



줄여서 넘겨준다, 귀로!

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트

목차

A table of contents

1

팀 구성 및 역할

2

프로젝트 배경

3

수행 절차 및 방법

4

수행 결과 및 서비스 시연

5

개발후기 및 느낀점





1

팀 구성 및 역할

1 팀 구성 및 역할

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

**이지현
(조장)**

데이터 수집(크롤링) 및 전처리, AI 서비스 구현,
HTML 작성, FastAPI 및 ngrok 구축, 발표자료 작성

명진현

데이터 수집(크롤링) 및 전처리, AI 서비스 구현,
HTML 작성, FastAPI 및 ngrok 구축, 발표자료 작성

유승희

데이터 수집(크롤링) 및 전처리, AI 서비스 구현,
HTML 작성, FastAPI 및 ngrok 구축, 발표자료 작성

신보영

데이터 수집(크롤링) 및 전처리, AI 서비스 구현,
HTML 작성, FastAPI 및 ngrok 구축, 발표자료 작성

프로젝트 배경

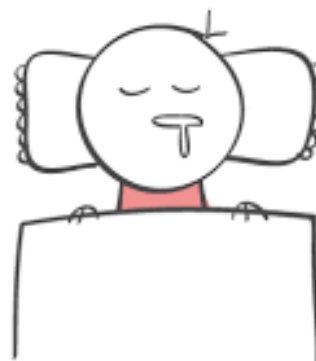
2



신문을
읽을까?

VS.

10분만 더
잘까?



요즘 이슈는 뭘까?
코로나? 지하철 이동권 시위?

#궁금한 소식이 있다면?

옷 입으면서,
밥 먹으면서,

요점만 듣는다!





원하는 언론사
원하는 키워드
단 3줄
읽어주는 서비스



A photograph of a yellow rectangular board leaning against a dark wooden bench. The bench is on a paved surface made of dark grey rectangular tiles. The board is positioned vertically, with its top edge resting on the bench and its bottom edge on the ground. The board has a small, light-colored rectangular label near its bottom left corner. The scene is brightly lit, casting shadows on the ground.

[줄여서 넘겨준다, 귀로!]

2 프로젝트 배경

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

1 > 프로젝트 목표

바쁜 삶을 살고, 효율적인 시간을 보내기 위한 현대인들을 위해 출근 준비나 이동 시 빠르게 뉴스를 접할 수 있도록, **간략히 요약한 뉴스를 AI 보이스를 통해 읽어주는 서비스**를 제공하고자 합니다.

2 > 주요 기능 및 기존 서비스와의 차별점

사용자가 자신이 원하는 **키워드**, **뉴스의 개수**, 뉴스를 검색할 **언론사**를 설정할 수 있습니다. 이를 통해 사용자가 원하는 뉴스를 찾는데 소요되는 **시간과 노력을 최소화** 해줍니다. 또한 요약된 내용을 AI 보이스로 읽어주어 다른 활동 중에도 뉴스의 내용을 습득할 수 있습니다.

3 > 부가 서비스 기능

기사의 요약문을 제공하여 내용을 다시 읽거나 저장할 수 있도록 하였고, **기사의 전문**을 읽을 수 있도록 **원문의 url** 을 제공합니다.



3

수행 절차 및 방법

3 수행 절차

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

구분	기간	활동
사전기획	3/17(목)	- 프로젝트 방향 및 주제 선정
기획안 작성	3/18(금)	- 기획안 작성 주제 및 내용 토의
데이터 크롤링 및 프로그램사용 서비스 구현	3/18(금)~3/30(수)	<div>- 5개의 언론사 네이버 뉴스 크롤링</div> <div>- postman / ngrok / HTML / naver CLOVA</div> <div></div>
포트폴리오 작성	3/29(화)~3/30(수)	-PPT 작성 및 수정 보완
프로젝트 발표	3/31(목)	- A반 2차프로젝트 발표

