



챗봇 프로젝트 04

Python

아나콘다 설치 – <https://www.anaconda.com/products/distribution>

설치 후 anaconda prompt 실행



Python

`conda --v` // 설치확인

`conda create -n chatbot python=3.8` // chatbot이라는 이름의 python 3.8버전 가상환경 생성

`conda env list` // 가상환경 리스트 확인

`conda activate chatbot` // chatbot 가상환경 활성화

// 라이브러리 설치

`pip install pandas`

`pip install sentence_transformers`

`pip install sklearn`

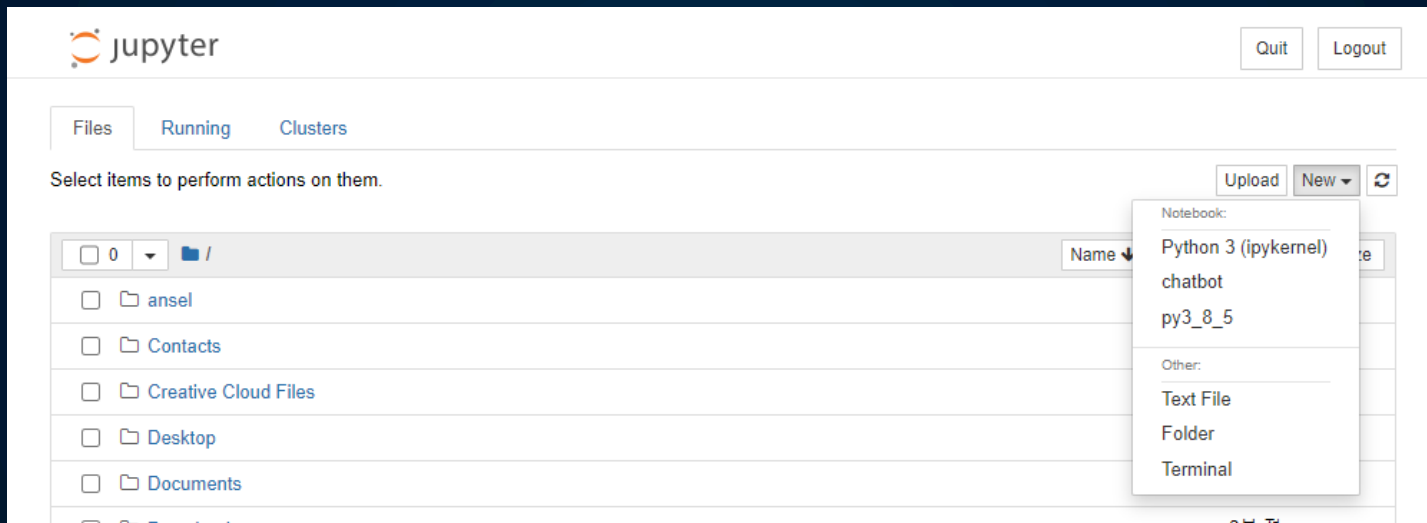
Python

```
pip install jupyter notebook // 가상환경에 주피터 노트북 설치
```

```
python -m ipykernel install -user -name chatbot --display-name chatbot  
// 주피터 노트북에 가상환경 연결하기
```

```
(chatbot) jupyter notebook  
// chatbot 가상환경이 선택된 상태에서 jupyter notebook 입력
```

Python



설치 성공 시 'chatbot' 가상환경 선택 가능

카카오 채널 가입

<https://business.kakao.com/info/kakaotalkchannel/>

추후 챗봇 연동을 위한 카카오톡 채널 가입

오픈빌더 심사통과를 위해
실제 서비스 처럼 채널을 개설

예시)

채널 이름 : 청년족발

채널 이름 : youngmzzokbal

소개글 : 족발 전문점 청년족발입니다.

채널 개설하기

카카오톡에서 고객을 만나는 방법! 지금 무료로 만들어 새로운 비즈니스를 경험하세요.

정보 입력 가이드

프로필 설정

관람 사이즈 : 640 x 640px / 지원파일 : jpg, jpeg (최대 10MB)

프로필 사진 첨부

배경 사진 첨부

채널 이름 *

검색용 아이디 *

소개글

부가 정보 설정

카테고리1 *

카테고리2

이전

확인

카카오 챗봇 관리자센터 참여 신청

i.kakao.com 접속

채널 URL : 생성한 카카오 채널의 URL 입력

신청사유

채널 생성과 동일하게 심사통과를 위해
실제 서비스 처럼 사유를 입력하기

예시)

식당 운영 정보(영업시간, 웨이팅) 등을
제공하기 위한 챗봇 구축

기업

개인

카카오톡 채널 홈 URL * ?

카카오톡 채널(구 플러스친구) 홈 URL을 입력해주세요

신청사유 *

신청 사유를 작성해주세요

네이버 톡톡 챗봇 신청

<https://partner.talk.naver.com/>

이미지, 프로필명 작성하여 챗봇 사용 신청

프로필 정보를 입력해주세요.

프로필 가져오기

연결된 서비스가 없습니다.

새로 작성하기

프로필 이미지 필수



파일찾기

파일유형 jpg,jpeg,png,gif / 최대파일크기 10MB

프로필명 필수

한글, 영문, 숫자, - (특수문자 하이픈) 포함

0자/20자

소개말

사용자의 간단한 소개말을 입력해 주세요.

0자/45자

홈페이지

http://

전화번호

전국대표

나머지 번호를 입력해주세요.

주소

고객이 찾아올 수 있는 주소를 입력해주세요.

주소찾기

심리 상담 봇 제작

심리상담 챗봇

당신:

전송

안녕하세요



기분이 풀릴 때까지 실컷 욕해도 좋아요. 제가 들어드릴게요.

저 너무 피곤해요



피곤하면 정말 아무 것도 하기 어렵죠. 잠깐 휴식을 취하는 건 어떨까요?

휴식을 취할 시간도 없어요



힘든 일이 있으신가 봐요. 저는 항상 여기에서 당신을 기다려요. 편할 때 말씀해주세요.

