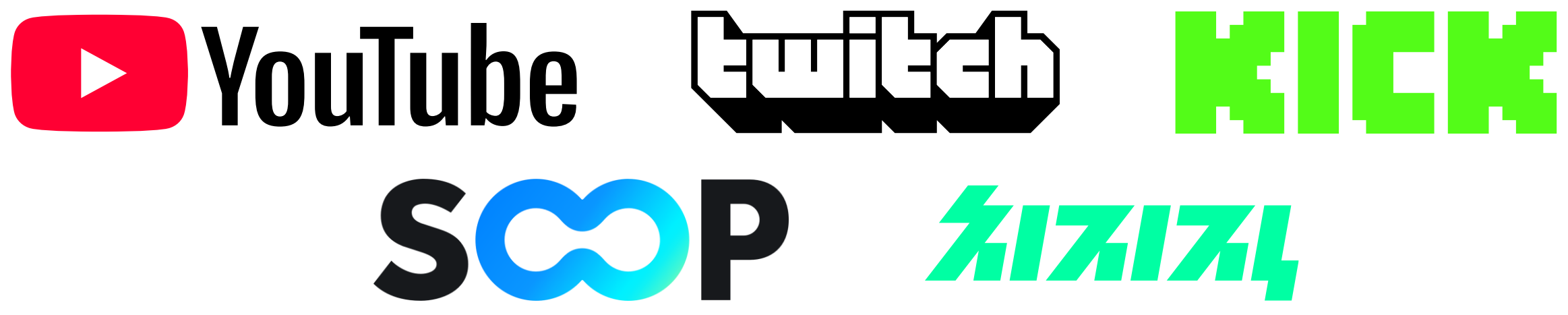


# 클러스터링 기반 실시간 스트리밍 방송 서비스 내 주요 채팅 선별 및 음성화 서비스

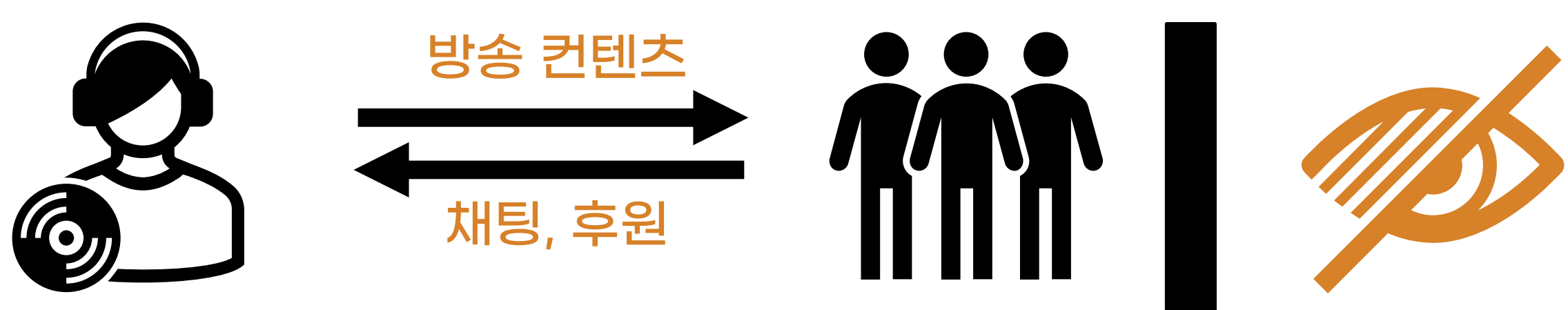
이름 양성호

지도교수 조현석

## 연구 배경



코로나 팬데믹 이후 실시간 스트리밍 방송 서비스는 전세계적으로 실내 여가 생활을 즐기는 이들의 주요한 여가 수단 제공 플랫폼 중 하나로 자리 잡음

