

# 자연어처리를 통한 온라인 학습 플랫폼 사용자 질의 답변 및 Word cloud를 활용한 키워드 시각화

유경록<sup>○</sup>, 정영섭<sup>\*</sup>

<sup>○</sup>한국교육개발원 미래교육연구본부,

<sup>\*</sup>충북대학교 컴퓨터공학과

e-mail: nox951021@kedi.re.kr<sup>○</sup>, ysjay@chungbuk.ac.kr<sup>\*</sup>

## Answering User Queries on Online Learning Platforms through Natural Language Processing and Keyword Visualization Using Word Cloud

Kyong Rok Yoo<sup>○</sup>, Young-Seob Jeong<sup>\*</sup>

<sup>○</sup>Korean Educational Development Institute,

<sup>\*</sup>Dept. of Computer Engineering, Chungbuk National University

### ● 요약 ●

최근 온라인 학습의 비중이 증가함에 따라 온라인 학습 서비스의 일부인 온라인 상담 부분도 비례하여 증가하고 있으며, 많은 상담량으로 인해 상담 서비스의 품질이 저하되고 답변의 속도, 효율성도 감소하는 문제가 발생한다. 국내 교육기관에서는 서비스 개선과 사용자 맞춤형서비스를 제공하기 위해 다양한 연구를 진행하고 있으며 민원을 처리하는 챗봇 등 자동 답변 서비스 도입을 추진하고 있다. 챗봇 및 자동 답변 서비스는 서비스 제공자 입장에서 저예산으로 단순한 질문에 대하여 신속하고 효율적인 서비스를 제공할 수 있으며 서비스 이용자는 즉각적인 답변과 유사한 답변 예시를 확인함으로써 질문을 빠르게 해결할 수 있는 장점이 있다. 국가 공공기관에서 제공하는 학습 서비스는 단순하고 반복적인 문의가 많고 정형적인 질의응답이 주로 등록이 되고 있다. 자동 답변 서비스는 이런 문제점을 해결할 수 있는 대안이 된다. 서비스 이용자가 등록한 문의를 기반으로 학습한 답변 서비스는 담당자의 반복된 업무처리 경감과 사용자의 답변감소, 일관된 답변 처리로 서비스 품질개선에 큰 영향을 줄 수 있다. 본 연구에서는 사용자의 질문에 효율적인 답변 및 민원 처리 서비스를 제공할 수 있는 방법을 제시하며, 관리자의 업무능력 향상과 효율성을 위해 기간별 키워드 빈도수를 계산하여 Word cloud를 생성하여 제공함으로써 사용자들에게 일정 기간 내 빈도수가 높은 키워드 관련 공지 및 안내를 할 수 있도록 한다.

**키워드:** 온라인 상담(online consulting), 민원처리(civil affair handling), 자동 답변서비스(automatic answering service), 단어구름(word cloud)

## I. Introduction

한국교육개발원은 한국교육이 당면한 제반 문제를 합리적으로 해결하는 혁신적인 교육체제를 개발함으로써 교육의 발전에 기여하는 국가교육 기관으로 학생들에게 오프라인과 온라인으로 학습을 제공하고 있다. 최근 코로나19로 인해 학생들의 교육이 비대면으로 대체 및 확대되었고 그에 따라 한국교육개발원에서는 학생들에게 온라인 학습 서비스 제공을 확대하는 추세이다. 홈페이지 사용자들은 오류 및 민원, 상담이 필요한 경우, 문의 게시판에 접속하여 문의 글을 작성할 수 있으며 이를 관리자가 직접 읽고 답변하는 방식으로 운영하

고 있다. 온라인 학습이 크게 확대됨과 동시에 온라인 상담 문의도 비례하게 증가하고 있으며 이에 따라 학습시스템 사용 문의, 시스템 불편 민원 반복적, 기술적 답변에 대한 담당자들의 업무 부담, 지연된 답변 등 문제점이 발생하였다.

