

---

# Software Requirements Specification (SRS)

작성자: 1821 (20180817\_조수현, 20192940\_강준규  
20201854\_성나영, 20211705\_이원호)  
생성일자: 2023. 04. 07.  
최종 갱신: 2023. 06. 13.  
버전: 1.1.1.

# 목차

목차	
1. 개 요	1
1.1 목적	1
1.2 관련부서/관계자	1
1.3 개발범위	1
2. 전체 개관	2
2.1 프로젝트 개요	2
2.2 Product Function	2
2.3 사용자 특성	3
2.4 소프트웨어 운영 환경	4
2.5 디자인과 배포 제약사항	4
2.6 가정과 의존성	4
2.7 데이터 요구사항	5
2.8 사용 시나리오	6
3. 인터페이스	7
3.1 사용자 인터페이스	7
3.2 소프트웨어 인터페이스	9
4. 기능 요구사항	10
4.1 postURL	10
4.2 postOption	10
4.3 showVideo	11
4.4 downloadSummaryVideo	11
4.5 clear	11
4.6 storageSourceVideo	12
4.7 executeSummaryModel	12
4.8 getOption	13
4.9 whisperX	13
4.10 modifySrt	14
4.11 convertText	15
4.12 summaryScript	16
4.13 cosSimilarity	16
4.14 clipEditor	17
4.15 storageSummaryVideo	17
5. 비기능 요구사항	18
5.1 성능 요구사항	18
5.2 소프트웨어 품질 요구사항	18

## 1. 개요

이 문서는 ChatGPT를 활용한 영상 요약 모델 소프트웨어 요구명세서로서 본 문서를 통해 Video Summary 소프트웨어의 기능을 구현하기 위함이다.

### 1.1 목적

본 문서는 Video Summary의 시스템 관리와 유지 보수를 위해 자세한 요구사항을 기술한 문서이다. 이 요구사항들은 Video Summary 소프트웨어의 설계와 확장의 기초가 된다.

### 1.2 관련부서/관계자

영상 요약 모델에 관심이 있는 사람일 경우, 4장 기능 요구사항에서 어떠한 과정으로 영상이 요약되는지 그 과정을 중점적으로 읽어보는 것을 추천한다.

ChatGPT의 활용성에 관심이 있는 사람일 경우, 3.3장 소프트웨어 인터페이스에서 영상을 요약하는 어떠한 과정에서 ChatGPT가 어떻게 활용되었는지, 그리고 어떠한 특성을 중심으로 사용되었는지 읽어보는 것을 추천한다.

본 모델을 활용해 추가적으로 개발하고자 하는 개발자의 경우, 2.5장 디자인과 배포 제약사항에서 설명하는 Library와 그 Version을 확인하고 2.6장 가정과 의존성에서 사용하는 API를 먼저 확인한 후, 4장의 기능 요구사항에 나와있는 Function들을 참고하는 것이 좋다.

### 1.3 개발범위

#### 1) 목적

해당 소프트웨어는 ChatGPT의 활용성을 확장하는 것을 목적으로 STT와 ChatGPT의 텍스트 요약 기술, NLP 유사도 분석을 이용한 영상 요약 기술을 통해 사용자들이 영상을 쉽게 가공하도록 한다. 이를 통해 영상 편집에 대한 이해가 없어도 짧은 영상을 제작할 수 있도록 돕는다. 또한 자동으로 제공되는 자막을 통해 청각장애인 등 문화적 약자의 영상 접근성을 높인다.

#### 2) 하드웨어 제한 사항

해당 소프트웨어는 Http와 Web를 통해 서비스를 제공하기에 Web에 접근할 수 있도록 Web Browser를 이용할 수 있는 하드웨어가 필수적이며, 온라인 상태에서만 해당 소프트웨어를 이용할 수 있으므로, 네트워크 연결이 가능한 하드웨어가 필수적이다.

#### 3) 특징

Video Summary는 플랫폼 종속적인 영상을 요약하는 것으로 시작하지만, 후에는 플랫폼과 독립된 모든 영상으로 타겟 범위가 확장될 수 있다.

