챗봇을 이용한 방언 번역기

오세민[O], 문미경[*]
[O]동서대학교 컴퓨터공학부,
[*]동서대학교 컴퓨터공학부
e-mail: qqweewq@naver.com[O], mkmoon@dongseo.ac.kr[*]

Regimal Translator Using Chatbot

Se-Min Oh[O], Mikyeong Moon[*]
[O]Division. of Computer Engineering, Dongseo University,
[*]Division. of Computer Engineering, Dongseo University

● 요 약 ●

지역여행이나 타 지역 방문 등 해당 지방 방언을 사용하는 사람과 대화를 할 시, 단어의 뜻을 이해하지 못해 대화를 진행하기 어려운 경우가 종종 발생한다. 이러한 문제점을 해결하기 위하여 본 논문은 챗봇에게 음성이나 단어, 문장을 입력하면 표준어로 번역 및 뜻을 알려주는 방언 번역기 시스템을 제안한다. 이 시스템은 카카오톡과 연동되어 누구나 이용 가능하며, 사용자가 발화한 단어를 입력하면 챗봇이 해당 내용을 표준어로 번역하여 해당 방언의 뜻, 내용, 사용지역을 알려주며, 예문이 존재 시, 해당 방언이 들어간 예문을 보여준다. 또한 문장을 작성할 시 해당 문장을 표준어로 번역된 문장으로 바꾸어주고 시각화하여 사용자는 이를 보고 상대방의 의도를 파악할 수 있도록 도와준다.

키워드: 방언(Regimal), 챗봇(Chatbot), 번역(Translation)

I. Introduction

‘방언’은 하나의 언어에서 ‘갈라져 나온 언어’를 뜻하는 말로 언어의 하위계념이다. 어느 나라 언어든 완전히 동일한 언어는 존재하지 않다. 하나의 언어라도 지역적인 차이에 의해 생기는 방언을 ‘지리적 방언’, 사회적인 이유로 생기는 방언을 ‘사회적 방언’ 이라고 한다. 보통 ‘방언’이라고 하면 흔히 ‘지리적 방언’을 생각할 정도로 ‘지리적 방언’은 대표적인 종류이며, 이러한 차이로 인해 서울에서 생활하다 타지역으로 내려온 사람이 해당 지역의 방언을 사용하는 사람과 대화를 할 때, 처음에는 이해하기 어려워 그 자리에서 무슨말인지 물어보거나 나중에 주변 지인들에게 물어보기도 한다.

현 시대는 실제 세상과 디지털 세상이 기술과 데이터로 연결되고 통합되어, 모든 분야에서 새로운 기술을 만들어내는 시대가 되었다. 그 중에서 ‘언어’를 기반으로 한 변화 중 ‘챗봇’이 있는데, 챗봇은 사용자가 작성한 내용의 의도를 파악하여 답변을 자동으로 진행하게 해주는 서비스 이다.

본 논문에서는 발화한 방언 내용을 표준어로 번역하여 카카오톡 채널을 통해 실시간으로 뜻, 내용, 사용지역을 알려주는 챗봇 시스템에 대해 기술한다.

II. Preliminaries

[1]의 연구에서는 표준어로 작성된 문장을 경상도, 전라도에서 사용하는 지역에 맞게 방언 문장으로 번역하는 연구를 진행하였는데, 표준어 문장을 각각의 단어로 나눈 후 단어마다 단어 사전에 있는 후보 방언단어와 BoW 사이의 코사인 유사도를 측정하여 점수를 부여한다. 이중 가장 점수가 높은 방언단어를 표준어 단어와 변환하게 되며 이 과정을 문장이 끝날 때 까지 반복한다.

[1]의 연구에서는 경상도, 전라도와 같은 지역에 제한을 둔 것과 달리, 본 논문에서는 지역을 가리지 않고 사용이 가능하며 방언-표준어간 번역도 가능하다.

III. The Proposed Scheme

1. 방언-표준어간 변환 방법

방언-표준어간 번역하는 방식으로는 두가지 방법이 있다. 첫 번째로는 우리말샘 API를 이용하여 해당 데이터를 수집하는 방식과 두 번째로는 홈페이지를 크롤링하여 데이터를 수집하는 방식이다.

