

캡스톤디자인 최종 발표

AI 챗봇을 활용한 온라인 강의 웹 서비스

20171612 박상현

20197132 주준하 담당교수 : 임경태 20171768 남승완

한밭대학교 컴퓨터공학과

<u>1. 프로젝트 개요</u>

<u>2. 전체 시스템 구성</u>

<u>3. 주요 기능</u> <u>4. 차별점</u>

Contents.

5. 설계 방법

기계독해(MRC)

데이터셋

모델

<u>6. 시스템 결과</u>

강의 업로드

PDF 기계독해

WIKI 기계독해

7. 한계 및 개선방향

데이터셋

WIKI depth

키워드

모델 성능의 한계

8. 진행도

9. 팀원 역할

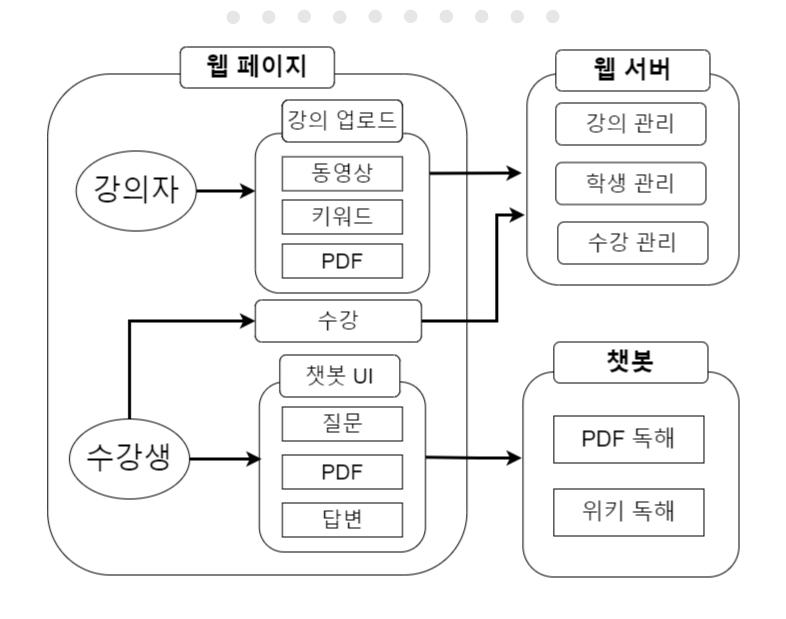
10. 데모

한밭대학교 컴퓨터공학과

프로젝트 개요



온라인 강의 환경에서 강의 내용에 대한 질의응답을 수행할 수 있는 기계독해(MRC) 챗봇을 제공. 수강생들은 학습효율 up! 온라인 강의 서비스



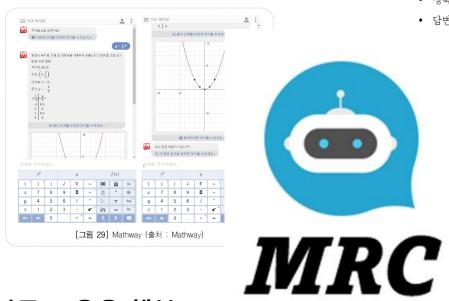
주요 기능

웹 기능	강의 업로드, 강좌 목록, 학습 페이지, 챗봇 페이지
PDF MRC(기계독해)	강의 자료(교안)를 업로드하고, 챗봇은 PDF의 내용을 바탕으로 기계독해를 하여 수강자에게 답변을 제공함
위키피디아 MRC (기계독해)	사용자의 질문을 분석하고, 관련된 문서를 위키에 탐색하여 적절한 답변을 사용자에게 제공함

차별점

☐ Mathway

- 대화형 인터페이스를 갖춘 수학 문제풀이 서비스
- 웹서비스로 별도의 설치 프로그램 없음
- 선택한 단원 및 유형에 따라 수식, 그래프, 도형 입력 등을 지원하고 그에 따른 풀이/해설 제공



□ Jill Watson(질 왓슨)

