

컴퓨터공학종합설계 서면 보고서

학생 팀별 작성용

과제 정보	
프로젝트명	비대면 환경에서의 효과적인 교육 및 비즈니스를 위한 개인 맞춤형 융합 콘텐츠 생성 기반 기술
팀명	김이홍조
협력 기업명	주식회사 유니크유엑스

프로젝트	
프로젝트 진행 내용	<p>1. 협력 기업 멘토님 측에 질문지 정리 후 전달</p> <p>가. 주제 관련 질문(비대면 환경에서의 교육, 비즈니스와 자막기능 개발의 연관성)</p> <ul style="list-style-type: none">✓ 비대면 환경의 핵심 소통 역량은 영상이며, 이로 교육 및 비즈니스 분야에 연관되어있음. 이러한 영상 자료에 대해 자막을 이용해 다양한 정보를 매칭하고자 함. 결론적으로 자막 기술이 비대면 교육 및 비즈니스에 필요한 기술이라고 생각하고 있음. <p>나. 최종 산출물 관련 질문(최종 목표 및 성능)</p> <ul style="list-style-type: none">✓ HTML5 플레이어에서 사용가능한 오버레이 자막 뷰어 및 자막 에디터✓ 한 영상에 다양한 자막이 보여져야 함 <p>다. 개발 환경 및 기능 질문</p> <p>1) 자막 외 사용자 관리 기능이 필요한가?</p> <ul style="list-style-type: none">✓ 기존 사이트와 연동 예정. 자막 관리기능은 필요 <p>2) 플레이어는 어떤 것을 사용하는가? 이에 따른 지원 동영상/자막 포맷은?</p> <ul style="list-style-type: none">✓ HTML5 표준 웹플레이어, 이에 제공하는 동영상/자막 포맷이면 가능 <p>3) 개발의 기준이 되는 환경은?</p> <ul style="list-style-type: none">✓ 크롬기반 웹(chrome, edge 등) <p>4) 처리해야하는 영상의 범위는?</p> <ul style="list-style-type: none">✓ 실시간, 비실시간 모두 가능✓ 실시간의 경우 임의의 파일명을 주고 자막 생성 >>영상 제작후 자막 일치✓ 비실시간 경우 웹상의 영상, 로컬 영상 모두 적용 <p>2. 구현해야하는 기술 정의</p> <p>가. 자막 생성/삽입 : 사용자가 원하는 시간(지점)에 원하는 텍스트(자막)을 입력</p> <p>나. 자막 저장/불러오기/수정 : 사용자가 생성한 자막 파일을 서버에 저장. (사용자가 가지고 있어도 되는데 원하는 키워드 검색을 위해서는 서버에 저장하는 것이 낫다고 판단) 동일한 영상에 대한 자막들끼리 분류한다거나 어떻게 저장할 것인지 모색 필요. 저장된 자막에 대해 수정도 가능하도록</p> <p>다. 한번에 자막파일 여러 개 불러오기 : 같은 영상에 대해 여러 사용자가 입력한 자막을 한번에 보고 싶은 경우를 고려하여 탭(자막을 영상 옆쪽 별도의 공간에 위치시켜, 탭으로 여러 자막파일을 나눠서 보여주고 사용자가 탭을 선택해서 열람 가능하도록)을 구현하거나, 동영상 화면에 여러 개의 자막을 한 번에 띄우는 방</p>

법 고려

라. 자막에서 원하는 키워드를 검색 : 사용자가 원하는 키워드 (ex. 속제 등등)를 검색하면 해당 키워드를 포함하는 자막들을 보여줌

마. 영상에서 원하는 키워드를 포함하는 부분 재생 : 위의 '라'에서 검색한 자막을 누르면 영상의 해당 부분으로 이동하여 영상 시청 가능

3. HTML5 기반 동영상 플레이어 기술 조사

가. VideoJS

- 1) 오픈소스
- 2) JS 및 CSS를 사용하여 제작됨
- 3) 데스크톱 모바일 지원, 플래시 어느정도 호환 가능
- 4) 플러그인 지원

나. JW player

- 1) 유료
- 2) 반응이 빠르고 사용자 정의가 가능한 HTML5 동영상 제공
- 3) 유튜브의 대체 옵션으로도 사용 가능
- 4) 광고 파트너십이나 소셜 네트워킹 툴 등의 부가 기능 제공