챗봇 대화 스크립트 데이터

AI Hub – <https://aihub.or.kr/>

회원가입 후

웹니스 대화 스크립트 데이터셋 다운로드

■ 신촌 세브란스 병원 정신건강 상담 기록 가공

○ 정신건강 상담 기록 데이터 총 16,104명 (진단받은 환자 군: 14,221명, 잠재적 환자 군: 1,883명)

Date	Informants	PI
2010-03-03	환자 본인	5~6년 전부터 닫혀 있는 공간이나 꽉 막혀 있는 공간이 무서워 밤에 잘 때 문을 열어놓고 잤다고 함...2009년 12월에 남동생과 사별..천식이 있어 스트레스를 받거나 공기가 나쁘거나 하면 숨이 막히는 느낌이 들기는 하나 그러나 오늘 정도로 심했던 적은 없었다고 함...환자는 사무실에서 서류 작업을 주로 하며 잊그게 까지도 일은 나갔다고 함...금일 백내장 수술을 받으려고 drep 을 하기 위해 얼굴에 포를 덮자 숨이 안 쉬어지면서 죽을 것 같은 느낌이 들었다고 함..수술의가 보기에 손도 떨리고 창백해 보였다고 전해 들음. ..5분여 정도 상기 증상 지속되다가 수술방에서 나오자마자 진정이 되었다고 함...아직도 답답한 느낌은 남아 있다고.

○ 상담기록 데이터를 문장단위로 분리 후 문장 내 담긴 대화 의도 분류 및 단어 추출 (30% 완료)

문장 단위 대화 분리

- 5~6년 전부터 닫혀 있는 공간이나 꽉 막혀 있는 공간이 무서워 밤에 잘 때 문을 열어놓고 잤다고 함...
- 2009년 12월에 남동생과 사별..
- 천식이 있어 스트레스를 받거나 공기가 나쁘거나 하면 숨이 막히는 느낌이 들기는 하나 그러나 오늘 정도로 심했던 적은 없었다고 함...
- 환자는 사무실에서 서류 작업을 주로 하며 잊그게 까지도 일은 나갔다고 함...
- 금일 백내장 수술을 받으려고 drep 을 하기 위해 얼굴에 포를 덮자 숨이 안 쉬어지면서 죽을 것 같은 느낌이 들었다고 함..
- 수술의가 보기에 손도 떨리고 창백해 보였다고 전해 들음.
- 5분여 정도 상기 증상 지속되다가 수술방에서 나오자마자 진정이 되었다고 함...
- 아직도 답답한 느낌은 남아 있다고.



대화 의도 분류

- 증상/수면양호/
- 배경/실패/대화
- 증상/자신감상실/버거움/업무
- 원인/군대/제대/실망
- 배경/육아
- 배경/건강/수술/유방암
- 현재상태/수면/그럭저럭
- 치료이력/복약/공황장애
- 증상/피해망상/헛소문/담배
- 증상/공황장애/증상악화
- 증상/긴장/상황회피
- 자가치료/한약
- 증상/외로움/과대망상/왕따
- 배경/성경체성/통생애
- 치료이력/공황장애/인지치료
- ...

단어 추출

- 가슴
- 가슴두근거림
- 가슴떨림
- 가슴통증
- 가정
- 가정불화
- 가족상담
- 갈등
- 감
- 감기
- 감수성
- 감정조절이상
- 갑상선 기능 항진증
- ...

인공지능 모델 공유 사이트

<https://huggingface.co/>

ko-sroberta-multitask 검색

jhgan/ko-sroberta-multitask 모델 사용

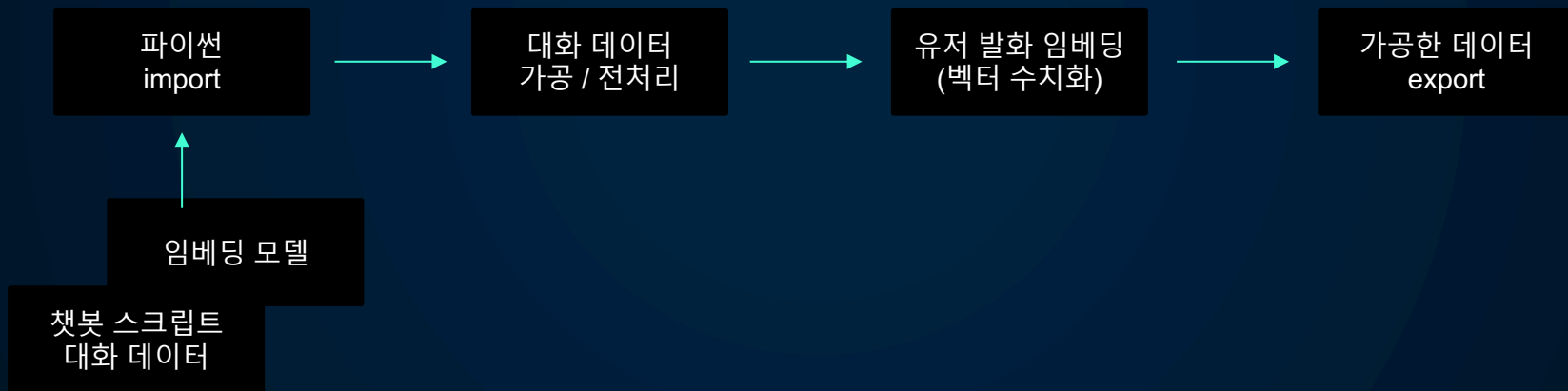


streamlit 설치

추가 라이브러리 streamlit / streamlit-chat

```
pip install streamlit  
pip install streamlit-chat
```

스크립트 데이터 가공



1. 심리 상담 봇 모델 구현 및 streamlit을 활용한 웹 어플리케이션 구현
2. Aihub의 다른 대화 데이터를 사용해
3. 대화 스크립트를 직접 작성하여 대화모델 구축

dataset.ipynb

```
import pandas as pd
from sentence_transformers import SentenceTransformer
from sklearn.metrics.pairwise import cosine_similarity
```


SentenceBERT 모델 로드

```
model = SentenceTransformer('jhgan/ko-sroberta-multitask')  
  
sentences = ["안녕하세요?", "한국어 문장 임베딩을 위한 버트 모델입니다."]   
embeddings = model.encode(sentences)  
  
print(embeddings)
```

```
[[-0.37510434 -0.7733841  0.5927711  ...  0.5792351  0.32683483  
  -0.6508966 ]  
 [-0.09361703 -0.18191542 -0.19230805 ... -0.03165817  0.30412528  
  -0.26793632]]
```

데이터셋 로드

웰니스 대화 스크립트 데이터셋

<https://aihub.or.kr/opendata/keti-data/recognition-language/KETI-02-006>

```
df = pd.read_csv('wellness_dataset_original.csv')  
df.head()
```