본 연구에서는 공공서비스의 품질 개선을 위해 사용자 측면의 기능만 고려하는 것이 아닌, 서비스 관리자의 업무능력 향상과 효율성을 높이는 점을 고려한다. 사용자 만족도 향상을 위해서는 사용자의 질의문과 유사한 게시글을 찾아주는 기능을 제공하는 한편, [6]에서와

같이 관리자 업무능률 향상을 위하여 해당 기간 내 가장 많이 등록된 키워드 10개를 Word cloud로 생성하여 사용자들이 최근 어떤 키워드로 궁금해하고 문의하는지 파악하여 공지 또는 개선할 수 있도록 한다. 본 연구에서 제안하는 서비스 개선 방법의 유효성 파악을 위해 한국교육개발원 사업담당자 15명의 관리자에 대한 시스템 서비스 만족도 조사를 실시하였다. 조사의 결과는 만족도 82.66%, 도움이 되는 정도는 73.33%로 서비스 개선 효과가 있는 것으로 나타났다.

## II. Related Works

현재까지 국내 기업, 공공기관, 학교 등 대부분의 기관에서 사용자의 문의 및 민원을 해결하고자 1:1문의 게시판, 콜센터를 운영하고 있다. 게시판과 콜센터는 사용자에게 관리자의 직접적인 확인을 통해서만 답변을 받을 수 있다는 단점이 있어 이를 해결하고자 24시간 즉시 답변이 가능하고 사용자에게 다양한 해결책을 제시해주는 챗봇을 도입하고 있다.

자연어처리 기술을 공공서비스 품질 개선에 적용하려 시도한 과거 연구들이 있었는데, 사용자의 질의문(query)과 유사한 기존의 게시글을 찾아서 제공해주는 챗봇 시스템을 제공하거나 [1], 오픈된 챗봇 제작 도구들을 활용하여 공공서비스용 챗봇을 개발하는 과정 일체에 관하여 기술하는 연구가 있었다 [2]. 또한 SIRI, Alexa와 같은 음성 인식 기반 지능형 가상 비서(Intelligent Virtual Assistance)나 텍스트 기반의 챗봇[3]과 Q&A 게시판 등을 이용한 질의응답 서비스는 지식 정보를 교류하는 사용자 참여 서비스로, 사용자의 질문에 대한 다른 사용자들의 답변이 축적된 지식 데이터에 대한 검색 기능을 제공하는 서비스에 대한 연구가 있었다 [4].

지능형 가상비서는 미리 정의된 콘텐츠나 서비스에 특화된 시스템으로서, 사용자가 원하는 서비스를 적시에 제공하고 일정 관리 등과 같은 인간 비서가 수행하는 업무를 효율적으로 대행할 수 있는 지능형 시스템이다. 지능형 가상비서는 무선네트워크 속도가 개선되고, 음성 인식 성능이 향상되기 시작하고, 클라우드 컴퓨팅이 등장과 새로운 웹서비스들이 매우 많이 등장하면서 실현 가능해졌다 [5]. 이러한 연구들은 자연어처리 기술(NLP)을 활용하여 텍스트사이의 similarity를 계산함으로써 관련 문서를 제공한다는 점에서 본 연구 내용과 유사한 측면이 있으나, 과거 연구들은 서비스를 이용하는 사용자를 위한 기능에만 집중하고 있었다는 점에서 본 연구와의 차이점이 있다. 가령, 사용자가 원하는 답변을 정확하고 빠르게 찾을 수 있도록 돕기 위한 다양한 기법들(예: Collaborative filtering)이 시도되었으나, 앞선 연구들은 관리자들이 업무 효율을 증대시키기 위한 시도는 비교적 소홀했다고 볼 수 있다.

