

텍스트 유사도 기반의 대학 입시 상담 챗봇

이세훈*, 차현석^o, 전찬호*, 백영태**

^o인하공업전문대학 컴퓨터시스템과

**김포대학교 멀티미디어과

e-mail: seihoon@inhac.ac.kr^o, sk4549131@naver.com*, jcho109@daum.net*, hanna@kimpo.ac.kr**

College Admissions Consultation Chatbot based on Text Similarity

Se-Hoon Lee*, Hyun-Suk Cha^o, Chan-Ho Jeon*, Yeong-Tae Baek**

^oDept. of Computer Systems and Engineering, Inha Technical College

**Dept. of Multimedia, Kimpo University

요약

본 논문에서는 입시상담을 위한 챗봇 시스템을 텍스트 유사도 기반으로 개발하였다. 텍스트를 인지하여 답변을 제공해 주는 방식이며 실시간을 요하는 데이터들은 크롤링한 데이터를 가공을 한 후 사용자에게 대담을 해주고 사용자가 답변에 얼마나 좋은 정보인지 체크하여 그에 맞는 답변을 내어 준다. 사용자의 텍스트를 인식하는 것은 텍스트 유사도를 이용하여 정확하게 인지하고 사용자의 질문과 답변을 서버 DB에 저장을 하여 비슷한 질문이 있을 경우 저장된 답변과 평점을 이용하여 답변을 제공한다.

키워드: 챗봇(Chatbot), 텍스트 유사도(text similarity), 입시상담(Entrance Counseling)

I. Introduction

대부분의 대학에서는 챗봇을 이용한 입시상담이나 질문을 사람이 처리하고 있다. 아마도 챗봇에 대한 신뢰성이 떨어지기 때문이 아닐까 생각한다. 그 이유를 학교 입학팀과 상담을 통해서 알게 된 것은, 우려하는 것이 “잘못 된 정보를 주면 학교에 타격이 크다”였으며, 그런 이유에서 챗봇 도입에 신중하다.

그러나, 신입생 및 재학생과 우리 대학의 입학 목적을 잡는 학생들을 대상으로 입학상담이나 학교에 관한 답변을 해주는 챗봇을 만들면, 인건비를 줄이는 동시에 입학 문의 학생들의 불편함도 해소할 수 있을 것으로 기대한다. 본 연구에서는 입시 시즌에 집중적으로 전화등을 통해 질문하고 있는 내용을 중심으로 질의, 응답이 가능한 챗봇을 개발한다. 입학관련 및 면접관련 질문들이 많은데 그 이유는 수시 및 정시 시즌이 되면 지원자가 입시에 궁금한 것을 전화로 물어보기 때문에 입학 시즌에는 상담원 인원이 부족해서 일반생을 고용하여 상담을 해준다는 말을 들었다. 우리는 이 고질적인 문제를 해결하고자 한다.



Fig. 1. Flow of Chatbot system

II. Implementation of the Chatbot

챗봇 시스템은 Fig.1과 같이 스마트폰 앱에서 질문을 서버로 전송하고, 서버에서는 받은 질문의 한글 단어간 유사도를 분석하여 그에 맞는 대담을 파싱하고, 파싱한 값을 스마트폰 앱으로 전송한다. 챗봇 서버는 플라스크와 MySQL 데이터베이스를 이용하였다.

1. Levenshtein Distance

한글 검색어 간 유사도를 측정하기 위해서는 평가기준이 필요하다. 그 평가기준이 될 수 있는 것이 Levenshtein 함수인데 이 함수는 문자열을 UCS4로 변환한 뒤에 거리값을 계산해서 유사도를 측정하는

역할을 한다. 여기서 거리값은 문자열 매트릭스 내에서 두 단어 사이의 특정 단어를 다른 단어로 바꾸는데 필요한 최소 편집 횟수를 말한다.

	ϵ	x	y	a
ϵ	0	2	7	13
a	7	8	8	7
b	8	9	9	8
x	11	8	12	11

Fig. 2. Distance Operating principles

2. Web Parsing

Python library인 requests와 BeautifulSoup을 사용하여 실시간 크롤링을 구현하였다.

