

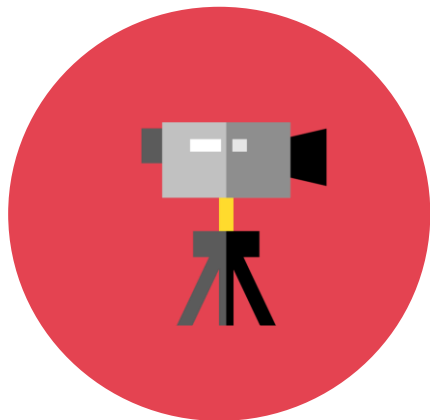
1인 방송 하이라이트 편집

Assistance Tool

1조 안진마

신상훈 김연수 송성재
박형준 허진선 윤정연





Full Video

(1시간 ~ 10+α 시간)



플랫폼 내 Chat

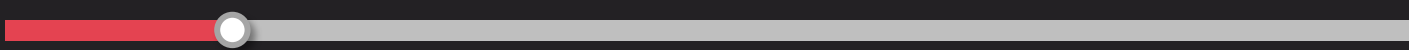


영상 내 소리

분석



하이라이트 추천 지점등 지
표화된 자료 제공

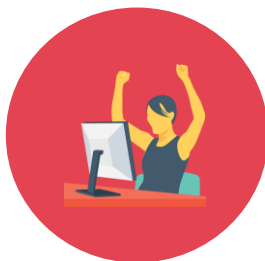


2 진행 상황 - 데이터 구축

URL



채팅로그



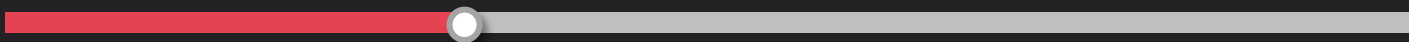
활동량



영상 내 소리



인공지능

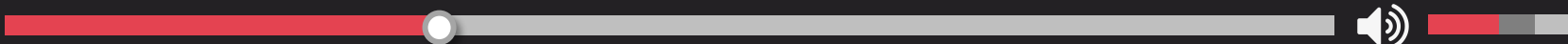


2 진행 상황 - 데이터 구축



활동량

- 인공지능에 사용하기 위해 라벨링 데이터가 필요
- 데이터셋 부족, 직접 구축 시도
- 정확도, 데이터셋 구축 등의 문제



2 진행 상황 - 데이터 구축

BeautifulSoup

JSON



Requests



활동량



인공지능



채팅로그



영상 내 소리

MoviePy



