

Final Project Presentation

VIEW:ROJECT

TEAM: 비수 오현지 김수빈 유승연 최태영

브레인 스토밍 / 프로젝트 개요

시장성장가능성/외부 환경의 변화

활용 프로그램 / 분석모델 / VIEW:PROJECT소개

웹UI / VIEW:PROJECT기능 구현

기대효과 및 활용방안 / 향후 발전방향

역할/소감



#01

프로젝트 기획

브레인 스토밍

“ 사람이 굳이 하지 않아도 되는 일을 자동화 시킴으로써
사람은 좀 더 의미 있는 일을 할 수 있지 않을까? ”

자동 회의록 작성 프로그램

입시

운세

여행

쇼핑몰

웹툰

스포츠

로또

반려동물

옷(패션)

숙박

유튜브

인테리어

주식

카페 입지 추천

음식

레시피

게임

헬스

자동차

취업

자동차

자격증

음악

컴퓨터

교통

연애

식물





VIEW:ROJECT

: 프로젝트를 보여주다

- 1 음성 인식을 기반으로 회의록 자동 작성
- 2 회의 내용을 한 눈에 파악할 수 있는 워드 클라우드 제공
- 3 안건, 내용, 결론 형식의 회의 내용 세부적으로 요약
- 4 회의록 간편 저장 및 관리





회의록 자동 작성 기능에 대한 수요 증가

‘연중 회의만 8,000회’ 서울시, 인공지능으로
회의록 받아 적는다

(출처: Decenter, 2019)

신입사원이 해야 하는 귀찮은 그 일,
AI가 대신 해준다?

(출처: 네이버 포스트, 2019)

지자체, 올해 AI 예산 늘리고 산업 육성 본격화

(출처: 메트로서울, 2020)

행정안전부, 4차 산업혁명 대비 ‘음성인식 다자간
회의록 시스템 사업’ 추진

(출처: 국민일보, 2018)

학폭처리 인력부족 ‘발등에 불’ ...
‘AI 속기시스템’ 꺼낸 인천교육청

(출처: 경인일보, 2020)



Daglo



‘한글’ 분석의 어려움으로 성공적인 서비스 사례X

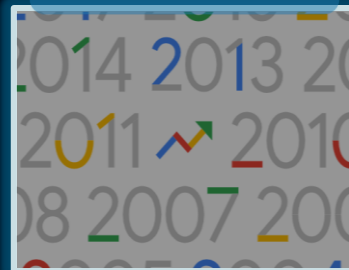


사회적 배경



- ✓ AI 관련 산업 육성 본격화
- ✓ AI 음성인식 관련 벤처기업과 협약
- ✓ AI 음성인식 기술을 접목한 아이디어에 대한 관심 증가

시대적 배경



- ✓ 회의록 자동 작성의 필요성 인식
- ✓ 서류 작성과 편집에 드는 인력과 시간에 대한 회의적 견해 증가
- ✓ 서류 작성 시간을 줄이고 이를 다른 업무에 투자

기술적 배경



- ✓ 화자 분리, 화자 구분 등 음성인식 관련 기술에 관한 연구의 성과
- ✓ Konlpy, Texttrank 등 텍스트 데이터 분석 모델 개발