- 조지아공과대학교(조지아테크)의 지식 기반 인공지능 과정의 온라인 조교 역할을 하는 AI 챗봇
- 기계학습을 통해 정확도를 97%까지 끌어 올리면서 모든 질문을 처리
- 정확도가 97% 이상이라고 판단될 때만 답변 제공
- 답변을 받은 학생들은 인공지능 컴퓨터인지 인지하지 못했을 정도로 높은 품질 제공

비스

분야		교육
제공목적		온라인 조교 이 그를 모이
제공기능		일상대화 및 학사장 어떤 과목의
이용대상자		^{강의 수강생} 강의 환경에서 라도
개인화여부		이 경기 단경에서 너그
터페이스	질의	^{텍스트} 위키피디아와
	답변	텍스트
개발방식		IBM 인공지능 wats A 및 교인,
모바일 알림연동		<u> </u>
특이사항		강의계획서를 통해
		답변을 제공한다.
		답민들 세증인다.

기존 교육용 챗봇

- 특정 분야에서만 사용가능 (ex. 수학, 온라인 조교)

Mrc(기계독해) - 기계가 주어진 지문과 질문을 이해하여 답변을 하는 태스크

korquad - KorQuAD1.0는 한국어 위키백과 문서에 대해 크라우드소싱을 통해 만든 70,000+개의 질문-답변 쌍을 포함한 대규모 데이터셋

데이터셋

모델

기계독해(MRC)

웹

Question

```
"id": "6548850-0-0",
"question": "임종석이 여의도 농민 폭력 시위를 주도한 혐의로 지명수배 된 날은?"
```

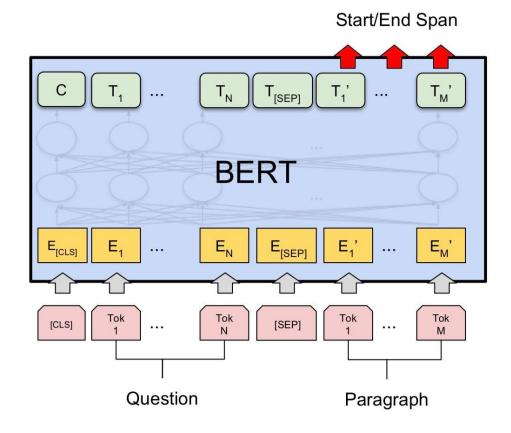
Answer

```
"answers": [
{
    "text": "1989년 2월 15일",
    "answer_start": 0
  }
],
```

Context

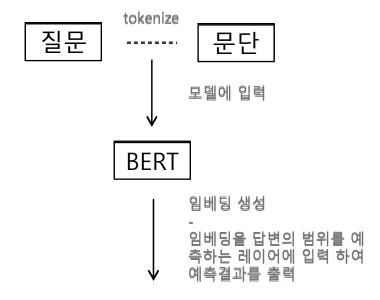
"context": "1989년 2월 15일 여의도 농민 폭력 시위를 주도한 혐의(폭력행위등처벌에관한법률위반)으로 지명수배되었다. 1989년 3월 12일 서울지방검찰청 공안부는 임종석의 사전구속영장을 발부받았다. 같은 해 6월 30일 평양측전에 임수경을 대표로 파견하여 국가보안법위반 혐의가 추가되었다. 경찰은 12월 18일~20일 사이 서울 경희대학교에서 임종석이 성명 발표를 추진하고 있다는 참보를 입수했고, 12월 18일 오전 7시 40분 경 가스총과 전자봉으로 무장한 특공조 및 대공과 직원 12명 등 22명의 사복 경찰을 승용차 8대에 나누어 경희대학교에 투입했다. 1989년 12월 18일 오전 8시 15분 경 서울청량리경찰서는 호위 학생 5명과 함께 경희대학교 학생회관 건물 계단을 내려오는 임종석을 발견, 검거해 구속을 집행했다. 임종 석은 청량리경찰서에서 약 1시간 동안 조사를 받은 뒤 오전 9시 50분 경 서울 장안동의 서울지방경찰청 공안분실로 인계되었다."