	구분	유저	챗봇	Unnamed: 3
0	감정/감정조절이상	제 감정이 이상해진 것 같아요. 남편만 보면 화가 치밀어 오르고 감정 조절이 안되요.	감정이 조절이 안 될 때만큼 힘들 때는 없는 거 같아요.	NaN
1	감정/감정조절이상	더 이상 내 감정을 내가 컨트롤 못 하겠어.	저도 그 기분 이해해요. 많이 힘드시죠?	NaN
2	감정/감정조절이상	하루종일 오르락내리락 롤러코스터 타는 기분이에요. 그럴 때는 밥은 잘 먹었는지, 잠은 잘 잤는지 체크해보는 것도 좋아요.		NaN
3	감정/감정조절이상	꼭 롤러코스터 타는 것 같아요.	NaN	NaN
4	감정/감정조절이상	롤러코스터 타는 것처럼 기분이 왔다 갔다 해요.	NaN	NaN

전처리

```
df = df.drop(columns=['Unnamed: 3'])  
df.head()
```

	구분	유저	챗봇
0	감정/감정조절이상	제 감정이 이상해진 것 같아요. 남편만 보면 화가 치밀어 오르고 감정 조절이 안되요.	감정이 조절이 안 될 때만큼 힘들 때는 없는 거 같아요.
1	감정/감정조절이상	더 이상 내 감정을 내가 컨트롤 못 하겠어.	저도 그 기분 이해해요. 많이 힘드시죠?
2	감정/감정조절이상	하루종일 오르락내리락 롤러코스터 타는 기분이에요. 그럴 때는 밥은 잘 먹었는지, 잠은 잘 잤는지 체크해보는 것도 좋아요.	
3	감정/감정조절이상	꼭 롤러코스터 타는 것 같아요.	NaN
4	감정/감정조절이상	롤러코스터 타는 것처럼 기분이 왔다 갔다 해요.	NaN

```
df = df[~df['챗봇'].isna()]
```

```
df.head()
```

	구분	유저	챗봇
0	감정/감정조절이상	제 감정이 이상해진 것 같아요. 남편만 보면 화가 치밀어 오르고 감정 조절이 안되요.	감정이 조절이 안 될 때만큼 힘을 때는 없는 거 같아요.
1	감정/감정조절이상	더 이상 내 감정을 내가 컨트롤 못 하겠어.	저도 그 기분 이해해요. 많이 힘드시죠?
2	감정/감정조절이상	하루종일 오르락내리락 롤러코스터 타는 기분이에요.	그럴 때는 밥은 잘 먹었는지, 잠은 잘 잤는지 체크해보는 것도 좋아요.
15	감정/감정조절이상/화	평소 다른 일을 할 때도 비슷해요. 생각한대로 안되면 화가 나고...그런 상황이 지속되...	화가 폭발할 것 같을 때는 그 자리를 피하는 것도 좋은 방법이라고 생각해요.
16	감정/감정조절이상/화	예전보다 화내는 게 과격해진 거 같아.	정말 힘드시겠어요. 화는 남에게도 스스로에게도 상처를 주잖아요.

```
df.loc[0, '유저']
```

'제 감정이 이상해진 것 같아요. 남편만 보면 화가 치밀어 오르고 감정 조절이 안되요.'

```
model.encode(df.loc[0, '유저'])
```

```
array([-4.80606496e-01, -2.94869184e-01,  4.37899917e-01, -6.40137970e-01,
        3.28670703e-02, -3.42647523e-01, -5.47482781e-02,  1.73056666e-02,
       -4.08220798e-01, -5.06033957e-01, -1.68733552e-01, -3.98677200e-01,
       -1.24776550e-01, -9.71538275e-02, -1.65286899e-01,  5.72609063e-03,
        6.13691993e-02, -1.91312209e-01,  2.53917545e-01, -5.85019588e-01,
       -2.84425557e-01, -2.32034832e-01, -3.27080727e-01,  6.72975853e-02,
       -1.65062283e-05, -4.72336203e-01, -3.60021919e-01,  2.91879803e-01,
       -6.63861275e-01, -3.10574532e-01,  5.79524934e-01, -3.11722696e-01,
        1.47698438e-02, -2.12172642e-01,  2.22058162e-01, -1.73828930e-01,
       -3.78458261e-01, -4.20398444e-01, -2.38218784e-01,  6.38706833e-02,
       -1.15304120e-01, -2.44564101e-01, -5.00228345e-01,  1.68355256e-01,
       -6.58360362e-01, -8.91942143e-01, -6.26957238e-01, -3.21965545e-01,
       -7.05358624e-01,  3.71447295e-01, -5.45803249e-01,  7.76298568e-02,
        1.88864848e-01,  8.68886681e-01,  5.88688814e-01,  4.88886888e-01,
```

유저 대화내용 인코딩

```
df['embedding'] = pd.Series([[]] * len(df)) # dummy
df['embedding'] = df['유저'].map(lambda x: list(model.encode(x)))
df.head()
```

	구분	유저	챗봇	embedding
0	감정/감정조절이 상	제 감정이 이상해진 것 같아요. 남편만 보면 화가 치밀어 오르고 감 정 조절이 안되요.	감정이 조절이 안 될 때만큼 힘을 때는 없는 거 같아요.	[-0.4806065, -0.29486918, 0.43789992, -0.64013...
1	감정/감정조절이 상	더 이상 내 감정을 내가 컨트롤 못 하겠어.	저도 그 기분 이해해요. 많이 힘드시죠?	[-1.1561579, -0.14506263, 0.29490355, -0.67394...
2	감정/감정조절이 상	하루종일 오르락내리락 롤러코스터 타는 기분이에요.	그럴 때는 밥은 잘 먹었는지, 잠은 잘 잤는지 체크해보는 것 도 좋아요.	[-0.6652002, -0.081268094, 1.0945568, 0.105792...
15	감정/감정조절이 상/화	평소 다른 일을 할 때도 비슷해요. 생각한대로 안되면 화가 나고...그 런 상황이 지속되...	화가 폭발할 것 같을 때는 그 자리를 피하는 것도 좋은 방법 이라고 생각해요.	[-0.76790583, 0.4652072, 0.52850723, -0.507604...
16	감정/감정조절이 상/화	예전보다 화내는 게 과격해진 거 같아요.	정말 힘드시겠어요. 화는 남에게도 스스로에게도 상처를 주 잖아요.	[-0.20277739, -0.37413904, 0.04053206, -0.8624...