2 진행 상황 - 휴리스틱한 방법 검증



채팅로그

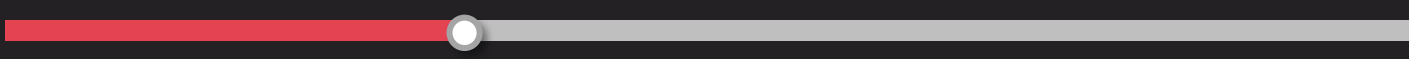


영상 내 소리

1. 풀 영상에서 데이터 다운로드

2. 각 데이터에서 상위 n 개 지점 추출

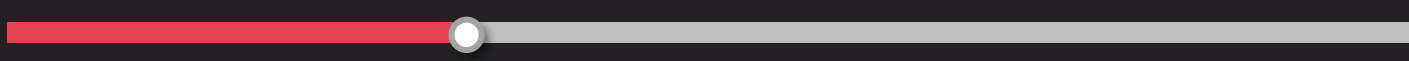
3. 추출된 $2n$ 개의 지점이 하이라이트 영상에 존재
하는지 확인



2 진행 상황 - 휴리스틱한 방법 검증

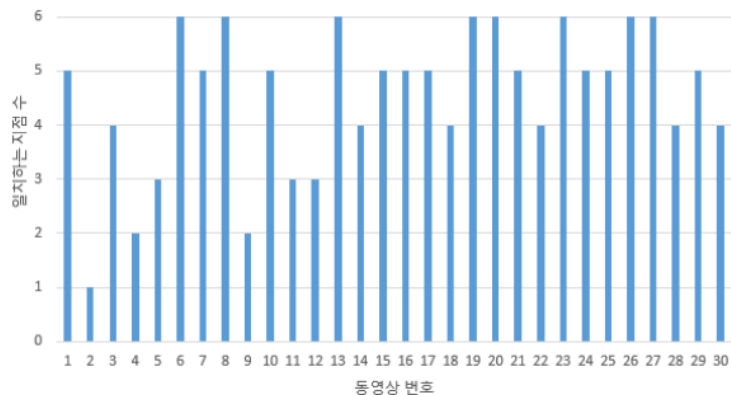
	3개 추출	5개 추출	10개 추출
Actual/Predict <input type="checkbox"/>	True	True	True
True	273(75.8%)	367(61.2%)	456(38.0%)
False	87(24.2%)	233(38.8%)	744(62.0%)

60개의 영상에 대한 검증 결과

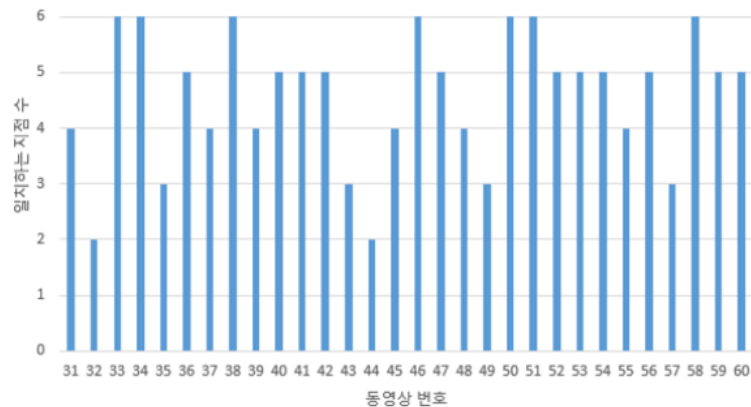


2 진행 상황 - 휴리스틱한 방법 검증

1-30번 동영상 결과



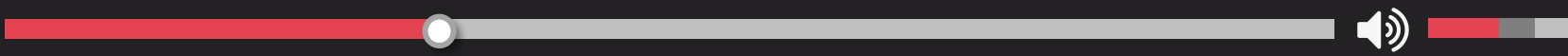
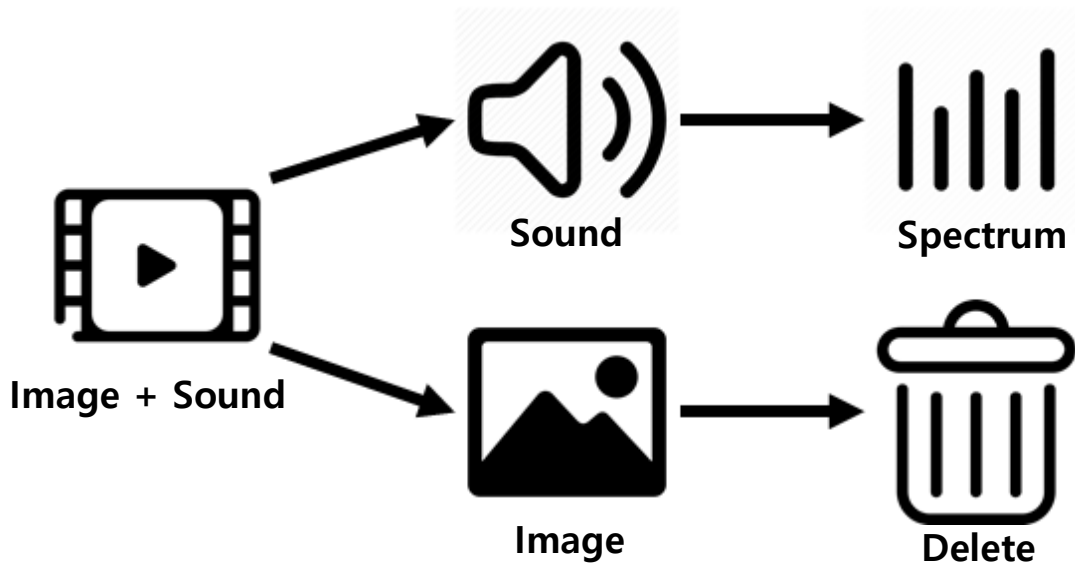
31-60번 동영상 결과



