뭐든 내게 물어봐

#회의록 #Closed Domain QA #Extractive QA #RoBERTa

NLP-09 (MNM 조)

boostcamp aitech

TABLE OF CONTENTS

- 1. Introduction
- 2. Data
- 3. Reader & Retrieval
- 4. Serving
- 5. Conclusion

Appendix.

1. Introduction

안녕하세요 MNM 팀입니다

AI의 A부터 I까지 모든 것을 경험할 준비가 된 열정 가득한 사람들이 모인 팀입니다

<김태일> github.com/detailTales



<문찬국> github.com/nonegom



<이재학> github.com/wogkr810



재학> <한나연>





<하성진>

github.com/maxha97



주제 선정 기준

저희는 최종 프로젝트 주제 선정 시 3가지 기준을 정했습니다.

- 내가 겪은 어려움/불편함을 해결할 수 있는가?
- 부스트 캠프에서 배운 내용을 활용해 AI 프로젝트의 전 과정을 모두 경험할 수 있는가?
- 주어진 기간(3주) 내에 완성할 수 있는가?

프로젝트 소개

주제

회의록을 활용한 Closed Domain Question Answering (CDQA)

문제 정의

클로바 노트* 등을 활용해 회의록을 쉽게 기록할 수 있게 됐지만, 정보 검색이 어렵다는 문제 발견

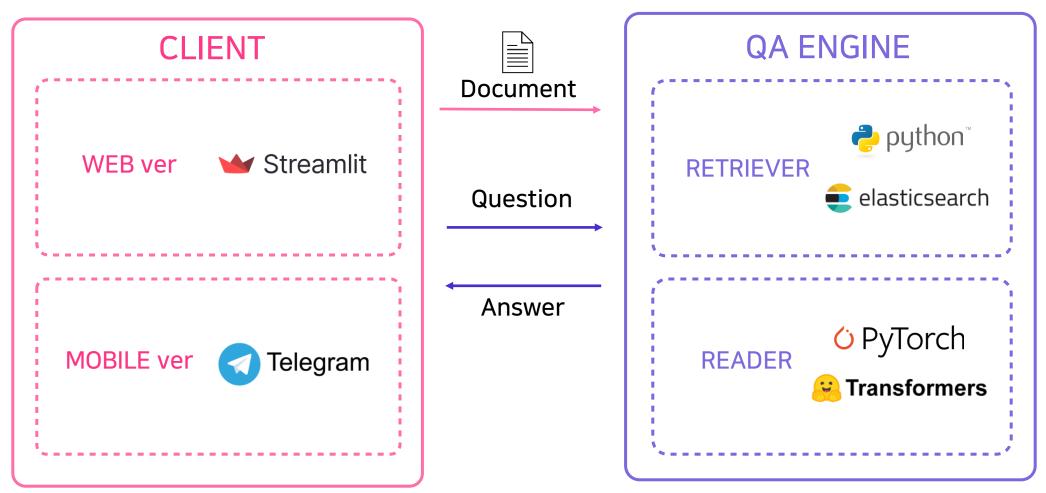
개발 목표

• 사용자의 회의록 코퍼스에서 궁금한 질문을 주고받을 수 있는 회의록 QA 모델 제작

*클로바노트: 네이버 하이퍼 클로바에서 제공하는 Speech To Text 서비스

CDQA 란?

특정 도메인의 데이터베이스에서 질의응답을 수행하는 task



boostcamp aitech

2. Data

어떤 데이터셋을 사용했는가?

클로바 노트 회의록의 경우 학습에 필요한 회의록의 절대적인 양이 부족

데이콘 회의 녹취록 요약 경진대회용 의회 데이터 사용 결정

- 저작권 문제가 해결된 데이터
- 안건 별로 구분되어 태깅 용이
- 회의의 형식을 띄고 있으며 질문 유형화 가능
 - 다양한 회의에서 공통적으로 나올 수 있는 질문으로 유형화

데이터 예시

★★ DACON | SON File

id: 1000 title: "제207회 완주군의회(임시회) 제 1 차 본회의회의록" region: "완주" num_agenda: 4

- ▼ context:
- - 0: "의석을 정돈하여 주시기 바랍니다. 성원이 되었으므로 제207회 관련 사항에 대한 보고가 있겠습니다. 의사팀장은 보고하여 주시기
 - 1: "의사팀장 이은미 입니다. 의회관련 사항을 보고 드리겠습니다. 조 제2항의 규정에 의거 서남용 의원님 외 4인의 의원님으로부터 니다. 다음은 의안접수 및 회부사항입니다. 먼저 의원발의 안건입 향토문화유산 보호 및 관리 조례안, 완주군 싸움소 육성 지원에 결 로부터 제출된 안건입니다. 삼례문화예술촌 설치 및 운영조례 일본 배부해 드린 유인물을 참조하여 주시기 바랍니다. 이상 보고를 마
- ► AGENDA 2:
- ▶ AGENDA_3:
- ► AGENDA 4:
- ▼ label:
- ▼ AGENDA 1:
- ▶ evidence:
 - summary: "제207회 완주군의회 임시회 제1차 본회의 개의 선포."
- ► AGENDA 2:
- ► AGENDA 3:



haystack annotation Tool

"3_제207회 완주군의회 임시회는 Question:

누가 요구했어"?

Annotation Document:

@제207회 완주군의회(임시회) 제 1 차 본회의회의록@

의석을 정돈하여 주시기 바랍니다. 성원이 되었으므로 제 207회 완주군의회 임시회 제1차 본회의 개의를 선포합니 다. 먼저 의사팀장으로부터 의회 관련 사항에 대한 보고가 있겠습니다. 의사팀장은 보고하여 주시기 바랍니다.