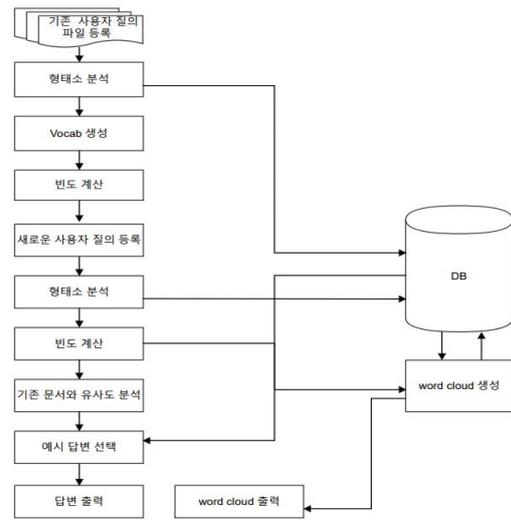


Fig. 1. Overall process of the proposed system.

## III. Method

본 연구에서는 Fig. 1 와 같이 공공서비스 게시글이 Database (DB)에 적재되어있고 이에 대한 접근이 가능한 상황을 가정한다. 제안하는 시스템은 DB에 적재된 자연어 게시글을 대상으로 자연어처리 기술을 적용함으로써 Word cloud를 생성하는 모듈과 유사 게시글을 찾아주는 모듈로 구성되어있으며, 자연어 데이터 전처리 파트는 공유하도록 설계되어 있다.

DB에 적재된 게시글들은 {제목, 내용, 작성자, 일시} 정보를 포함하고 있으며, 자연어 전처리 파트는 게시글의 ‘내용’만을 사용한다. 이는 게시글의 제목을 의미없이 작성하는 경우가 빈번하게 발생한다는 사실과 게시글의 제목과 내용이 반드시 일치하지는 않는다는 점을 반영하기 위함이다. 중복된 게시글을 제거한 후, 게시글 ‘내용’에 대하여 형태소 분석을 적용함으로써 게시글 내용은 형태소 및 Part-of-Speech (PoS) 태그의 나열로써 표현이 가능해진다.

임의의 게시글  $d$  에  $N$ 개의 형태소가 포함되어있을 때 해당 게시글에 대한 형태소 분석에 의한 결과물은  $\{(m_1, p_1), (m_2, p_2), \dots, (m_N, p_N)\}$  로써 나타낼 수 있으며, 이 때  $m_i$ 는  $i$ 번째 형태소,  $p_i$ 는  $i$ 번째 형태소에 대한 PoS 태그를 의미한다. 이처럼 형태소 및 PoS 태그 쌍의 나열로 표현된 게시글에 대하여  $p_i$  가 Noun, Adjective, Verb, Adverb 에 속하는  $m_i$  들만 추려내도록 하였으며, 이에 대한 vocabulary 를 구축하였다. 그 후, 1-gram Bag-of-Word (BoW) 방식에 기반하여 각 게시글을 Document-Morpheme Matrix (DMM) 으로 변환한다. 게시글의 개수가  $D$ 개, vocabulary 크기가  $V$ 라고 할 경우, DMM 은  $D \times V$  크기의 행렬이 되며, DMM 을 이루는 값들은 각 형태소가 게시글에서 등장한 빈도를 나타낸다.

Word cloud 모듈은 전처리 파트에 의해 변환된 DMM 을 사용함으로써 빈도수 기반으로 단어 군집을 시각화해준다. Word cloud는 관리자들에게 최근 게시글들에 내재된 사용자들의 needs 에 대한 트렌드를 파악하기 쉽도록 해줌으로써 업무능률 향상에 기여하게 된다.

유사 게시글 검색 모듈은 사용자로부터 입력된 임의의 새로운 게시글 제목을 입력으로 취한다. 새로운 게시글의 내용이 아닌 제목을 입력으로 취하는 이유는, 사용자가 내용 입력을 마치기 전에 미리 유사한 게시글을 제공하도록 함으로써 사용자가 불필요한 내용 작성을 하지 않도록 하기 위함이다. 물론, 사용자가 작성한 게시글의 제목이 내용과 반드시 일치하지는 않겠으나, 제목과 관련 깊은 게시글이 추천되는 기능의 존재를 인지한 사용자들은 추후 서비스를 이용할 시, 게시글 제목을 신중하게 작성하게 될 것이라고 기대할 수 있을 것이다. 새로운 게시글 제목을  $d_{TITLE}$  이라고 한다면, 이에 대하여 앞서 설명한 전처리 파트를 적용함으로써 V-dimensional frequency vector 를 얻게 된다.