데이터 전체를 인코딩하기 때문에 시간이 오래 소요됨

```
df.to_csv('wellness_dataset.csv', index=False)
```

데이터프레임을 csv 파일로 변환

간단한 챗봇

```
text = '요즘 머리가 아프고 너무 힘들어'  
embedding = model.encode(text)
```

생성한 데이터로 간단한 챗봇 테스트 해보기

```
df['distance'] = df['embedding'].map(lambda x: cosine_similarity([embedding], [x]).squeeze())
```

```
df.head()
```

	구분	유저	챗봇	embedding	distance
0	감정/감정조절이 상	제 감정이 이상해진 것 같아요. 남편만 보면 화가 치밀어 오르고 감정 조절이 안되요.	감정이 조절이 안 될 만큼 힘들 때는 없는 거 같아요.	[-0.4806065, -0.29486918, 0.43789992, -0.64013...	0.448967
1	감정/감정조절이 상	더 이상 내 감정을 내가 컨트롤 못 하겠어.	저도 그 기분 이해해요. 많이 힘드시죠?	[-1.1561579, -0.14506263, 0.29490355, -0.67394...	0.490199
2	감정/감정조절이 상	하루종일 오르락내리락 롤러코스터 타는 기분이에요.	그럴 때는 밥은 잘 먹었는지, 잠은 잘 잤는지 체크해보는 것도 좋아요.	[-0.6652002, -0.081268094, 1.0945568, 0.105792...	0.352131
15	감정/감정조절이 상/화	평소 다른 일을 할 때도 비슷해요. 생각한대로 안되면 화가 나 고...그런 상황이 지속되...	화가 폭발할 것 같을 때는 그 자리를 피하는 것도 좋은 방법이라고 생각해요.	[-0.76790583, 0.4652072, 0.52850723, -0.507604...	0.422284
16	감정/감정조절이 상/화	예전보다 화내는 게 과격해진 거 같아요.	정말 힘드시겠어요. 화는 남에게도 스스로에게도 상처 를 주잖아요.	[-0.20277739, -0.37413904, 0.04053206, -0.8624...	0.315118


```
answer = df.loc[df['distance'].idxmax()]

print('구분', answer['구분'])
print('유사한 질문', answer['유저'])
print('챗봇 답변', answer['챗봇'])
print('유사도', answer['distance'])
```

구분 증상/편두통

유사한 질문 요즘은 머리가 한쪽만 지그시 누르는 것처럼 무겁고 아파요.

챗봇 답변 으으, 머리가 아프면 정말 힘들죠. 그 마음 정말 이해해요.

유사도 0.8296288251876831

chatbot.py

chatbot.py 코드는 vscode를 통해 작성.
(jupyter notebook에서 작성하는 경우 확장자 차이로 인해 오류 발생)

```
import streamlit as st
from streamlit_chat import message
import pandas as pd
from sentence_transformers import SentenceTransformer
from sklearn.metrics.pairwise import cosine_similarity
import json
```

```
@st.cache(allow_output_mutation=True)
def cached_model():
    model = SentenceTransformer('jhgan/ko-sroberta-multitask')
    return model

@st.cache(allow_output_mutation=True)
def get_dataset():
    df = pd.read_csv('wellness_dataset.csv')
    df['embedding'] = df['embedding'].apply(json.loads)
    return df

model = cached_model()
df = get_dataset()
```

언어해석모델, 대화데이터파일을 캐싱, 실행할 때 마다 매번 불러오는 것이 아닌,
최초 실행 시 한번만 불러오고 해당 파일을 계속해서 사용가능

```
st.header('심리상담 챗봇')
```

심리상담 챗봇

streamlit 함수 st.header로 페이지 상단 제목 출력

```
if 'generated' not in st.session_state:  
    st.session_state['generated'] = []  
  
if 'past' not in st.session_state:  
    st.session_state['past'] = []
```

유저와 챗봇의 대화 내역을 저장하기 위한 배열을 생성

```
with st.form('form', clear_on_submit=True):  
    user_input = st.text_input('당신: ', '')  
    submitted = st.form_submit_button('전송')
```

유저가 채팅 내용을 입력할 수 있는 채팅 입력 폼을 만들고,
입력받은 내용과, 전송 버튼의 클릭 여부를 전달받는 변수를 생성

streamlit 실행

anaconda prompt(window) / terminal(mac)
cd (python 파일과 csv 데이터 파일이 있는 폴더로 경로 이동)

예시) (chatbot) C:\Users\Nextop(사용자이름)\chatbot

streamlit run chatbot.py 입력

최초 실행시 email 입력 필요 (이메일 인증 절차는 없음)

심리상담 챗봇

당신:

전송

정상 실행 시, 챗봇 레이아웃이 출력.

오류 발생 시 체크

1. streamlit, sklearn 등 라이브러리 설치 여부 확인
2. prompt / terminal 에서 폴더 경로 설정을 올바르게 했는지 확인
3. chatbot.py 파일 단순 오타가 있는지 확인


```
if submitted and user_input:
    embedding = model.encode(user_input)

    df['distance'] = df['embedding'].map(lambda x: cosine_similarity([embedding], [x]).squeeze())
    answer = df.loc[df['distance'].idxmax()]

    st.session_state.past.append(user_input)
    st.session_state.generated.append(answer['챗봇'])
```

유저가 내용을 입력하고, 전송버튼을 눌렀을 경우 실행.
유저가 입력한 문장을 인코딩하여
데이터프레임의 임베딩된 값과 비교하여 유사도를 판별.
유사도가 가장 높은 열의 '응답'을 answer에 저장.
그 후 유저와 챗봇의 대화 내용을 저장하는 배열에 각각 해당 내용을 저장

```
for i in range(len(st.session_state['past'])):
    message(st.session_state['past'][i], is_user=True, key=str(i) + '_user')
    if len(st.session_state['generated']) > i:
        message(st.session_state['generated'][i], key=str(i) + '_bot')
```

대화 내용이 저장된 배열에 마지막으로 저장된 내용을 대화창에 출력해준다

1. 심리 상담 봇 모델 구현 및 streamlit을 활용한 웹 어플리케이션 구현
2. Aihub의 다른 대화 데이터로 교체하기
3. 대화 스크립트를 직접 작성하여 챗봇에 적용