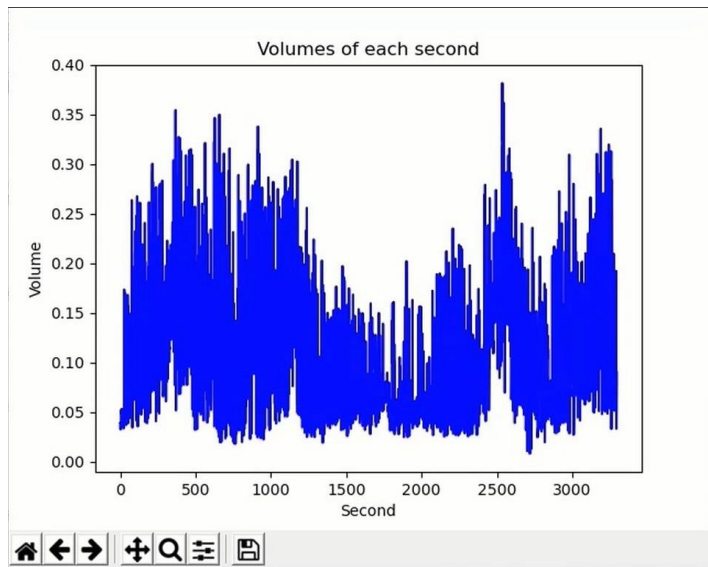
60개의 영상에 대한 검증 결과



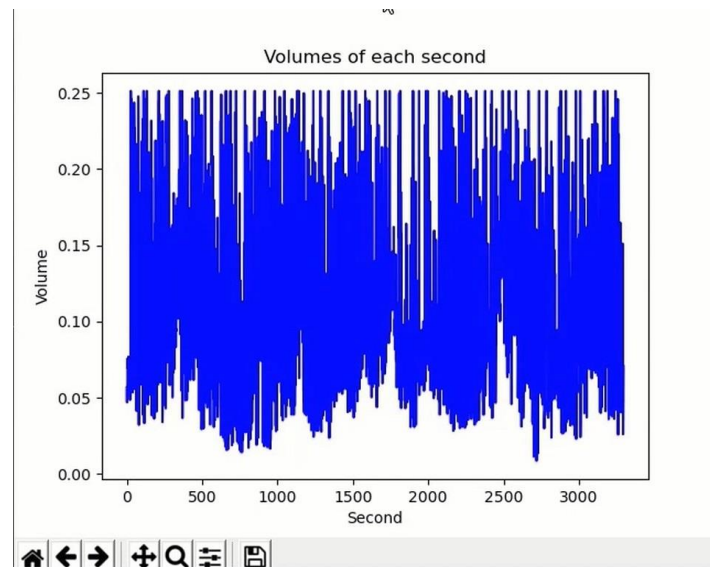
2 진행 상황 - 음성 분석 모듈



2 진행 상황 - 음성 분석 모듈



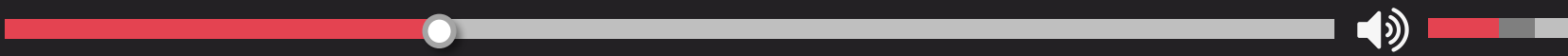
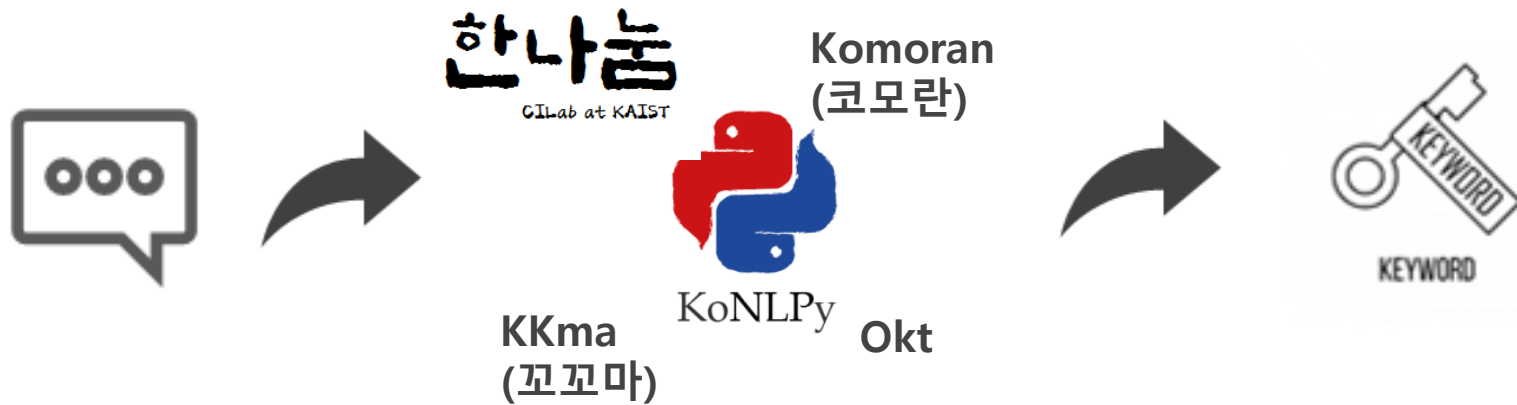
평준화 전