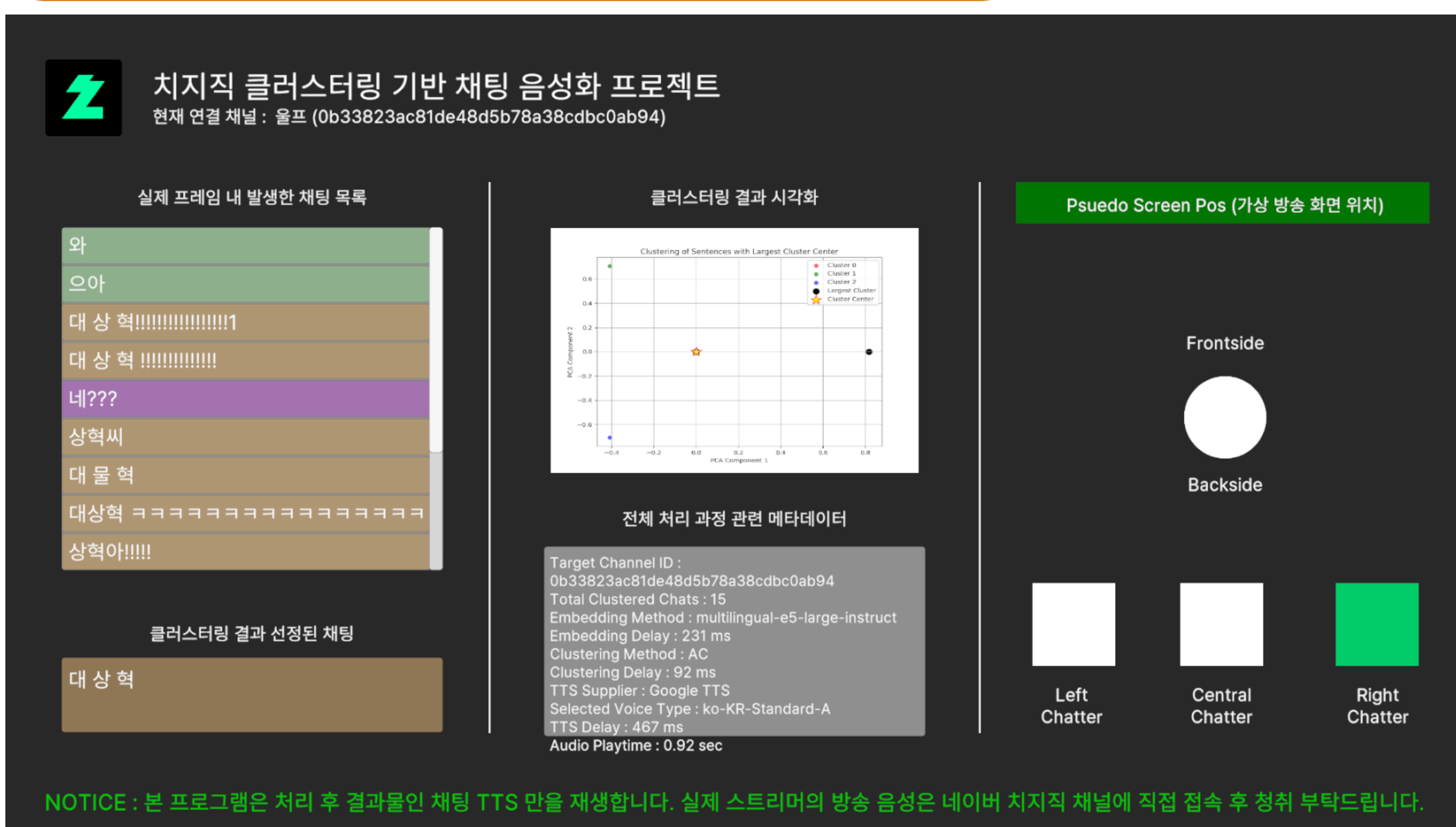


이러한 플랫폼의 시청자들은 스트리머의 방송 콘텐츠를 통해 대리 만족을 느끼며 채팅 및 도네이션 등의 수단을 통해 스트리머와 소통하고 유대감을 형성함