다른 데이터셋 활용하기

한국어

소상공인 고객 주문 질의-응답 텍스트

무인 상점에서 고객의 질의에 답변 가능한 고객 응대 학습용 AI 텍스트 데이터

#소상공인

상점

주문

질의응답















무인상점

챗봇

유형

텍스트

소상공인 고객 주문 질의 - 응답 텍스트 데이터셋 다운로드

 가구인테리어_train.csv	2021-10-27 오후 5:56	Microsoft Excel ...	89,026KB
 건강_train.csv	2021-10-27 오후 5:56	Microsoft Excel ...	24,790KB
 기타_train.csv	2021-11-16 오후 5:25	Microsoft Excel ...	15,906KB
 디지털가전_train.csv	2021-10-27 오후 5:58	Microsoft Excel ...	120,893KB
 병원_train.csv	2021-10-27 오후 5:59	Microsoft Excel ...	16,227KB
 뷰티_train.csv	2021-11-16 오후 5:27	Microsoft Excel ...	38,543KB
 생활잡화_train.csv	2021-10-27 오후 6:01	Microsoft Excel ...	107,562KB
 슈퍼_train.csv	2021-10-27 오후 6:01	Microsoft Excel ...	30,083KB
 식품_train.csv	2021-10-27 오후 6:02	Microsoft Excel ...	69,791KB
 음식점_train.csv	2021-11-16 오후 6:03	Microsoft Excel ...	33,202KB
 의류_train.csv	2021-11-16 오후 5:35	Microsoft Excel ...	57,427KB
 출산육아_train.csv	2021-11-16 오후 5:37	Microsoft Excel ...	61,273KB
 카페_train.csv	2021-11-16 오후 5:39	Microsoft Excel ...	20,460KB
 패션_train.csv	2021-11-16 오후 5:41	Microsoft Excel ...	29,650KB

다양한 업종의 질의응답 데이터를 csv 형태로 제공

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	IDX	발화자	발화문	카테고리	QA번호	QA여부	감성	인텐트	가격	수량	크기	장소	조직	사람	시간	날짜	상품명	상당번호	상당내순번
2	1	s	그래서 그 병원		8 q	m		제품_일반_확인										5455	15
3	2	c	네 병원		8 a	m		제품_일반_확인										5455	16
4	3	s	일반 밴드 병원		16 q	m		제품_일반_확인									밴드	5711	33
5	4	c	네 병원		16 a	m		제품_일반_확인										5711	34
6	5	s	건선이네도 병원		43 q	m		제품_일반_확인										5711	87
7	6	c	네 다른 데 병원		43 a	m		제품_일반_확인										5711	88
8	7	c	건선이라고 병원		47 q	m		제품_일반_확인										5711	99
9	8	s	네 그런 걸 병원		47 a	m		제품_일반_확인										5711	100
10	9	s	근데 그럴 병원		54 q	m		제품_일반_확인										5711	115
11	10	c	네 병원		54 a	m		제품_일반_확인										5711	116
12	11	s	보이는 곳 병원		54 q	m		제품_일반_확인										5711	117
13	12	c	아 아니요 병원		54 a	m		제품_일반_확인										5711	118
14	13	s	사람으로 병원		5 q	m		제품_일반_확인									사료	5582	7
15	14	c	아 그럼 고 병원		5 a	m		제품_일반_확인									고기 간식	5582	8
16	15	s	네 그게 제 병원		6 q	m		제품_일반_확인										5582	9
17	16	c	그럼 어떡 병원		6 a	m		제품_일반_확인										5582	10
18	17	s	훈련을 조 병원		7 q	m		제품_일반_확인									간식	5582	11
19	18	c	그럼 지금 병원		7 a	m		제품_일반_확인										5582	12
20	19	s	지금은 사 병원		8 q	m		제품_일반_확인									사료 간식	5582	13
21	20	c	그렇게 해 병원		8 a	m		제품_일반_확인									간식	5582	14
22	21	s	어쩔 수 없 병원		9 q	m		제품_일반_확인										5582	15
23	22	c	어떡해 바 병원		9 a	m		제품_일반_확인										5582	16
24	23	s	네 강아지 병원		12 q	m		제품_일반_확인									간식 고기	5582	21
25	24	c	신경을 쓰 병원		12 a	m		제품_일반_확인										5582	22
26	25	s	보이는 곳 병원		64 q	m		제품_일반_확인										5751	126
27	26	c	아 아니요 병원		64 a	m		제품_일반_확인										5751	127
28	27	s	신장이 안 병원		13 q	m		제품_일반_확인										3640	25
29	28	c	치료받고 병원		13 a	m		제품_일반_확인										3640	26
30	29	c	만성이에 병원		6 q	m		제품_일반_확인										2098	11
31	30	s	그렇죠 그 병원		6 a	m		제품_일반_확인										2098	12
32	31	s	인공침가 병원		39 q	m		제품_일반_확인									과자	5138	65
33	32	c	오 그럼 영 병원		39 q	m		제품_일반_확인										5138	66
34	33	s	제일 중요 병원		39 a	m		제품_일반_확인										5138	67
35	34	s	요새 막 간 병원		22 q	m		제품_일반_확인										3037	40

해당 csv 데이터에서 챗봇 개발을 위해 필요한 데이터는 “발화문” 질문, 답변이 하나의 열에 섞여 있어서 이것을 별개의 열로 구분해주는 작업이 필요

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	IDX	발화자	발화문	카테고리	QA번호	QA여부	감성	인텐트	가격	수량	크기	장소	조직	사람	시간	날짜	상품명	상담번호	상담내순
2	113405	c	붕대 안 썼 병원		1	q	m	교환 반품 환불_방법_질문									붕대	1	1
3	113406	s	네, 붕대 인 병원		1	a	m	교환 반품 환불_방법_질문									붕대	1	2
4	84127	c	붕대 안 썼 병원		2	q	m	교환 반품 환불_방법_요청									붕대	2	1
5	84128	s	네, 붕대 인 병원		2	a	m	교환 반품 환불_방법_요청									붕대	2	2
6	84419	c	붕대 안 썼 병원		4	q	m	교환 반품 환불_방법_요청									붕대	4	1
7	84420	s	네, 환불해 병원		4	a	m	교환 반품 환불_방법_요청										4	2
8	83769	c	붕대 안 쓸 병원		6	q	m	교환 반품 환불_방법_요청									붕대	6	1
9	83770	s	그럼요, 환 병원		6	a	m	교환 반품 환불_방법_요청										6	2
10	113261	c	붕대 안 쓸 병원		7	q	m	교환 반품 환불_방법_질문									붕대	7	1
11	113262	s	그럼요, 환 병원		7	a	m	교환 반품 환불_방법_질문										7	2
12	84321	c	붕대 안 쓸 병원		9	q	m	교환 반품 환불_방법_요청									붕대	9	1
13	84322	s	네, 환불해 병원		9	a	m	교환 반품 환불_방법_요청										9	2
14	83907	c	붕대 안 쓸 병원		10	q	m	교환 반품 환불_방법_요청									붕대	10	1
15	83908	s	네, 붕대 흰 병원		10	a	m	교환 반품 환불_방법_요청									붕대	10	2
16	84209	c	붕대 안 썼 병원		11	q	m	교환 반품 환불_방법_요청									붕대	11	1
17	84210	s	네, 안 쓰 병원		11	a	m	교환 반품 환불_방법_요청										11	2
18	84301	c	붕대는 안 병원		12	q	m	교환 반품 환불_방법_요청									붕대	12	1
19	84302	s	알겠습니다 병원		12	a	m	교환 반품 환불_방법_요청									붕대	12	2
20	84407	c	기본 제공 병원		13	q	m	교환 반품 환불_방법_요청									붕대	13	1
21	84408	s	네, 붕대 흰 병원		13	a	m	교환 반품 환불_방법_요청									붕대	13	2
22	84093	c	붕대는 안 병원		14	q	m	교환 반품 환불_방법_요청									붕대	14	1
23	84094	s	네, 환불 기 병원		14	a	m	교환 반품 환불_방법_요청										14	2
24	84757	c	붕대는 안 병원		15	q	m	교환 반품 환불_방법_요청									붕대	15	1
25	84758	s	네, 안 쓰 병원		15	a	m	교환 반품 환불_방법_요청										15	2
26	82218	c	심장사상충 병원		17	q	m	제품 가격 질문									심장사상충	17	1

