

# 12 경상도 사투리 번역 웹/앱 서비스

소속 정보컴퓨터공학부

분과 A

팀명 내마리그마리다

참여학생 김도영, 김윤중, 황승찬

지도교수 이도훈

## 과제 개요 및 목표

### 과제 개요

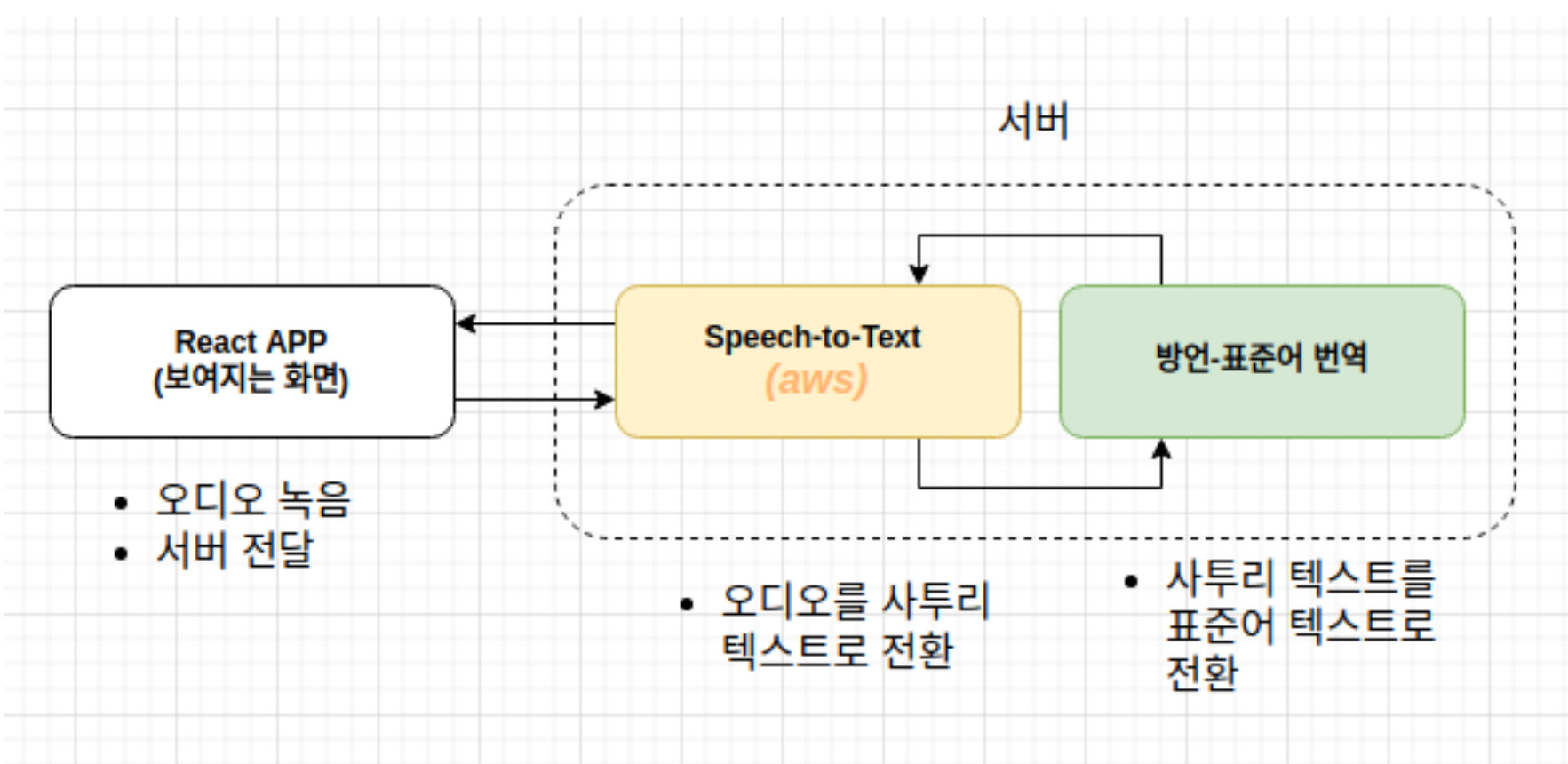
동일한 언어를 사용하는 나라 안에서, 지역마다 다른 언어 특성으로 소통에 불편함이 생기는 경우가 있다.

### 과제 목표

웹/앱 서비스를 제공하여, 경상도 사투리 번역 기능으로 지역 간의 소통을 원활하게 해준다.