#03

프로젝트 개발과정

26

FEBRUARY
FRIDAY



View:roject



#03

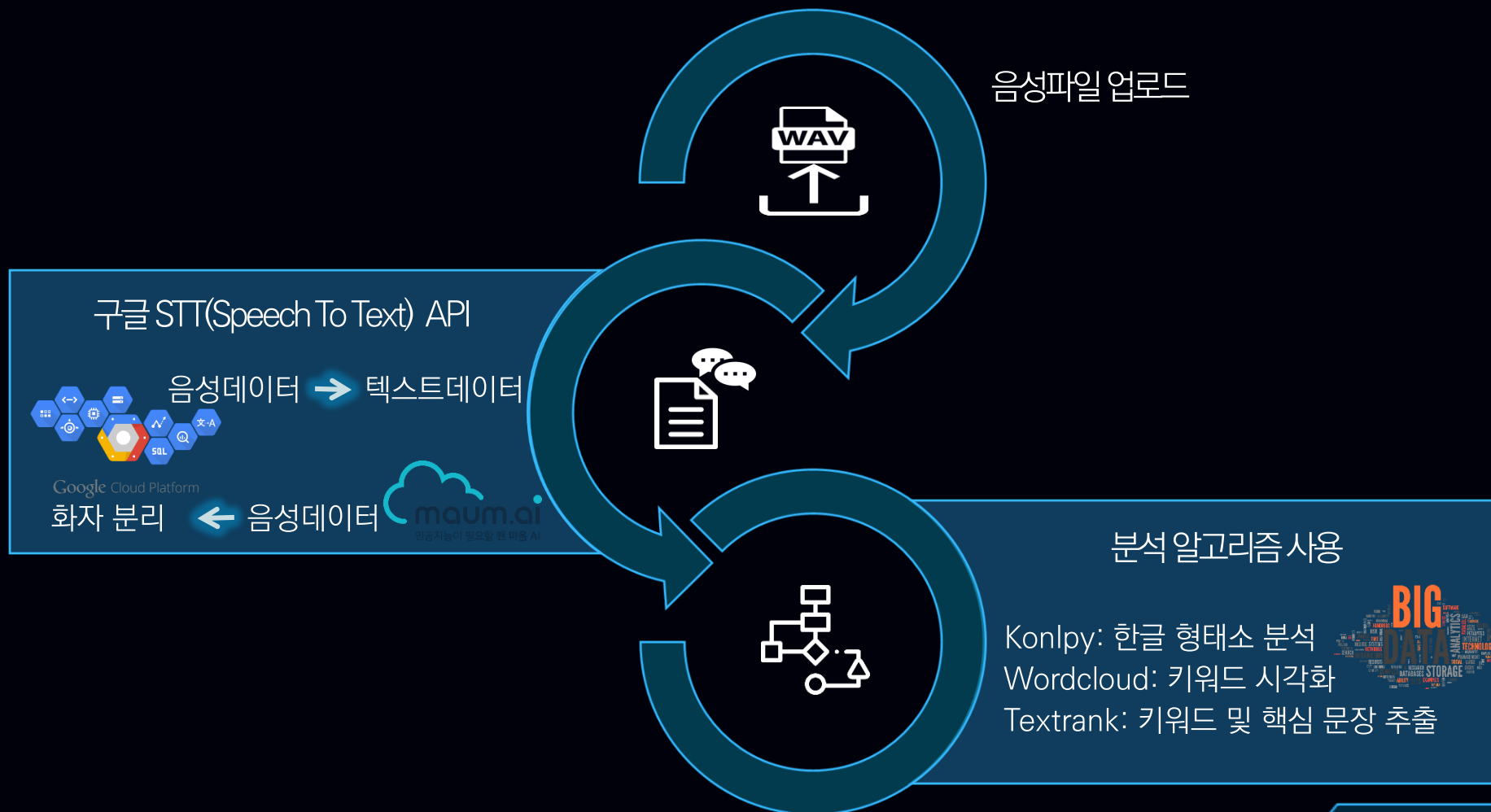
프로젝트 개발과정

26

FEBRUARY
FRIDAY

VIEW:ROJECT 실행 과정





```
#화자 분리
for i in range(len(cut2_df)):
    cut2_audio = cut2_df['cut2_audio'][i]

    # API 요청
    response = transcribe_audio(cut2_audio)

    #fname4 = cut2_df['cut2_audio'][i].split('.')[0].split('/')[-1]

    for result in response.results:
        alternative = result.alternatives[0]

    if i%2 == 0:
        text = 'A : '+ alternative.transcript + '&#10;&#10;'
        print('A : '+ alternative.transcript)
        fileName = 'result'
        f = open('textResult/result.txt', 'a')
        f.write(text)
        f.close()
    elif i%2 == 1:
        text = 'B : '+alternative.transcript + '&#10;&#10;'
        print('B : '+alternative.transcript)
        fileName = 'result'
        f = open('textResult/result.txt', 'a')
        f.write(text)
        f.close()
```

마음(maum) API 사용 : 화자 분리

A : 지금부터 사람의 불시착 국가 코로나 관련 대책 회의를 시작하도록 하겠습니다 오늘의 안건은 보건용마스크 및 손소독제 긴급 공급 조정조치 추진 계획입니다 손에진 대통령님 원 폐렴으로 불리는 신종 코로나 바이러스 감염증 이빠지는 속도만큼이나 마스크 가격도 오르고 있습니다

B : 맞습니다 이에 따라 최근 보건용 마스크 수급에 대한 국민의 불안과 우려가 커지고 있습니다 홈페이지와 공식 블로그 등에 개설한 보건용마스크 손 소독제 매점매석 행위 등 신고 센터에는 소비자들의 신고가 쏟아지고 있어요