원본 데이터 시트 이름 hospital_train 으로 변경.
필터 기능으로 전체 데이터를 상담번호 기준 오름차순으로 정렬

새로운 시트 생성.

A열 - idx (인덱스 순번 1000개 까지 생성)

B열 - 유저

C열 - 챗봇

	A	B	C
1	idx	유저	챗봇
2	1		
3	2		
4	3		
5	4		
6	5		
7	6		
8	7		
9	8		
10	9		
11	10		
12	11		
13	12		
14	13		
15	14		
16	15		
17	16		
18	17		
19	18		
20	19		
21	20		
22	21		
23	22		
24	23		
25	24		
26	25		

데이터 추출 시 참조할 열

A열 - idx

B열 - 발화자

c = 손님 / s = 가게

C열 - 발화문

오고 간 대화 내용

F - QA여부

Q = 질문 / A = 답변

S - 상담내순번

해당 QA에서 이뤄진 대화의 순번
(예시)

안녕하세요 이거 환불하고 싶어요 (1)

네 어떤 물건인가요(2)

옷을 샀는데 사이즈가 안 맞아요(3)

네 배송비는 개인부담 입니다(4)

	A	B	C
1	idx	유저	챗봇
2	1	봉대 안 썼으면 환불 가능해요?	네, 봉대 안 쓰셨으면 환불 가능합니다.
3	2	봉대 안 썼는데 환불 가능한가요?	네, 봉대 안 쓰셨다면 환불 가능합니다.
4	3		
5	4	봉대 안 썼는데 환불해 주실 수 있어요?	네, 환불해드릴 수 있습니다.
6	5		
7	6	봉대 안 쓸 건데 환불해 주실 수 있죠?	그럼요, 환불해드릴 수 있습니다.
8	7	봉대 안 쓸 건데 환불해 주실 수 있는 건가요?	그럼요, 환불해드릴 수 있어요.
9	8		
10	9	봉대 안 쓸 건데 환불해 주시겠어요?	네, 환불해드리겠습니다.
11	10	봉대 안 쓸 건데 환불해 주실래요?	네, 봉대 환불해드릴게요.
12	11	봉대 안 썼으니가 환불되는 거 맞죠?	네, 안 쓰셨으면 환불됩니다.
13	12	봉대는 안 쓸 거니까 환불해 주세요.	알겠습니다, 봉대 환불해드릴게요.
14	13	기본 제공되는 봉대는 안 쓸 건데 환불 되죠?	네, 봉대 환불 됩니다.
15	14	봉대는 안 쓸 건데 환불 가능한 거 맞나요?	네, 환불 가능한 것 맞습니다.
16	15	봉대는 안 쓰면 환불 가능한 거 맞죠?	네, 안 쓰시면 환불 가능한 것 맞습니다.
17	16		
18	17	심장사상중 예방약 1만원 넘어요?	네, 1만원 넘습니다.
19	18		
20	19	심장사상중 예방약 비용은요?	1만 5천원입니다.
21	20	심장사상중 예방약 비용이 어떻게 되나요?	12,500원이예요.
22	21	심장사상중 예방약 비용은 어떻게 돼요?	심장사상중 예방약 비용은 12,500원입니다.
23	22		
24	23	심장사상중 예방약 받으러 왔는데, 얼마예요?	심장사상중 예방약은 12,500원이예요.
25	24	심장사상중 예방약 받으러 왔는데, 얼마죠?	12,500원입니다.
26	25	심장사상중 예방약 받으러 왔는데, 얼마가요?	심장사상중 예방약 비용은 12,500원이예요.
27	26	심장사상중 예방약 받으러 왔는데, 얼마 내야 돼요?	12,500원 내시면 됩니다.
28	27	심장사상중 예방약 받으러 왔는데, 얼마 내면 되나요?	12,500원 내시면 돼요.
29	28	심장사상중 예방약 받으러 왔는데, 얼마 낼까요?	12,500원 내주시면 됩니다.
30	29	심장사상중 예방약 받으러 왔는데, 결제 금액은요?	결제 금액은 12,500원입니다.
31	30	심장사상중 예방약 받으러 왔는데, 결제 금액이 어떻게 돼요?	결제 금액은 12,500원이예요.