3 수행 방법

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

STEP 1

언론사, 키워드, 기사 개수 입력 >> 언론사별 기사 크롤링 및 저장

```
1 import os
2 import sys
3 import urllib.request
4
5
6 client_id = "yQk2NXsWHcmq7BptZLf7"
7 client_secret = "vUkizuIOIB"
8 encText = urllib.parse.quote("대통령")
9 url = "https://openapi.naver.com/v1/search/news?query=" + encText # json 결과
10 # url = "https://openapi.naver.com/v1/search/news.xml?query=" + encText # xml
11 request = urllib.request.Request(url)
12 request.add_header("X-Naver-Client-Id", client_id)
13 request.add_header("X-Naver-Client-Secret", client_secret)
14 response = urllib.request.urlopen(request)
15
16 rescode = response.getcode()
17 if(rescode==200):
18     response_body = response.read()
19     new = response_body.decode('utf-8')
20     file = open('news.txt', 'w') #
21     file.write(new) # 파일에 문자열 저장
22     file.close()
23     print(response_body.decode('utf-8'))
24 else:
25     print("Error Code:" + rescode)
```

크롤링을 시작합니다.

```
0%|          | 0/3 [00:00<?. ?it/s]C:\Users\WPublic
WDocuments\WESTsoft\CreatorTemp\ipykernel_20648\1155766905.py:152: DeprecationWarning: find_element_by_* commands are deprecated. Pleas
e use find_element() instead
    table = browser.find_element_by_xpath('//ul[@class="list_news"]')
C:\Users\Wjhjh3\Wanaconda3\lib\site-packages\selenium\webdriver\remote\webelement.py:393: UserWarning: find_element_by_* commands are de
precated. Please use find_element() instead
    warnings.warn("find_element_by_* commands are deprecated. Please use find_element() instead")
100%|██████████| 3/3 [00:03<00:00, 1.06s/it]
```

브라우저를 종료합니다.

데이터프레임 변환

엑셀 저장 완료 | 경로 : C:\Users\Wjhjh3\WPJT_02\네이버뉴스_본문_3개_코로나_2022-03-22_16시19분.xlsx

결과물의 일부

	title	url	text
1	코로나19 신규 확진 35만3,980명..사망 384명	https://imnews.imbc.com/news/2022/society/arti...	사진 제공: 연합뉴스중앙방역대책본부는 오늘 0시 기준 코로나19 신 규 확진자가 35...
2	"코로나19 사태 속 저소득층 물가부담, 고소득 층 1.4배"	https://imnews.imbc.com/news/2022/econo/articl...	사진 제공: 연합뉴스코로나19 사태 속에서 물가 상승에 따른 부담이상 대적으로 저소득...
3	축구대표팀 소집된 백승호, 코로나19 확진..원두 재 대체 발탁	https://imnews.imbc.com/news/2022/sports/artic...	사진 제공: 연합뉴스이란과의 월드컵 최종예선을 앞둔 축구대표팀에 코로나19 비상이 결...

3 수행 방법

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

STEP 2

크롤링 된 기사 원문 내용 >> 3줄 요약 (CLOVA SUMMARY SERVICE)

```
1 import requests
2 import json
3
4 url = "https://naveropenapi.apigw.ntruss.com/text-summary/v1/summarize"
5
6 document = {"title": "'하루 2000억' 판 커지는 간편송금 시장",
7             "content": "간편송금 이용금액이 하루 평균 2000억원을 넘어섰다. 한"}
8 option = {"language": "ko",
9           "model": "news",
10          "tone": 2,
11          "summaryCount": 3}
12 msg = {'document':document, 'option':option}
13 payload = json.dumps(msg)
14
15
16 headers = {
17     'X-NCP-APIGW-API-KEY-ID': 'qj40wl5n0t',
18     'X-NCP-APIGW-API-KEY': '1xzXEYgwlsbv9RQxd1onUSByiyvtBDPFV9NEY1mv',
19     'Content-Type': 'application/json'}
20
21 response = requests.request("POST", url, headers=headers, data=payload)
22
23 print(response.text)
```

{"summary": "간편송금 이용금액이 하루 평균 2000억원을 넘어섰습니다. 한은은 카카오 페이, 토스 등 간편송금 서비스를 제공하는 업체 간 경쟁이 심화되면서 이용규모가 크게 확대됐다고 분석했습니다. 국회 정무위원회 소속 바른미래당 유의동 의원에 따르면 카카오페이, 토스 등 선불전자지급서비스 제공업체는 지난해 마케팅 비용으로 1000억원 이상을 지출했습니다."}




3 수행 방법

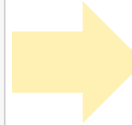
빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

