|  |
| --- |
| **1. 주제**  lms 강의 자막 생성 및 요약 프로그램  **분반, 팀, 학번, 이름**  나반, 1팀, 20241973, 이성원 |

|  |  |
| --- | --- |
| **2. 요약**  현재 lms에서는 자막 생성 기능을 지원하지 않고 있다. 이로 인해 청각장애인 학생은 학습시 불편함을 겪을 수 있으며, 비장애인 학생들도 학습시 집중력이 떨어질 수 있다는 문제가 존재한다. 따라서 본 제안서에서 제안하는 프로그램은 lms 강의에 실시간 자막과 내용 요약 기능을 제공함으로써, 이와 같은 문제를 해결하는 것을 목표로 한다.  주요 기능으로는 실시간 자막 생성, 전체 자막 파일 제공, 강의 내용 요약 기능이 있으며, 파이썬과 PyTorch를 이용해 구현된다. 자막 생성 기능은 Whisper 모델을 기반으로 한 커스텀 STT 모델을, 강의 내용 요약 기능은 KoBART 모델을 사용해 구현할 예정이다. 또한, VAC를 통해 강의 영상에서 실시간으로 출력되는 소리를 사용함으로써 기존 프로그램의 한계였던 사용자 인증과 영상 다운로드 문제를 해결할 수 있다.  본 프로그램을 통해 청각장애인 학생이 학습시 겪는 문제점을 해결하고, 학생들의 학습 효율을 높일 수 있을 것으로 기대된다. | **3. 대표 그림**  - 커스텀 STT 모델을 통해 자막 생성 및 기록  - 기록된 자막을 바탕으로 텍스트 요약 모델을 통해 강의 내용 요약    그림 1. 실시간 자막 생성 기능    그림 2. 강의 내용 요약 기능 |

|  |
| --- |
| **4. 서론**  현재 숭실대의 lms에선 자막 생성 기능을 지원하지 않고 있다. 2022학년도 2학기, 대학측에서 일부 과목을 대상으로 ‘강의 영상 자막 제공 시범 사업’을 진행했으나, 현재 대부분의 강의에선 자막이 제공되지 않고 있는 상황이다.  이로 인해 생기는 문제점은 크게 두 가지이다. 첫 번째 문제는 청각장애인 학생의 경우 학습에 불편함을 겪을 수 있다는 점이다. 교수님의 수업 방식에 따라 시각적 자료보다 직접 말로써 설명하는 방식이 주가 되는 경우가 있는데, 이 경우 청각장애인 학생은 제공되는 자료만으로 수업의 모든 내용을 파악하기 어려울 수 있다. 두 번째 문제는 강의를 듣는 학생들의 집중력이 떨어질 수 있단 점이다. 온라인 강의는 그 특성상, 교수님 • 타 학생 • 주변 환경 등과 상호작용하며 참여할 수 있는 오프라인 수업에 비해 수강자의 집중도가 떨어질 수밖에 없다. 특히, 선술했던 시각적 자료보다 구두 설명의 비중이 높은 강의의 경우 이러한 문제는 더욱 두드러지며, 자신도 모르는 새에 중요한 내용을 놓칠 가능성이 있다.  이러한 문제점을 해결하기 위해, 본 제안서에서는 세 가지의 주요 기능을 지닌 프로그램을 제안한다. 첫 번째 기능은 실시간 자막 생성 기능으로, 강의 영상 근처에 실시간으로 자막을 생성하여 보여줌으로써 청각장애인의 학습을 돕고, 강의의 전달력을 높여 학생들의 집중력을 향상시킬 수 있을 것이다. 두 번째와 세 번째 기능은 전체 자막 파일 제공 및 강의 내용 요약 기능으로, 강의에서 놓쳤을지도 모르는 중요한 내용의 확인과 전체적인 내용의 복습을 돕는다.  또한, 본 프로그램은 기존의 자막 생성 • 요약 프로그램의 문제점을 해결하였다. 기존 프로그램의 사용 방식은 공개된 영상의 링크를 업로드하는 방식과 영상을 다운로드한 후 프로그램에 업로드하는 방식, 두 가지로 나뉜다. 전자의 경우 사용자 인증이 필요한 사이트, 즉 lms와 같은 사이트에서 사용할 수 없고, 후자의 경우 매번 영상을 다운로드 받아야하기에 번거롭다는 단점이 있다. 따라서 본 프로그램에서는 시스템에서 출력되는 소리를 사용하여 사용자 인증 없이 사용을 가능케 하고, 동영상 옆에 새 창을 띄워 실시간으로 자막을 생성함으로써 영상의 다운로드 없이도 사용이 가능하도록 하고자 한다. |