유저 열에 추출할 발화문의 참조 조건

발화자 - c(손님)
QA여부 - q(질문)
idx - idx일치여부
상담내순번 - 1

챗봇 열에 추출할 발화문의 참조 조건

발화자 - s(가게)
QA여부 - a(답변)
idx - idx일치여부
상담내순번 - 2

	A	B	C
1	idx	유저	챗봇
2		1 봉대 안 썼으면 환불 가능해요?	네, 봉대 안 쓰셨으면 환불 가능합니다.
3		2 봉대 안 썼는데 환불 가능한가요?	네, 봉대 안 쓰셨다면 환불 가능합니다.
4		3	
5		4 봉대 안 썼는데 환불해 주실 수 있어요?	네, 환불해드릴 수 있습니다.
6		5	
7		6 봉대 안 쓸 건데 환불해 주실 수 있죠?	그럼요, 환불해드릴 수 있습니다.
8		7 봉대 안 쓸 건데 환불해 주실 수 있는 건가요?	그럼요, 환불해드릴 수 있어요.
9		8	
10		9 봉대 안 쓸 건데 환불해 주시겠어요?	네, 환불해드리겠습니다.
11		10 봉대 안 쓸 건데 환불해 주실래요?	네, 봉대 환불해드릴게요.
12		11 봉대 안 썼으니깐 환불되는 거 맞죠?	네, 안 쓰셨으면 환불됩니다.
13		12 봉대는 안 쓸 거니까 환불해 주세요.	알겠습니다, 봉대 환불해드릴게요.
14		13 기본 제공되는 봉대는 안 쓸 건데 환불 되죠?	네, 봉대 환불 됩니다.
15		14 봉대는 안 쓸 건데 환불 가능한 거 맞나요?	네, 환불 가능한 것 맞습니다.
16		15 봉대는 안 쓰면 환불 가능한 거 맞죠?	네, 안 쓰시면 환불 가능한 것 맞습니다.
17		16	
18		17 심장사상충 예방약 1만원 넘어요?	네, 1만원 넘습니다.
19		18	
20		19 심장사상충 예방약 비용은요?	1만 5천원입니다.
21		20 심장사상충 예방약 비용이 어떻게 되나요?	12,500원이에요.
22		21 심장사상충 예방약 비용은 어떻게 돼요?	심장사상충 예방약 비용은 12,500원입니다.
23		22	
24		23 심장사상충 예방약 받으러 왔는데, 얼마예요?	심장사상충 예방약은 12,500원이에요.
25		24 심장사상충 예방약 받으러 왔는데, 얼마죠?	12,500원입니다.
26		25 심장사상충 예방약 받으러 왔는데, 얼마가요?	심장사상충 예방약 비용은 12,500원이에요.
27		26 심장사상충 예방약 받으러 왔는데, 얼마 내야 돼요?	12,500원 내시면 됩니다.
28		27 심장사상충 예방약 받으러 왔는데, 얼마 내면 되나요?	12,500원 내시면 돼요.
29		28 심장사상충 예방약 받으러 왔는데, 얼마 낼까요?	12,500원 내주시면 됩니다.
30		29 심장사상충 예방약 받으러 왔는데, 결제 금액은요?	결제 금액은 12,500원입니다.
31		30 심장사상충 예방약 받으러 왔는데, 결제 금액이 어떻게	결제 금액은 12,500원이에요.

FILTER 함수

=FILTER(배열, 포함, [if_empty])

배열, 범위에서 원하는 조건으로 데이터를 조회하는 함수

배열 : 조회(필터링)할 배열 또는 범위

포함 : 조회(필터링)의 조건

[if_empty] : 찾는 값이 없을 때 표시할 값

B2열 입력

=FILTER(hospital_train!C:C,(hospital_train!R:R=A2)*
(hospital_train!S:S=1)*(hospital_train!B:B="c")*(hospital_train!F:F="q"),")

C2열 입력

=FILTER(hospital_train!C:C,(hospital_train!R:R=A2)*
(hospital_train!S:S=1)*(hospital_train!B:B="c")*(hospital_train!F:F="q"),")

실행결과

	A	B	C
1	idx	유저	챗봇
2	1 봉대 안 썼으면 환불 가능해요?		네, 봉대 안 쓰셨으면 환불 가능합니다.

정상 출력 시

1000번째 열까지 자동 채우기

다른 이름으로 저장

hospital_dataset(사용 데이터에 따라 설정)

CSV UTF-8 형식으로 저장

1069	이 약 먹을 때 술 마시면 안 되죠?	네, 술 절대 드리지 마세요.
1070	이 진통제 한 번에 두 알까지 먹을 수 있는 거 맞죠?	네, 1회 2알까지 복용 가능합니다.
1071	라식 후에도 클감 있는 인공눈물 사용해도 괜찮은 건가요?	가급적이면 클감 없는 제품 사용해주시는 게 좋아요.
1072	생리식염수 제일 작은 용량은 몇 밀리로 나왔죠?	지금 있는 것 중에선 20밀리짜리 20개입으로 나온 게 제일 작아요.
1073		
1074		
1075	이지엔시스 프로 하루 최대 복용량이 몇 알이나 되죠?	최대 4알 초과 안 하도록 드셔야 해요.
1076	혈압약이 떨어졌는데요	네네
1077	인사돌 알약 색깔 주황색인 거 맞죠?	네, 주황색 알약 맞습니다.
1078	알약 너무 큰 건 잘 못 먹어서 그런데 베크롬씨는 알약 크기 다른 약에 비해서 별로 큰 편은 아니에요.	
1079		
1080	이지엔시스 스트롱도 해열 기능 있는 거죠?	아뇨, 스트롱은 성분이 달라서 해열 기능은 없어요.
1081	탁센 이브는 일반이랑 뭐가 달라요?	생리 기간 붓기를 완화시켜주는 성분이 들어있어요.
1082		
1083	당뇨 검사할 때 이 약 먹는다고 얘기 해야 하나요?	네, 당뇨 검사시에 미리 복용 사실 알려드려야 하는 약이에요.
1084		
1085		
1086	이 진통제에 들어 있는 파마프로펜이란 성분은 뭔가요?	부종이나 경련을 예방하기 위해 들어간 성분이에요.
1087	텍시부프로펜이라는 건 이부프로펜이란 어떨게 다르죠?	이부프로펜의 진통효과는 강화하고 부작용은 줄인 게 텍시부프로펜이에요.
1088	이 장용캡슐이라는 건 캡슐이 장에 가서 녹는다는 뜻이죠?	네, 위에서 안 녹고 장에서 녹게 만든 제품이에요.
1089	이 숙취해소제 안에 몇 포 들어 있는 거죠?	이 제품은 한 팩에 5포 들어 있어요.
1090		
1091	인공눈물 이 두 가지 중에 어떤 게 클감이 더 강해요?	이 진한 녹색이 더 클감이 강한 제품이에요.
1092	이 방수 밴드 너무 잘 떨어지는데 접착력 좋은 제품은 없어요?	이 제품이 잘 안 떨어지고 방수도 잘 되고 좋아요.
1093	요즘 눈이 영 침침한데 괜찮은 영양제 있을까요?	눈 영양제 찾으시면 이 제품이 루테인 함량도 높고 편찮아요.
1094	손목 보호대 중에 서포트 좀 확실하게 되는 걸로 추천 좀 해주실래요?	손목 보호대 중에서는 이 제품이 서포트하는 힘이 좋아서 추천해요.
1095	목 아플 때 육칸디가 괜찮은가요?	육칸디보다 목에 뿌리는 게 더 나아요.
1096		
1097	효과 좀 빠르고 간독성 없는 진통제 추천해주세요.	그럼 이 진통제 추천해드릴게요.
1098	쓴맛 안 나는 어린이용 해열제 뭐가 괜찮아요?	어린이용 해열제 중에서는 이 제품이 제일 괜찮아요.
1099	요새 구내염이 너무 자주 생기는데 괜찮은 영양제 좀 추천해주세요.	구내염 때문에 고민이시면 이 제품 추천드릴게요.
1100	먹어도 속 안 쓰린 진통제 좀 있을까요?	이 진통제는 위장 부담 없이 나온 제품이에요.