STEP 3

3줄 요약문 >> AI 음성 서비스(CLOVA VOICE)로
요약된 n개의 기사별 오디오 파일 생성 후 최종오디오파일 생성

```
1 import os
2 import sys
3 import urllib.request
4 client_id = "qj40wl5n0t"
5 client_secret = "1xzXEYgwlsbv9RQxd1onUSByiyvtBDPFV9NEY1mv"
6 encText = urllib.parse.quote(response.text)
7 data = "speaker=njihun&volume=0&speed=0&pitch=0&format=mp3&text=" + encText;
8 url = "https://naveropenapi.apigw.ntruss.com/tts-premium/v1/tts"
9 request = urllib.request.Request(url)
10 request.add_header("X-NCP-APIGW-API-KEY-ID", client_id)
11 request.add_header("X-NCP-APIGW-API-KEY", client_secret)
12 response = urllib.request.urlopen(request, data=data.encode('utf-8'))
13 rescode = response.getcode()
14 if(rescode==200):
15     print("TTS mp3 저장")
16     response_body = response.read()
17     with open('1111.mp3', 'wb') as f:
18         f.write(response_body)
19 else:
20     print("Error Code:" + rescode)
```

 voice0
 voice1
 voice2



```
9 @app.post('/topic', response_class=HTMLResponse)
10 async def get_topic(request: Request, press_nm: str = Form(...), keyword:
11     news_df = process_a(press_nm, keyword, ct)
12     news_list = process_b(news_df)
13     news_list = clean_news(news_list)
14     audio1 = open("voice0.mp3", "rb").read()
15     audio2 = open("voice1.mp3", "rb").read()
16     audio3 = open("voice2.mp3", "rb").read()
17     audioJoin = audio1 + audio2 + audio3
18     audioFinal = open("./static/audioFinal.mp3", "wb").write(audioJoin)
19
20     return templates.TemplateResponse("resultnews2.html", {"request": request})
21
```

 audioFinal

3 수행 방법

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

STEP 4

웹서비스를 구축하기 위해 Fast API를 활용

```
1 app=FastAPI()
2 templates = Jinja2Templates(directory="templates")
3 app.mount('/static', StaticFiles(directory='static'), name='static')
4
5 @app.get('/', response_class=HTMLResponse)
6 async def read_topic(request: Request):
7     return templates.TemplateResponse("inputnews2.html", {"request": request})
8
9 @app.post('/topic', response_class=HTMLResponse)
10 async def get_topic(request: Request, press_nm: str = Form(...), keyword: str = Form(...), ct: int = Form(...)):
11     news_df = process_a(press_nm, keyword, ct)
12     news_list = process_b(news_df)
13     news_list = clean_news(news_list)
14     audio1 = open("voice0.mp3", "rb").read()
15     audio2 = open("voice1.mp3", "rb").read()
16     audio3 = open("voice2.mp3", "rb").read()
17     audioJoin = audio1 + audio2 + audio3
18     audioFinal = open("./static/audioFinal.mp3", "wb").write(audioJoin)
19
20     return templates.TemplateResponse("resultnews2.html", {"request": request, 'press_nm': press_nm, 'keyword': keyword, 'ct': ct})
21
```

