

캡스톤디자인 최종발표

뉴스분석과 키워드 추출

뉴스마스터

20184412 박현규
20203156 김하늘
20204498 이아영
20203307 정은선
20184422 김태원

목차

개요 및 필요성

어떤 환경에서 작업

개발환경

어떤 환경에서 작업

프로젝트 구현방법

프로젝트 대표 구현방법

프로젝트 구조

프로젝트 구상도

진행과정

크롤링, wordtovec, chatbot
구현결과

최종결과

챗봇 형태로 구현

기대효과

어떤 프로그램을 만들것인가

개요 및 필요성

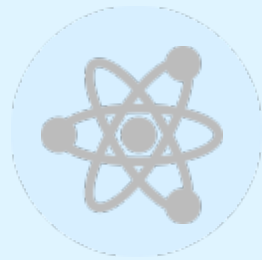
트렌드 및 정보 파악



**뉴스 데이터를 분석하여
챗봇 형태로 소비자에게 뉴스를 전달해주자!**

개요 및 필요성

트렌드 및 정보 파악



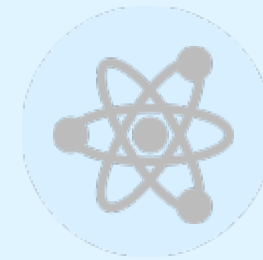
뉴스 소비의 양 변화

인터넷과 모바일 기기의 보급으로 뉴스에 접근하는 방법이 다양화되었으며, 많은 양의 정보가 사용자에게 제공됩니다. 이에 따라 뉴스를 효율적으로 소비하고 필요한 정보를 추출하는 도구의 필요성이 커졌습니다.



뉴스 분석 기술의 발전

자연어 처리, 머신러닝, 딥러닝 등의 기술의 발전으로 뉴스 분석과 관련된 다양한 기능이 개발되었습니다. 이러한 기술을 활용하여 뉴스를 요약, 키워드 추출, 트렌드 분석 등을 수행할 수 있으며, 사용자에게 가치 있는 정보를 제공할 수 있습니다.



사용자 경험의 개선

사용자들은 정보에 대한 빠른 접근과 필요한 내용의 신속한 파악을 원합니다. 기존의 뉴스 웹사이트나 앱은 사용자들이 원하는 정보를 찾기 위해 수많은 기사를 살펴봐야 하는 번거로움이 있습니다. 이에 대한 사용자 경험을 개선하고 편의성을 제공하기 위해 개인화된 챗봇 형식으로 정보를 제공하는 프로젝트가 필요하게 되었습니다.

프로젝트 수행절차 단계, 목표



개발 단계

1. 목표설정
2. 데이터 수집
3. 데이터 전처리
4. 뉴스 분석 알고리즘 개발
5. 챗봇 개발

목표는 사용자에게 신속하고 정확한 뉴스 요약,
주요 이벤트 파악, 키워드 추출, 트렌드 분석 등의 가치 있는 정보를 제공하는 것

개발환경

어떤 환경에서 작업



뉴스의 동향 파악 향상
의사 결정에 대한 정보 효과적 제공

python™



kakao i developers

GitHub



프로젝트 구현 방법

주요 기술들



크롤링(crawling)

웹사이트(website), 하이퍼링크(hyperlink), 데이터(data), 정보 자원을 자동화된 방법으로 수집, 분류, 저장하는 것.