다. Kaltura HTML5 Video Player

- 1) 오픈 소스 (무료)
- 2) 유연성, 사용자 정의, 플러그인 등 매우 용이함
- 3) Flash와의 호환성이 떨어짐.
- 4) 교육용 플랫폼으로 사용이 됨

라. FlowPlayer

- 1) 오픈소스
- 2) 스트리밍, 유튜브 비디오 재생 지원
- 3) JS api 지원
- 4) 사용시 포맷 변환 필요
- 5) 100% 커스터마이징 가능. 스킨 자막 재생속도 조절 가능.

마. Wistia

- 1) 비즈니스용 html5 동영상 플레이어
- 2) 시청자 환경에 맞게 동영상 재생 가능 (접속 속도 등)
- 3) 쉽고 간단한 코드

바. 위안플레이어

- 1) 여러 브라우저 호환 가능.
- 2) 북마크 기능
- 3) 백그라운드 재생 가능
- 4) JS api 제공

4. 자막 관련 기술 조사

가. 자막 파일의 종류 조사

- 1) smi : 자막의 내용만 저장
- 2) srt : 자막의 시작 시간과 끝 시간과 같은 시간 정보가 함께 저장

3) WebVTT : 파일 안에 시간 정보가 같이 저장

나. <track> 태그 조사

- 1) 외부 자막 파일을 연결을 할 수 있는 태그
- 2) 사용법 : <track kind = “자막 종류” src=“파일 경로” srclang=“사용한 언어” label=“제목” default>
- 3) kind : 자막의 종류를 지정. 사용할 수 있는 값들은 다음과 같고 생략할 경우 subtitles로 간주됨 (subtitles, captions, description, chapters, metadata)
- 4) src : 텍스트 트랙 데이터의 주소를 지정
- 5) srclang : 사용한 언어를 지정. en나 ko처럼 표기
- 6) label : 트랙의 제목을 지정
- 7) default : 기본 트랙을 나타낼 것을 지정

다. SST(Speech-To-Text) 기술 조사

- 1) 음성을 텍스트로 변환하는 기술
- 2) 동기 인식 : 오디오 데이터를 SST API로 보내고, 해당 데이터를 인식, 모든 오디오가 처리된 후 결과를 확인, 인식 요청 대상은 길이가 1분 이하인 오디오 데이터로 제한
- 3) 비동기 인식 : 오디오 데이터를 SST API로 보내고, 장기 실행 작업을 시작. 이 작업을 사용하여 주기적으로 인식 결과를 폴링할 수 있음. 최대 480분 길이의 오디오 데이터에 사용.
- 4) 스트리밍 인식 : 양방향 스트림에 제공되는 오디오 데이터를 인식. 스트리밍 요청은 마이크에서 라이브 오디오 캡처 용도와 같은 실시간 인식 용도로 설계됨. 오디오 캡처 중간에 중간 결과를 제공하므로, 사용자가 계속 말하는 중에도 결과를 표시할 수 있음
- 5) 오픈소스