## 2. 전체개관

### 2.1 프로젝트 개요

기존 시스템인 ChatGPT와 여러 API를 활용하여 AI 영상 요약 모델을 구현하고, 인터페이스를 통해 사용자에게 영상 요약 서비스를 제공하는 것이다.

Video Summary는 사용자에게 영상의 링크를 받아 ChatGPT의 도움을 받아 요약 영상을 제공하는 프로그램이다. STT 모델을 활용해 얻은 영상의 스크립트를 ChatGPT를 통해 요약하고, 기존 스크립트와 요약 스크립트의 비교를 통해 영상을 요약하여 이를 제공한다.

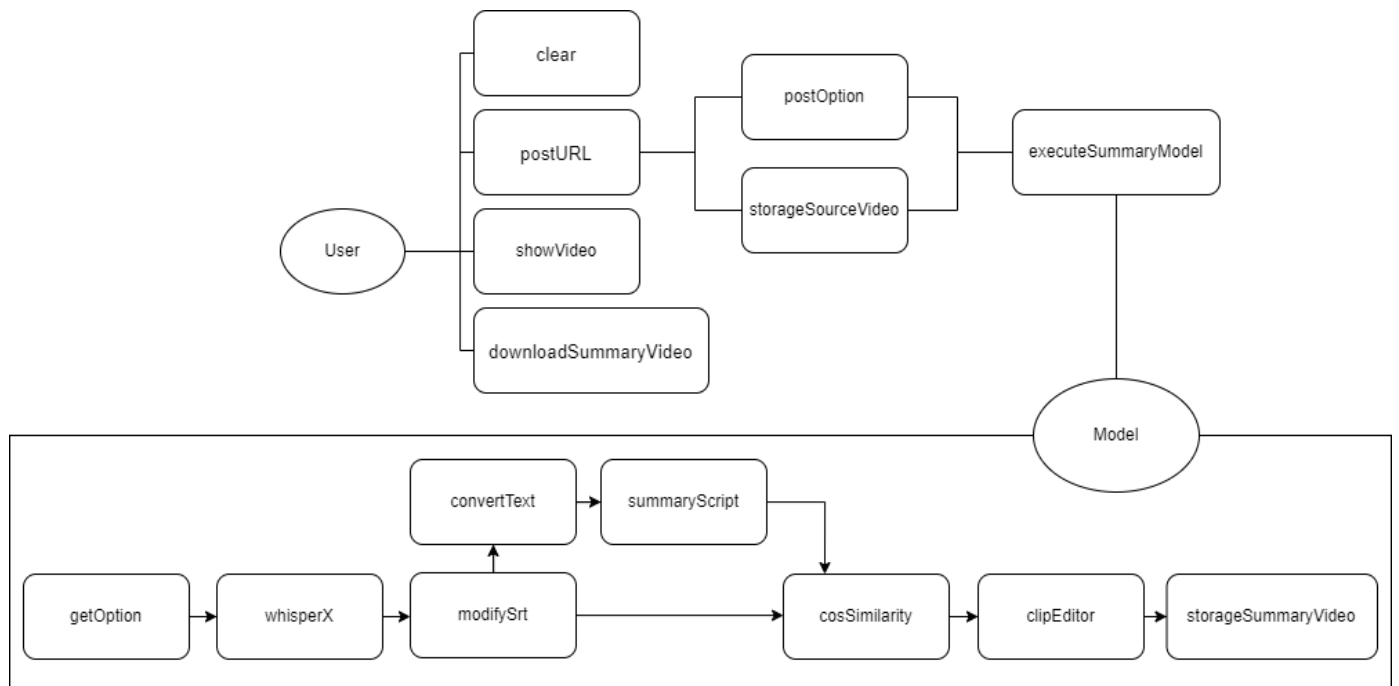
### 2.2 Product Function

#### 1. Function Lists

구분	함수명
Client	postURL
	postOption
	showVideo
	downloadSummaryVideo
	clear
Server	storageSourceVideo
	executeSummaryModel
Model	getOption
	whisperX
	modifySrt

	convertText
	summaryScript
	cosSimilarity
	clipEditor
	storageSummaryVideo

## 2. Function Diagram



## 2.3 사용자 특성

본 소프트웨어는 영상을 요약하고자 하는 모든 사용자들을 대상으로 한다.