평준화 후



2 진행 상황 - 채팅 키워드 추출



2 진행 상황 - 채팅 키워드 추출

◦ Hannanum

	명사	출연 빈도수
1	ㅋㅋㅋㅋㅋ	510
2	ㅋㅋ	279
3	기훈	240
4	ㅋㅋㅋ	192
5	ㅋㅋㅋㅋ	178
6	얼음	137
7	돼지	191
8	우유	80
9	기후	79
10	가현	73

◦ Kkma

	명사	출연 빈도수
1	얼음	264
2	기훈	220
3	기훈이	167
4	가현	164
5	돼지	142
6	우유	129
7	기후	112
8	커피	93
9	오늘	85
10	술가락	47

◦ Komoran

	명사	출연 빈도수
1	기후	226
2	얼음	252
3	가현	238
4	돼지	135
5	우유	124
6	커피	91
7	오늘	84
8	유하	72
9	김정은	62
10	달고나	56

◦ Okt

	명사	출연 빈도수
1	기훈	276
2	얼음	169
3	가현	173
4	돼지	145
5	우유	118
6	기후	102
7	그냥	98
8	오늘	90
9	커피	82
10	유하	74

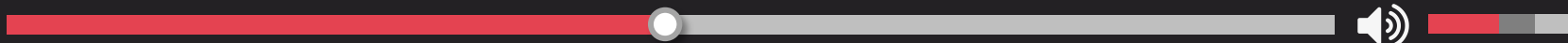
AfreecaTV BJ킹기훈의 채팅로그를 사용하여 테스트한 결과



2 진행 상황 - 채팅 키워드 추출

형태소 분석기	소요시간(초)
Hannanum	11
Kkma	132
Komoran	8
Okt	15

명사 추출에 소요된 시간



2 진행 상황 - 채팅 키워드 추출

형태소 분석기	결과
Hannanum	오늘레전드컨텐츠, 들고왔땅
Kkma	오늘, 오늘레전, 컨텐츠, 왔땅
Komoran	추출된 명사 없음
Okt	오늘, 레전드, 컨텐츠

테스트 문장 1: 오늘레전드컨텐츠 들고왔땅

형태소 분석기	결과
Hannanum	질혜님생일파티안가
Kkma	파티, 질혜님생일, 생일, 안가, 안가세요, 세요
Komoran	생일, 파티, 가세
Okt	질혜, 생일, 파티

테스트 문장 2: 질혜님생일파티안가세요



2 진행 상황 - 채팅 긍정/부정 분류



KoNLPy



감정분석

- 영상의 특정 지점에서 시청자가 보이는 반응을 지표화?
- 네이버 영화리뷰 데이터셋을 사용하여 모델 학습
- 새로운 데이터셋 구축
- 토큰나이저 모델을 바꿔가며 성능 측정