텍스트 마이닝(text mining)

특정 목적에 맞게 유의미한 정보를 추출하는 분석 및 처리 과정

챗봇(chat bot)

채터봇(Chatterbot)은 음성이나 문자를 통한 인간과의 대화를 통해서 특정한 작업을 수행하도록 제작된 컴퓨터 프로그램이다

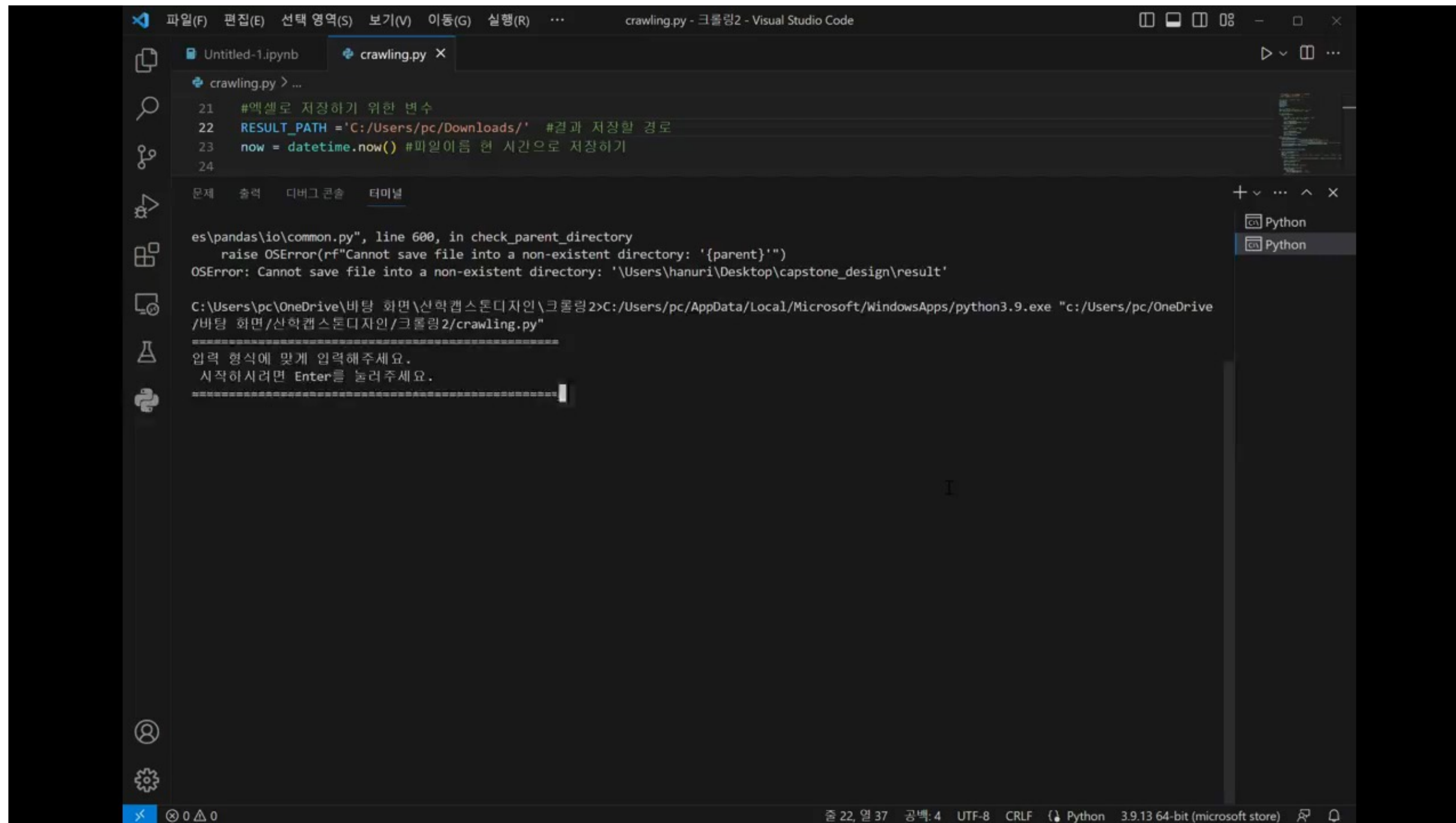
프로젝트 구조

구상도



진행과정

크롤링 실행결과



```
21 #엑셀로 저장하기 위한 변수
22 RESULT_PATH = 'C:/Users/pc/Downloads/' #결과 저장할 경로
23 now = datetime.now() #파일 이름 현 시간으로 저장하기
24

es\pandas\io\common.py", line 600, in check_parent_directory
    raise OSError(rf"Cannot save file into a non-existent directory: '{parent}'")
OSError: Cannot save file into a non-existent directory: '\Users\hanuri\Desktop\capstone_design\result'

C:\Users\pc\OneDrive\바탕 화면\산학캡스톤디자인\크롤링 2>C:/Users/pc/AppData/Local/Microsoft/WindowsApps/python3.9.exe "c:/Users/pc/OneDrive
/바탕 화면/산학캡스톤디자인/크롤링 2/crawling.py"

압력 형식에 맞게 입력해주세요.
시작하시려면 Enter를 눌러주세요.
=====
```



크롤링 실행 영상

네이버 뉴스 데이터를 가져와 검색어를 입력받
고 관련도순,최신순,오래된순으로 csv파일에 저
장 아래는 저장결과

로컬 디스크 (C:) > 사용자 > pc > 다운로드

다운로드 검색

이름	수정한 날짜	유형	크기
오늘			
crawling_result	2023-06-06 오후 4:38	Microsoft Excel 실효...	9KB

진행과정

크롤링 실행결과

크롤링

	date	title	source	contents	link				
0	1시간 전	LGU+ 반려동물	ZDNet Korea	LG유플러스는	https://zdnet.co.kr/view/?no=20230517085639				
1	23시간 전	반려동물 축적	뉴스시스언론사	기사내용 요약	http://www.newsis.com/view/?id=NISX20230516_0002304551&cID=10401&pID=10400				
2	23시간 전	반려동물 보살	매일경제언론	이종식 반려	https://www.mk.co.kr/article/10737007				
