



국민대학교
전자정보통신대학
컴퓨터공학부

캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	실시간 화상통화 번역 시스템
팀 명	LTE
문서 제목	중간 보고서

Version	1.31
Date	2019-04-16

팀원	박 지훈 (조장)
	공 병민
	최 창규
	고 준규
	박 귀환
	오 석현
지도교수	이 재구 교수



프로젝트 제안서		
프로젝트 명	실시간 화상통화 번역 시스템	
팀 명	LTE	
Confidential Restricted	Version 1.31	2019-APR-16

CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 "LTE"를 수행하는 팀 "LTE"의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 "LTE"의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

문서 정보 / 수정 내역


Filename	프로젝트제안서-LTE.docx`
원안작성자	박지훈, 공병민, 최창규, 고준규, 박귀환, 오석현
수정작성자	박지훈, 공병민

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2019-04-15	박지훈	1.0	최초 작성	
2019-04-15	고준규	1.1	수정	오타 및 잘못 입력된 부분 수정
2019-04-15	공병민	1.2	수정	향후 추진 계획 수정
2019-04-16	최창규	1.3	수정	건의사항 추가
2019-04-16	박귀환	1.31	수정	오타 수정 및 검토

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	프로젝트 제안서		
	프로젝트 명	실시간 화상통화 번역 시스템	
	팀 명	LTE	
	Confidential Restricted	Version 1.31	2019-APR-16

목 차

1	프로젝트 목표	4
2	수행 내용 및 중간결과	5
2.1	계획서 상의 연구내용	5
2.2	수행내용	5
3	수정된 연구내용 및 추진 방향	7
3.1	수정사항	7
4	향후 추진계획	8
4.1	향후 계획의 세부 내용	8
5	고충 및 건의사항	9

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	프로젝트 제안서		
	프로젝트 명	실시간 화상통화 번역 시스템	
	팀 명	LTE	
	Confidential Restricted	Version 1.31	2019-APR-16


1 프로젝트 목표

본 프로젝트는 다국적 간의 영상통화에서 언어의 차이로 인한 의사소통의 불편함을 해소를 목적으로 한다. 사용자의 영상에서 음성을 추출한 후 음성 인식 기능을 통해 Script로 변환한다. 이렇게 변환된 Script는 각각의 사용자들이 설정한 언어에 맞게 번역된 Script로 각각의 사용자의 화면에 표시하여 다국적 간의 의사소통을 원활하게 한다.

1. 1.3명 이상의 화상 통화 서비스를 제공하는 것을 목표로 한다.
2. 영상에서 사용언어를 다국어(한국어, 영어, 일본어)로 번역하여 제공한다.
3. 기존의 서비스와 다르게 설정 단계를 어렵지 않게 하여 접근성을 높인다.

ex) 간편한 채팅 Room 생성, 간편한 ID 등록 등

4. 각각의 사용자에게 보여지는 번역 Script를 채팅방 형식으로 기록을 남겨서 이를 통해 의사소통을 하는데 어려움을 해소한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	프로젝트 제안서		
	프로젝트 명	실시간 화상통화 번역 시스템	
	팀 명	LTE	
	Confidential Restricted	Version 1.31	2019-APR-16

2 수행 내용 및 중간결과

2.1 계획서 상의 연구내용

서로 다른 언어를 사용하는 3인이 화상통화 서비스를 이용할 수 있게 한다. 각 사용자는 접속할 때 아이디와 사용언어를 입력한다. 화상통화를 이용 시 입력한 사용언어로 번역된 자막이 화면에 떠 있게 된다. 이를 통해 다국적 사용자간 언어의 소통이 원활히 할 수 있다.

계획서 상의 진도는 speech to text 기능 구현이다. 웹서버 환경이 구현이 되고 사용자 간의 화상통화가 가능한 상황에서 사용자가 말한 음성을 인식해 text로 변환하는 것을 계획했었다.