2 진행 상황 - 채팅 긍정/부정 분류

```
Preprocessing
Model define and train
2020-04-20 23:01:42.792775: I tensorflow/stream_exec
2020-04-20 23:01:43.050767: I tensorflow/core/common
name: GeForce GTX 960 major: 5 minor: 2 memoryClockR
pciBusID: 0000:01:00.0
2020-04-20 23:01:43.208535: I tensorflow/stream_exec
2020-04-20 23:01:43.227322: I tensorflow/core/common
2020-04-20 23:01:43.241622: I tensorflow/core/platfo
2020-04-20 23:01:43.311320: I tensorflow/core/common
name: GeForce GTX 960 major: 5 minor: 2 memoryClockR
pciBusID: 0000:01:00.0
2020-04-20 23:01:43.319165: I tensorflow/stream_exec
2020-04-20 23:01:43.324533: I tensorflow/core/common
2020-04-20 23:01:51.448072: I tensorflow/core/common
2020-04-20 23:01:51.454392: I tensorflow/core/common
2020-04-20 23:01:51.457189: I tensorflow/core/common
2020-04-20 23:01:51.558203: I tensorflow/core/common
ice: 0, name: GeForce GTX 960, pci bus id: 0000:01:0
2020-04-20 23:01:53.866980: I tensorflow/stream_exec
accuracy: 0.84936
```

Okt 토큰나이저를 사용한 모델

```
Preprocessing
Model define and train
2020-04-20 23:24:49.044689: I tensorflow/stream_ex
2020-04-20 23:24:49.178280: I tensorflow/core/comm
name: GeForce GTX 960 major: 5 minor: 2 memoryCloc
pciBusID: 0000:01:00.0
2020-04-20 23:24:49.192208: I tensorflow/stream_ex
2020-04-20 23:24:49.203156: I tensorflow/core/comm
2020-04-20 23:24:49.212021: I tensorflow/core/plat
2020-04-20 23:24:49.221338: I tensorflow/core/comm
name: GeForce GTX 960 major: 5 minor: 2 memoryCloc
pciBusID: 0000:01:00.0
2020-04-20 23:24:49.238090: I tensorflow/stream_ex
2020-04-20 23:24:49.253282: I tensorflow/core/comm
2020-04-20 23:24:56.890323: I tensorflow/core/comm
2020-04-20 23:24:56.896674: I tensorflow/core/comm
2020-04-20 23:24:56.899407: I tensorflow/core/comm
2020-04-20 23:24:56.946989: I tensorflow/core/comm
ice: 0, name: GeForce GTX 960, pci bus id: 0000:01
2020-04-20 23:24:58.843789: I tensorflow/stream_ex
accuracy: 0.82922
```

SPM 토큰나이저를 사용한 모델



2 진행 상황 - 채팅 긍정/부정 분류

[엠브로 반가워요🍌🍌🍌❤️]는 부정
[슬기님 안녕하세요]는 부정
[브로~~~~~]는 부정
[오~]는 부정
[슬기님은 엠브로님이랑같이일하시나요]는 부정
[유튜브 잘보고있는데]는 부정
[오~~]는 부정
[오 슬기님 안녕하세요!!!!!!]는 부정
[슬기님 하이 좋음]는 부정
[둘이케미 좋음]는 긍정
[공방인가]는 부정
[오!! 자꾸옆쪽을보시지]는 부정
[옆쪽에 사람들이많대요]는 부정
[생방인가요]는 부정
[브로 ~ 슬기 ~]는 긍정
[안녕하신교]는 부정



긍정 1 부정 0 중지 q:0
ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
쓰는 돈이 뭘 문제야 될놈될이지 삼 ㅋㅋㅋ
긍정 1 부정 0 중지 q:0
ㅋㅋㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
ㅋㅋㅋㅋㅋㅋ 배달의민족이지
긍정 1 부정 0 중지 q:1
전투민족...ㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
전투의민족 ㅋㅋㅋ
긍정 1 부정 0 중지 q:1
오늘 올라온 영상 썸네일 실한가 아직 안봤는데 ㅋㅋㅋ
그저 1 부정 0 중지 q:1