3	19시간 전	구팡, 반려동물	뉴스1	구팡은 반려	https://www.news1.kr/articles/5048505				
4	20시간 전	양평군, 동물	연합뉴스	경기 양평군은	https://www.yna.co.kr/view/AKR20230516103100061?input=1195m				
5	20시간 전	반려동물 수의	데일리한국	사진= 구팡 저	https://daily.hankooki.com/news/articleView.html?idxno=957529				
6	19시간 전	서대문구, 반려	이투데이	이투데이=김기	https://www.etoday.co.kr/news/view/2249586				
7	7시간 전	임실에서 반려	동아일보	국내 반려동물	https://www.donga.com/news/Society/article/all/20230516/119316963/1				
8	20시간 전	미펫, 반려동물	이데일리	이번 미펫 가	http://www.edaily.co.kr/news/newspath.asp?newsid=02876566635609576				
9	18시간 전	충북대 수의대	뉴스1	충북대학교는	https://www.news1.kr/articles/5048523				
10	20시간 전	미펫, 반려동물	이데일리	이번 미펫 가	http://www.edaily.co.kr/news/newspath.asp?newsid=02876566635609576				
11	2시간 전	친환경·안전	뉴스시스	특히 어린아	http://www.newsis.com/view/?id=NISX20230516_0002305458&cID=10433&pID=13000				
12	1일 전	염록스 백신	코메디닷컴	정부는 의심	https://kormedi.com/?p=1590405				
13	18시간 전	대구보건대, 노컷	뉴스	대구보건대 본	https://www.nocutnews.co.kr/news/5944446				
14	1일 전	바이오노트, 노컷	뉴스	점점 높아지	http://moneys.mt.co.kr/news/mwView.php?no=2023051608494965186				
15	1일 전	아이·반려동물	한국경제TV	특히 어린아	http://www.wowtv.co.kr/NewsCenter/News/Read?articleId=A202305160414&t=NN				
16	1일 전	"반려동물과"	레이디경향	더불어 동물	https://lady.khan.co.kr/khlady.html?mode=view&artid=202305160726011&code=11				
17	3시간 전	"배를 많이	할	국민일보	구팡이 반려	https://news.kmib.co.kr/article/view.asp?arcid=0018267618&code=61171811&cp=nv			