INPUT HTML

함수 process_a

step1 언론사별 기사 크롤링 및 저장

함수 process_b

step2&3 기사 요약 후 오디오파일 생성

함수 clean_news

요약된 기사 전처리

RESULT HTML

3 수행 방법

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

STEP 5

INPUT HTML & RESULT HTML

```
1 %%writefile templates/inputnews2.html
2 <!DOCTYPE html>
3 <html>
4 <head>
5 <meta charset="UTF-8">
6 <title>뉴스 즐넘귀</title>
7 <style>
8 body {background-color: lightblue;}
9 .white {background-color: white;}
10 </style>
11 </head>
12
13 <body>
14
15 <h1 style="text-align: center;"><strong>👂 즐넘, 귀!</strong></h1>
16 <p style="text-align: center;">요즘 많이 바쁘신가요?</p>
17 <p style="text-align: center;">즐넘귀는 바쁜 삶을 살고 있는 여러분을 위해 이동 시 빠르게 뉴스를 접할 수
18 <p style="text-align: center;">간략히 요약한 뉴스를 보이스를 통해 읽어주는 서비스를 제공합니다. 🗣️</p>
19 <p>&nbsp;</p>
20
21 <form action="/topic" method = 'post'>
22 <h3><strong><label for="press_nm">검색할 언론사:</label><br></h3>
23 <p>현재 연합뉴스, MBC, 매일경제, SBS, KBS, JTBC 검색 기능을 지원합니다.</p>
24 <input type="text" id="press_nm" name="press_nm" value="KBS"><br>
25 <h3><strong><label for="keyword">검색할 키워드:</label><br></h3>
26 <input type="text" id="keyword" name="keyword" value="코로나"><br>
27 <h3><strong><label for="ct">수집 뉴스의 수(숫자만 입력) :</label><br></h3>
28 <input type="int" id="ct" name="ct" value="3"><br><br>
29 <input type="submit" value="줄여서 넘겨줘, 귀로!">
30 </form>
31
32 <p>&nbsp;</p>
33 <p><strong>주요 서비스<strong>👂 즐넘, 귀!</strong></h1>
12 <h3 style="text-align: center;">뉴스를 읽어드립니다 🗣️</h3>
13 <p>&nbsp;</p>
14 <h3><strong>검색할 언론사 : {{ press_nm }}</strong></h3>
15 <p>&nbsp;</p>
16 <h3><strong>검색할 키워드 : {{ keyword }}</strong></h3>
17 <p>&nbsp;</p>
18 <h3><strong>수집할 뉴스 개수 : {{ ct }}</strong></h3>
19 <p>&nbsp;</p>
20 <h3>뉴스</h3>
21 <audio style="display: block; margin-left: auto; margin-right: auto;" controls autoplay>
22 <source src= {{ fn }} type="audio/mpeg">
23 Your browser does not support the audio element.
24 </audio>
25 <h3><strong>요약 내용 : </strong></h3>
26 <p>상세한 기사 내용은 하단 링크를 참조하세요.</p>
27 <p>&nbsp;</p>
28 <p>{{ news_list[0].replace('Wn', '') }}</p>
29 <p>{{ news_list[1].replace('Wn', '') }}</p>
30 <p>{{ news_list[2].replace('Wn', '') }}</p>
31 <p>{{ news_list[3].replace('Wn', '') }}</p>
32 <p>{{ news_list[4].replace('Wn', '') }}</p>
33 <p>{{ news_list[5].replace('Wn', '') }}</p>
34
35 </body>
36 </html>
37
```

3 수행 방법

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

STEP 6

위 전과정 함수화 및 결과물(오디오파일) >> FastAPI와 ngrok를 활용하여 서비스 구현

```
1 ngrok_tunnel = ngrok.connect(8000)
2 print('Public URL:', ngrok_tunnel.public_url)
3 nest_asyncio.apply()
4 uvicorn.run(app, host='0.0.0.0', port=8000)
```