A : 그러니까요 특히 온라인 마켓 등에서 마스크를 주문했는데 일방적으로 취소 당했다고 불만을 토로하는 내용이 다수입니다 일단은 식약처 관계자들이 신고내용을 검토해 사실 여부를 확인하고 사실로 확인되면 조사에서 적절한 조치를 할 수 있도록 지시에 두었습니다

B : 그렇군요 저 또한 마스크 유통시장을 원활하게 하기 위해 먼저 매점매석 군대를 빨리 조치해야 한다고 생각해요 현빈 씨는 이번 사건의 말으면 서 어떤 부분이 가장 문제라는 생각이 들었나요

A : 일단 마스크가 국내 부족한 가장 큰 이유 중 하나가 밀수출이라고 생각했습니다 따라서 마스크와 손소독제의 대량 방출 차단을 위해 수출통관 관리도 대폭 강화해 한다고 생각합니다

B : 그렇군요 이제는 마스크까지 밀수출 이란이 마스크 미스트레 대해 고민할 줄이야 정말 상상도 못 했어요 요즘 시국이 정말 좋지 않네요

A : 알겠습니다 운영 중인 합동 단속반에 경찰 관세청 인원의 추가 투입해 단속 인원을 150명으로 드리겠습니다

B : 맞습니다 유통 문제도 정말 심각해요 수요증가에 따라 가격이 올라가는 것은 당연한 일이지만 문제는 공급 맞아 왜곡시키는 자들이 있다는 점이 예요 만났을 가리지 않고 수천만원 많게는 수십억원 현금을 바이어들이 마스크 생산 공장 앞으로 모여들고 있다고 합니다

```
In [36]: # speech-to-text
def transcribe_gcs(gcs_uri):

    client = speech.SpeechClient()
    audio = types.RecognitionAudio(uri=gcs_uri)
    config = types.RecognitionConfig(
        encoding=enums.RecognitionConfig.AudioEncoding.LINEAR16,
        sample_rate_hertz=16000, audio_channel_count=2,
        language_code='ko-KR',
        enable_automatic_punctuation=True)

    operation = client.long_running_recognize(config, audio)
    response = operation.result()

    return response
# speech-to-text 함수 실행
response = transcribe_gcs("gs://test_p1/"+file_id)

text2 = ''
for result in response.results:
    text2 += result.alternatives[0].transcript

text2 : '지금부터 사람의 불시착. 국가 코로나 관련 대책 회의를 시작하도록 하겠습니다. 오늘의 안건은 보건용마스크 및 손소독제 긴급 수급 조정조치 추
```

진 계획입니다. 손예진 대통령님 원 폐렴으로 불리는 신종 코로나 바이러스 감염증 이빠지는 속도만큼이나 마스크가격 떠오르고 있습니다. 있습니다. 이에 따라 최근 보건용마스크 수급에 대한 국민의 불안과 우려가 커지고 있습니다. 홈페이지와 공식 블로그 등에 개설한 보건용마스크 손소독제 매점매석 행위 등 친구 센터에는 소비자들의 신고가 쏟아지고 있어요. 그러니까요. 특히 온라인 마켓 등에서 마스크를 주문했는데 일방적으로 취소 당했다고 불만을 토로하는 내용이 다수입니다. 일단은 식약처 관계자들이 신고내용을 검토에 사실 여부를 확인하고 사실로 확인되면 조사에서 적절한 조치를 할 수 있도록 지시에 두었습니다. 잘했습니다. 아주 빠른 대응방안이 없다고 생각해요. 감사합니다. 마스크 매점매석 관련하여 문제를 해결하다 보니 이를 막을 수 있는 강력한 법적 조치가 필요하다는 생각이 들었습니다. 그렇군요. 저 또한 마스크 유통 시장을 원활하게 하기 위해 먼저 매점매석 문제를 빨리 조치 해야 한다고 생각해요. 현빈 씨는 이번 사건을 맡으면서 어떤 부분이 가장 문제라는 생각이 들었나요? 일단 마스크가 국내 부족한 가장 큰 이유 중 하나가 수출이라고 생각했습니다. 따라서 마스크와 손소독제의 대량 방출 차단을 위해 수출통관 관리도 대폭 강화해 한