그러나, 게임 화면에 집중하는 스트리머와 노안, 장애 등으로 인한 시각장애인 및 시력 약자, 화면 앞을 이탈한 사용자 등 '채팅을 볼 수 없는 환경의 이용자'들은 위 특징을 누리기 어려움

본 프로젝트는 텍스트 임베딩 및 클러스터링, 음성화 기술을 통해 대상 사용자가 제약 상황을 극복할 수 있도록 도움으로써 위 서비스들이 제공하는 실시간 양방향 소통의 이점을 일반 시청자들과 유사한 수준으로 누릴 수 있도록 하는 것을 목표

## 결과 및 분석



프로젝트의 산출물인 위 Unity Engine 기반 클라이언트를 통해 실시간 스트리밍 방송과 함께 클러스터링 과정의 정보를 확인하며 음성화 채팅을 경험해 볼 수 있다. 공간 음향을 적용하여 전방의 방송 음향과 구분되는 채팅 음성이 제공함으로써 다중 음향을 구분해야 하는 청자의 피로도를 낮추고자 하였다.

본 프로젝트를 통해 시각을 충분히 활용할 수 없는 상황의 시청자에게 보다 나은 접근성을 제공하는 방법을 탐색하였다. 추후 STT/LLM 등 NLP 관련 기술들을 추가 적용한다면 듣는 것의 단방향 시청에서 더 나아가 채팅 참여 등 진정한 양방향 소통을 제공하는 접근성 제공 수단으로 발전시킬 수 있을 것이다.

## 오픈소스 URL

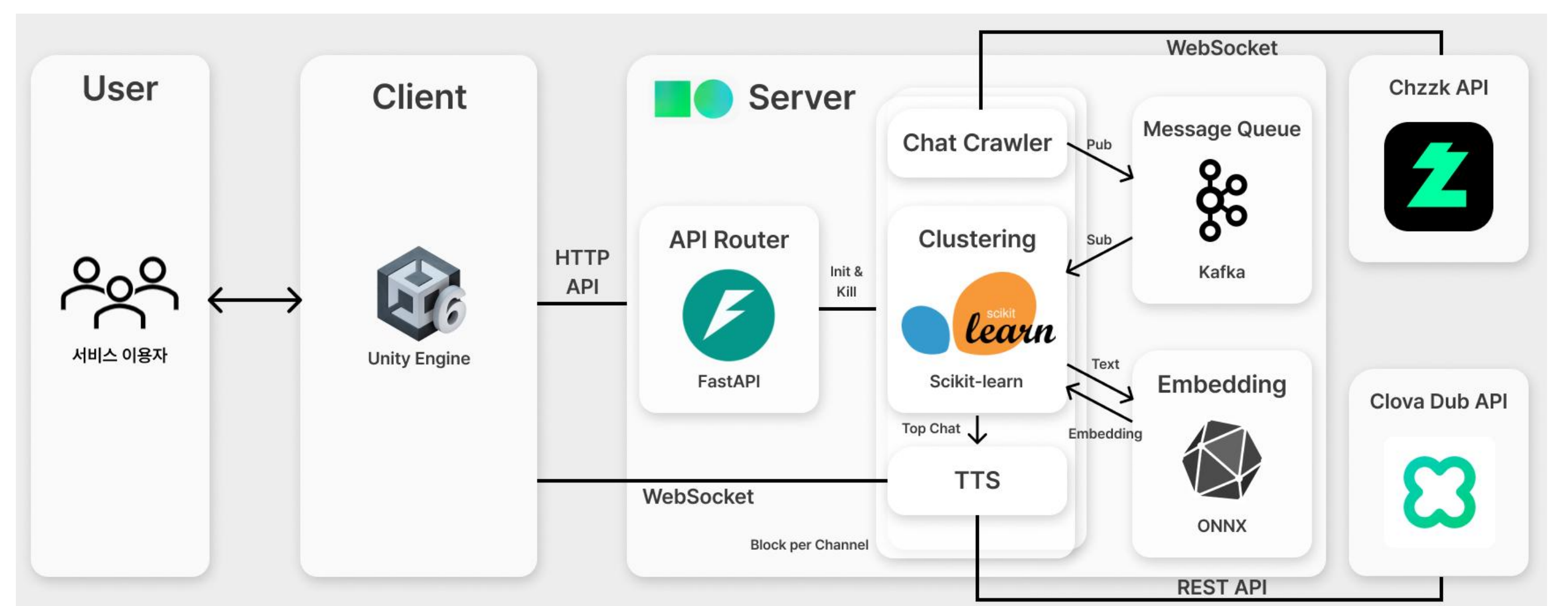
Github Repo: <https://github.com/SyngSHY/SDP2402-CLIENT>