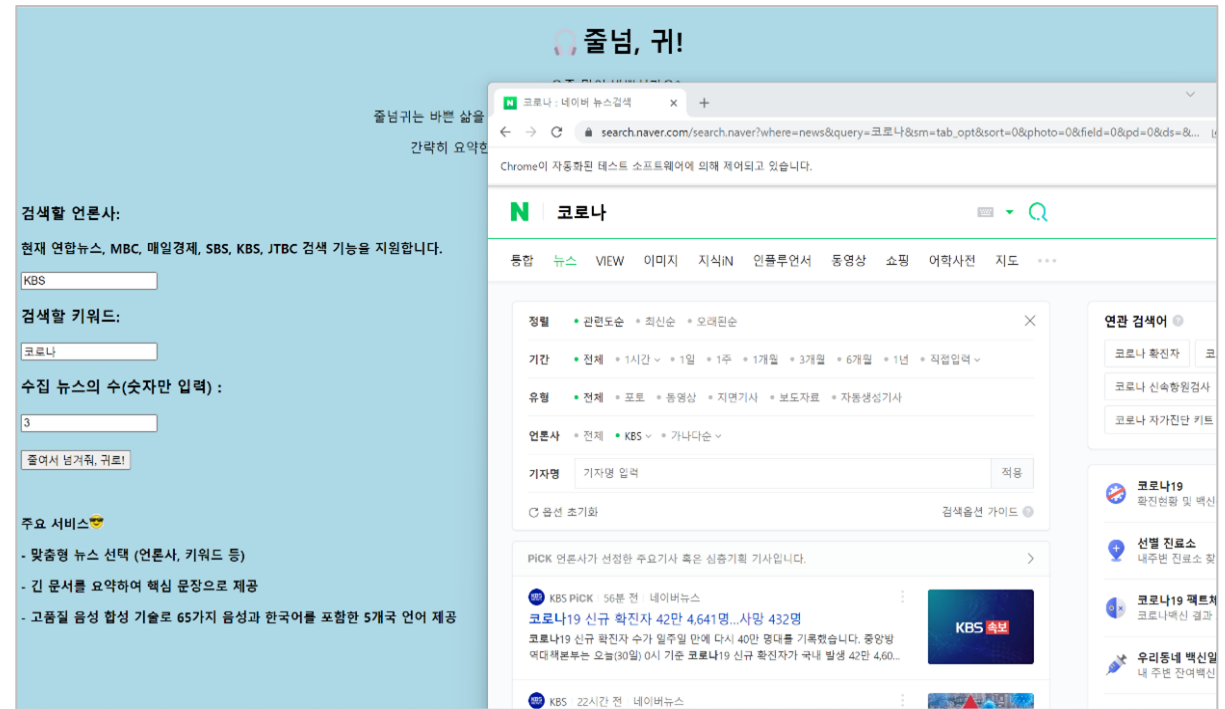
INFO: Started server process [12868]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on <http://0.0.0.0:8000> (Press CTRL+C to quit)

Public URL: <http://e336-211-44-6-54.ngrok.io>

INFO: 211.44.6.54:0 - "GET / HTTP/1.1" 200 OK
INFO: 1.254.179.100:0 - "GET / HTTP/1.1" 200 OK
INFO: 211.44.6.54:0 - "GET / HTTP/1.1" 200 OK
INFO: 211.44.6.54:0 - "GET /favicon.ico HTTP/1.1" 404 Not Found
INFO: 211.44.6.54:0 - "GET / HTTP/1.1" 200 OK
본문 크롤링에 필요한 함수를 로딩하고 있습니다...

검색할 언론사 : KBS

ngrok를 사용하여 네트워크 터널(Public URL) 생성



터널링 프로그램(Public URL 실행)으로 서비스가 작동하는 것을 확인


수행 결과 및 서비스 시연

4

4 수행 결과 및 서비스 구현

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

INPUT.HTML

 줄넘, 귀!

요즘 많이 바쁘신가요?

줄넘귀는 바쁜 삶을 살고 있는 여러분을 위해 이동 시 빠르게 뉴스를 접할 수 있도록,
간략히 요약한 뉴스를 보이스를 통해 읽어주는 서비스를 제공합니다.😊

검색 언론:

현재 연합뉴스, MBC, 매일경제, SBS, KBS, JTBC 검색 기능을 지원합니다.

KBS

검색 키워드:

코로나

검색할 뉴스 개수(숫자만 입력) :


3

줄여서 남겨줘, 귀로!

주요 서비스👉

- 맞춤형 뉴스 선택 (언론사, 키워드 등)
- 긴 문서를 요약하여 핵심 문장으로 제공
- 고품질 음성 합성 기술로 65가지 음성과 한국어를 포함한 5개국 언어 제공

RESULT.HTML

 줄넘, 귀!

뉴스를 읽어드립니다 😊

검색할 언론사 : KBS

검색할 키워드 : 코로나

수집할 뉴스 개수 : 3

뉴스

0:32 / 2:04

요약 내용 :

상세한 기사 내용은 하단 링크를 참조하세요.