## 2.4 소프트웨어 운영환경

본 소프트웨어는 Http 통신과 Web으로 구성된 소프트웨어이고 Linux 환경에서 운영되고 있다. client, server, model 부분으로 나누어져 서비스를 독립적으로 운영하였다. client와 server는 사용자의 접속과 요청을 처리하기 위해 항상 네트워크에 연결된 상태로 유지되어 있다. 또한 Node.js의 프로세스 매니저인 “PM2”를 통해 사용자의 요청과 접속상태를 실시간으로 모니터링하고 있다.

## 2.5 디자인과 배포 제약사항

서버는 Oracle Cloud 서비스를 사용하였으며 Node.js와 Express 웹 프레임워크로 제작되었다. 또한 본 서버의 저장공간은 100GB이다. 서버에서는 Video Summary Model을 실행시키기 위해 python 3버전이 설치되어 있고, FFmpeg 라이브러리가 설치되어 있다.

Video Summary Model에서는 Python을 토대로 ChatGPT와 STT 기술을 사용한다. 이 때, STT 모델로는 WhisperX를 사용한다. 사용하는 라이브러리로는 pypinyin, MoviePy, FFmpeg가 있다.

## 2.6 가정과 의존성

본 소프트웨어에서는 영상의 STT 과정을 위해 WhisperX 2.0 모델을 사용하는데, 이에 따라 Python 라이브러리의 의존성 문제로 각 라이브러리의 Version 관리를 필요로 한다.