데이터셋 모델 기계독해(MRC) 웹

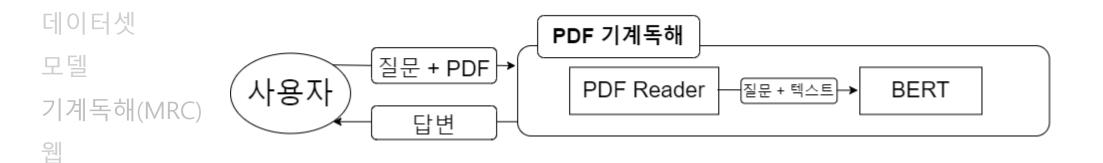


질의응답의 핵심인 언어모델.

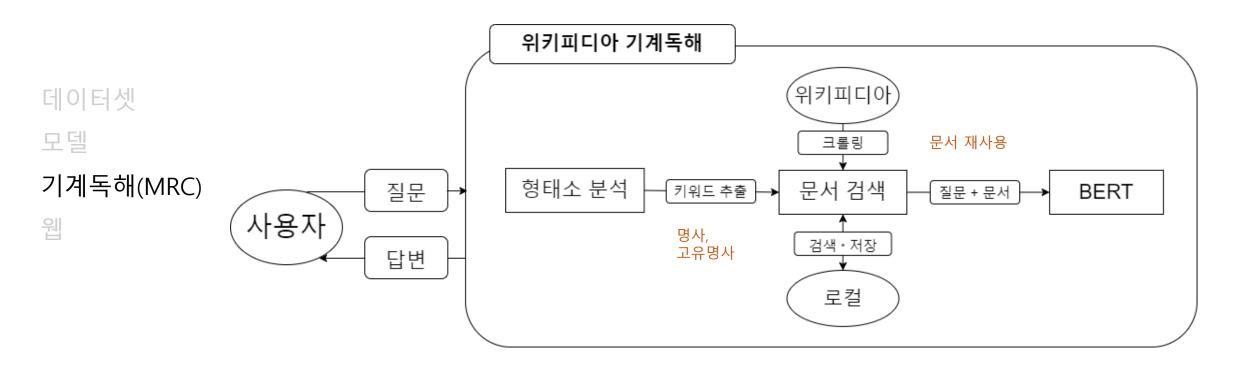
구글에서 공개한 사전 훈련 모델 BERT를 korquad 데이 터 셋에 파인 튜닝하여 질 의응답 모델로 사용



1. 강의자료(PDF) 질의응답, 강의자료(교안)이 있을때.



2. 위키피디아 문서 질의응답 (강의자료가 없을 시)