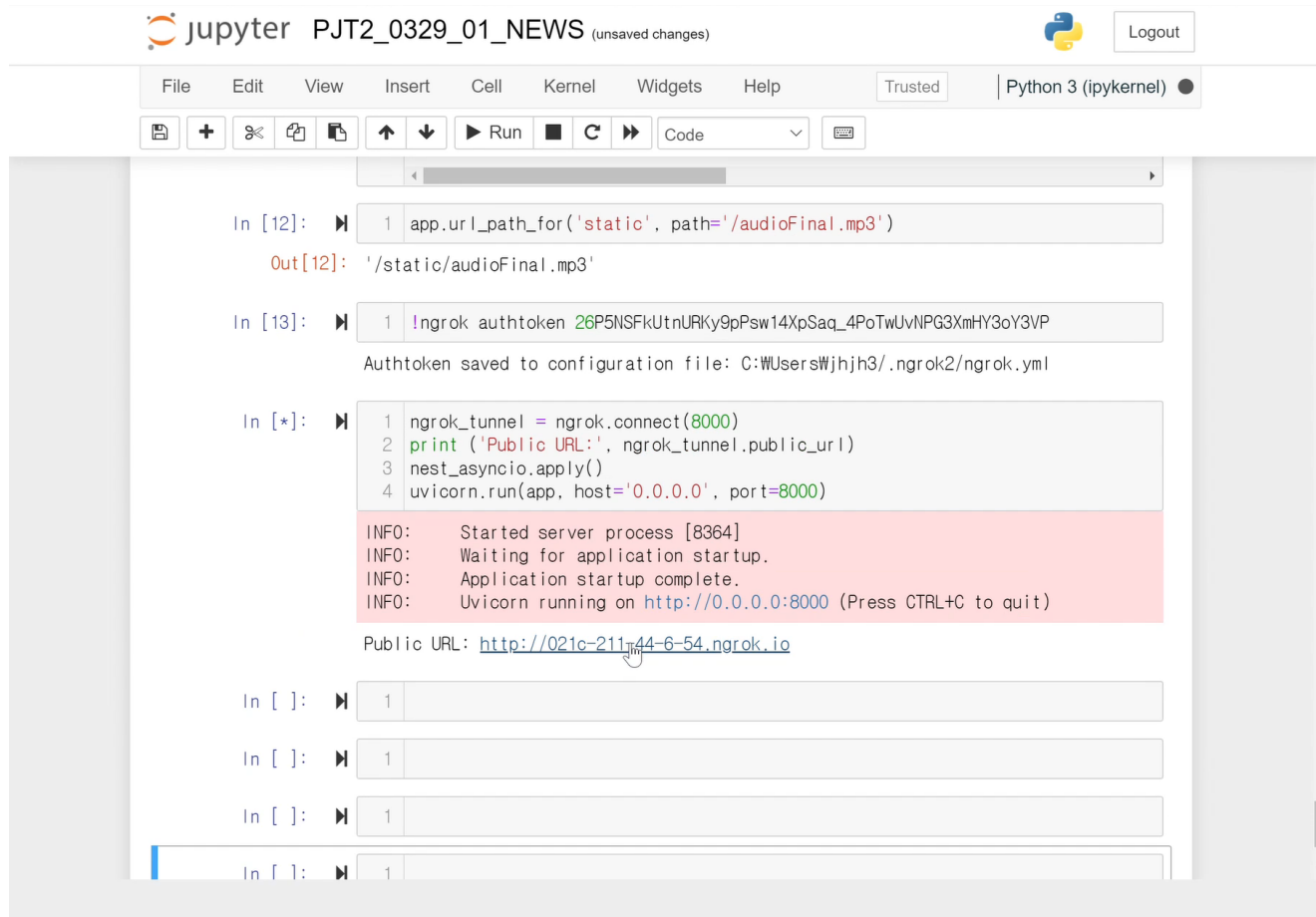
1번째 뉴스입니다. 모레(31일)부터 5~11세 소아 대상 코로나19 백신 접종이 시작되는 가운데, 식품의약품안전처가 화이자사의 소아용 백신을 국가출하승인했습니다. 국가출하승인은 백신이 시중에 유통되기 전에 제조단위(로트)별로 국가가 검정시험 결과와 제조사의 제조·시험 결과를 종합적으로 평가해 품질을 한 번 더 확인하는 제도입니다. 식약처는 5-11세용 코미나티주에 대한 국가 검정 시험과 제조사의 제조·시험 자료를 검토한 결과, 효과성과 안전성, 품질이 확인돼 국가출하승인을 결정했다고 설명했습니다.