18	20시간 전	영진전문대	동	한국대학신문	곽지현 애니	https://news.unn.net/news/articleView.html?idxno=546739			
19	14시간 전	영진전문대	동	NSP통신	곽지현 애니	https://www.nspna.com/news/?mode=view&newsid=637572			
20	1분 전	LGU+, 반려	동물IT비즈뉴스	염상필 LG유플	https://www.itbiznews.com/news/articleView.html?idxno=99270				
21	47분 전	마이펫닥터,	뉴스시스	그러면서 "마	https://www.sisunnews.co.kr/news/articleView.html?idxno=183634				
22	17시간 전	대구보건대	이베리타스알파	대구보건대는	http://www.veritas-a.com/news/articleView.html?idxno=457194				
23	17시간 전	충북대, 반려	충북일보	오는 21일 청	https://www.inews365.com/news/article.html?no=765124				
24	1일 전	서대문구, 반려	전국매일신문	이 사업은 반	https://www.jeonmae.co.kr/news/articleView.html?idxno=957858				
25	19시간 전	파주시-파주시	동부일보	파주시는 다	http://www.joongboo.com/news/articleView.html?idxno=363592368				
26	23시간 전	구팡, 반려	동물디스커버리	▲ 구팡, 반	http://www.discoverynews.kr/news/articleView.html?idxno=1012002				
27	20시간 전	구팡, 수의사	디지털투데이	수의사가 실	http://www.digitaltoday.co.kr/news/articleView.html?idxno=476559				
28	18시간 전	제주에서 무	뉴스제주	제주에서 무	http://www.newsjeju.net/news/articleView.html?idxno=389031				
29	16시간 전	대구보건대	본	뉴스프리존	대구보건대학	http://www.newsfreezone.co.kr/news/articleView.html?idxno=469723			
30	19시간 전	서울호서직업	뉴스	서울호서직업	http://www.kukinews.com/newsView/kuk202305160168				
31	20시간 전	반려동물	강남구민일보	반려동물의	http://www.thegangnamtimes.co.kr/news/articleView.html?idxno=58016				