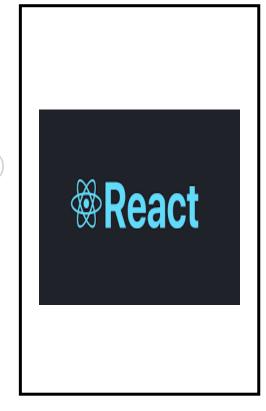




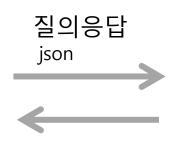


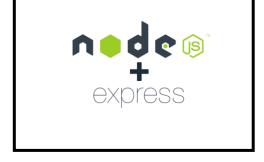
DB

데이터셋 모델 기계독해(MRC) 웹



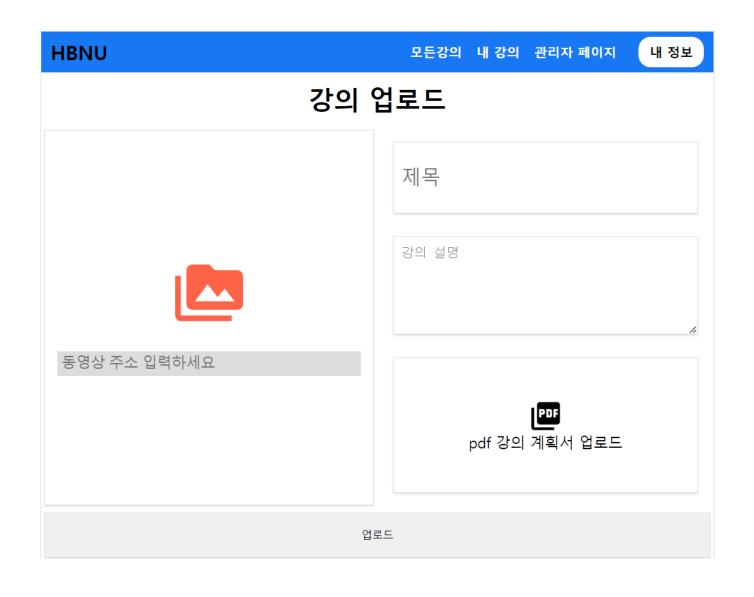












HBNU

모든강의 내 강의 관리자 페이지

내 정보

강의 목록



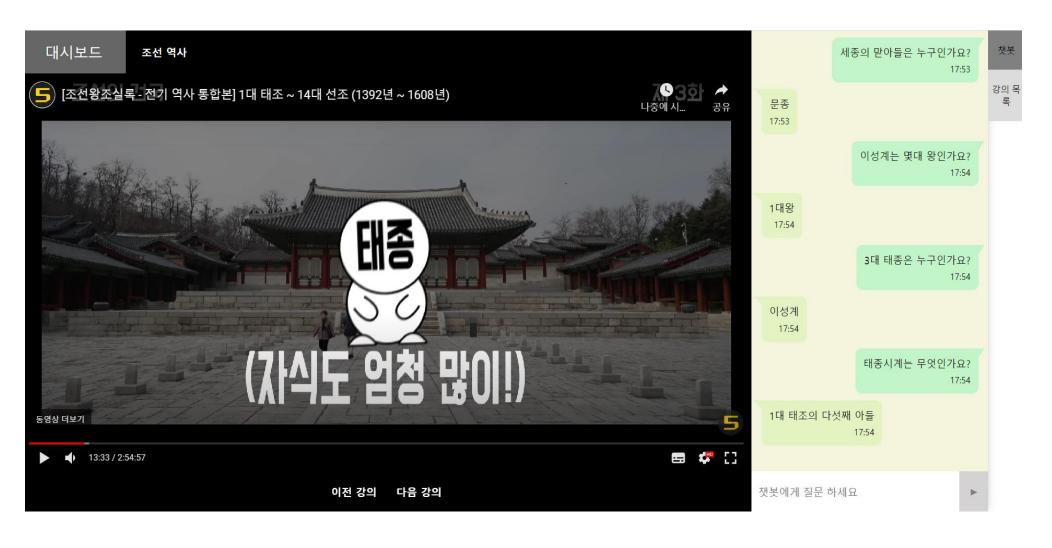
조선역사와 조선왕조실록



인공지능2 인공지능의 현재와 미래



인공지능 개요



강의 업로드

강의 목록

PDF 기계독해

WIKI 기계독해



>>> model.fileMrc('세종의 맏아들은 누구인가요?', '예시_pdf2_조선_왕_계보.pdf')
''조선 왕 2개보1대 태조 : 이성계, 조선의 제 1대왕2대 정조 : 1대 태조의 아들, 제 1차 왕자의 난으로 세자로 책립3대 태종 : '대 태조인 아들, 제 1차 왕자의 난으로 세자로 책립3대 태종 : '대 태조인 다중다. '대 태조인 대전째 아들. 왕권을 강하기 위해 사병을 없애고 관직 제도를 개혁함4대 세종 : 3대 태종의 셋째 아들. 훈민정음을 창제함. 정치치 사회, 경제, 문화, 과학, 음악 등 전 분야를 발전시켜 조선의 황금시대를 연 성군.5대 문종 : 4대 세종의 맏아들 성품이 온호하고 학문을 좋아한 왕.'] 조선 왕 계5보대 태조 : 이성계, 조선의 제 1대왕2대 정조 : 1대 태조의 아들, 제 1차 왕자의 난으로 세자로 책립3대 태종 : 1대 태조의 다섯째 아들, 왕권을 강화하기 위해 사병을 없애고 관직 제도를 개혁함4대 세종 : 3대 태종의 셋째 아들. 훈민정음을 칭 대조의 다섯째 아들, 왕권을 강화하기 위해 사병을 없애고 관직 제도를 개혁함4대 세종 : 3대 태종의 셋째 아들. 훈민정음을 칭 생품이 온화하고 학문을 좋아한 왕.