<https://news.kbs.co.kr/news/view.do?ncd=5426719&ref=A>

2번째 뉴스입니다. 코로나19 국내 발생 첫 해인 2020년 직장인 1인당 평균 대출액이 전년보다 10% 넘게 증가해 역대 가장 큰 폭으로 늘었습니다. 기업 규모별로 보면 대기업 임금근로자의 평균대출이 7,624만 원으로 가장 높았고, 중소기업 임금근로자의 평균대출액인 3,941만 원보다 1.9배인 것으로 나타났습니다. 종사자 규모별로 보면 300인 이상 기업 근로자의 평균대출은 6,123만 원으로 가장 높았

4 수행 결과 및 서비스 구현

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조



```
jupyter PJT2_0329_01_NEWS (unsaved changes) Python 3 (ipykernel) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted | Python 3 (ipykernel)

In [12]: 1 app.url_path_for('static', path='/audioFinal.mp3')
Out[12]: '/static/audioFinal.mp3'

In [13]: 1 !ngrok authtoken 26P5NSfkUtnURKy9pPsw14XpSaq_4PoTWUvNPG3XmHY3oY3VP
Auth token saved to configuration file: C:\Users\Wjhjh3\.ngrok2\ngrok.yml

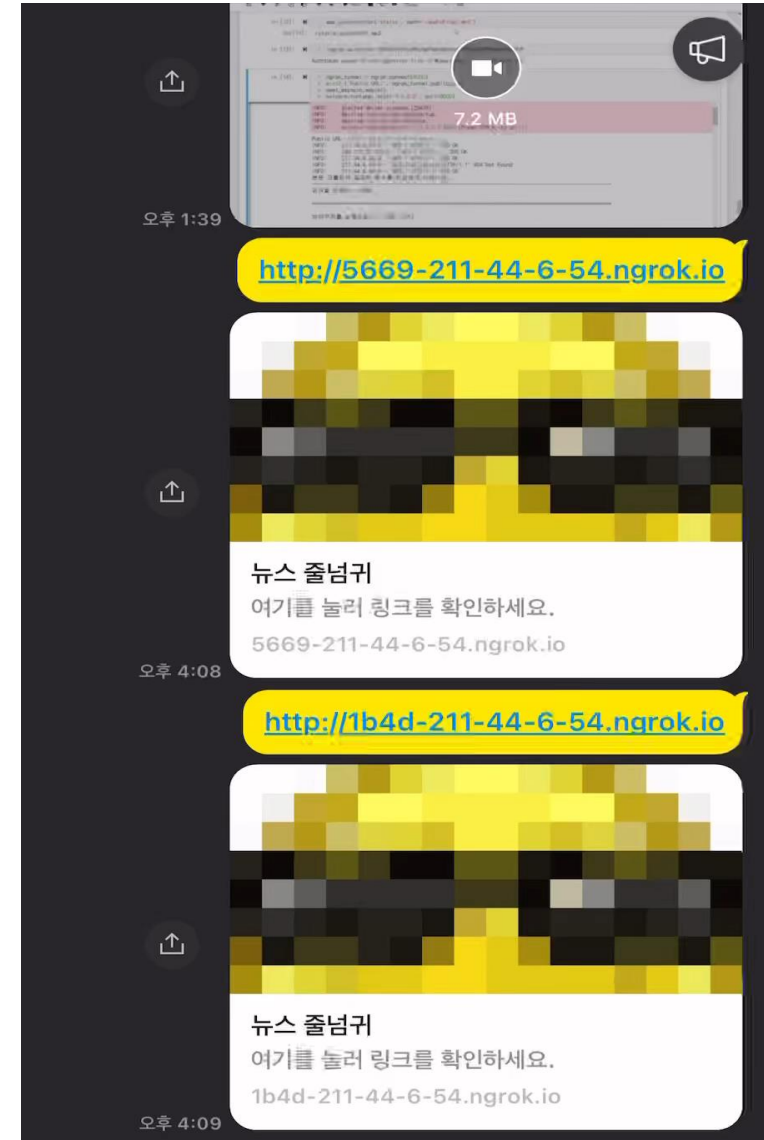
In [*]: 1 ngrok_tunnel = ngrok.connect(8000)
2 print('Public URL:', ngrok_tunnel.public_url)
3 nest_asyncio.apply()
4 uvicorn.run(app, host='0.0.0.0', port=8000)

INFO: Started server process [8364]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)

Public URL: http://021c-211-44-6-54.ngrok.io

In [ ]: 1
In [ ]: 1
In [ ]: 1
In [ ]: 1
```