구글 STT(Speech To Text) API 사용

#03

프로젝트 개발과정

```
# komoran_tokenizer : 형태소분석기 선언, 정의
komoran = Komoran()
def komoran_tokenizer(sent):
    words = komoran.pos(sent, join=True)
    words = [w for w in words if ('/NN' in w or '/XR' in w or '/VA' in w or '/VV' in w)]
    return words

# komoran_tokenizer : 형태소분석기 사용
summarizer = KeysentenceSummarizer(
    tokenize = komoran_tokenizer,
    min_sim = 0.3,
    verbose = False
)
```

```
# 워드클라우드에 불필요한 단어 없애기
# 형태소 분리
okt = Okt()

# 불용어 처리
stopwords = pd.read_csv("./stopwords/stopwords.txt")
stopwords = stopwords['불용어'].tolist()

okt_morphs = okt.pos(text2)
Noun_words = okt.nouns(text2)

for word, pos in okt_morphs:
    if pos == 'Noun':
        Noun_words.append(word)

unique_Noun_words = set(Noun_words)
for word in unique_Noun_words:
    if word in stopwords:
        while word in Noun_words: Noun_words.remove(word)

# 워드 클라우드로 표현
FONT_PATH = 'c:/Windows/Fonts/nanum/NanumBarunGothic.ttf'
noun_text = ''
for word in Noun_words:
    noun_text = noun_text + ' '*word

fig = plt.figure()
wordcloud = WordCloud(font_path=FONT_PATH, background_color = 'white', max_words = 70, width = 480, height = 280).generate(noun_text)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
```

Konlpy 모델 사용

Komoran: 한글 형태소 분석

Okt: 형태소 분리

Wordcloud 사용

: 불용어 처리 후
Keyword 추출

#03

프로젝트 개발과정

```
app = Flask(__name__)  
  
|  
@app.route('/fileupload', methods = ['GET', 'POST'])  
def upload_file():  
    if request.method == 'POST':
```

Flask 사용

: JSP와 연동하여 분석 값 전송

```
# fileName titleNm finalNm summaNm  
return redirect('http://localhost:8082/Meeting/showMinute?path='+fileName + '&path2='+titleNm+ '&path3='+finalNm+ '&path4='+summaNm  
  
if __name__ == "__main__":  
    from werkzeug.serving import run_simple  
    run_simple("localhost", 9000, app)
```

* Running on <http://localhost:9000/> (Press CTRL+C to quit)

빠른 진행을 위해 4번째 안건으로 이동하겠습니다
지금 단과대별로 IT 준비하고 계실텐데 성경 대화 그 중대에서 기숙사 입주 날짜가 매매하다고 하셨어요
있기 때문에 확정이 되는 대로 여러분이 신입생 분들께 바로 공지할 수 있도록 최대한 빠르게 공지하도록 하겠습니다.





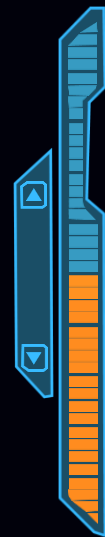
#04



시연 영상

26

FEBRUARY
FRIDAY



#05

프로젝트 방향

26

FEBRUARY
FRIDAY



다자간 화자 구분



음성인식 정확도



안전 구분



사업분야의 확장



#05

프로젝트 방향

26

FEBRUARY
FRIDAY

기대 효과



회의 내용을 한눈에 확인

문서작성 및 편집시간 단축

회의 자체에 집중 → 생산성 향상

남은 시간을 다른 업무에 효율적으로 투자

활용 방안



편리하고 효율적인
업무 인수인계 가능



다양한 분야의 기업에서 활용 가능



텍스트 분석

텍스트 마이닝에서 근접성 데이터의 군집화 기법 연구 = A Study on Clustering Methods for Proximity Data in Text Mining 정민지 발행사항 부산 : 부산대학교 대학원, 2017

음성 분석

Google STT(speech-to-text) API
Mindslab 화자 분리 API

웹 디자인

HTML5 UP! Responsive HTML5 and CSS3 Site Templates html5up.net

<https://m.blog.naver.com/PostView.nhn?blogId=rebehayan&logNo=90063635604&proxyReferer=https%3A%2F%2Fwww.google.com%2F>

#06

팀원 소개

26

FEBRUARY
FRIDAY

최태영

경영학부
음성 분석(텍스트 마이닝)
발표/PPT/문서작성

김수빈

자율전공학부(4년)
음성 분석(텍스트 마이닝)
서버 기능 구현

유승연

불어불문학과
웹 디자인
HTML/JSP

오현지(팀장)

경영학부

음성 분석(STT)
서버 기능 구현
DB설계

비전공자 수준