2.2 수행내용

현재 화상통화 부분을 제외한 채팅방 서버를 구현하였다. 사용자에게 랜덤한 아이디를 부여한 뒤 방을 생성하거나 이미 생성되어 있는 방에 입장하여 채팅을 할 수 있게 구현하였다. 또한 방을 생성 시 비밀번호를 설정할 수 있도록 하여 회의 인원 이외에 허락되지 않은 인원은 들어올 수 없도록 구현하였다.


채팅은 사용자의 마이크를 통해 말을 하면 annyang api(speech to text api)를 통해 인식을 한 뒤, 이를 서버에 보낸다. 서버에서는 구글의 translate api를 통해 보낸 사용자의 말을 3개의 언어(한국어, 영어, 일본어)로 번역하게 된다. 번역된 결과를 각 사용자의 채팅방에 보여준다. 또한 번역 결과는 서버내 mongo 디비에 저장되어 방을 나가더라도 다시 들어왔을 때 이전의 대화를 보여주는 역할을 한다.

계획과 현재 상황을 비교해보면 화상통화 부분이 구현되어 있지 않다. 이는 원래의 계획(화상통화 화면에 자막을 달아주는 것) 과 다르게 사용자 간의 대화를 처음부터 기록하여 log형식의 채팅방으로 보여주기 위해 채팅방을 구성하고 있기 때문이다. log 형식의 기록을 서버에 저장하는 것뿐만 아니라 사용자에게 보여주어 기존의 자막 형식보다 가독성을 높이게 되고, 회의, 통화를 끝낸 뒤 회의, 통화 내용을 한눈에 파악하기도 용이하다는 장점이 있다. 또한 네트워크 속도상 음성 쿼리 스케줄링이 잘못되어 먼저 말한 질문이 대답보

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	프로젝트 제안서		
	프로젝트 명	실시간 화상통화 번역 시스템	
	팀 명	LTE	
	Confidential Restricted	Version 1.31	2019-APR-16

다 늦게 나오더라도 채팅방에서 눈으로 보고 맥락을 파악할 수 있어 스케줄링의 한계를 보완할 수 있다는 장점이 존재한다.

또한 STT기능 뿐 아니라 translate 기능 역시 구현 완료하여 인식된 결과물이 input으로 들어오고 이를 영어와 일본어로 번역 가능하다. 하지만 사용자마다의 사용언어를 입력받는 기능을 만들지 못하였기 때문에 인식은 한국어만 가능하고, 인식된 결과물(한국어)와 번역된 결과물(영어, 일본어)를 함께 채팅방에 보여주는 형식으로 구현되어 있다. 이는 인식과 번역의 테스트 과정에서 얼마나 정확한지 눈으로 확인하기 위해 우선적으로 만들었다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	프로젝트 제안서		
	프로젝트 명	실시간 화상통화 번역 시스템	
	팀 명	LTE	
	Confidential Restricted	Version 1.31	2019-APR-16

3 수정된 연구내용 및 추진 방향


3.1 수정사항

구글 음성인식 api 를 이용하여 speech to text를 구현하려 했지만, 사용해본 결과, 클라이언트 측에서 마이크의 권한을 가져와 소리를 인식할 방법을 찾지 못하였다.

그래서 음성인식 서비스를 다른 방면으로 조사해 보았고, annyang api를 사용하여 클라이언트 측에서 마이크 권한을 요청할 수 있었다. 이로써 클라이언트에서 인식한 소리를 서버에 전송하여 서버에서 구글 translate api를 통해 번역한 뒤, 클라이언트에 다시 전송하는 작업을 수행할 수 있었다. 서버는 Google Cloud Platform에서 vm 인스턴스를 생성하여 구축하였다. 처음 서버는 http 프로토콜을 이용하여 구축하였는데, 보안상의 이유로 웹 브라우저 상에서 마이크 권한 설정 자체를 차단시켰다. 따라서 서버를 https 프로토콜로 전환하였고, ssl 인증서는 일단 구현의 목적이 크기 때문에 자체 인증서를 발급받았다.

또한 원래는 socket으로 통신하려고 하지 않고, 일일이 서버가 계속해서 각각의 클라이언트에서 인식한 결과가 있는지 확인하는 구조로 구현하려 했으나, 서버 성능 부하를 우려하여 웹 socket을 사용하였다.