의사팀장 이은미 입니다. 의회관련 사항을 보고 드리겠습 니다. 먼저 제207회 완주군의회 임시회 소집경위를 보고 드리겠습니다. 지방자치법 제45조 제2항의 규정에 의거 서남용 의원님 외 4인의 의원님으로부터 임시회 집회 요 구가 있어 지난 8월6일 집회 공고를 하고 오늘 집회를 갖 게 되었습니다.



▼ ■ origin_data.csv

	category	3
	Document_id	1000
	question	'제 207회 완주군의회 임 시회는 누가 요구했어?'
	Answers	{'answer_start :183, ' text' : 서남용 의원님 외 4인의 의원님}
		@제207회 완주군의회(임시회) 제 1 차 본회의 회의록@
	context	의석을 정돈하여 주시기 바랍니다







뭐든 내게 물어봐(MNM)

데이터 구축 절차

질문 유형화

회의록에서 나올 수 있을 만한 범용적인 질문으로 유형화

질문유형	답변 유형	질문 패턴	질문 예시
When(기간)	기간(날짜, 시간)	<사건>의 기간은?	특별위원회의 현지확인은 언제 하지?
When(일시)	일시(날짜, 시간)	<사건>는 언제 진행되는가?	음성군 행정기구설치조례 일부 개정 조례안은 언제 제출 되었어?
Who(능동주, 행동)	사람(들)	<행동>을 수행한 사람은 누구인가?	5분 자유발언을 한 사람은 누구야?
Who(피동주, 역할)	사람	<역할>을 맡게 된 사람은 누구인가?	특별예산결정위원장을 맡게 된 사람은 누구야?
Why(행동, 의견)	명사구/기타	<행동>을 한 이유는 무엇인가?	제 212회 완주군의회 2제2차본회의를 휴회하는 이유는 뭐야?
What(지표)	숫자(금액, 단위)	<사건>의 나온 <지표>는 얼마인가?	2014 년도 예비비 지출 승인의 건의 표결 결과 찬성은?
What(회의)	명사구	<사건>를 다루고 있는 회의는?	2014 년도 예비비 지출 승인의 건을 다룬 회의는?

데이터 구축 절차

데이터 가이드라인 제작

데이터 태깅의 가이드라인과 FAQ 작성

'회의록 Question Answering' 데이터셋 구축 가이드라인

본 가이드라인은 회의록 Question Answering(이하 QA) 데이터셋 구축 가이드라인에 대해 다루고 있습니다. Annotation 에 참여하기 전에 반드시 숙지해주시기 바랍니다.

1. 데이터 구축 가이드

1.1. 데이터 구축 개요

Question Answering 태스크 자연어 처리 분야에서 인간이 제기하는 질문에 자동으로 대답하는 시스템을 구축하는 것과 관련이 있습니다. 이번 데이터셋 구축 작업은 회의록 데이터셋을 활용한 기계독해 모델 생성을 위한 지문(Passage) - 질문(Question) - 답변(Answer) 데이터셋을 구축한다.

1.2. 문제 정의

1,2,1, 임무정의

경제 작업을 거친 국회 회의목을 대상으로 작업자들이 지문(Passage) - 질문(Question) - 답변(Answer)으로 구성된 1500 개의 기계 독해용 질의용답 데이터셋을 구축한다. 이 중 기계독해 모델의 성능 항상을 위해 질문 중 답이 없는 경우를 포함한다.

공정명	건수(비율)	데이터셋 구성
정답이 있는 질문	1350 건(90%)	지문-질문-정답 데이터셋으로 작성
정답이 없는 질문	75~150 건(5~10%)	지문-질문- 빈칸 데이터셋으로 작성

1-3, 획득-정제 절차

원시 데이터는 문서 요약을 위해 가공된 의회 데이터를 사용하기로 했다. '완주', '음성', '청주' 의회 데이터 중 구축 작업 소요시간과 데이터 주제 간 통일성을 위해 크기를 고려해 '완주'의 회의 데이터 165 건을 안건별로 나눠 861 건의 지문을 작업 데이터로 정했다.

기계 독해 데이터셋은 일정 길이의 텍스트로 이루어진 지문, 지문에 내포된 정보를 묻는 질문,

그에 대한 답변으로 구성된 데이터셋으로, 수집된 원시데이터는 다음과 같은 정제 절차를 거쳐 회의에 대한 정보가 담긴 지문으로 변환되어야 한다.

4. Annotation FAQ

Q1, 간결체의 질문은 어떤 것을 의미하는가?

A1. 실제 대화할 때 사용하는 어투를 생각하면 된다. 시리나 빅스비와 같은 가상비서에 질문하는 느낌으로 생각하면된다.

(잘못된 예시) 완주군 싸움소 육성 지원에 관한 조례안을 다룬 회의는 무엇인가? (올바른 예시) 완주군 싸움소 육성 지원에 관한 조례안을 다룬 회의는 뭐야? / 뭘까?

Q2, no_answer 질문은 얼만큼 생성해야 하는가?

A2. no_answer 질문은 지문 8 개마다 1 개씩 만든다는 느낌으로 생성을 해 대략 5-10%의 개수를 맞추고자 한다.

Q3, title 로 제시된 회의 이름을 태깅할 때는 '회의록'도 포함해야 하는가?

A3. 회의 이름을 title(제목)에서 태깅하는 경우에는 회의 이름만 태깅한다.

(잘못된 예시) @제 207 회 완주군의회(임시회) 제 1 차 본회의회의록@

(올바른 예시) @제 207 회 완주군의회(임시회) 제 1 차 본회의회의록@

Q4. 회의 이름의 경우 title 말고, 본문에 등장하는 경우도 있는데 이때는 어느 위치에 태기의 하며 되느가?

