Google Cloud Platform을 활용한 음성 인터페이스 설계 및 구현

최동욱, 이지훈, 문남미
호서대학교
dwchoi95@gmail.com, monaminino@gmail.com, moon.nammee@gmail.com

Design and Implementation of Voice Interface Based on Google Cloud Platform

DongWook Choi, JiHoon Lee, NamMee Moon
Hoseo University

요약

음성 인터페이스는 Google Cloud Platform에서 제공하는 STT(Speech To Text)와 TTS(Text To Speech) API를 이용하여 구현한다. 주요 명령어로는 읽기, 쓰기, 실행, 종료, 검색이 있으며, 이 중 검색은 JSoup과 Selenium 크롤링 라이브러리를 사용하여 구현하였다. 사용자가 마이크를 통해 말을 하면 STT를 통해 텍스트로 반환되어 Java 언어로 구현된 시스템을 통해 명령어를 해석하여 PC를 제어한다. 본 시스템은 장애인들이 전자기기를 사용하는데 많은 편리함을 줄 것으로 예상하며 나아가 장애인뿐만 아니라 비장애인들도 활용 가능할 것으로 기대한다.

1. 제작 동기

전자기기의 사용량이 증가함에 따라 소셜 네트워크와 같은 다양한 플랫폼이 등장하고 디지털 디톡스 (Digital Detox)라는 신개념이 생길 정도로 사용량이 증가하였다[1]. 2016년 국내 스마트폰 가입률은 90.6%이며, 이는 계속해서 증가하고 있다[2]. 더불어, 장애인에게 보조기기는 자신의 장애를 수용하기까지 오랜 시간이 걸리고 자신의 장애가 겉으로 드러나는 것을 매우 싫어하는 특징에 따라 추가적인 다른 보조기기의 사용을 꺼린다.[3]

이러한 문제점을 기반으로 장애인이 전자기기를 사용하는 데에 일어나는 불편함을 해소하고자 여러 가지 보조 기기들이 출시되고 있으며, 이런 보조기기의 연구개발 및 지원방안에 관한 연구가 활발히 진행되고 있다.[4][5][6]

이런 동향에 따라 장애인들이 모바일뿐만이 아닌 PC 환경에서도 원활한 작업이 가능하도록 Google Cloud Platform에서 제공하는 Google Speech API를 활용하여 음성 인식 인터페이스를 설계 및 구현하였다.

2. 설계 및 구현

본 연구에서 제안한 음성 인터페이스는 Window 환경에서 이클립스, GCP(Google Cloud Platform), MySQL을 이용하여 시스템을 구현하였다. 음성 인터페이스를 기본 구성은 STT, Command, TTS 이렇게 3가지 부분으로 나누어지며 GUI를 통해 사용자가 사용하기 편리하게 개발하였으며 로그인 기능과 프로그램 경로 설정을 위해 DB를 연동하였다.

그림1은 본 시스템의 구성도이다. 사용자가 회원가입 후 로그인을 하면 음성 인식을 할 수 있게 된다. 그러면 사용자는 마이크를 통해 말을 하고 마이크를 통해 인식된 음성 데이터는 인코딩 과정을 거쳐 STT API로 보내진다. STT API에서는 인코딩된 음성 데이터를 Text 형태로 변환해준다. 변환된 Text는 구현된 명령 알고리즘을 통해 어떤 명령인지 해석하여 프로그램을 실행, 종료, 읽기, 쓰기, 검색 등의 실행 가능한 데이터 형태로 바뀌어 PC를 제어한다. 예를 들어 '울트론, 카카오톡 켜 줘'와 같이 말을 하면 '켜줘'라는 동사와 '카카오톡'이라는 명사를 인식하여 카카오톡을 PC 화면에 띄워준다. 또한, '카카오톡이 실행됩니다.'와 같이 제어 결과를 TTS API를 통해 MP3 파일로 변환해주어 스피커를 통해 출력해주기도 한다.