구체적인 구현 내용은 클라이언트에서 소리를 가지고 인식된 결과가 있다면 바로 socket을 이용하여 서버에게 전달한다. 서버에서는, 디비에 저장된 현재 접속한 사람의 언어를 바탕으로 번역을 실시하여 번역된 결과를 클라이언트에 보내게 된다. 또한 이 서비스의 원래목표는 동시 접속 가능한 사람 수가 최대 4명이었지만, 서버에서 5개 정도의 언어까지 번역을 처리하는데 시간차이가 없었고, socket을 통해 데이터 교환이 이루어지기 때문에 동시 접속 가능한 사람 수를 더 늘릴 수도 있을 것 같다. 이 프로젝트의 목표인 실시간에 가깝게 하려면, 소리의 유사도, 즉 인식된 결과의 유사도를 검사하는 횟수량을 조절하거나, 인식된 결과물이 아닌, 시간을 기준으로 소리를 가져와야 하는데, 이는 정확성과의 tradeoff 관계이기 때문에 차후에 더 고민해 볼 예정이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	프로젝트 제안서		
	프로젝트 명	실시간 화상통화 번역 시스템	
	팀 명	LTE	
	Confidential Restricted	Version 1.31	2019-APR-16

4 향후 추진계획

4.1 향후 계획의 세부 내용

현재까지 만든 서비스에 화상 통화 기능을 구축하는 것을 우선으로 한다. 화상 통화 api는 여러가지가 있으나 webRTC api를 사용할 것이 유력한데, 이유는 WebRTC로 구성된 프로그램들은 별도의 플러그인이나 소프트웨어 없이 p2p 화상회의 및 데이터 공유가 가능하기 때문이다. 현재는 카카오톡처럼 대화방을 만들 수 있어서 여러 집단이 서비스를 이용할 수 있지만, 화상 서비스까지 구현한 후, 웹 통신이 느려지게 된다면 없앨 예정이다. 여기까지가 필수 향후 계획이다.

성능적인 부분에서 수행해 볼 것은 구글 speech to text api를 클라이언트 측에서 사용할 수 있는 방법을 연구하여 사용해 볼 예정이다. 로컬에서 테스트한 결과로는 딱히 음성인식 결과가 크게 차이나진 않지만 그래도 google api인 만큼 사용해볼 가치가 있다고 생각한다. 그리고 더 우선적으로 수행할 것은 google translation api에 커스텀 모델을 사용해볼려 하는데, 아무래도 번역이다 보니, 일상적인 대화와 회의나 연설 같은 전문적인 대화를 똑같은 어체로 번역한다. google translation api는 커스텀 모델을 학습시켜 그것을 이용해 번역할 수 있는데, 일상 대화적인 가벼운 어체를 사용하는 모델과 보다 무거운 어체를 사용하는 모델을 만들어서 사용해볼 생각이다.

마지막으로, 현재는 각 사용자가 필요한 번역을 제공하는 것이 아니라, 모든 사용자가 같은 번역 결과를 본다. 예를 들어, 한국인 일본인 미국인이 대화하고 있다면 한국인이 말한 음성 결과는 일본어와 미국어로 번역되고, 그것이 통째로 브라우저에 뿌려지는데, 가능하다면 자신에게 필요한 언어만 보일 수 있도록 구현하고자 한다. 이를 구현하려면 각 사용자들마다 서로 다른 웹페이지를 가지고 있어야 하는데, 효율적으로 동적 웹페이지를 구성하기 위해 좀 더 구현에 관해 생각해 보아야 한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	프로젝트 제안서		
	프로젝트 명	실시간 화상통화 번역 시스템	
	팀 명	LTE	
	Confidential Restricted	Version 1.31	2019-APR-16

5 고충 및 건의사항

Speech to text api를 사용할 경우 시끄러운 환경에서 음성인식을 하는데 어려움을 갖는다.
 그로 인해서 향후 결과발표 시연 시 어려움을 겪을 것으로 예상된다. 이에 따라 조용한
 장소에서 시연하는 것을 희망합니다.