가) Kaldi - <https://github.com/kaldi-asr/kaldi>

나) CMU Sphinx - <https://cmusphinx.github.io/wiki/download/>

다) Mozilla - <https://github.com/mozilla/DeepSpeech>

라) VideoBookmarkingService -

<https://github.com/reemaRaven/VideoBookmarkingService>

라. Google의 Speech to text 조사

- 1) 자동 음성 인식을 위해서 Google의 최고급 딥러닝 신경망 알고리즘을 적용했고, 125개가 넘는 언어 및 방언을 지원하는 음성 인식 기능을 가지고 있다.
- 2) 음성 적응 : 음성 인식 맞춤설정을 통해 힌트 제공으로 분야별 용어 및 많이 쓰이지 않는 단어를 텍스트로 변환하고 특정 단어 또는 어구의 텍스트 변환 정확도를 향상시킬 수 있습니다. 클래스를 사용해서 음성으로 인식된 숫자를 주소, 연도, 통화 등으로 자동 변환할 수 있습니다.
- 3) 분야별 모델 : 분야별 품질 요구사항에 따라 최적화된 음성 제어, 전화 통화, 동영상 텍스트 변환에 맞게 학습 모델 옵션을 선택적으로 사용할 수 있습니다. 예를 들어 Google의 향상된 전화 통화 모델은 8khz 샘플링 레이트로 녹음된 전화 통화와 같이 전화에서 유래된 오디오에 맞게 조정되어 있습니다.

4) 음성 인식 스트리밍 : API가 애플리케이션의 마이크로부터 스트리밍되거나 사전 녹음된 오디오 파일로부터 전송되는(인라인 또는 Cloud Storage 사용) 오디오 입력을 처리할 때 음성 인식 결과를 실시간으로 수신할 수 있습니다.

5) API 다운 주소 : <https://cloud.google.com/speech-to-text?hl=ko>

마. 관련 프로그램 조사

1) Youtube API

가) 동영상 관련 가장 큰 플랫폼으로, 플레이 바 등에 관련된 정보 존재

나) 기본적으로 iframe player API를 사용

다) caption 기능과 search 관련 내용이 존재

2) subtitle edit

가) 자막 생성, 조정, 싱크 조절등에 매우 효과적인 edit 프로그램

나) 무료로 사용이 가능

다) 170개 이상의 자막 포맷을 지원하며, 대략 30개 언어로 이용이 가능

3) SubtitleCreator

가) 동영상 제작용 프로그램이 따로 있어서 영상에 바로 오버레이하면서 편집가능

나) 영상 환경에서 자막의 색상, 동기화 편집등이 가능

4) 북마크 및 단축키 설정 기술 - Video Bookmarks Controller

가) Video progress bar에 북마크 표시 기능

나) 생성된 북마크를 URL로 관리를 할 수 있다.

다) 구간 반복 기능

라) 속도 조절 기능

마) 단축키 기능

바) github - https://github.com/chiubeta/video_bookmarker

5. 특허 기술 조사

가. 북마크 기술 관련 특허

1) 광고 재생에 사용되는 기술, 사용자의 북마크 데이터를 저장소에 저장하는 방식

2) 해당 특허는 광고에 특화되어 있으나 전체적인 분야로 사용이 가능

나. 태그 기술 관련 특허

1) 키워드 검색을 통해 생성된 자막에서 키워드를 검색해서 해당 프레임으로 이동할 하는 기술

2) 해당 특허 흐름도

다. 자막의 자동 생성과 영상과 싱크 조절 특허

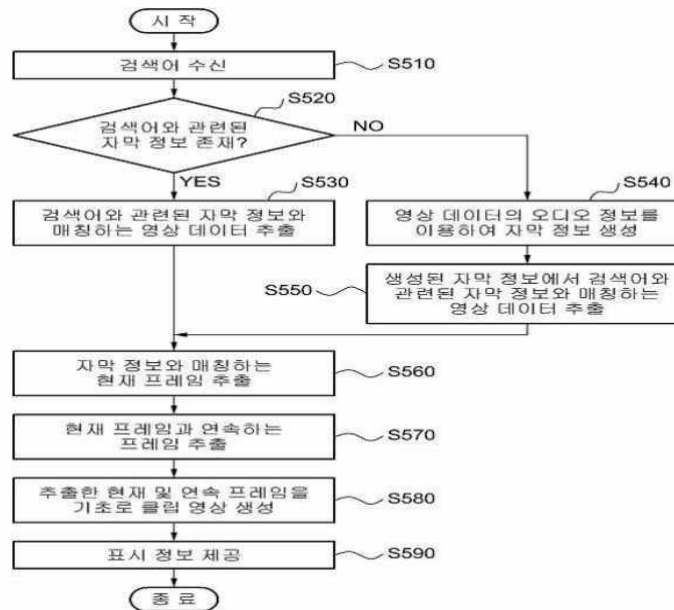
1) 자막과 영상의 싱크를 스미스-워터맨(smith-waterman) 알고리즘을 활용하여 자동으로 조절