<주요 라이브러리 Version>

PyTorch : 1.13.1

torchaudio : 0.13.1

torchvision : 0.14.1

ffmpeg-python : 0.2.0

CUDA : 11.4

<가정>

### 1. 인터넷

온라인 상태가 아닐 경우 기능 수행이 어려움

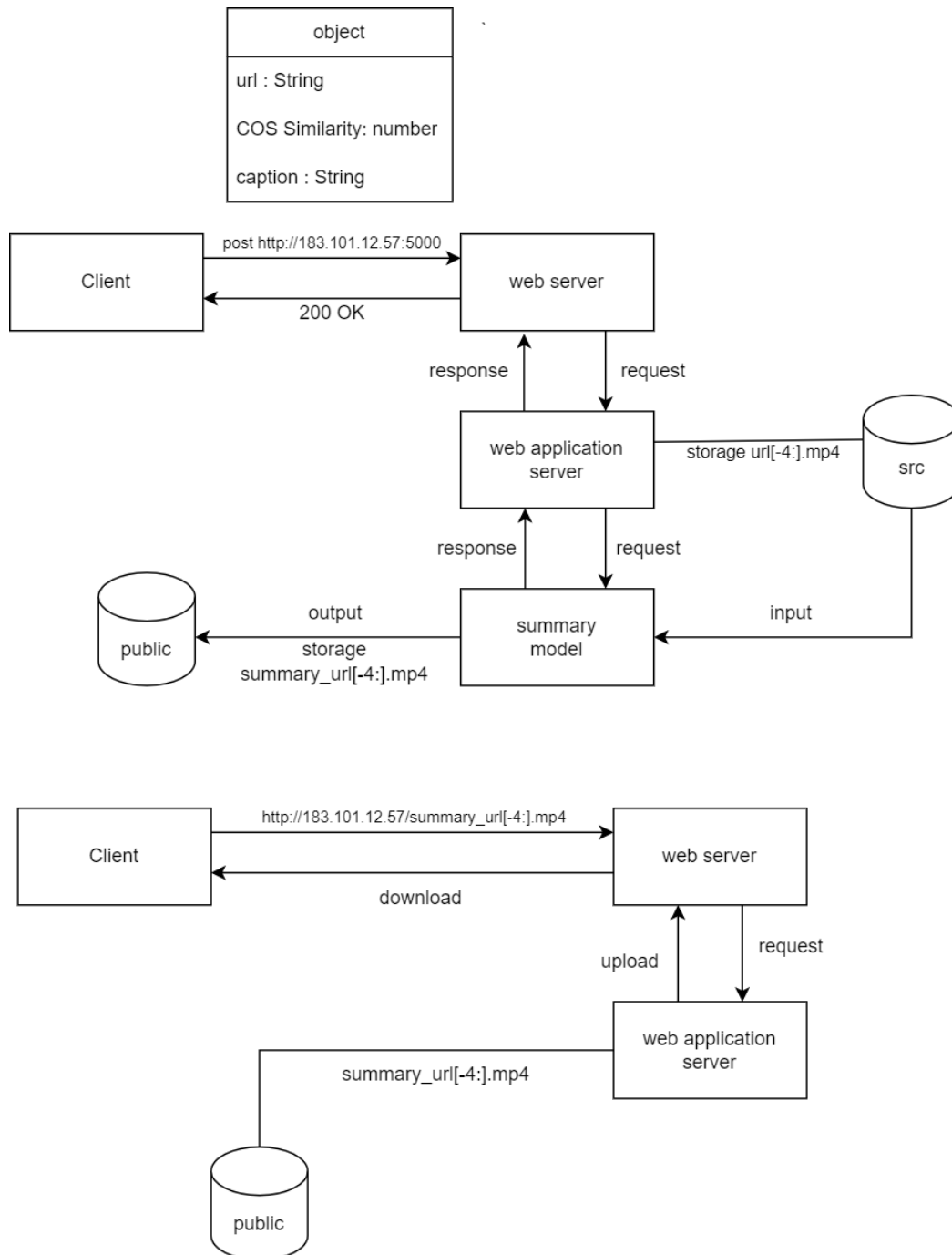
### 2. 의존성

영상 요약 모델은 WhisperX 2.0 모델과 ChatGPT API에 기반을 둬

ChatGPT API의 사용방안에 대해서는 OpenAI 사사의 <https://platform.openai.com/overview>를 참고하는 것이 좋다.

## 2.7 데이터 요구사항

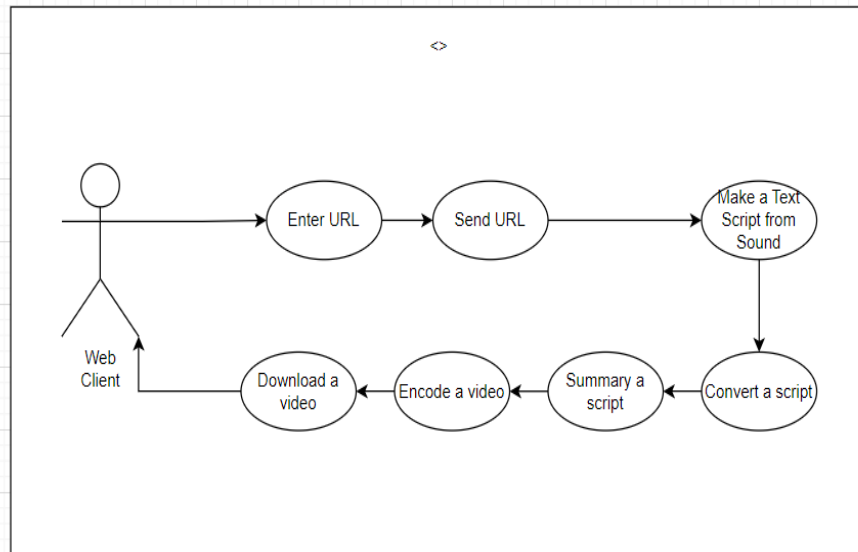
프로세스 순서



## 2.8 사용 시나리오

이 시스템을 설명할 수 있는 사용자 관점에서의 **dramatic** 시나리오.

1. 영상 URL 입력
  - 사용자가 Youtube 영상 URL과, 성능을 위한 옵션을 입력한다.
2. 영상 다운로드
  - 요약된 영상을 사용자가 다운로드 한다.

