유튜브에 올라갈 법한 영상의 자막을 생성하는 프로그램을 만들기 위해, 기존 음성인식들과 다르게 여러 최신 유행어, 밈(meme), 비속어 등이 들어간 dataset 구성

**[음성인식 부분]**

* Pytube를 사용하여 유튜브 영상 추출 후, 스크립트 작성하여 dataset 생성
* 실제 영상에는 잡음이 많이 섞이게 되는데 이를 개선하기 위해 카카오 엔터프라이즈에서 제시한 SE-Conformer 모델로 학습

(<https://kakaoenterprise.github.io/papers/interspeech2021-se-conformer>)

* Kospeech github 참조

(<https://github.com/sooftware/kospeech#introduction>)

**[자연어 처리 부분]**

* 가장 최신 데이터까지 포함 되어있는 Korean unsmile dataset 사용

(<https://github.com/smilegate-ai/korean_unsmile_dataset>)

* 많이 사용하는 모델인 Kcbert 모델을 사용하여 학습
* 프로젝트를 진행하며 dataset이 업데이트 되지 않을 경우 직접 크롤링을 통해 dataset수집 예정

**[개발 부분]**

* 영상 편집 프로그램, 윈도우 GUI프로그램 개발 시 C# 기반 유니버설 Windows 사용
* 더 자세한 부분은 음성인식과 자연어처리를 우선적으로 진행한 후 결정