모델 test : 잘못된 분류

새로운 데이터셋 구축



2 진행 상황 - 채팅 긍정/부정 분류

긍정 1 부정 0 중지 q:0
ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
쓰는 돈이 뭘 문제야 월놈월이지 샴 ㅋㅋㅋ
긍정 1 부정 0 중지 q:0
ㅋㅋㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
ㅋㅋㅋㅋㅋㅋ 배달의민족이지
긍정 1 부정 0 중지 q:1
전투민족...ㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ
긍정 1 부정 0 중지 q:1
전투의민족 ㅋㅋㅋ
긍정 1 부정 0 중지 q:1
오늘 올라온 영상 썸네일 실한가 아직 안봤는데 ㅋㅋㅋ
긍정 1 부정 0 중지 q:1

수정된 데이터셋



[답변 부탁드립니다]는 긍정
[엠브로형 잘생겼어요]는 긍정
[2+1도아닌데]는 긍정
[사진도찌금]는 긍정
[ㅇㅇㅁㅇ ?]는 긍정
[스레기님 못생긴 이적닭았어요]는 부정
[빠연이]는 긍정
[저는요 ?]는 긍정
[1+2 인 듯]는 긍정
[이적보단 잘생겼죠]는 긍정
[아]는 긍정
[저 살짝 이현우닭음]는 긍정
[그냥 우유를 더넣으면되지 —]는 부정
[아닌데]는 긍정
[슬기님이 훨잘생김]는 긍정
[은주님 까불지 마세요 디집니다]는 부정

새로운 긍부정 결과



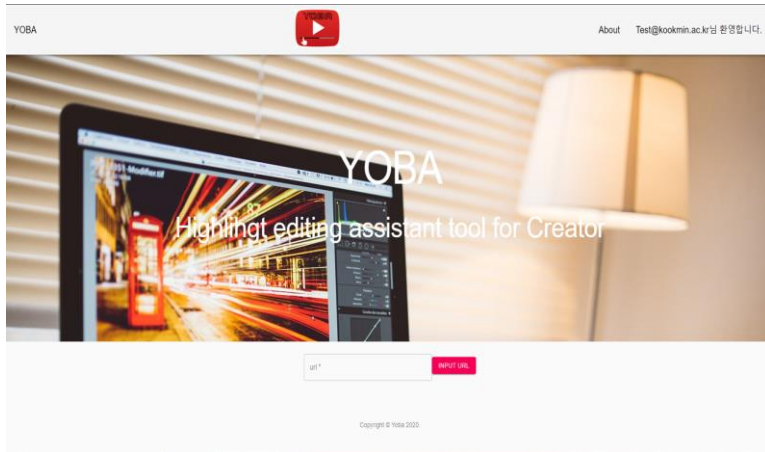
2 진행 상황 - 채팅 긍정/부정 분류

	Okt	SPM
Accuracy	71.38%	86.03%
Time	47.7s	47.0s

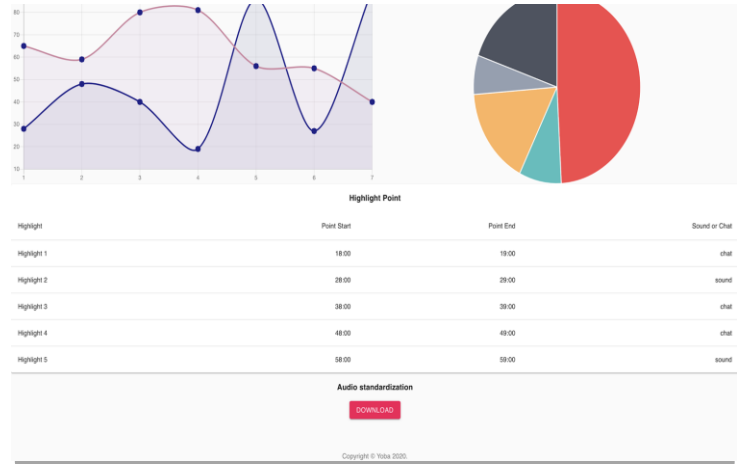
토크나이저 성능 비교



2 진행 상황 - 프론트엔드



메인화면



기본화면





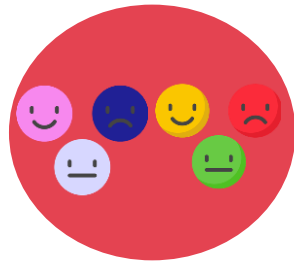
백엔드 연결



채팅 긍/부정 정확도 향상



분석한 데이터 시각화



다양한 감정 분류



THANK YOU!