2) 흐름도

3) 학습 영상파일에서 영상데이터와 음성데이터, 및 자막데이터를 분리하는 분리부

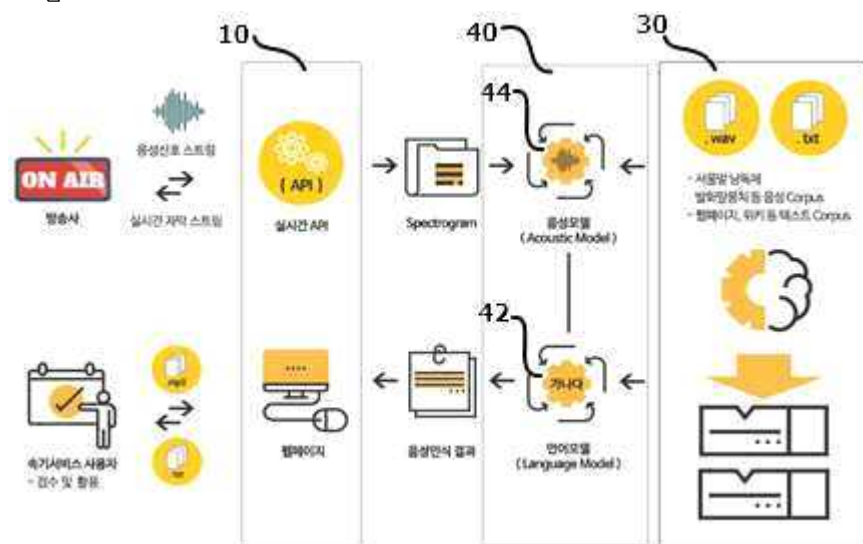
4) 분리부에서 분리된 음성추출부에서 추출된 음성을 텍스트화 하는 텍스트변환부

5) 텍스트변환부에서 생성된 텍스트데이터와 분리부에서 분리된 자막데이터의 시간차이를 분석하는 비교분석부 비교분석부에서 분석된 시간차를 스미스-워터맨(Smith-Waterman) 알고리즘을 기반으로 자막데이터와 텍스트 데이터의 싱크를 조절하는 맞춤조절부



라. 음성인식 엔진을 이용한 실시간 지능형 자막화 서비스 제공 시스템 특허

- 1) 음성 인식 엔진을 기반으로 실시간 음성신호를 인식하고 이를 자막으로 실시간 생성하여 제공하는 음성인식 엔진을 이용한 실시간 지능형 자막화 서비스
- 2) 흐름도



6. 그 외에 자연어 처리, 자막 처리에 관련된 opensource 등

- 가. <https://github.com/scopeInfinity/Video2Description> : 자연어 처리모델, 문장을 원형으로 바꾸어줌, 비디오를 보고 이에 대한 내용을 자막을 생성
- 나. <https://github.com/vijayvee/video-captioning> : 자동자막기능으로 python 사용. Tensorflow를 이용한 자연어 처리 등
- 다. <http://caption.devco.kr/> : VVT 포맷으로 자막 파일을 형성해주는 웹사이트

	<p>라. https://github.com/vietank/video-subtitle : 영상을 보여주면서 영상이 자막에 오버레이 되면서 옆에는 자막 탭이 생성되고 자막을 클릭하면 영상의 해당 지점으로 이동이 가능</p>
회의내용	<p>#선행 기술 조사서 작성</p> <ul style="list-style-type: none"> - 선행 기술 조사서를 작성을 하기 위해서 기술 요소 정의와 같은 기본적인 내용을 먼저 정리 및 역할분담 - 작성을 위해 기술 요소에 따라 조사했던 내용보다 많은 특허 조사가 필요하다 판단 <p>#차주 진행 사항 결정</p> <ul style="list-style-type: none"> - 선행 기술 조사서를 쓰기 위해서 특허 부분에 대한 자료조사 - 자료조사를 바탕으로 선행기술조사서를 작성 - 팀별 발표 준비를 위한 PPT 제작