A4. 회의 이름을 태강하는 경우 title 혹은 본문 둘 다 상관없다. 하지만 본문에 회의 이름이 등장하는 경우를 권장한다.

Q5, 사건에 해당하는 '△△△<mark>조례안과 □□□제안에 대한 심사</mark>'처럼 여러가지를 다루는 사건의 경우 어떻게 태강해야 하는가?

A5. 여러가지를 다루는 사건이더라도 그 길이가 너무 길지 않고(공백 포함 80 자 이상). 의미상 이어진다면 하나로 태강해도 된다.

Q6, '<mark>5 분 사전 발언, 감표위원</mark>과 같은 단어는 회의 특성 상 많이 나오는 것 같은데 어떻게 진무용 생성해야 하는가?

A6. 다양한 회의에서 나올 수 있을만한 용어에 대해서는 회의를 특정지을 수 있을만한 정보를 앞에 표시해줘야 한다.

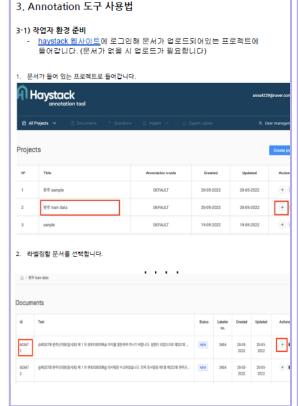
(예시 1) △△△내용으로 5분 사전발언을 수행한 사람은 누구야?

(예시 2) ♦♦♦의원이 ♦♦♦건으로 제안설명을 한 회의는 무슨 회의야?

(예시 3) ☆☆☆ 회의에서 5분 사전발언을 한 사람은 누구야?

데이터 태깅







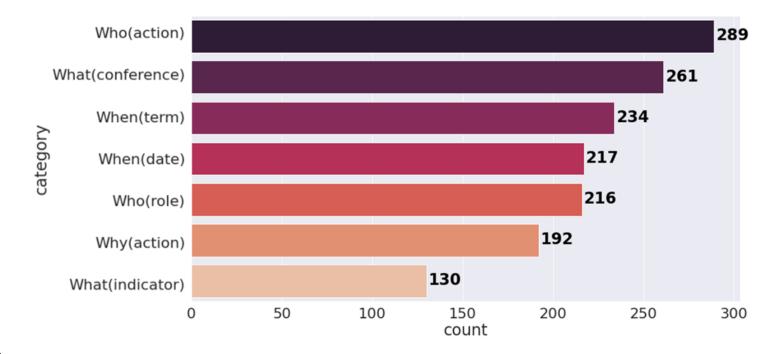
데이터 구축 결과

Data Info

구축: 165개의 완주군 회의록 데이터를 Agenda별로 나눈 861개의 지문

개수: 지문(Context) - 질문(Question) - 답변(Answer) 형식의 MRC 데이터 1539개

질문 유형별 분포



boostcampaitech

데이터 전처리 및 증강

Data Preprocess

Mislabel & 이상치 데이터 제거

질문 내 개행 문자 제거

답변 내 좌우 공백 제거

Tokenizer 적용후 데이터 수정

No Answer 답변 변경

Data Augmentation

Example

Query Back Translation(일본어)

Train & Val & Test

Train: 1246 (81%)

Val: 139 (9%)

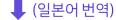
Test: 154 (10%)

No answer Data

Answer: 1429 (93%)

No Ans: 110 (7%)

"제178회 완주군의회 임시회 회기는 언제로 결정됐어?"



"제178회 완주군 의회의 임시 회의는 언제로 정해졌니?"

No Answer 데이터 증강(Regex)

"제 (n)회 완주군의회 2차 본회의는 언제 진행 돼?"

"제 (n+10)회 완주군의회 2차 본회의는 언제 진행 돼?"

Train & Val & Test

Train: 1443 (83%) Val: 139 (8%)

Test: 154 (9%)

No answer Data

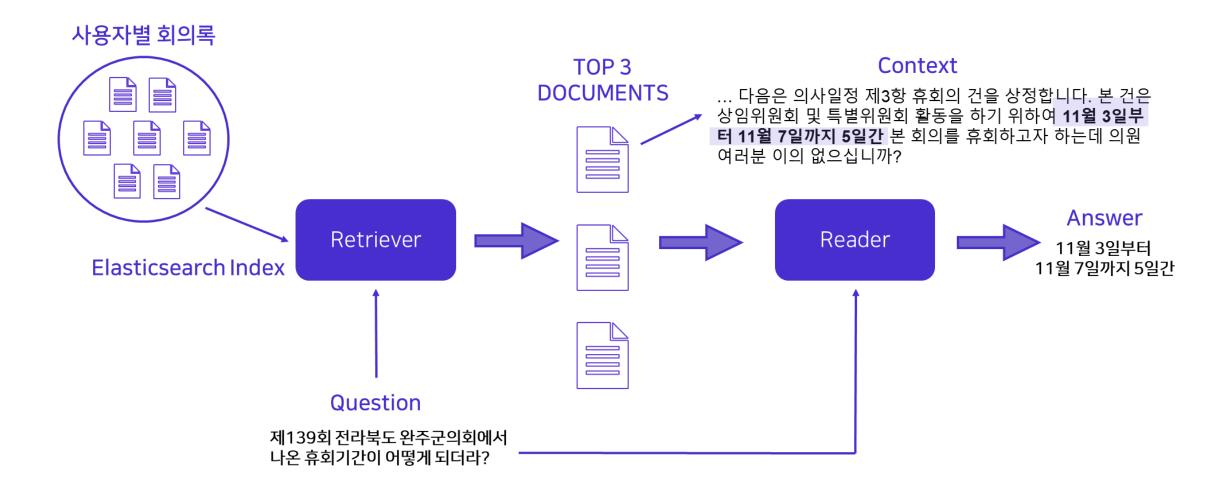
Answer: 1519 (88%) No Ans: 217 (12%)