>> model.fileMrc('이성계는 몇대 왕인가요?', '에시_pdf2_조선_왕_계보.pdf') ''조선 왕 계보1대 태조 : 이성계, 조선의 제 1대왕2대 정조 : 1대 태조의 아들, 제 1차 왕자의 난으로 세자로 책립3대 태종 : 1 내 태조의 다섯째 아들. 왕권을 강화하기 위해 사병을 없애고 관직 제도를 개혁함4대 세종 : 3대 태종의 셋째 아들. 훈민정음을 방제함. 정치, 사회, 경제, 문화, 과학, 음악 등 전 분야를 발전시켜 조선의 황금시대를 연 성군.5대 문종 : 4대 세종의 맏아들. 성품이 온화하고 학문을 좋아한 왕 '] 조선 왕 계보1대 태조 : 이성계, 조선의 제 1대왕2대 정조 : 1대 태조의 아들, 제 1차 왕자의 난으로 세자로 책립3대 태종 : 1대 태조의 다섯째 아들. 왕권을 강화하기 위해 사병을 없애고 관직 제도를 개혁함4대 세종 : 3대 태종의 셋째 아들. 훈민정음을 창 제함. 정치, 사회, 경제, 문화, 과학, 음악 등 전 분야를 발전시켜 조선의 황금시대를 연 성군.5대 문종 : 4대 세종의 맏아들. 성품이 온화하고 학문을 좋아한 왕.

'조선 왕 계보1대 태조 : 이정계, 조선의 제 1대왕2대 정조 : 1대 태조의 아들, 제 1차 왕자의 난으로 세자로 책립3대 태종 : 1 내 태조의 다섯째 아들. 왕권을 강화하기 위해 사병을 없애고 관직 제도를 개혁함4대 세종 : 3대 태종의 셋째 아들. 훈민정음을 방제함. 정치, 사회, 경제, 문화, 과학, 음악 등 전 분야를 발전시켜 조선의 황금시대를 연 성군.5대 문종 : 4대 세종의 맏아들. 성품이 온화하고 학문을 좋아한 왕.'] 단선 왕 계보1대 태조 : 이성계, 조선의 제 1대왕2대 정조 : 1대 태조의 아들, 제 1차 왕자의 난으로 세자로 책립3대 태종 : 1대 태조의 다섯째 아들. 왕권을 강화하기 위해 사병을 없애고 관직 제도를 개혁함4대 세종 : 3대 태종의 셋째 아들. 훈민정음을 창 네함. 정치, 사회, 경제, 문화, 과학, 음악 등 전 분야를 발전시켜 조선의 황금시대를 연 성군.5대 문종 : 4대 세종의 맏아들. 영품이 온화하고 학문을 좋아한 왕.







1. 데이터 셋과 모델이 실제 질문을 처리하기에 적합하지 않음.

2. 질문이 복잡하거나 추론이 필요한 내용에 대해서 잘 답변하지 못함.

데이터셋 구성에서 정답의 유형이 대부분 단답형, 문장형 답변을 필요로 하는 질문에 대해서 잘 답변하지 못함.

데이터셋

WIKI depth

키워드

모델 성능의 한계

```
"id": "6548850-0-0",
"question": "임종석이 여의도 농민 폭력 시위를 주도한 혐의로 지명수배 된 날은?"
```

Answer

Question

```
"answers": [
{
    "text": <mark>"1989년 2월 15일",</mark>
    "answer_start": U
}
],
```

Context

"context": "1989년 2월 15일 여의도 농민 폭력 시위를 주도한 혐의(폭력행위등처벌에관한법률위반)으로 지명수배되었다. 1989년 3월 12일 서울지방검찰청 공안부는 입종석의 사전구속영장을 발부받았다. 같은 해 6월 30일 평양측전에 임수경을 대표로 파견하여 국가보안법위반 혐의가 추가되었다. 경찰은 12월 18일~20일 사이 서울 경희대학교에서 임종석이 성명 발표를 추진하고 있다는 협보를 입수했고, 12월 18일 오전 7시 40분 경 가스총과 전자봉으로 무장한 특공조 및 대공과 직원 12명 등 22명의 사복 경찰을 승용차 8대에 나누어 경희대학교에 투입했다. 1989년 12월 18일 오전 8시 15분 경 서울청량리경찰서는 호위 학생 5명과 함께 경희대학교 학생회관 건물 계단을 내려오는 입종석을 발견, 검거해 구속을 집행했다. 임종석은 청량리경찰서에서 약 1시간 동안 조사를 받은 뒤 오전 9시 50분 경 서울 장안동의 서울지방경찰청 공안분실로 인계되었다."