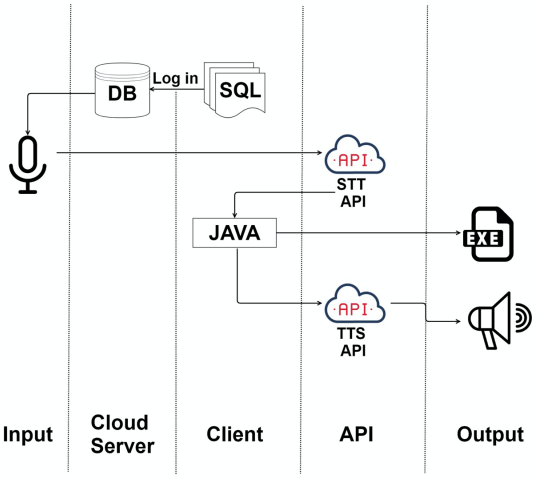

그림 1 음성 인터페이스 구성도

2.1 STT(Speech To Text)

사용자의 음성을 마이크를 통해 인식해야 하므로 사람의 음성 주파

수 중 음성 인식에서 가장 많이 사용되는 16000Hz[7]로 Audio Formating을 한 뒤 Linear PCM 인코딩을 사용하여 음성을 인식한다. 음성은 실시간으로 계속 입력받아야 하므로 Google STT API에서 streamingMicRecognize를 선택하였다. 그리하여 실시간으로 음성을 텍스트로 변환 받았다.

2.2 Command

사용자는 '울트론. 인터넷 켜 줘'와 같이 3형식의 문장으로 말을 하므로 텍스트 전처리 과정을 통해 고유명사, 주어, 동사 등으로 분류하여 명령 알고리즘으로 보내주었다. 명령 알고리즘은 해당 문장을 해석하여 어떤 명령인지 파악하며 만일 일치하는 명령이 없다면 사용자에게 다시 말할 것을 권유한다. 명령어는 크게 실행, 종료, 읽기, 쓰기, 검색 등 총 5개로 나뉜다. 첫 번째, 실행 명령어는 주어에 따라 실행할 프로그램을 선택하고 해당 프로그램의 exe파일을 찾아 실행하여 준다. 프로그램이 실행되면 프로세스의 주소를 리스트에 저장하여 두었다가 종료 명령을 사용할 때 사용한다. 두 번째, 종료 명령어는 실행 명령어와 반대로 주어에 따라 종료할 프로그램을 실행된 프로그램 리스트에서 찾아 task kill을 사용하여 종료하고 리스트에서 지워준다. 세 번째, 읽기 명령어는 사용자가 읽을 파일명을 말하면 해당 파일명을 찾아 파일 안의 내용을 가져와 TTS를 통해 스피커로 출력해준다. 네 번째, 쓰기 명령은 메모장, 엑셀, 한글, 워드 등의 편집할 문서를 말하면 해당 문서에 사용자의 음성을 text로 변환해주어 적어준다. 문서 편집이 끝나면 '저장'이라고 외쳐서 파일을 저장한다. 파일을 저장할 때 사용자가 파일명을 말로써 지정할 수 있다. 마지막으로 검색 명령어는 Selenium과 JSoup을 사용하여 개발하였다. 검색하기 위해 크롬에서 검색할 내용을 크롤링하여 보여주거나 검색 결과를 TTS를 통해 스피커로 출력하도록 하였다.

2.3 TTS(Text To Speech)

스피커를 통해 음성을 출력해주기 위해 Google TTS API를 사용하였는데 TTS API의 경우 텍스트를 MP3의 음성파일 형태로 만들어주었다. 그래서 음성을 출력하기 위해 JLayer 라이브러리를 사용하여 MP3를 재생해주었다.