PC.ver



MOBILE.ver

4 수행 결과 및 서비스 구현

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

1 > 장점 및 기대 효과

>

- 불특정 다수를 대상으로 하는 기존 매체와 달리 사용자가 원하는 키워드의 뉴스에 빠르게 접근 가능.
- 요약 서비스를 통해 신속하게 뉴스의 개요를 파악할 수 있음.
- 동시에 해당 뉴스의 URL을 제공해 상세한 뉴스를 읽는 것을 선택할 수 있음.
- 시각 정보를 청각 정보로 변환해 제공함으로써 운전 등 시각 의존도가 큰 다른 행동과 병행할 수 있도록 함.
- 직장인 등 정보에 민감하면서도 시간을 효율적으로 사용해야 하는 사람들이 활용하기 좋을 것으로 기대됨.

2 > 향후 발전/개선 방안

>

- 크롤링하는 언론사의 다양화 및 해외 뉴스 제공
- 모르는 시사 용어 검색 챗봇 구현
- 사용자 음성을 인식해 키워드를 검색하는 기능 추가
- 오늘의 날씨 등 간략한 생활 정보 병행 제공



5

개발 후기 및 느낀점

5 개발 후기 및 느낀점

빅데이터 기반 지능형 서비스 개발 A반 2차 프로젝트 2조

이지현

이번 프로젝트를 진행하면서 정말 많은 것을 배웠다. 초반에는 생각만큼 구현되지 않는 부분도 있었고, 수업에서 배웠던 내용이라고 해도 확실하게 이해되지 않았던 부분도 있어서 어려움이 많았다. 프로젝트 결과물을 만들어 내기 위해 선생님께도 다시 한번 여쭙보고 조원들의 도움을 받아 부족함 점을 채워 나아갔다. 이번 결과물을 통해 Fast API, ngrok, html 등 할 수 있는 부분, 웹서비스, AI 서비스 구현 등 만들어 낼 수 있는 부분이 늘어났다. 이 부분만큼은 어디서든 할 수 있다는 자심감이 생겼다는 것이 가장 만족스럽고 뿌듯하다.

유승희

이번 프로젝트는 직접 서비스를 구현해본다는 점에서 기대도 크고 걱정도 됐다. 초반에는 수업 이후로 처음 사용해보는 Fast API와 클로바 API가 어렵고 오류도 많아 힘들었지만 팀원들과 오류를 해결해 나가고 보완해 나가는 과정에서 배우는 점이 많았다. 우리 조는 오류나 문제를 함께 해결하려는 팀워크가 멋지고 좋았다. 팀 프로젝트를 통해서 많은 것을 배울 수 있어 뜻깊은 시간이었다. 실제 눈으로 보이는 결과물이 있어 정말 뿌듯하다. 스스로 부족한 부분에 대해서도 알게 되어 향후 공부에 보완점으로 삼아야겠다.

신보영

기획단계에서 생각했던 서비스를 전부 다 구현하는 것이 마음먹은 대로 쉽게 되지는 않았던 부분이 있었다. 프로젝트 과정에서 사용한 여러 도구들에 더 숙달되어 있었다면 시간이나 서비스의 다양함 등에서 더 뛰어난 결과물을 만들 수 있었을 것 같아 아쉬움이 남는다. 그러나 한편으로는 큰 틀에서 개발하고자 했던 서비스가 만들어져서 실제로 구동되는 것을 보니 신기하고 재미있어서 좋은 경험이었다. 이번 경험을 발판삼아 앞으로 더 좋은 서비스를 개발할 수 있도록 노력해야겠다. 또 프로젝트를 성실히 이끌어준 팀장과 팀원들에게 진심으로 감사드린다.

명진현

지금까지 공부했던 내용을 통해 첫 서비스 구현을 해보았다. 익숙치 않은 다양한 프로그램을 사용하여 서비스 구현에 완주한 것에 감사하다. 조금 더 여유가 있었다면 더 나은 서비스를 사람들에게 제공할 수 있겠다는 아쉬움과 자신감이 생겼다. 부족한점은 앞으로 더 보완하여 채우고, 더 나은 서비스를 만들고 싶다는 생각도 들었다. 팀원 다같이 고민하고 인터넷에 검색해 서로 도움을 주고받은 팀원들에게 감사하다.



줄여서 넘겨준다, 귀로!



Thanks!