|  |
| --- |
| **5. 본론**    그림 3. 시스템 개요  프로그램의 전체적인 작동 과정은 다음과 같다. 우선, 사용자가 본 프로그램을 작동시킨 후 lms에서 원하는 강의를 재생하면, 해당 강의의 오디오 출력이 가공되어 커스텀 STT 모델에 입력된다. 이를 통해 생성된 자막은 영상의 근처에 띄운 새 창에 표시될 예정이며, 프로그램에 지속적으로 기록되었다가 영상이 끝난 후 전체 자막 텍스트 파일을 만들 때 이용된다. 전체 자막 파일은 텍스트 요약 모델에 전달되어 요약된 후, 프로그램에 표시된다. 이렇게 만들어진 요약문의 파일과 전체 자막 파일은 프로그램에서 다운로드 받을 수 있다. 전체적인 코드는 파이썬으로 구현할 것이며, 모델의 구현엔 PyTorch를 사용할 예정이다.  자막 생성 기능에는 Whisper 모델을 베이스로 한 커스텀 STT 모델을 사용할 예정이다. Whisper는 OpenAI에서 개발한 ASR(Automatic Speech Recognition) 모델로, 실시간 STT 기능을 제공하며, 다양한 언어의 인식이 가능하다. 또한, 다른 모델에 비해 높은 제로샷 성능을 보이기에 다양한 분야의 강의에 사용해야하는 본 프로그램에 적합하다 판단하여 Whisper 모델을 베이직 아키텍쳐로 선정하였다. 본 프로그램의 경우에는 한국어에 대한 기능만 필요하기 때문에, 커스텀 모델은 한국어에 특화된 경량 모델로써 구축될 것이다. 데이터셋은 AI Hub의 ‘한국어 강의 음성’ 데이터를 이용할 예정이다. 강의 내용 요약 기능은 SKT의 KoBART 모델을 사용할 예정이다. KoBART는 페이스북에서 개발한 모델인 BART(Bidirectional Auto-Regressive Transformer)를 SKT-AI에서 한국어 데이터로 학습시켜 한국어 NLP에 최적화시킨 모델이다. 본 모델을 사용하여 테스트를 진행한 후, 필요성이 느껴진다면 AI Hub의 ‘요약문 및 레포트 생성 데이터’ 등의 데이터셋을 이용해 fine tuning할 예정이다.  자막 생성 기능은 사용자 인증 또는 영상의 다운로드 없이 사용할 수 있도록, 강의 영상에서 실시간으로 출력되는 소리를 사용할 것이며, 이를 위해 VAC(Virtual Audio Cable, 가상 오디오 케이블)를 이용할 예정이다. VAC는 본 프로그램의 사용자의 환경에 설치되어야 하며, 한 번 설치되면 자동으로 재생 장치와 녹음 장치에 가상 케이블이 추가되어 편리하게 사용할 수 있다. 다만, VAC라는 별도의 프로그램을 컴퓨터에 설치해야 한다는 점이 사용자에겐 번거롭게 느껴질 수 있으므로, VAC와 유사한 기능을 사용할 수 있는 다른 방안을 모색하여 가능하다면 추가적인 소프트웨어의 설치 없이 사용할 수 있도록 할 예정이다.  UI는 사용자 친화적인 GUI로 구성될 것이며, 프로그램의 배포 방식을 정한 후 그에 맞춰 제작될 예정이다. 본 프로그램을 웹을 통해 서비스할 경우 HTML • CSS • JavaScript를 통해 UI를 구성하고, 파이썬 기반 웹 프레임워크인 Flask를 사용할 예정이다. 만약 프로그램을 데스크톱 앱으로 만들어 배포할 경우 PyQt5를 이용해 UI를 제작하고, 패키징 도구인 pyinstaller를 이용할 예정이다. |

|  |
| --- |
| **6. 결론**  본 제안서에서는 lms 강의에 실시간 자막을 제공하고, 강의 내용을 요약하는 기능을 통해 청각장애인 학생이 학습시 겪는 문제점을 해결하고, 학생들의 학습 효율을 높일 수 있는 프로그램을 제안하였다. 주요 기능으로는 실시간 자막 생성, 전체 자막 파일 제공, 강의 내용 요약 기능이 있으며, 전체적으로 파이썬과 PyTorch를 통해 구현된다. 실시간 자막 생성 기능에는 Whisper 모델 기반의 커스텀 STT 모델을, 강의 내용 요약 기능에는 KoBART 모델을 사용할 예정이다. 또한, VAC를 통해 강의 영상에서 실시간으로 출력되는 소리를 사용함으로써 사용자 인증이나 영상 다운로드가 필요하던 기존 프로그램의 단점을 해결할 수 있다.  향후, 프로그램의 UI 설계 및 배포 방식을 결정해야 한다. 웹 기반으로 제공할 경우 Flask와 HTML, CSS, JavaScript를 사용할 계획이며, 데스크톱 앱으로 배포할 경우 PyQt5와 pyinstaller를 통해 구현할 예정이다. 또한, 현재 확보한 데이터셋에 대한 Whisper의 성능을 측정하고, Whisper에서 불필요한 기능을 제거한 한국어 특화 경량 모델을 제작하여 같은 데이터셋에 대한 성능을 측정하고 비교하여야 한다. |

**7. 출처**

[1] OpenAI, 「Introducing Whisper」, OpenAI 공식 홈페이지, <https://openai.com/index/whisper/> (2024.10.08)

[2] SKT-AI, 「KoBART」, SKT-AI github 페이지, <https://github.com/SKT-AI/KoBART> (2024.10.12)

[3] Eugene (Evgenii) Muzychenko, 「Virtual Audio Cable (VAC)」, VAC 공식 홈페이지, <https://vac.muzychenko.net/en/> (2024.10.10)