## 연구 진행 과정

### 대상 채팅 데이터

본 프로젝트 수행 및 서비스를 위해서는 대용량의 실시간 채팅 데이터가 필요하다. 클러스터링에 사용한 대상 채팅 데이터는 네이버 치지직(Chzzk) 서비스의 채팅 데이터를 이용하였다. 채널에 연결한 사용자들은 WebSocket을 통해 제공되는 채팅 데이터를 실시간으로 받아올 수 있다. 본 프로젝트에서는 클러스터링을 위해 채팅 데이터 중 메시지 본문만을 활용하였으며 이외 채팅 닉네임 등의 메타데이터는 제거하였다.

### 프로젝트 구조



본 프로젝트는 MSA(Micro-Service Architecture)를 채택하여 개발하였다. 클라이언트의 연결 요청을 처리할 각 세부 모듈을 분리하여 일부 채널의 모듈에서 Fail이 발생하더라도 전체 서비스에 영향이 미치는 것을 방지하였다.

각 모듈은 메시지 보호의 필요성에 따라 Kafka 메시지 큐를 통해 간접 메시징, 또는 HTTP API를 통한 직접 메시징으로 통신한다. 클러스터링의 경우 채팅 데이터 각각이 클러스터링 결과와 그 의미에 중대한 영향을 끼치기에 MQ를 거치게 되며 TTS의 경우 해당 순간이 지나면 메시지의 가치가 사라지기에 보호 필요성이 상대적으로 떨어져 HTTP API를 선택하였다.

서빙 인스턴스 환경은 네이버 클라우드 플랫폼을 채택하고 플랫폼 내 서버 및 Clova Voice를 활용하여 내부 통신 지연 시간 단축과 추후 서비스 확장 시 내부 연계성을 고려하였다.

### 임베딩 및 클러스터링

본 프로젝트 내 텍스트 임베딩을 위해 다양한 임베딩 모델을 시험한 결과 'multilingual-e5-large-instruct' 모델을 선택하였다. 해당 모델은 BERT 계열 다국어 임베딩 모델로 적은 메모리 사용량 및 작은 임베딩 차원 수로 고품질의 경쟁력 있는 임베딩 결과를 도출해냈다. 해당 모델 적용 시 ONNX 변환 모델로 CPU 환경에서 추론 비용의 최적화를 시도하였다.

클러스터링 알고리즘으로는 Agglomerative Clustering 알고리즘을 선택하여 적용하였다. 잘 알려진 K-means 알고리즘 대비 클러스터 간 밀도 차이를 적절히 반영할 수 있음에 주목하였으며, DBSCAN 계열 알고리즘과 비교하였을 때 샘플 포인트 수가 적을 때도 더 의미 있는 결과를 보여준다고 판단하였다. 밀도 차이를 바탕으로 더 큰 군집의 중심 채팅을 주요 채팅으로 판단하여 TTS 대상으로 전달하게 된다.

Github Repo QR ▶  
시연용 Web Client QR ▶▶