3. Reader & Retrieval

뭐든 내게 물어봐(MNM)

모델 구조



boostcamp aitech

Retriever 모델

Elasticsearch

- 문장 안의 키워드를 인덱스로 만들어 빠른 검색 가능
- 내용 전체를 인덱싱해 특정 단어가 포함된 문서 전문 검색 가능

RDB

Document id	Content
1	먼저 제207회 완주군의회 임시회 소집경위를 보고 드리겠습니다.
2	먼저 <mark>의사팀장</mark> 으로부터 의회 관련 사항에 대 한 <mark>보고</mark> 가 있겠습니다.
3	의사팀장은 <mark>보고</mark> 하여 주시기 바랍니다.

Elasticsearch

Term	Document
완주군의회	1
임시회	1
의사팀장	2, 3
보고	1, 2, 3

O(n) O(1)

Retriever 모델

질문 Q와 문서 D간의 유사도 계산

- TF-IDF : 너무 흔하지도 않지만, 너무 희귀하지도 않게
 - •TF(Term Frequency): 특정 문서에서 특정 단어의 등장 빈도
 - •IDF(Inverse Document Frequency): 전체 문서에서 특정 단어의 등장 빈도의 역수
- BM25: TF-IDF의 파라미터를 변경해 Best Match를 찾음

 $egin{align*} rac{ ext{Ed } D ext{Old } q_i ext{의 Term frequency}}{f(q_i,D)\cdot(k_1+1)} & ext{Ed } D ext{의 길이} \ \hline f(q_i,D)+k_1\cdot\left(1-b+b\cdotrac{|D|}{ ext{avgdl}}
ight) \ & ext{Term requency의 scaling } ext{조절} \ & ext{Ed } ext{Cl } ext{Old } ext{Normalization} \ & ext{(}k_1=5) & ext{Ed } ext{Cl } ext{normalization} \ & ext{(}b=1) \ \hline \end{pmatrix}$

평가지표 (Retriever)

• Retrieval accuracy: retriever의 성능 측정을 위한 metric

최종 retriever의 Retrieval accuracy

Accuracy와 Turn-Around Time*을 고려해 Top K 3으로 선정

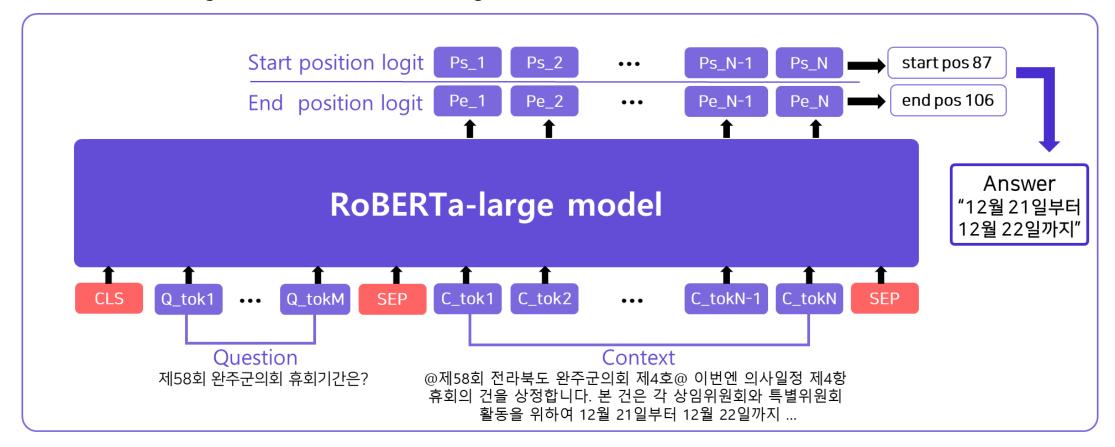
Topk	Retrieval accuracy	Turn-Around Time (150개 질문 평균)
1	0.90	2.04 s
2	0.95	2.18 s
3	0.98	2.30 s
4	0.99	2.44 s
5	1.00	2.55 s

*Turn-Around Time: 사용자가 파일의 처리를 부탁한 후 결과를 받아볼 때까지 걸리는 시간

Reader 모델

모델 선정

• RoBERTa-Large(MRC SOTA) 와 KoBigBird 모델(긴 문서 최적화) 성능 비교 후 RoBERTa 채택

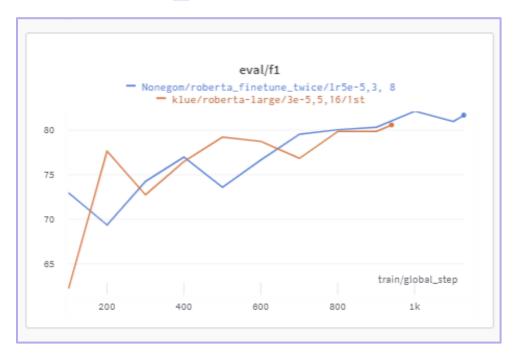


Reader 모델

2번의 Fine-tuning

- QA Task에 최적화시키기 위해 QA데이터셋(KorQuAD1.0)으로 한 번 학습
- 도메인에 최적화시키기 위해 회의록 데이터셋으로 한 번 더 학습

모델 validation 그래프



2번의 Fine-tuning 실험 결과

Topk	F1	
Roberta-large (Our Data only)	72.72	
Roberta-large (KorQuAD + Our Data)	74.13	+1.4