데이터셋

WIKI depth

키워드

모델 성능의 한계

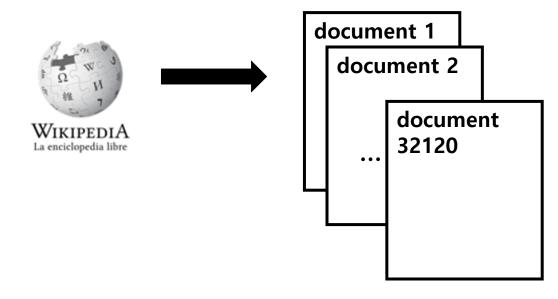
장문형 답변 개선 진행 중

>>> question = '챗봇은 무엇인가요?'
>>> context = '챗봇의 생성 및 구현은 여전히 인공 지능 및 기계 학습과 밀접한 관련이 있는 전망성이 밝은 개발 영역이다. 챗봇은 분명한 장점을 가지고 있지만, 몇 가지 중요한 제한이 있다. 그러나 이것은 확실하게 정해진 것이 아니므로 계속 변하고 있다. 또한, 챗봇을 만들때 흔히 하는 실패요인이 있으니 주의해야 한다. 출력 생성에 사용되는 데이터베이스가 고정되고 제한되어 있기 때문에 학습되지 않거나 저장되지 않은 질문은 챗봇이 처리를 실패할 수 있다. 챗봇은 언어 프로세스에 크게 영향을 받으며 액센트 및 실수와 같은 불규칙성으로 인해 제한될 수 있다. 챗봇은 동시에 여러 질문을 처리할 수 없으므로 사람과 대화할 때보다 효율성이 떨어질 수 있다. 챗봇은 학습경험을 위해 많은 양의 대화 데이터가 필요하다. 챗봇은 사용자와 이전의 대화를 되고 받아야 하는 비선형 대화를 처리하는데 어려움이 있다. 일반적으로 기존 서비스의 기술 주도적 변화에서 발생하는 것처럼 일부 소비자는 챗봇에 불편함을 느끼고 자신의 요청이 기계에 의해 처리되고 있음을 분명히 알 수 있다.'

' 첫봇은 언어 프로세스에 크게 영향을 받으며 액센트 및 실수와 같은 불규칙성으로 인해 제한될 수 있다', ' 챗봇은 동시에 0 네 질문을 처리할 수 없으므로 사람과 대화할 때 보다 효율성이 떨어질 수 있다']

Wiki depth : 탐색되는 문서가 매우 많을 경우 답변하는데 시간이 걸리는 성능 저하 문제가 야기된다.

데이터셋 WIKI depth 키워드 모델 성능의 한계

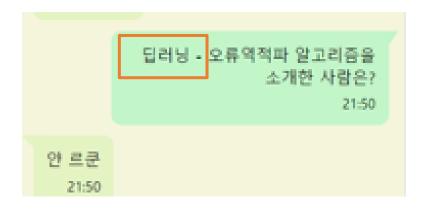


```
ave file : ./documents\인공지능(철확적 관점의 인공지능
ave file : ./documents\인공지능\강인공지능strong AI 범용인공지능 AGI과 약인공지능weak AI
ave file : ./documents\인공지능\강인공지능범용인공지능
ave file : ./documents\인공지능\약인공지능weak ai
ave file : ./documents\인공지능\강인공지능에 대한 철학적인 주장과 반론
            ./documents\인공지능\역사
            ./documents\인공지능\인공두뇌학과 초기 신경 네트워크
          ./documents\인공지능\게임 인공지능
            ./documents\인공지능\상징 추론과 논리 이론
           ./documents\인공자능\황금기19561974년
./documents\인공자능\작업들
            ./documents\인공지능\AI의 첫번째 암흑기19741980
save file : ./documents\인공자능\문제
            ./documents\인공지능\캠퍼스 전역의 비판들
           ./documents\인공지능\대샙트론과 연결망의 어무운 시대
./documents\인공지능\깔끔이 논리 프롤로그와 전문가 시스템
./documents\인공지능\지저분이 프레임과 스크립트
ave file : ./documents\인공지능\Boom 19801987
           ./documents\인공자능\돈은 되돌아온다 5세대 프로젝트
./documents\인공자능\신경망 이론의 복귀
            ./documents\인공지능\AI의 두번째 암흑기 19871993
ave file : ./documents\인공자능\인공자능의 겨울
save file : ./documents\인공지능\몸통을 갖는 것의 중요성 Nouvellle AI and embodied reasor
save file : ./documents\인공지능\AI 2018현재
```