## 과제 구성



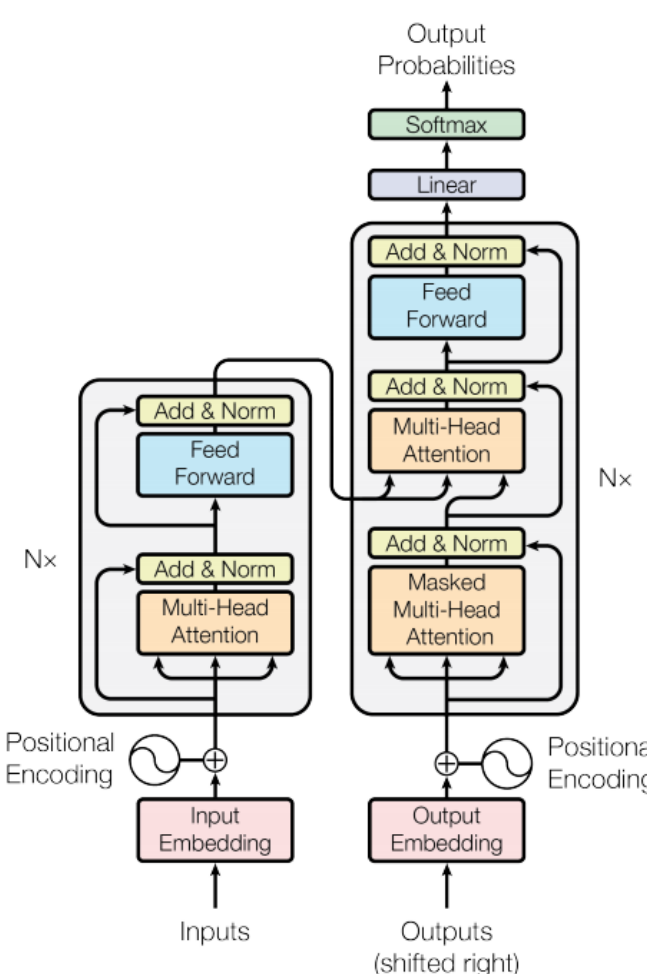
1. 사용자는 경상도 사투리를 웹/앱에 녹음 또는 입력한다.
2. 녹음된 음성은 STT를 통해, 사투리 텍스트로 변환된다.
3. 딥러닝을 사용하여 사투리를 표준어로 번역한다.
4. 번역된 사투리는 웹/앱 화면으로 출력된다.

## 과제 내용



```
f = open("Saturi_file.csv", 'r', encoding='utf-8-sig')
Saturi_word = csv.reader(f)
f.close()
#사투리 단어 우선도 추가
speech_contexts = []
scw_dict = {}
for row in Saturi_word:
    phrases = row[0]
    boost = (int)(row[1])
    if not boost in scw_dict:
        scw_dict[boost] = list()
    scw_dict[boost].append(phrases)
for key in scw_dict:
    speech_contexts_element = {"phrases": scw_dict[key],
    "boost": key}
    speech_contexts.append(speech_contexts_element)
(생략)
config = {
    "speech_contexts": speech_contexts,
    "sample_rate_hertz": 44100,
```

A1	A	B	C
16	안	50	
17	카트나	30	
18	해	10	
19	말해	10	
20	말해	10	
21	마강거로	10	
22	말노	10	
23	흔시합거로	10	
24	말만	10	
25	말해	10	
26	이해해	10	
27	어대요	10	
28	아나	10	
29	나노	10	
30	나노	10	
31	아로	10	
32	하나	10	
33	하구로	10	
34	하나	10	
35	아이고	10	
36	말만	10	
37	아이다	10	
38	진부해	10	
39	진부해	10	
40	그라해라	10	
41	말만	10	
42	안키로	10	
43	말노	10	
44	말노	10	
45	말노	10	
46	외그라노	10	
47	외그라노	10	
48	외그라노	10	
49	외그라노	10	
50	안키	10	
51	도가자	10	
52	말만	10	
53	말다 크마해라	10	
54	가임사다	10	
55	말만	10	



구글 STT가 음성을 텍스트로 변환

딥러닝을 활용한 NMT기반의 번역

웹/앱 화면

## 향후 방향

한국어 환경의 음성 인식 기능이 부족한 API 대신, 음성 인식 프로그램을 자체 개발하여, 한국어 음성 인식율을 높일 것이다.

유저들이 사용한 데이터를 바탕으로 파인튜닝하여 모델을 학습하는 방법을 사용하면 질 높은 데이터셋이 만들 것이다.

경상도 사투리뿐만 아니라, 타 지역의 사투리 모델까지 추가하여, 전국 기반의 사투리를 번역해주는 서비스로 거듭날 것이다.