과거의 게시글들이 모두 V-dimensional frequency vector 이고 새로운 게시글 제목  $d_{TITLE}$  또한 동일한 vector 형태를 갖추었으므로, 이들 사이에 임의의 유사도 함수  $f(d_{TITLE}, d_i)$  를 적용함으로써  $i$ 번째 게시글  $d_i$  의 내용과 게시글 제목  $d_{TITLE}$  간의 유사도 점수를 계산할 수 있다. 계산된 점수들 중에서 가장 높은 점수를 가진  $k$ 개의 게시글 목록을 취하여 사용자에게 제시하도록 한다.

Table 1. Data statistics.

Item	Value
Total # documents	4,062
Total # vocabulary words	3,855
Average # words per sentences	6.37



Fig. 2. Word cloud that analyzes user queries.

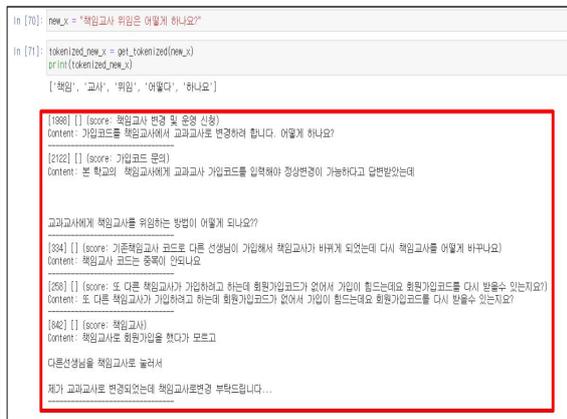


Fig. 3. Automatic response service.

## IV. Experiments

2016년 1월부터 2023년 10월까지 한국교육개발원의 1:1문의 게시글에 등록된 사용자(학생, 교사)들의 홈페이지 이용 문의, 민원, 오류, 학습 관련 질문 데이터를 활용하였다. 총 4,062개의 사용자 질의 텍스트로부터 vocabulary 내 3,855개 단어를 생성할 수 있었으며, 이를 Table 1에서 요약해주고 있다. 사용자 질의 텍스트에는 중복이 종종 발견되었다. 예를 들어 전학 관련된 질의문의 경우, ‘전학 처리는 어떻게 하나요?’ 등을 포함하여 122개, 교과 배정 관련된 질의문은 ‘교과 배정은 어디서 받나요?’를 포함하여 314개, ‘학기 기간은 언제까지 인가요?’와 동일한 질의문은 166개가 중복되어 있음을 확인하였다.

사용자 질문 데이터에 대한 형태소 분석을 통해 얻은 PoS 태그가 ‘명사’에 해당되는 것들만 추려서 Fig. 2 과 같이 시각화를 하였다. 이를 통해, 학습 과정과 관련된 특정 ‘회차’에 대한 ‘신청’ 문의, 강의 참여를 위한 ‘아이디’ 관련 문의가 지속적으로 발생하고 있음을 알 수 있다. Fig. 3는 사용자의 게시글 제목에 대한 Top-3개의 유사 게시글들을 볼 수 있는 프로그램의 예시 화면을 보여주고 있는데, 이를 Word cloud와 함께 활용하게 되면 반복된 질의문이 줄어들게 됨으로써 사용자 질의문의 토픽 변화 양상을 관리자가 확인할 수 있게 될 것이다.



Fig. 4. Satisfaction survey questions of the proposed method.