3. 구현 결과

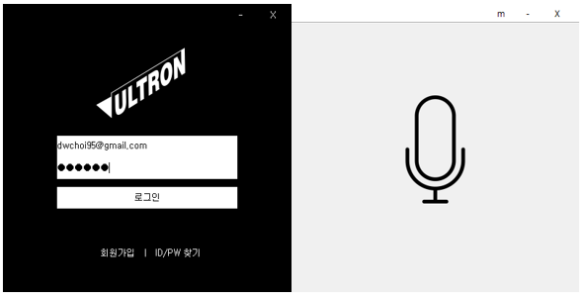

그림 2 음성 인터페이스 로그인 및 대기 화면

그림 2는 울트론 음성 인터페이스의 사용을 위해 사용자가 로그인 후 인터페이스가 음성 인식을 시작하기 전 대기하는 화면이며, 사용자가 중앙의 마이크 화면을 클릭할 경우 음성 인식 서비스가 시작된다.

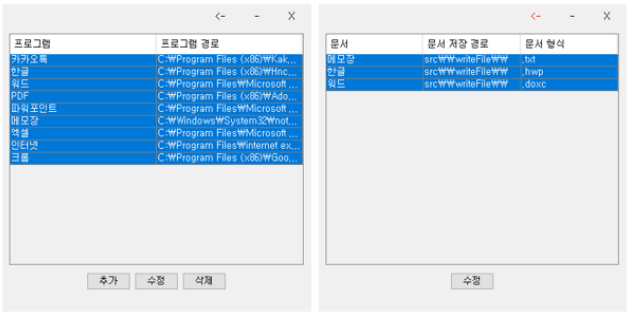

그림 3 실행 프로그램 경로 및 문서 저장 경로 화면

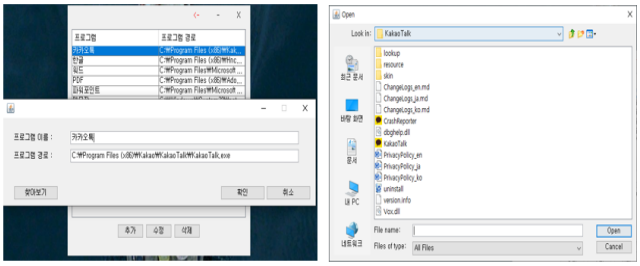

그림 4 실행 프로그램 경로 지정 과정

그림 3과 그림 4는 사용자가 실행 명령어를 전달할 때 실행할 수 있는 프로그램의 목록이며 하단의 추가 및 수정을 통해 실행 가능한 응용 프로그램의 목록 및 경로를 수정할 수 있다. 우측의 문서 저장 경로에서는 사용자가 쓰기 명령 수행을 원하면 한글, 워드, 메모장과 같은 문서 작성 프로그램에 전달하여 해당 문서 형식을 지정하고, 문서 저장 경로 또한 지정할 수 있다.

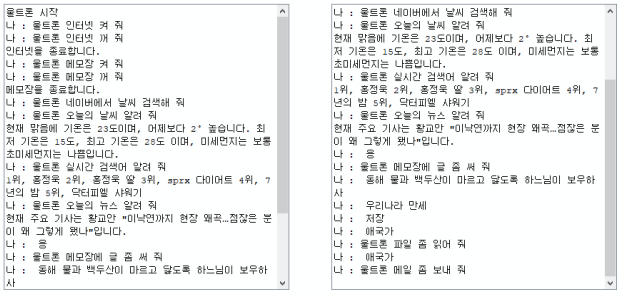

그림 5 음성 인터페이스 실행 로그 확인 화면

그림 6 Selenium 크롤링 결과 창

그림 5는 구현 결과 설명을 위해 테스트 시 사용한 명령어의 로그이

다. 사용자는 음성 인터페이스를 실행하면서 말한 명령어를 확인할 수 있으며, 이는 로그를 통해 확인할 수 있다. 위의 로그에서 볼 수 있는 바와 같이 사용자가 음성으로 즉각적인 응답을 원할 때 Jsoup 라이브러리를 사용하여 크롤링 된 데이터가 TTS를 통해 전달된다. 또한, 사용자가 주요 뉴스나 기사에 관해 묻고 해당 기사를 보여주길 원한다면 그림 6에서와같이 Selenium 라이브러리를 통해 Web browser를 띄워준다.