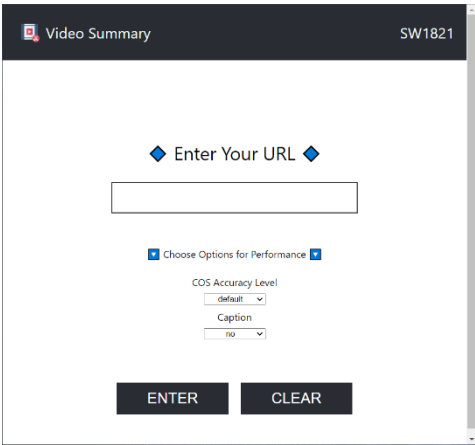




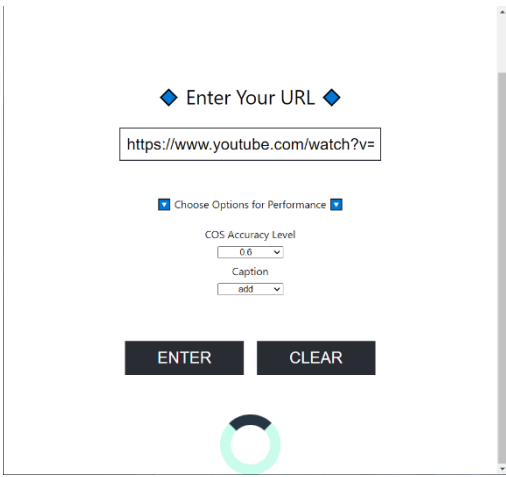
3. 인터페이스

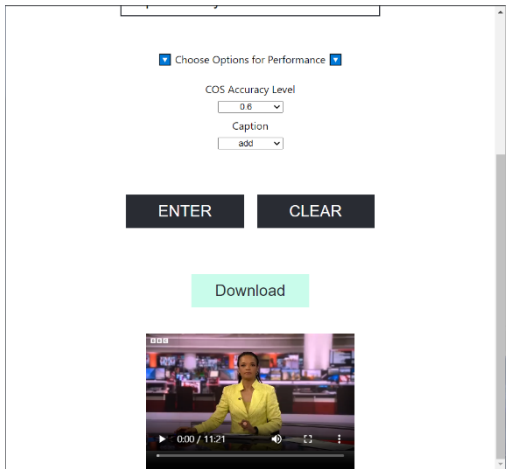
3.1 사용자 인터페이스

1. 화면 명세

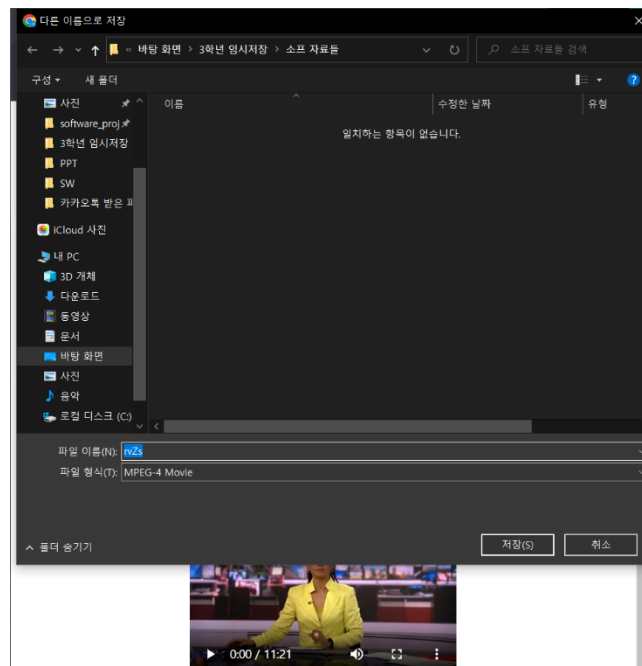
메인 화면	
	
사용되는 Function	postURL, clear

메인 화면	
	
사용되는 Function	postURL, clear, postOption, storageSourceVideo, executeSummaryModel

모델 실행 및 영상 대기 화면	
	
사용되는 Function	whisperX, modifySrt convertText, summaryScript cosSimilarity,clipEditor, storageSummaryVideo

요약 영상 생성 완료 및 영상 다운로드 화면	
	
사용되는 Function	showVideo, downloadSummaryVideo

## 요약 영상 생성 완료 및 영상 다운로드 화면



사용되는 Function

showVideo, downloadSummaryVideo

## 3.2 소프트웨어 인터페이스

다른 소프트웨어 컴포넌트에서 이 소프트웨어로 출입하는 데이터나 메시지 정의 API :

## &lt;whisperX&gt;

- whisperX를 통해 영상에 STT를 적용하여 srt 파일을 얻기 위해서는 터미널 명령어를 입력해야 한다. 명령어는 아래와 같다. 명령어의 인자로는 요약하고자 하는 영상이 저장된 경로가 입력된다.

```
whisperx 'path_originVideo'
```

## &lt;ChatGPT&gt;

- ChatGPT에는 whisperX와 modifySrt, converText를 거친 originScript.txt가 input으로 들어간다. 텍스트를 입력하면 기존에 입력된 "Summarize these text."와 합쳐져 ChatGPT에 Query로 들어간다.