표 1은 입력받은 방언을 표준어로 변환하는 함수로 DialogFlow에

서 전달받은 액션 값을 변수에 담아 해당 액션이 이름이 'regimal_Name'과 같을 시, 전달받은 값에 들어 있는 방언 단어를 searchRegimal 메서드로 넘겨준다. searchRegimal 메서드는 우리말 샘 API로 사전검색 Request 요청을 보내 해당 방언의 표준어, 뜻, 고유번호를 얻고 해당 고유번호로 다시 내용검색 Request 요청을 보내어 해당 방언을 사용하는 지역이 어느 지역인지를 찾는다. 그 후 원하는 데이터로 가공한 뒤, 카카오톡과 연동되어있는 채널로 그림 1과 같이 전송되어 보여준다.

Table 1. Regimal-Standard Translation Function

```
req = request.get_json(force=True)
action = req['queryResult']['action']
if(action == 'regimal_Name'):
name =req['queryResult']['parameters']
dialect_result=hello.searchRegimal(Regimal)
return {'fulfillmentText': ...}
```

데이터를 크롤링하여 수집하는 방법으로는 방언 문장을 번역하기 위해 본 논문에선 파이썬Selenium 모듈을 사용한다. 데이터를 수집하기위해 부산대학교에서 운영 중인 맞춤법 검사기 홈페이지에 번역할 방언문장을 입력하고 검사하기 버튼을 누르면 검사 결과페이지창으로 넘어가는데, 해당 페이지를 크롤링하여 표준어로 변환된 단어들을 가져와 정제한 후 카카오톡에 연동된 채널로 전송하여 보여준다.

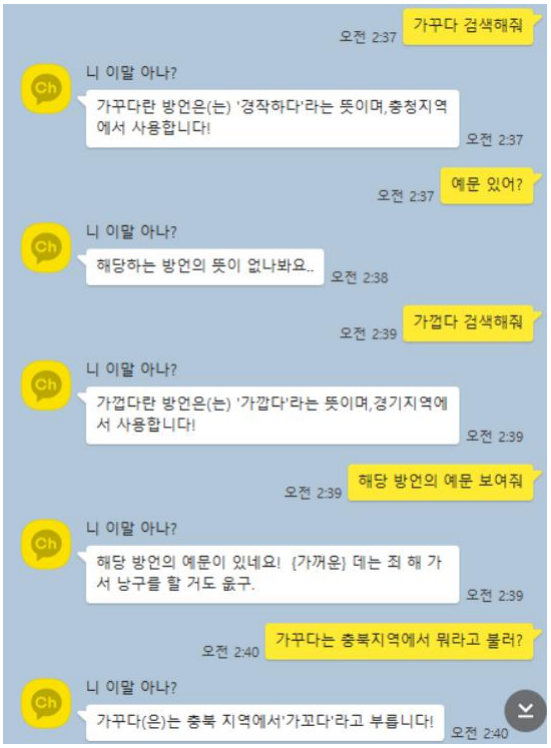

Fig. 1. Result Screen

2. 결과화면

그림 1은 사용자가 방언 단어를 입력 시에 해당 방언에 맞는 뜻과 사용하는 지역을 보여주는 결과화면이다. 해당 방언에 예문이 존재할 시, 해당 방언이 사용된 예문을 보여주고, 예문이 존재하지 않을 시엔 '해당 방언이 없나 봐요'라는 메시지를 출력한다. 또한 '랜덤'이라는 키워드를 입력하면 무작위로 선택된 5개의 방언단어를 보여준다. 마지막으로 표준어를 특정 지역에 맞추어 전달 시에 해당 지역의 방언으로 번역하여 알려주는 기능도 가지고 있다.

IV. Conclusions

본 논문에서 해당 지방 언어를 사용하는 사람과 대화 시 특정 방언을 이해하기가 힘들다는 점을 사용자가 발화한 단어를 분류한 후에 외부 API 연동을 통하여 원하는 데이터를 간추려 낸 후, 카카오톡 챗봇을 통해 사용자에게 표준어로 번역된 내용을 보여준다.

사용자는 표준어로 번역된 내용을 보며 상대방이 대화하려는 의도를 파악한 후 대화를 이어갈 수 있다.

REFERENCES

[1] Kiyoung Yoon, Jungyeon Sohn, Seonggyu Park, Hojae Lee, Chansong Jo, Yuri Lee, Myoung-Wan Koo(2019). 2019.12, 1406-1408 (3 pages)