실행결과

데이터셋 로드

헬니스 대화 스크립트 데이터셋

<https://aihub.or.kr/opendata/keti-data/recognition-laguage/KETI-02-006>

```
df = pd.read_csv('wellness_dataset_original.csv')
df.head()
```

	구분	유저	뜻	Unnamed: 3
0	감정/감정조절이상	제 감정이 이상해진 것 같아요. 남편만 보면 화가 지밀어 오르고 감정 조절이 안되요.	감정이 조절이 안 될 때만큼 힘들 때는 없는 거 같아요.	NaN
1	감정/감정조절이상			
2	감정/감정조절이상	하루종일		
3	감정/감정조절이상			
4	감정/감정조절이상	물?		

전처리

```
df = df.drop(columns=['Unnamed: 3'])
df.head()
```

	구분
0	감정/감정조절이상
1	감정/감정조절이상
2	감정/감정조절이상
3	감정/감정조절이상
4	감정/감정조절이상

유저 대화내용 인코딩

```
df['embedding'] = pd.Series([[] * len(df)]) # dummy
df['embedding'] = df['유저'].map(lambda x: list(model.encode(x)))
df.head()
```

	구분	유저	뜻	embedding
0	감정/감정조절이상	제 감정이 이상해진 것 같아요. 남편만 보면 화가 지밀어 오르고 감정 조절이 안되요.	감정이 조절이 안 될 때만큼 힘들 때는 없는 거 같아요.	[-0.4806065, -0.29486918, 0.43789992, -0.64013...
1	감정/감정조절이상	더 이상 내 감정을 내가 컨트롤 못 하겠어.	저도 그 기분 이해해요. 많이 힘들시죠?	[-1.1561579, -0.14506263, 0.29490355, -0.67394...
2	감정/감정조절이상	하루종일 오락가락내리막 롤러코스터 타는 기분이에요.	그럴 때는 밥은 잘 먹었는지, 잠은 잘 잤는지 체크해보는 것도 좋아요.	[-0.6652002, -0.081268094, 1.0945568, 0.105792...
15	감정/감정조절이상/화			
16	감정/감정조절이상/화			

```
df.to_csv('wellness_dataset.csv', index=False)
```

1. 기존에 작성한 심리상담 챗봇 코드에 데이터셋을 바꿔서 적용, 실행해보기

2. Aihub에서 다른 유형의 csv 대화 데이터를 받아서 전처리 가공 및 적용해보기

python으로 구현한 챗봇 텔레그램 연동

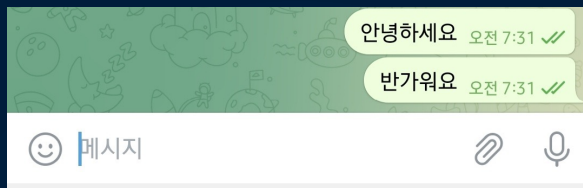
기존에 만들었던 텔레그램 봇 재사용(없을 경우 새로 생성)

bot 토큰 필요

<https://api.telegram.org/bot{봇토큰}/getUpdates>

ex) <https://api.telegram.org/bot6148368452:AAF-w4V91NLqst87DrBaJmMxK80M4p62HNI/getUpdates>

```
{"ok":true,"result":[]}
```

```
{
  "ok": true,
  "result": [
    {
      "update_id": 28530052,
      "message": {
        "message_id": 103,
        "from": {
          "id": 6095113280,
          "is_bot": false,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "language_code": "ko"
        },
        "chat": {
          "id": 6095113280,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "type": "private"
        },
        "date": 1682202658,
        "text": "\uc57c\uc548\ub155"
      },
      "update_id": 28530053,
      "message": {
        "message_id": 104,
        "from": {
          "id": 6095113280,
          "is_bot": false,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "language_code": "ko"
        },
        "chat": {
          "id": 6095113280,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "type": "private"
        },
        "date": 1682202659,
        "text": "\uc548\ub155"
      },
      "update_id": 28530054,
      "message": {
        "message_id": 105,
        "from": {
          "id": 6095113280,
          "is_bot": false,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "language_code": "ko"
        },
        "chat": {
          "id": 6095113280,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "type": "private"
        },
        "date": 1682202660,
        "text": "\ud558\uc774"
      },
      "update_id": 28530055,
      "message": {
        "message_id": 106,
        "from": {
          "id": 6095113280,
          "is_bot": false,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "language_code": "ko"
        },
        "chat": {
          "id": 6095113280,
          "first_name": "\uc720\ucbe48",
          "last_name": "\uc774",
          "type": "private"
        },
        "date": 1682202661,
        "text": "\ubb50\ud568"
      }
    ]
  ]
}
```

API 리퀘스트로 채널에 메시지 전송

```
import requests

token = ' 본인토큰입력 '
chat_id = 채팅id입력
bot_message = " hello "

url =
f"https://api.telegram.org/bot{token}/sendMessage?chat_id={c
hat_id}&text={bot_message}"
requests.get(url)
```

```
{'ok': True, 'result': [
  {'update_id': 28530020, 'message': {'message_id': 14, 'from': {'id': 6095113280, 'is_bot': False, 'first_name': '유빈', 'last_name': '이', 'language_code': 'ko'}, 'chat': {'id': 6095113280, 'first_name': '유빈', 'last_name': '이', 'type': 'private'}, 'date': 1682135197, 'text': '안녕하세요'},
  {'update_id': 28530021, 'message': {'message_id': 15, 'from': {'id': 6095113280, 'is_bot': False, 'first_name': '유빈', 'last_name': '이', 'language_code': 'ko'}, 'chat': {'id': 6095113280, 'first_name': '유빈', 'last_name': '이', 'type': 'private'}, 'date': 1682135198, 'text': '반갑습니다'}
],
```

텔레그램 api의 getUpdates를 통해 가져온 JSON 파일에서 'text' 항목 추출 필요

streamlit으로 구현했던 심리상담 봇의 코드를 일부 수정하는 것으로 쉽게 연동이 가능.