

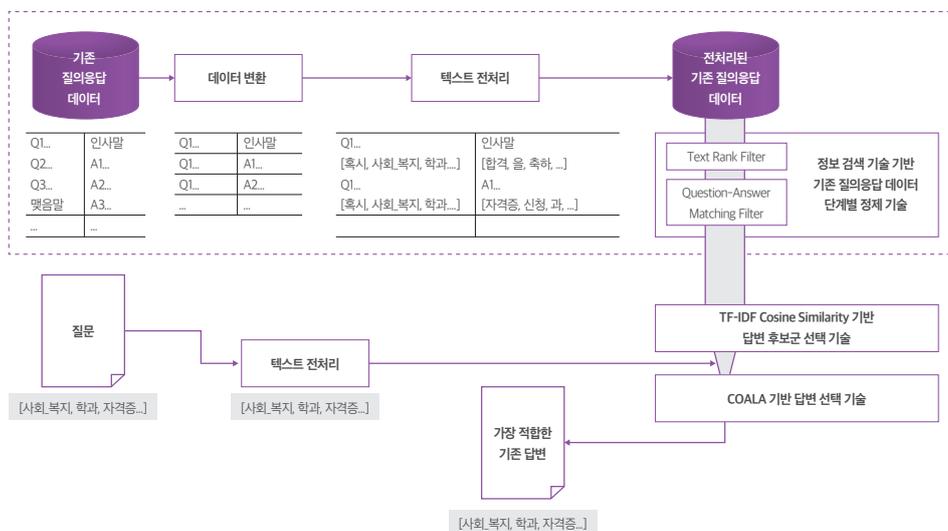
## Community QA Chatbot & ChatHelper

### 01. 개요

- 본 연구에서는 소량의 정제되지 않은 데이터로부터 자동으로 최신 챗봇 기술을 적용한 챗봇을 구축할 수 있도록 챗봇 기술 및 챗봇 구축 도구 개발을 진행하였습니다.
- (챗봇 기술 연구) 반복적인 질의 응답이 요구되는 다양한 분야에서 Q&A 게시판 데이터 등 정제되지 않은 소량의 질의 응답 데이터로부터 챗봇을 구축할 수 있는 챗봇 기술 연구를 진행하였습니다.
- (챗봇 구축 도구 개발) 배경 지식이 없는 일반 사용자가 쉽게 사용할 수 있는 챗봇 구축 도구의 개발을 목표로 하여 개발하였습니다.

### 02. 연구 내용

- 본 기술은 Q&A 게시판 데이터 등 소량의 정제되지 않은 데이터로부터 검색 기반 방법을 적용하여 딥러닝 기반 자동 질의 응답 시스템을 구축하였습니다.
- 챗봇 기술 개발 연구를 위해 서울 사이버 대학교의 입학상담 게시판 질의응답 데이터 1248건을 수집하였으며, 각 질의응답 데이터는 질문 게시물과 상담자가 답변한 답변 게시물의 쌍으로 구성됩니다.
- 본 연구에서 제안한 Community QA 챗봇의 구조는 1) 데이터 전처리 2) 질의응답 데이터 단계별 정제 3) 답변 후보군 선택 4) 딥러닝 기반 답변 선택 모듈과 같이 총 4개의 세부 모듈로 나뉩니다. 전체 모델 구조는 아래 그림과 같습니다.



- (데이터 전처리) 정제되지 않은 질의 응답 데이터를 챗봇 서비스에 적합한 형태로 전처리를 진행하였습니다. 실험에 사용한 게시판 데이터 등의 정제되지 않은 데이터는 하나의 질의응답 쌍에 여러 개의 질문 내용과 이에 대한 대답이 혼재되어 있을 수 있습니다. 반면 챗봇 서비스에서 질문자가 제시하는 질문은 하나의 질문 내용으로 이루어지며 이에 대한 답변을 제시해야 합니다.
- (질의 응답 데이터 단계별 정제) 1) Text Rank<sup>1</sup> 알고리즘을 통해 각 문서 내 문장의 중요도를 판단하여 각 답변의 중요도를 계산한다. 이에 따라, '메일로 답변 드립니다.'와 같이 다른 답변들과의 유사도가 낮은 무의미한 답변들을 제거하였습니다. 2) 질문과 답변의 TF-IDF Cosine Similarity를 계산한 결과가 일정 수준 이하일 경우 질문과 답변의 키워드가 일치하지 않는 것으로 판단하여 제거하는 선별 단계 또한 진행하였습니다.

<sup>1</sup>Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. In EMNLP 2004.

- (답변 후보군 선택) 주어진 질문과 기존 질의 응답 데이터의 질문의 TF-IDF Cosine Similarity를 계산하여 주어진 질문과 유사한 질문들을 선택하고, 이 질문들에 대한 기존 답변들을 답변 후보군으로 선택합니다.
- (딥러닝 기반 답변 선택 모듈) 딥러닝을 기반으로 하여 답변 후보 군 중에 질문과 가장 근접한 답변을 추론합니다. 딥러닝 모델로는 CQA에서 소량의 데이터로도 속도와 성능 면에서 뛰어난 COALA<sup>2</sup> 모델을 기반으로 하였습니다. COALA는 질문의 각 단어 토큰의 관련성을 포착하기 위해 N-gram과 CNN을 사용하여 Aspect Identification을 진행하였습니다.
- 텍스트 전체 문맥을 파악하는 대신 여러 부분의 지역적 문맥 정보를 파악하도록 모델링함으로써 긴 답변 텍스트에 대한 표현력을 높이고 요구되는 학습 데이터의 양을 절감하였습니다. 최종적으로 Relevance Matching, Aggregation과 같은 모듈을 통해 각 후보 답변의 관련성 점수를 계산합니다.
- 아래 예시는 본 연구 및 기술을 적용하여 구축한 자동 질의 응답 시스템의 결과 예시입니다.

The screenshot shows a light blue background with two white text boxes containing text. Each text box is preceded by a dark blue header box with white text. The first header is '학사편입 전형 기준이 궁금합니다.' and the text below it discusses admission criteria for domestic universities. The second header is '심리상담학부 수업 온라인인가요' and the text below it discusses online classes at Seoul National University of Education.

**학사편입 전형 기준이 궁금합니다.**

학사편입 전형은 국내의 4년제 대학을 졸업(예정)하고 학사학위를 취득하거나 또는 법령에 의하여 이와 동등이상의 학력이 있다고 인정되면 지원자격에 학사편입전형으로 지원하실 수 있습니다.(2016.12.06) [더보기>>](#)

**심리상담학부 수업 온라인인가요**

2. 서울사이버대학교는 시험과 수업이 모두 온라인으로 진행됩니다. 다만, 학과별 특성에 따라 대면수업, 실습수업, 오프라인 특강 및 학과활동, 동아리 활동, 스터디모임 등 다양한 오프라인 콘텐츠가 진행되고 있습니다. 특히 상담심리학과는 경우 대면수업 및 실습 수업이 오프라인으로 진행되며, 본교 서울사이버대학교는 전국에 심리상담센터를 두고 있어, 심리상담학부 재학생과 졸업생의 실무능력 강화를 위한 실습 및 수련기관 역할을 하고 있습니다. 본교 상담심리학과 홈페이지에 상담심리학과의 수업 내용 및 프로그램 확인해보시길 바랍니다.  
[http://counsel.iscu.ac.kr/\(2017.02.10\)](http://counsel.iscu.ac.kr/(2017.02.10)) [더보기>>](#)

<sup>2</sup>Rücklé et al. (2019). COALA: A neural coverage-based approach for long answer selection with small data. In AAAI 2019.