평가지표 (Retriever+Reader)

- Exact Match (EM): 모델의 예측과, 실제 답이 정확하게 일치할 때만 점수 부여
 - 예측이 김상식이고 답이 김상식 의원일 때 EM score 는 0점
- F1 score: EM과 다르게 precision과 recall을 계산하여 부분 점수를 제공
 - 예측이 김상식이고 답이 김상식 의원일 때 F1 score 는 precision이 1, recall이 0.5이므로 0.67
- F1 score을 주요 평가지표 선정

The definition of true positive (TP), true negative (TN), false positive (FP), false negative (FN)

	Tokens in Reference	Tokens Not in Reference
tokens in candidate	TP	FP
tokens not in candidate	FN	TN

• 최종 모델의 평가 결과

F1	EM
79.7866	64.2857

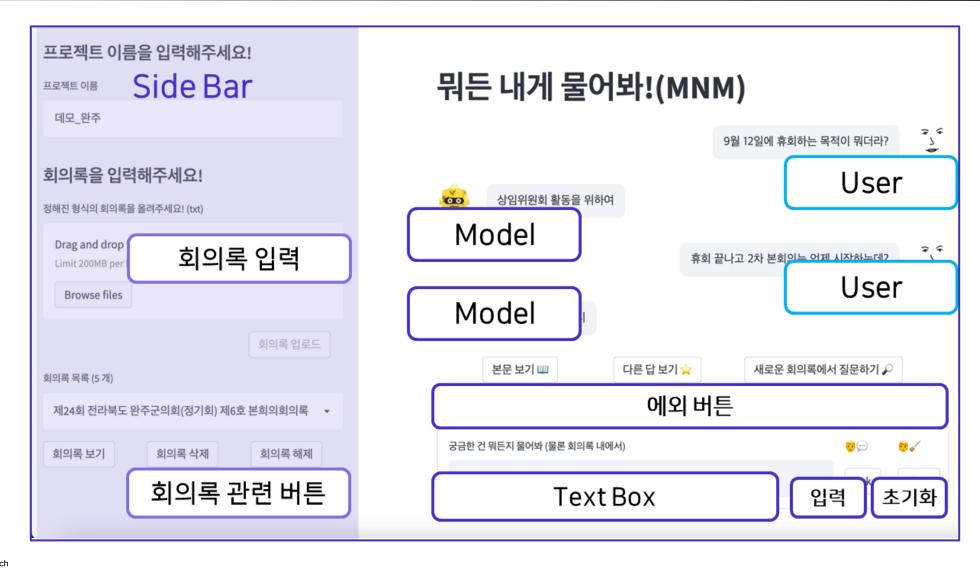
$$Precision = \frac{num(same_token)}{num(pred_tokens)}$$

$$Recall = \frac{num(same_token)}{num(groud_tokens)}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

4. Serving

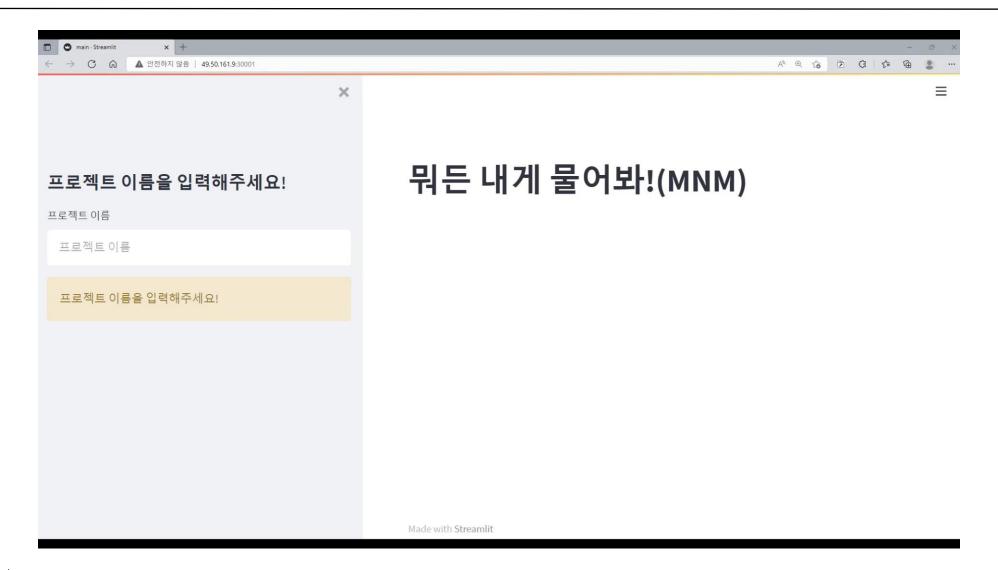
웹 서비스 레이아웃



boostcamp aitech

뭐든 내게 물어봐(MNM)

시연영상

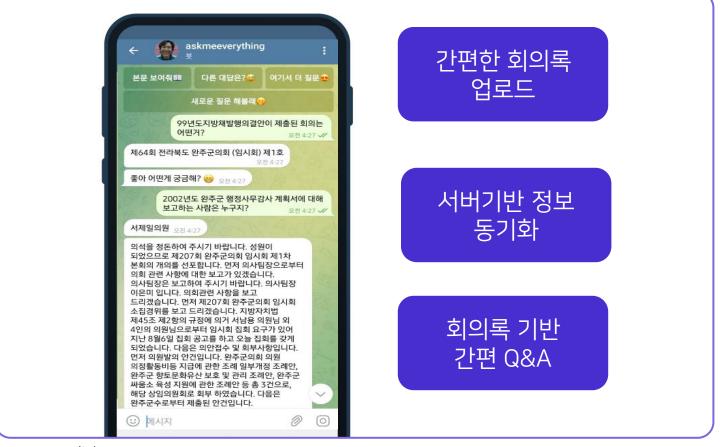


boostcamp aitech

뭐든 내게 물어봐(MNM)