크롤링 코드를 실행한 결과, 기사가 올라온 시간, 기사 제목, 기사 본문, 기사 링크를 추출하는 csv파일이 출력됨 뉴스크롤링을 사용하기 위해 html파일을 추가적으로 봐야했음. html에 익숙하지 않아서 xPath와 같은 경로의 의미를 인터넷으로 공부함

진행과정

텍스트마이닝 실행결과



반려동물	98
동물	25
쿠팡	21
닥터	21
서비스	20



기사본문의 명사들을png파일으로 명사추출
왼쪽사진은 반려동물 뉴스기사 검색 실행결과
NLTK를 활용하여 키워드 추출을 하려고했으나
한글지원이안되어 막혔고, konlpy 를 찾아 키워
드추출
한글 형태소 분석을 지원하는 konlpy 패키지에
서 okt모듈을 사용하여 크롤링한 뉴스기사 중 본
문의 명사를 뽑아 워드클라우드에 사용함

진행과정 link 제공 실행결과

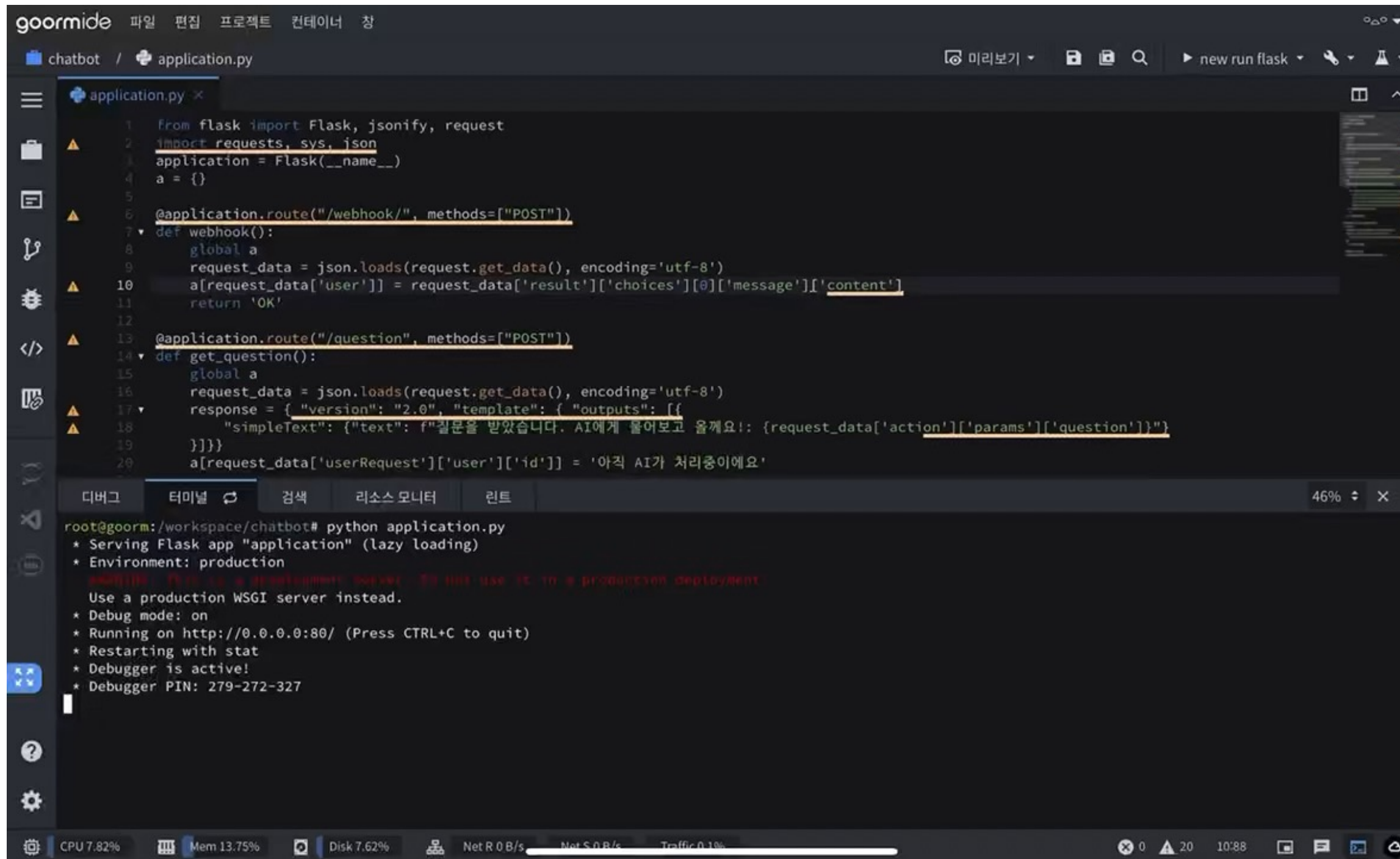
```
(py36) C:\python_test>python links.py
83      https://www.xportsnews.com/article/1716975
Name: link, dtype: object
58      https://sports.donga.com/article/all/20230426/119022124/1
Name: link, dtype: object
63      http://enews.imbc.com/News/RetrieveNewsInfo/380327
Name: link, dtype: object
```

검색한 기사의 url이 실행결과로 뜸

정보를 가져와서 단어 많은순로 url 결과를 출력

텍스트마이닝에서 뽑은 명사를 모두 링크를 추출하는데 사용하기에는 빈도수가 높은 5개 정도의 단어를 제외하고는 본문에 한 두번 정도 나오기 때문에 필요하지 않다고 느낌