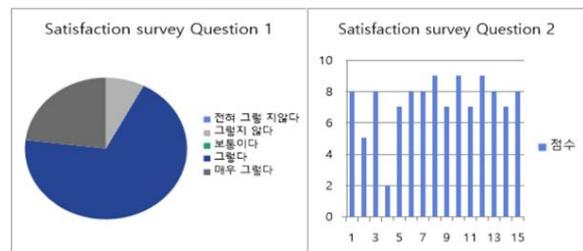


Fig. 5. Satisfaction survey of the second question.

한국교육개발원의 사업담당자 및 시스템 관리자 15명을 선정하여 프로그램 사용 후 Fig. 4 와 같이 서비스 만족도 조사를 실시하였다. 프로그램 만족도에 관한 문항, 장점 및 단점 개선사항과 기타 의견으로 문항을 구성하여 질의하였다.

Fig. 5는 조사 결과를 요약하여 보여주고 있는데, 프로그램 사용 만족도는 82.66%였고, 도움이 되는 정도는 73.33%라는 결과를 얻었다. 기타 의견으로는 Word cloud 사용은 관리자 입장에서 어떤 특정 질문이 어떤 시기에 물리는지 파악이 가능하고 이를 통해 사전에 홈페이지 팝업 안내 또는 공지사항 안내를 통해 선제적 대응이 가능하며 가장 많이 등록된 키워드를 한눈에 파악 할 수 있었다는 의견이 있었으며, 반복적 질문에 대해 똑같은 답변을 하지 않고 사용자에게 빠른 답변을 제공할 수 있다는 의견이 있었다. 프로그램 개선 의견으로는 1차로 사용자의 질문을 받은 후 2차로 키워드를 사용자에게 제시하여 선택하게 하는 프로세스를 통하여 답변의 정확도를 높일 수 있을 것이라는 의견이 있었다.

## V. Conclusions

본 연구에서는 Word cloud와 사용자 질문에 대한 유사 게시글을 자동으로 제공하는 프로그램을 개발하여 한국교육개발원 관리자들에게 설문조사를 실시하였고, 본 프로그램이 관리자들을 만족시키고 도움을 줄 수 있음을 보였다. 본 연구에서는 한국교육개발원의 일부 데이터만을 사용하여 실험하였으나, 향후 긴 기간 동안 축적된 전체 게시글을 대상으로 개인정보 및 식별정보를 마스킹하여 보다 정확한 검증이 가능할 것으로 기대한다.

## ACKNOWLEDGEMENT

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2020R111A3053015). 이 논문은 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5A2A03052880)

## REFERENCES

- [1] Jin-Hyun Kim, Seong-Wook Jeong, Sin-Jae Kang. "자연어 처리와 오픈소스 SW를 활용한 학사민원 챗봇 시스템." 예술인문사회융합멀티미디어논문지 8.3 (2018): 907-914.
- [2] Kim Gye-Soo, & Kwak jiyoung (2023). Development of AI chatbot to improve public service quality and increase administrative satisfaction -Focused on Tourism AI Chatbot-. The Academy of Customer Satisfaction Management, 25(1), 79-96.
- [3] Dong-ah Park (2017). A Study on Conversational Public Administration Service of the Chatbot Based on Artificial Intelligence. Journal of Korea Multimedia Society, 20(8), 1347-1356.
- [4] Ki-Dong Ryu, Jong-Pil Park, Young-min Kim, Dong-Hoon Lee, & Woo-Je Kim (2019). Development of AI-based Real Time Agent Advisor System on Call Center - Focused on N Bank Call Center -. Journal of the Korea Academia-Industrial cooperation Society, 20(2), 750-762.
- [5] 이현아, 유동현, "Q&A 문서의 검색 결과 요약을 활용한 질의응답 시스템," 정보처리학회논문지. 소프트웨어 및 데이터 공학 3.4 (2014): 149-154.
- [6] Susan Ramlo (2011). 워드 클라우드를 사용하여 Q 방법론의 데이터와 결과를 시각적으로 보여주기. Journal of Human Subjectivity, 9(2), 95-108.