Telegram 챗봇

• 모바일로 쉽게 접근 가능한 텔레그램 앱 기반 챗봇 구현





5. Conclusion

결론

회의록을 활용한 QA 모델 생성

- 정보검색에 어려움이 있던 내부문서를 활용한 QA 모델 생성
- TTS 서비스에서 생성된 데이터를 활용할 수 있는 새로운 방향 제시

새로운 데이터셋 제시

• 회의록 QA 데이터셋 가이드라인을 만들어, 추후 Closed Domain QA 모델 생성에 도움

boostcampaitech

뭐든 내게 물어봐(MNM)

29

발전 방향

추가 데이터 수집

- 잘 맞추지 못하는 질문 유형들에 대한 성능 개선
- 일반화된 데이터를 통한 학습 필요
- 질문 유형의 다양화 필요
- 화자 구분된 데이터로 모델 학습

boostcamp^{aitech}

Q & A

Appendix

32

Appendix 목차

- 질문 유형 선정 기준
- 데이터 증강
- 모델 와이어 프레임
- 워크플로우
- Retrieval 모델 파라미터 실험 결과
- 2차 파인튜닝 효과 입증 자료

boostcamp aitech

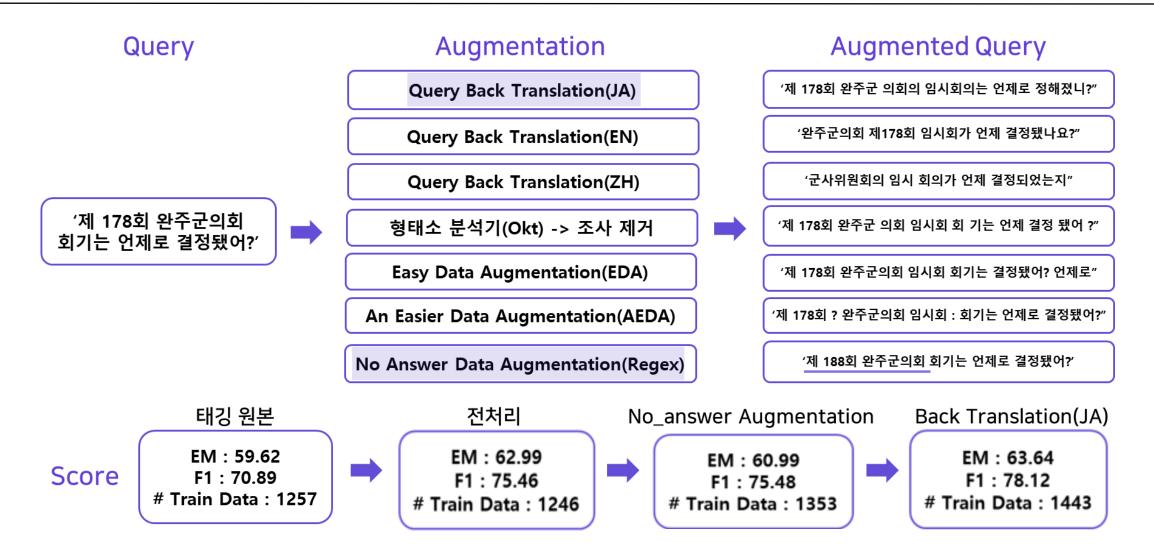
Appendix - 질문 유형 선정 기준

- 실제 사용자가 회의록을 검색할 때 물을 것 같은 질문
- 의회 회의 뿐만 아니라 추후 다른 회의에도 적용시킬 수 있는 질문

질문유형	답변 유형	질문 패턴	질문 예시	실제 회의 대응 질문 예시
When(기간)	기간 (날짜, 시간)	<사건>의 기간은?	특별위원회의 현지확인은 언제 하지?	MRC 대회 기간은?
When(일시)	일시 (날짜, 시간)	<사건>는 언제 진행되는가?	음성군 행정기구설치조례 일부 개정 조 례안은 언제 제출되었어?	Wrapup 리포트 작성은 언제까지지?
Who(능동주, 행동)	사람(들)	<행동>을 수행한 사람은 누구인가?	5분 자유발언을 한 사람은 누구야?	Ensemble을 구현한 사람은 누구지?
Who(피동주, 역할)	사람	<역할>을 맡게 된 사람은 누구인가?	특별예산결정위원장을 맡게 된 사람은 누구야?	모더레이터를 할 사람은 누구야?
Why(행동, 의견)	명사구/기타	<행동>을 한 이유는 무엇인가?	제 212회 완주군의회 2제2차본회의를 휴 회하는 이유는 뭐야?	Currciculum Learning 실험을 한 이유는 뭐였지?
What(지표)	숫자 (금액, 단위)	<사건>의 나온 <지표>는 얼마인가?	2014 년도 예비비 지출 승인의 건의 표결 결과 찬성은 몇 명?	Reader 모델 2차 실험에 대한 F1- score는 얼마야?
What(회의)	명사구	<사건>를 다루고 있는 회의는?	2014 년도 예비비 지출 승인의 건을 다 룬 회의는?	MRC 프로젝트 랩업리포트를 다룬 회의는 무슨 회의야?

boostcampaitech

Appendix - 데이터 증강



Appendix – 데이터 증강(No_answer)

실험 동기

- Top-k를 올리면 retrieval accuracy는 오르지만 F1과 EM 은 떨어지는 현상을 발견
- 답이 없는 회의록에서 찾은 예측의 확률이 더 높았기 때문이라고 판단
- 특정 회의록 내에서 답변할 수 없는 질문이 주어 졌을 경우 no_answer를 예측할 수 있게 증강 시도