진행과정 챗봇 실행시 터미널



The screenshot shows the Goorm IDE interface. The top bar includes the Goorm logo and navigation links: 파일 (File), 편집 (Edit), 프로젝트 (Project), 컨테이너 (Container), and 창 (Window). The breadcrumb path is chatbot / application.py. The main editor displays the code for application.py, which is a Flask application with two endpoints: /webhook and /question. The /webhook endpoint uses the requests library to fetch data from an external API. The /question endpoint returns a JSON response with a version, template, and a simple text message. The terminal at the bottom shows the command root@goorm:/workspace/chatbot# python application.py and its output, which includes Flask's startup messages, environment information, and the server's address (http://0.0.0.0:80/).

```
1 from flask import Flask, jsonify, request
2 import requests, sys, json
3 application = Flask(__name__)
4 a = {}
5
6 @application.route("/webhook/", methods=["POST"])
7 def webhook():
8     global a
9     request_data = json.loads(request.get_data(), encoding='utf-8')
10    a[request_data['user']] = request_data['result']['choices'][0]['message']['content']
11    return 'OK'
12
13 @application.route("/question", methods=["POST"])
14 def get_question():
15     global a
16     request_data = json.loads(request.get_data(), encoding='utf-8')
17     response = { "version": "2.0", "template": { "outputs": [{
18         "simpleText": { "text": f"질문을 받았습디다. AI에게 물어보고 올께요!: {request_data['action']['params']['question']}"
19     } ] } }
20     a[request_data['userRequest']['user']['id']] = '아직 AI가 처리중이에요'
```

```
root@goorm:/workspace/chatbot# python application.py
* Serving Flask app "application" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Running on http://0.0.0.0:80/ (Press CTRL+C to quit)
* Restarting with stat
* Debugger is active!
* Debugger PIN: 279-272-327
```

ai에게 질문을받고 대답을 이끌어나가는 과정에서 터미널의 변화

질문을 입력받으면 /question, 답을 입력 받으면 /ans, openai에서 질문을 받게되면 /webhook 함수가 실행되며 http 상태코드가 나옴

openai를 사용할때마다 결제를 해야하는데 그걸 몰라서 다른 데에서 방법을 갈구함 asynclia 사용량에 제한이 있어 한도 추가에 시간이 오래 걸림

최종결과

챗봇형태로 구현



뉴스봇 앱 오전 11:56

wordcloud ▼



반려동물 관련 검색 워드클라우드를 보니 반려, 건강,치료와 같은 키워드 들이 보임

최종결과

챗봇형태로 구현



뉴스봇 앱 오전 11:13

링크는 <https://www.yna.co.kr/view/AKR20230525138300017?input=1195m> 입니다.

연합뉴스

[게시판] 그린벳-경기도수의사회, 반려동물 건강증진 협력 | 연합뉴스

▲ 반려동물 헬스케어 기업 그린벳은 경기도수의사회와 반려동물 건강 검진 활성화를 위한 업무협약을 체결했다고 25일 밝혔다. 양측은 협약에 따라 ...