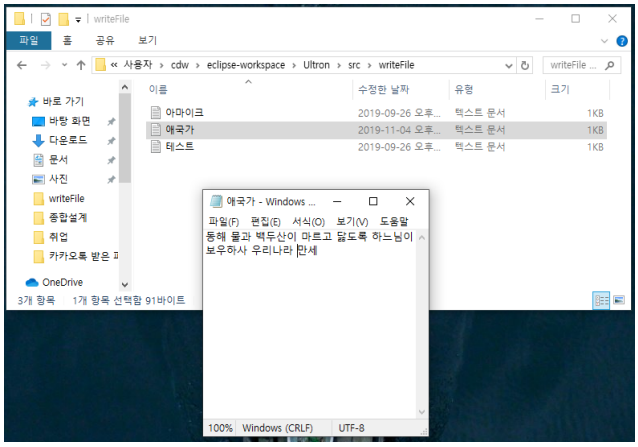

그림 7 쓰기 명령 결과 화면

쓰기 명령어와 같은 메모 기능은 사용자가 지정한 경로와 확장자를 통해 그림 7에서 보이는 바와 같이 저장된다.

4. 기대효과

다수의 응용 프로세스들은 키보드 또는 마우스 혹은 터치스크린을 활용하여 인터페이스가 이루어져 있으며, 이는 장애인이 전자기기를 사용하는 데에 많은 불편함이 있다. 장애인이 전자기기를 사용하는 데 필요로 하는 보조기기는 키보드나 마우스처럼 기본적으로 제공되는 인터페이스를 제외한 별도로 출시되며, 추가적인 비용을 지출하거나 복잡한 절차를 거쳐 지원받아야 한다. 장애인이 전자 보조 기기를 이용하면서 느껴지는 거부감을 위 음성 인터페이스로 해소할 수 있으며, PC를 사용하는 데 불편함을 획기적으로 줄일 수 있다.

더 나아가 인터페이스에 예측 알고리즘을 추가하여 사용자의 관심 정보에 따라 검색어를 추천해주거나 필요로 하는 검색 결과가 필요 없는 정보인지 분류해주는 기능 또한 추가할 수 있다. 이런 기능들을 추가함으로 장애인뿐만이 아닌 비장애인도 음성 인터페이스를 사용할 수 있다.

ACKNOWLEDGEMENT

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.NRF-2017R1A2B4008866).

참고문헌

[1] 최효영, 변신철, 박형규, 장문선, 곽호완 (2019). 스마트폰 사용량 조절에 면대면 및 메시지 피드백 개입의 효과. 재활심리연구, 26(3), 153-169
[2] 문종훈, 전민재, 송이슬 (2019). 스마트폰 과사용이 청소년의 건강과 학업에 미치는 영향. 한국엔터테인먼트산업학회논문지, 13(2), 177-186
[3] 오익표, 백아름, 권진아, 박홍진, 손상옥, 최혁진 (2016). 저시력 장애인을 위한 보조기기 개선 방안에 대한 연구. 한국HCI학회 학술대회, 198-205
[4] 김태용, 강정배 (2016). 장애인 보조기기 연구개발 지원방안. 한국재활복지공학회 학술대회 논문집, 6-7
[5] 김정 (2015). 고령자 및 장애인을 위한 보조 기술. 한국정밀공학회지, 32(10), 843-843
[6] 허다경, 이병권, 이소정, 남양지, 권준한, 송병섭 (2015). 청각장애인을 위한 소리정보전달 보조기기 개발. 한국재활복지공학회 학술대회 논문집, 18-20
[7] 김지환 (2019), 딥러닝 기반 음성인식, 정보과학회지, 37(2), 9-15.