	Retrieval accuracy	F1	EM
top-k 1	0.83	72.52	62.98
top-k 5	0.98	70.35	58.44

실험 방법

- 구축한 데이터셋 내 Context와 Matching되지 않는 Query 데이터를 정규표현식을 통해 생성
- 질문 유형별 데이터 개수의 분포를 고려하여 Random Sampling을 통해 총 107개의 No Answer Data 증강

실	험	결	고	ŀ

• Top – k 를 늘려 retrieval accuracy를 올리면서도 F1을 유지

	원본 데이터	증강 데이터		
context	제 178회 완주군의회	제 178회 완주군의회		
query	제 178회 완주군의회 회기는 언제로 결정됐어?	제 188회 완주군의회 회기는 언제로 결정됐어?		
answer	5월 15일	no_answer		

top-k: 1 EM : 62.99 F1 : 75.46

Train Data : 1246

top-k: 3 EM : 60.99 F1 : 75.48

Train Data: 1353

뭐든 내게 물어봐(MNM)

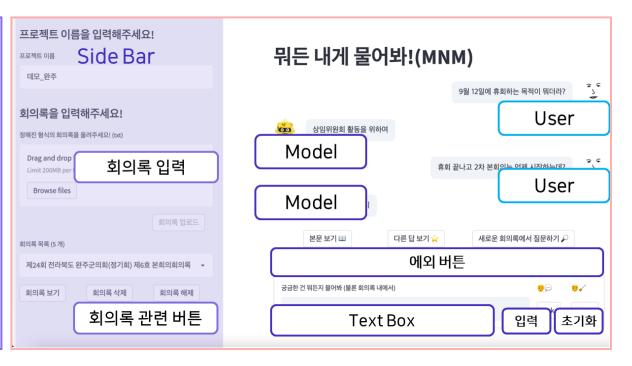
Appendix - 모델 와이어 프레임

- 초기 와이어 프레임에서 사이드바를 추가하여 직관적인 구조로 변경
- 사용자에게 구어체로 질문을 입력 받기 위한 '챗봇'레이아웃 도입

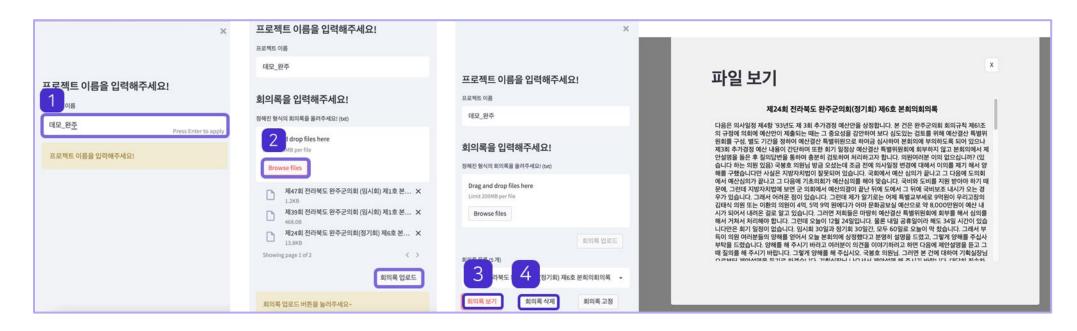
변경 전

8 회의록 추가 5/30 결산 검토 🗏 9 5/29 완주 총회 🖫 5/28 의장 선출 📋 우호 친선 교류 재정안은 왜 제안 되었을까? 세계화 정보화 시대를 맞이하여 날로 급증하는 국제교류 수요와 무한 회의록 폴더를 업로드 해주세요 경쟁 시대의 파고에 능동적으로 대처함은 물론, 1998년부터 현재까 지 우호친선교류 및 문화-농업 등 다방면에서 활발한 교류를 추진해 온 태주시와의 자매결연을 체결함으로써 상호 대등한 입장에서 교류 가 가능한 모든 분야에 걸쳐 교류를 추진하기 위하여 지방자치법 제35 조의 규정에 의거 사전 의결을 받고자 하는 것입니다." 5/28 의장 선출 4/2 결산 의원장 선출 ⊞ "제94회 본회의 제1차(2000.05.22.)" 다른 답변 살펴보기 검색된 회의록으로 이동 해당 회의록에서 추가 질문 회의 내용 본문 팝업 창

변경후

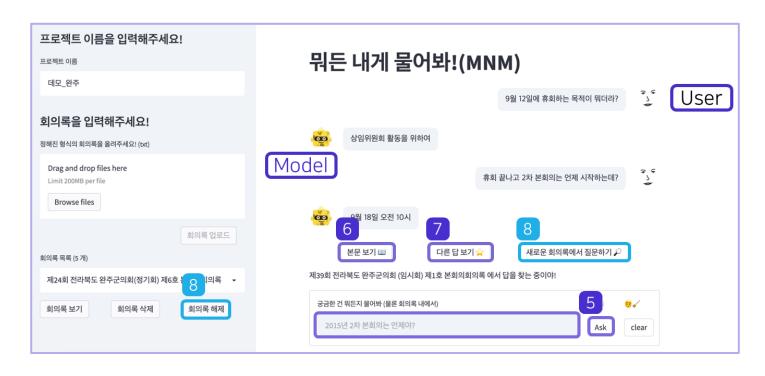


Appendix - 모델 와이어 프레임(1)



- 1 프로젝트 이름을 입력
- 2 Browse files 버튼을 눌러 업로드할 회의록을 선택한 후, 회의록 업로드 버튼을 눌러 회의록을 업로드
- 3 팝업창을 통해 현재 선택한 회의록의 본문 전체를 확인
- 4 업로드한 회의록을 선택해 삭제 가능