5월 25일 (406kB) ▾



반려동물 관련기사인 건강검진 활성화 업무협약 체결
뉴스기사가 뜸

최종결과

챗봇형태로 구현

뉴스봇 ✓

태원 3

+ 책갈피 추가

오늘 >



B I \mathcal{E}

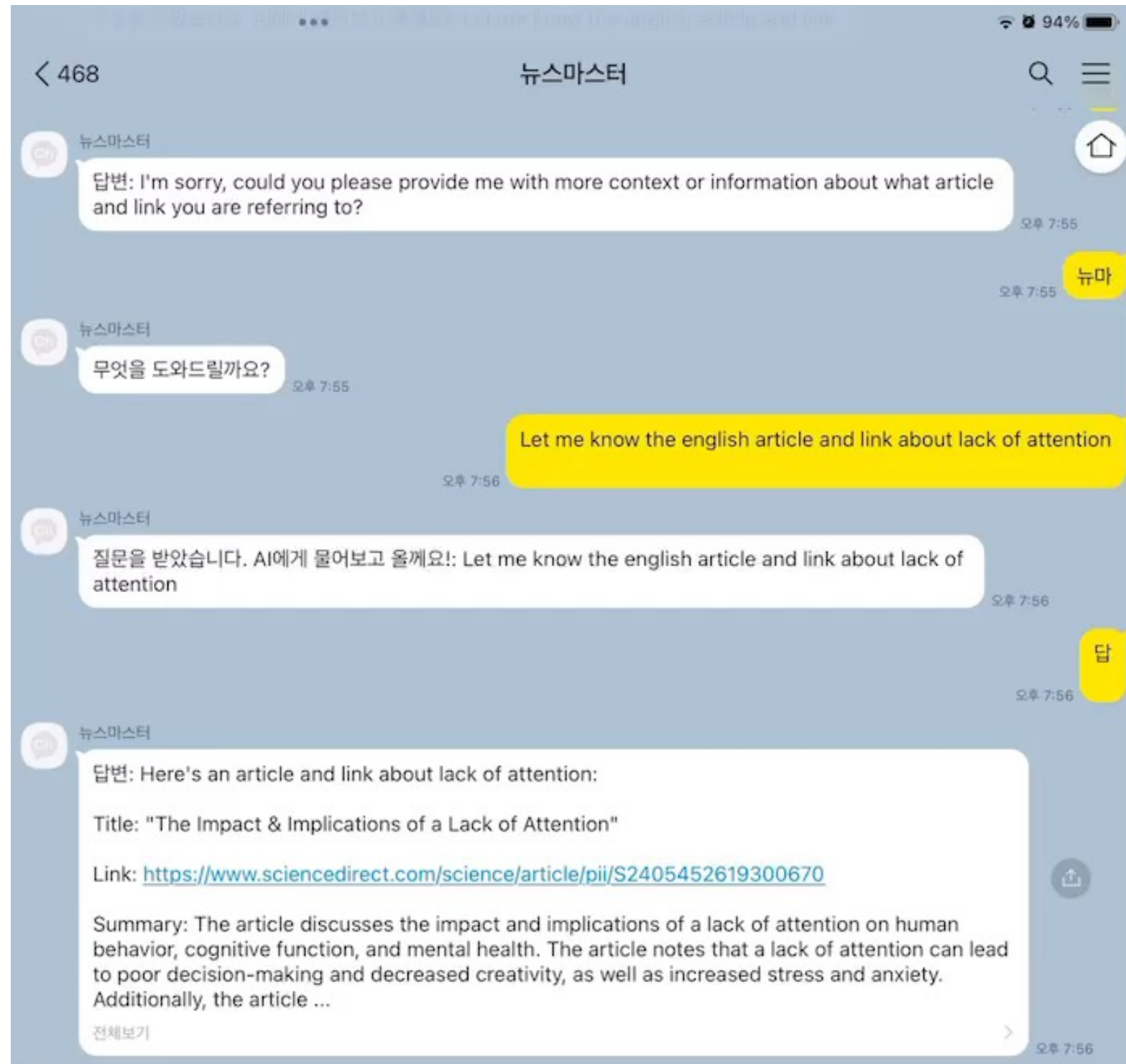
#뉴스봇에 메시지 보내기

+ Aa ☺ @ 📺 🎤 📝

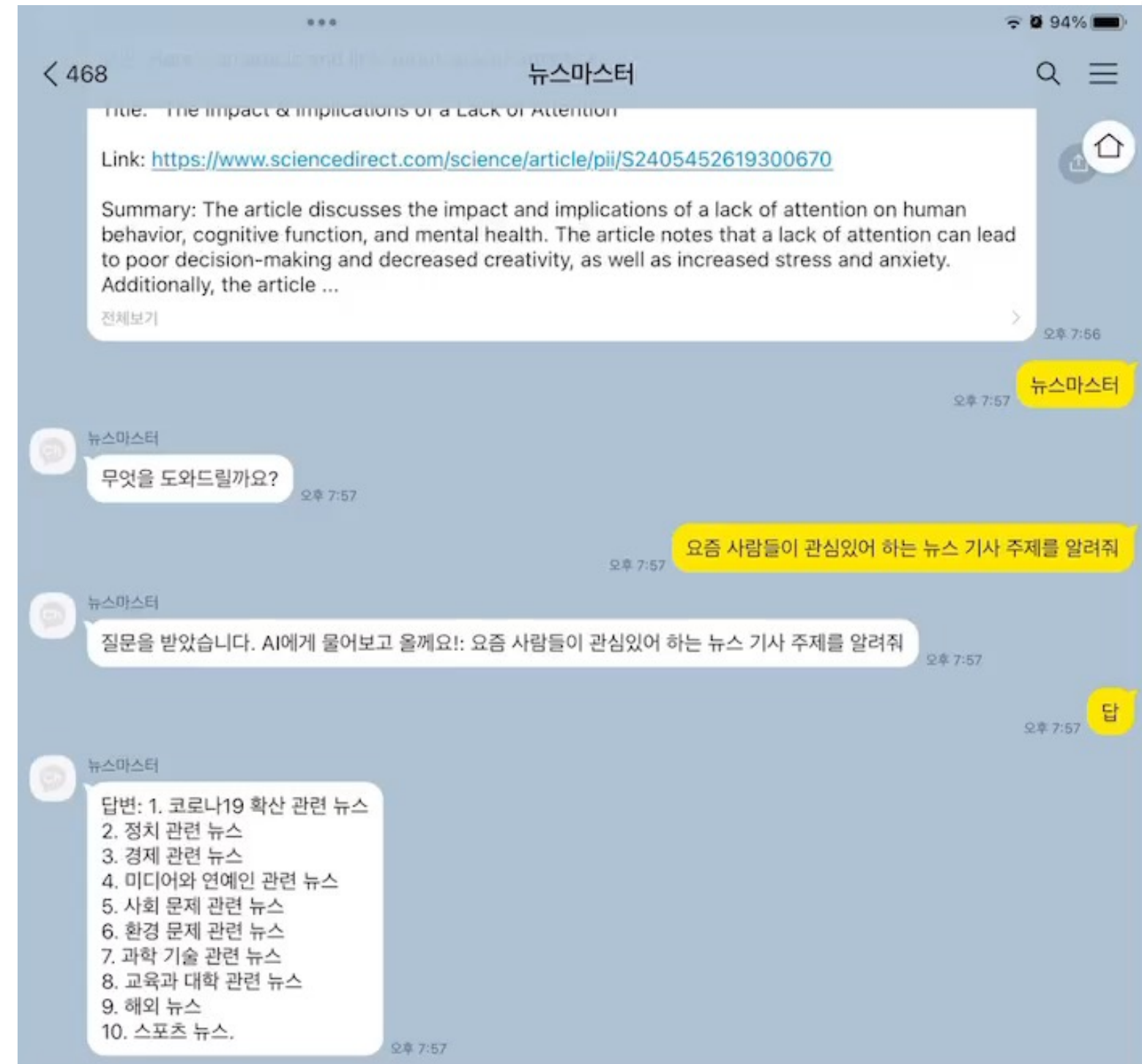


최종결과

카카오 ai 챗봇



+ 챗봇에게 메시지 보내기



+ 챗봇에게 메시지 보내기

최종결과

카카오 ai 챗봇



프로젝트 기대효과

우리는 어떤 프로그램을 만들 것인가



사용자의 질문에 대한 즉각적인 답변을 제공할 수 있음. 이를 통해 사용자는 원하는 정보를 쉽게 검색하고 얻을 수 있음

사용자들은 챗봇을 통해 신속하고 정확한 뉴스 요약, 키워드 추출, 트렌드 분석 등의 정보를 얻을 수 있음

뉴스를 쉽게 제공해 사람들의 사회 인지도가 향상된다

효율적인
뉴스 소비

정보 검색
용이성

인지도
향상



감사합니다

질문이 있다면 말씀해주세요.