텍스트유사도를 통해서 나온 결과 값을 이용하여 미리 정의된 URL에 값에 맞는 미리 지정된 주소 변수(String)에 추가하고 requests로 html 소스를 읽어서 soup.select로 필요한 부분의 지정한 태그를 사용하여 필요한 값만 파싱하여 사용자에게 전달해 주는 방식이다.

III. Experiment

질문 유사도 실험에서는 Python distance 라이브러리를 사용하여 텍스트간의 유사도를 체크하여 가장 비슷한 텍스트를 찾아서 그에 맞는 최적화된 답변을 제공해주는 방식으로 구현하여 실험하였다. Fig. 1에서 4번째 문장 단어가 가장 유사함을 보여주고 있다.

구글의 음성인식 기술인 STT를 사용하여 안드로이드에서 구현하였다. 오디오를 텍스트로 변경하고 노이즈 캔슬링을 시켜서 잡음을 없애주므로 소음이 심한 장소에서도 음성 인식률이 높다.

```

질문 : 수시1차 일반 추천 컴퓨터시스템과 경쟁률 알려줘
word_list=["영어 특기자", "중국어 특기자", "일본어 특기자", "일반 주간",
"전문 주간", "일반 야간", "전문 야간", "특기자", "동아촌", "수급자"]

distance값 : [25, 25, 24, 21, 23, 22, 24, 26, 26, 25]
    
```

Fig. 2. Text Similarity

Flask를 사용하여 서버 API를 개발해서 앱에서 AsyncTask방식으로 stream post를 사용하여 서버로 텍스트를 보내면 만들어둔 API에서 질문을 인식하여 그에 맞는 답변들을 다시 앱으로 전송한다.

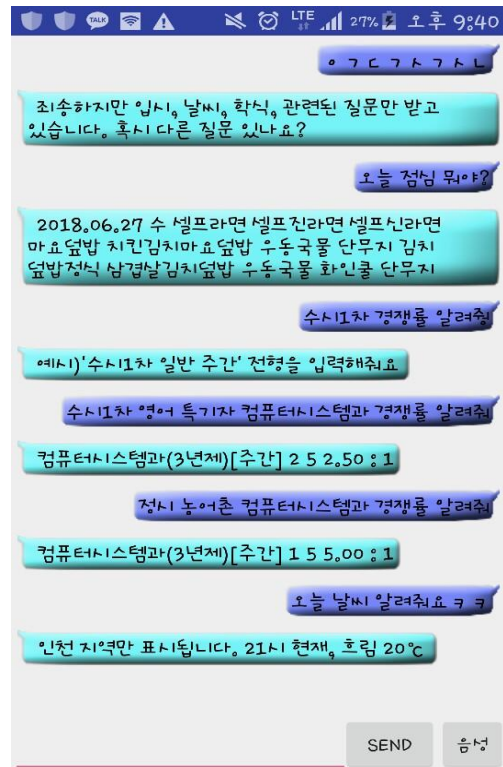


Fig. 3. Entrance Counseling Dialog

실시간을 요하는 그 외의 유사한 질문은 전에 대답했던 답변의 데이터들이 이용해서 답변을 제공한다. 카카오톡에서도 저장된 데이터로 고객상담을 한다고 한다. 그래서 질문과 답변을 DB(Mysql)에 저장을 시킨다.

IV. Conclusion

본 연구에서는 입시 시즌에 집중적으로 문의가 있는 내용을 중심으로 대화식 입시 상담 챗봇을 개발하였다. 개발한 챗봇은 한글 단어간의 유사도를 통해 Levenshtein를 통해 하며, 플라스크 프레임워크를 이용하였다. 유사도를 통해 나온 결과를 미리 정의된 사이트를 실시간 클로링해서 원하는 정보를 얻을 수 있는 구조를 개발하였다.

REFERENCES

[1] Levenshtein Distance, https://www.python-course.eu/levenshtein_distance.php
 [2] BeautifulSoup, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>