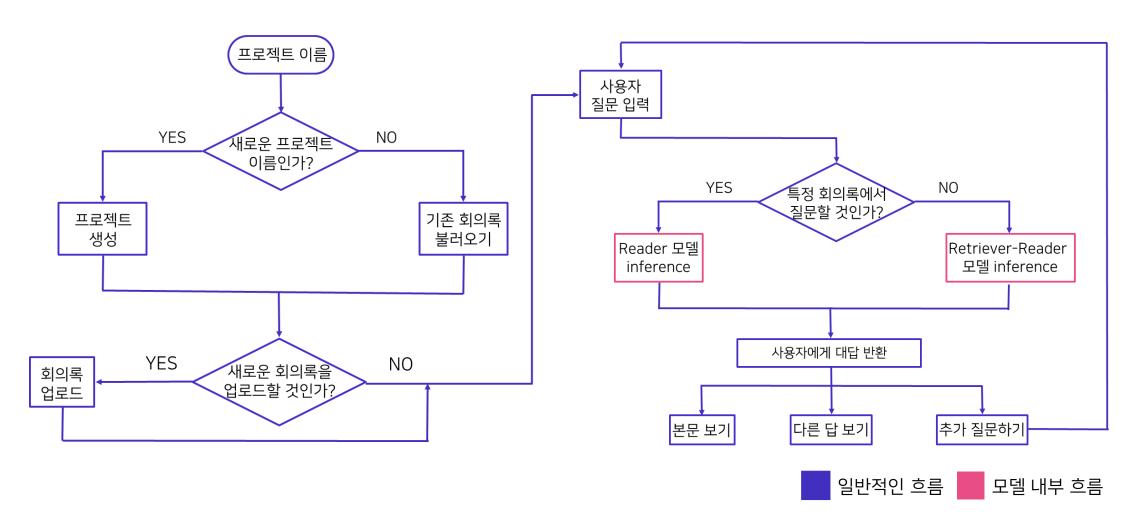
Appendix - 모델 와이어 프레임(2)



- 5 검색 칸에 질문을 입력한 후 Ask 버튼을 눌러 질문
- 6 모델이 찾은 답을 회의록 본문에서 확인
- 7 사용자가 원하는 답이 아닐 경우 다른 답 확인
- 8 답변이 나온 회의록으로 고징시켜 추가 질문 가능 혹은 업로드한 전체 회의록 중 고정시켜 질문

뭐든 내게 물어봐(MNM)

Appendix - 워크플로우



Appendix - Retrieval 모델 파라미터 실험 결과

각 파라미터 종류별로 실험 결과 (변인통제 후 진행) *DFR: Divergence From Randomness으로다중 모델을 기반으로 하는 프레임워크이자 정규화 기법

• 유사도: Tf-idf, BM25, DFR*

• n-gram : (3) , (4), (3,4), (4, 5) (Unigram은 기본 포함)

• BM25 파라미터(b, k10): (b:1, k1:5), (b:0, k1:10). (b:1, k1:8), (b:0.5, k1:5)

유사도	Tf-idf	BM25	DFR
topk1	0.74	0.83	0.59
topk2	0.82	0.92	0.75
topk3	0.85	0.95	0.85
topk5	0.87	0.98	0.93

N-gram	(3)	(4)	(3,4)	(4,5)
topk1	0.85	0.85	0.80	0.80
topk2	0.94	0.93	0.92	0.92
topk3	0.95	0.96	0.94	0.94
topk5	topk5 0.98		0.98	0.98

BM25 파라미터	b:0.5,k1: 0	b:1,k1:5	b:0,k1:10	b:1, k1:8
topk1	0.52	0.89	0.52	0.88
topk2	0.62	0.96	0.62	0.95
topk3	0.71	0.97	0.71	0.97
topk5	0.85	0.98	0.85	0.98

파라미터 조합으로 실험한 결과

• 종류별 실험 결과과는 달리, 조합으로는 BM25, (3,4), (b:1, k1:5)가 좋았음

	topk1	topk2	topk3	topk4	topk5
BM25, (3), {b:1, k1:5}	0.89	0.96	0.97	0.97	0.98
BM25, (4), {b:1, k1:5}	0.88	0.96	0.97	0.97	0.99
BM25, (3,4), {b:1, k1:5}	0.90	0.95	0.98	0.99	1.0

Appendix - 2차 파인튜닝 효과 입증 자료

'삼성 SDS Techtonoic 2020' 행사 Track 1 발표 내용

 MRC 모델을 생성하기 위해, Monolingual 모델에 대해 유사 데이터 (NewsQA)를 이용하여 1차 파인튜닝을 진행 후 기존 데이터 (KorQuad)로 2차 파인튜닝을 진행해 성능 향상시킴

'<u>Fine-tuning Strategies for Domain Specific Question Answering</u> <u>under Low Annotation Budget Constraints (annon., ACL)</u>' 논문

- MRC 모델에 학습 데이터(Target QA)로 파인튜닝하기 전에, 먼저 1차적으로 SQuAD 데이터로 파인튜닝한 후 진행한 것이 MRC 모델에 성능 향상을 가져왔음
- 이는 4가지 종류의 서로 다른 도메인의 QA 데이터셋에 적용했을 때 모두 효과가 있음을 확인



	Covid-QA	CUAD-QA	Movie-QA	KG-QA
Target QA only	59.2	48.5	80.1	78.6
SQuAD + TargetQA	64.1	50.2	84.0	81.3

1600개 태깅한 TargetQA 데이터 기준, Macro-F1 성능지표