데이터셋 WIKI depth 키워드 모델 성능의 한계

키워드를 찾지 못하면, 위키피디아 탐색을 할 수 없기 때문에 원하는 답변을 얻을 수 없다.

- → 강의자가 키워드 등록(일부해결),
- → 사용자가 키워드를 포함하여 질문



데이터셋 WIKI depth 키워드 모델 성능의 한계

korQuAD 1.0 dev set	BERT fine tuning 모델	korBERT
exact match	60	86
F1 score	80	94

다른 모델 적용하는 방법 고려 중

위키피디아를 이용한 질의응답은 데이터 셋의 구성, 모델의 특성 등의 이유로 서비스에 이용하기에 **부적합**하다 판단.

데이터셋

WIKI depth

키워드

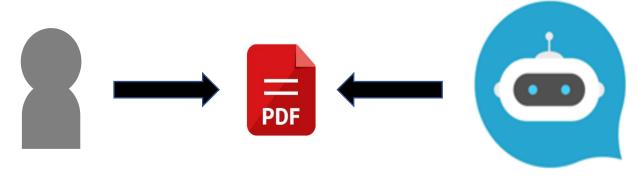
강의자가 강의 교안(pdf)를 업로드하는 방식으로 강의자료 기반의

질의응답 챗봇을 개선, 발전하는 것을 계획.

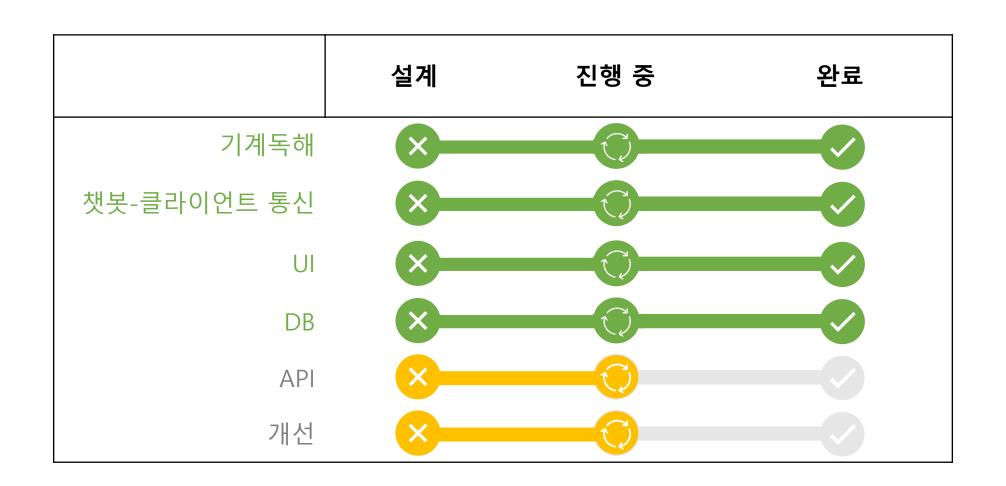
업로드된 pdf가 있을 때,

위키피디아에 비해 탐색 depth가 적다는 장점.

패텔방향의 한계



진행도



팀원 역할



박상현 UI REST API WEB



남승완 백엔드 서버 REST API DB



주준하 챗봇 서버 자연어처리 챗봇 모델링



Q&A

Q&A

감사합니다