는 것이 필요하다.

V. 결론 및 향후 연구

공공정보를 포함한 많은 정보들이 Open API를 통해 제공되지만, 외국인의 경우 언어적 문제로 인해 이를 직접적으로 활용하기 어렵다. 본 논문에서는 언어적 문제없이 국내 문화정보 활용할 수 있도록 온톨로지 기반 번역 매쉬업 시스템을 제안하였다. 문화정보 서비스를 제공하기 위해 서울시 모바일 공공정보 Open API를 이용하였고, 번역을 위해 구글 언어 API를 이용하였다. 문화정보의 번역과정에서 발생하는 번역 오류를 최소화하기 위해 문화정보 어휘 온톨로지를 모델링하고 각 클래스의 인스턴스들을 생성하였다.

논문에서 제안한 온톨로지 기반 공공정보 번역 매쉬업 시스템을 통해 정보를 번역한 결과 온톨로지 기반 시스템이 단순 번역 API를 활용했을 때보다 평균 17.8%의 정확도 향상을 보임을 알 수 있었다. 항목별 분석을 통해 장소와 주최자와 같이 고유한 표현이나 축약된 표현이 다수 포함된 경우에는 번역 API의 번역 정확도가 크게 떨어짐을 알 수 있었다. 제목의 경우 고유명사 및 축약 표현의 많은 부분을 온톨로지 인스턴스로 생성하지 못해 정확도가 장소와 주최자에 비해 높지 않았다. 하지만 제목의 경우 번역 API의 성능이 장소와 주최자의 경우보다 높음을 알 수 있었다. 실험 결과로부터 제안하는 시스템을 공공정보 제공시스템에 활용할 경우, 보다 정확한 번역을 통해 외국인에게 정확한 정보를 제공할 수 있게 된다. 증가하는 공공정보에 대해서는 온톨로지 클래스 확장하고 온톨로지 인스턴스를 추가적으로 생성함으로써 정보의 정확한 번역 서비스를 제공할 수 있다.

본 논문에서는 한국어와 영어에 대한 실험 평가만을 기술하였으나, 영어로의 번역이 정확하게 될 경우 다른 언어로의 번역이 또한 정확함을 알 수 있었다. 제안한 시스템에서는 문화정보에 대한 온톨로지가 작성되어 사용되고 있지만, 문화정보 이외의 도메인 분석을 통해 온톨로지가 확장될 경우 정확한 번역을 통해 공공정보 서비스가 가능할 것으로 보인다.

참고문헌

[1] M. Popovic, D. Vilar, E. Avramidis, and A. Burchardt, "Evaluation without references: IBM1 scores as evaluation metrics," Proceedings of the 6th Workshop on Statistical Machine Translation, pp. 99-103, July, 2011.

[2] SK. Choi, KY. Lee, YH. Roh, OW. Kwon, YG. Kim, "Customization Method for Commercialization of a Pattern-based English-Korean Machine Translation System," Journal of KIISE : Software and Applications, Vol. 39, No. 4, pp. 253-260, 2012.

[3] Open API, http://en.wikipedia.org/wiki/Open_API, 2007.

[4] P. Parycek, and M. Sachs, "Open Government - Information Flow in Web 2.0," European Journal of ePractice, Vol. 9, No. 3, pp. 1-12, Mar, 2010.

[5] Programmableweb, Mashups, Mashup Directory, http://www.programmableweb.com, 2011.

[6] SC. Hwang, HC. Kim, SJ. Kim, JS. Jeong, SJ. Kang, "A Study on the Modeling of Multilingual Information Search System using Open API," 09 Proceedings of the Korea Society for Industrial Systems Conference, pp. 129-132, 2009.

[7] SJ. Kim, SJ Kang, "Modeling and Implementation of Multilingual Meta-search Service using Open APIs and Ajax," Journal of the Korea Industrial Information System Society, Vol. 14, No. 5, pp. 11-18, Dec, 2009.

[8] S. K. Samanta, A. Achilleos, S. Moiron, J. C. Woods, and M. Ghanbari, "Automatic language translation for mobile SMS," International Journal of Information Communication Technologies and Human Development (IJICTHD), Vol. 2, No. 1, pp. 43-58, Mar, 2010.

[9] SK. Kim, C. Kim, HC. Jang, SJ. Yea, MY. Song, "Mash-up system for searching herb using herb ontology," Journal of Information Management, Vol. 39, No. 4, pp. 173-186, Dec, 2008.

[10] SK. Kim, KA. Kim, KH. Ahn, "Using Machine Translation Agent Based on Ontology Study of Real Tranlsation," 06 Proceedings of the Korea Inteligent Information System Society Conference, pp. 229-233, Jun, 2006.

[11] D. W. Embley, S. W. Liddle, D. W. Lonsdale, and Y. Tijerino, "Multilingual ontologies for cross-language information extraction and semantic search," Proceedings of the 30th international conference on Conceptual modeling, pp. 147-160, Apr, 2011.

[12] L. Zhuhadar, O. Nasraoui, R. Wyatt, and E. Romero, "Multi-language Ontology-based Search Engine," Proceedings of 3rd international conference on Advances in Computer-Human Interactions, pp. 13-18, Feb, 2010.

[13] M. Kumar, and R. Vig, "Multilingual Context Ontology Rule Enhanced Focused Web Crawler," Journal of Advances in Information Technology, Vol. 1, No. 1, pp. 21-25, Feb, 2010.

[14] Jena Framework, http://jena.sourceforge.net/, 2007.

[15] AY. Shin, "Research of Epic on Drama by Gun-Sam, Lee - Center of Gravity the "Gungmul It Sa Ob ni Da"," The Journal of Drama, pp. 131-149, Jun, 2007.

저 자 소 개

오 경 진

2006: 인하대학교
컴퓨터정보공학과 공학사
2008: 인하대학교
컴퓨터정보공학과 공학석사
2008~현재: 인하대학교
컴퓨터정보공학과 박사과정
관심분야: 시맨틱 웹, 개인화, 데이터마이닝
Email : okjkillo@eslab.inha.ac.kr

권 기 영

2009: 평생교육진흥원
컴퓨터과학과 이학사
2011: 인하대학교
컴퓨터정보공학과 공학석사
1995~현재: 인천광역시
감사관실
관심분야: 온톨로지, 매쉬업, 데이터베이스
Email : aura@korea.kr

홍 명 덕

2008: 서울디지털대학교
컴퓨터공학과 공학사
2011: 인하대학교
컴퓨터정보공학과 공학석사
2011~현재: 인하대학교
컴퓨터정보공학과 박사과정
관심분야: 추천 시스템, 시맨틱 웹, 군집 지능, 메타휴리스틱
Email : hmdgo@eslab.inha.ac.kr

조 근 식

1982: 인하대학교
전자계산학과 공학사
1985: Queens Colleg/CUNY M.A.
컴퓨터공학과 공학석사
2004: City University of New York
컴퓨터공학과 공학박사
현 재: 인하대학교
컴퓨터정보공학과 교수
관심분야: 인공지능, 시맨틱 웹, 전문가시스템, 지능형에이전트시스템
Email : gsjo@inha.ac.kr