---

## 4. 기능 요구사항

---

### 4.1 <postURL>

---

#### Purpose

*URL*을 *Server*로 보내기 위함이다.

---

#### inputs

*Youtube URL*

Youtube URL이 아니면 오류 메시지를 출력한다.

---

#### outputs

*Youtube URL*에 해당하는 영상 *Data*

Type : .mp4

Destination : /server/src

---

### 4.2 <postOption>

---

#### Purpose

사용자가 입력한 *option*을 *Model*로 보내기 위함이다.

---

#### inputs

*cosSimilarity, caption*

---

#### outputs

*Response message = 200(OK)*

### 4.3 <showVideo>

---

#### Purpose

요약 영상을 웹 클라이언트에게 제공하기 위함이다.

---

#### inputs

요약 영상 *data* 주소

영상이 없다면 웹 클라이언트에게 영상을 표현해주지 않는다.

---

#### outputs

요약 영상 *Data* 화면 표시

### 4.4 <downloadSummaryVideo>

---

#### Purpose

사용자가 요약 영상을 저장할 수 있도록 하기 위함이다.

---

#### inputs

요약 영상 *Data* 주소

영상이 없다면 다운로드 버튼이 활성화되지 않는다.

---

#### outputs

요약 영상 *Data*

### 4.5 <clear>

---

#### Purpose

사용자가 입력한 내용을 *Reset*하기 위함이다.

---

#### inputs

*button click event*

**outputs**

---

-

---

**4.6 <storageSourceVideo>****Purpose**

---

*Client* 가 요청한 영상을 *src* 폴더에 저장하기 위함이다.

**inputs**

---

*Youtube URL*

URL의 경로가 잘못된 경우 Client로 Response message = 404 (Not Found) 전송

**outputs**

---

*Response message = 200(OK)*

---

**4.7 <executeSummaryModel>****Purpose**

---

*Server*에서 *AI* 영상 요약모델에 소스를 제공하고 모델을 작동시키기 위함이다.

**inputs**

---

*url , cosSimilarity, caption*

**outputs**

---

*AI* 영상 요약모델의 결과인 *summaryVideo\_<video\_name>.mp4*

Destination : /server/public

## 4.8 <getOption>

### Purpose

*Client*가 보낸 *option*을 받아 *Model*에 적용하기 위함이다.

### inputs

*cosSimilarity, caption*

argv로 1번째 인자로 cosSimilarity(0.1, 0.2, 0.3)를, 2번째 인자로 caption(add, no)을 입력받는다. 모두 string 형태로 받아온 후 cosSimilarity는 float 형태로 Casting 한다.

### outputs

*opt, <video\_name>.mp4*

opt와 script 전역변수를 설정하고 cosSimilarity는 opt로, caption은 script로 저장한다. script가 "add"일 시, 원본 영상에 자막을 추가하고 이를 <video\_name>.mp4로 대체한다.

## 4.9 <whisperX>

### Purpose

STT 모델인 *whisperX*에 영상을 *Input*하여 타임스탬프와 자막이 담긴 *.srt* 파일을 추출해내기 위함이다.

### inputs

*Server*에서 *src* 폴더에 저장한 *mp4 Video*(의 경로)

### outputs

*<video\_name>.srt*

자막 파일은 원본 비디오 파일명과 자막파일 확장자의 이름을 가진다. Line 별로 Index가 생기고 Index마다 start, end, text의 속성을 가진다. pysubs2 라이브러리로 srt 파일을 로딩 시, 이를 확인할 수 있으며 또한 해당 속성으로의 접근이 가능하다.

Ex)

```
4
00:00:33,063 --> 00:00:38,246
The journal says the White House needs to expel the Russian ambassador and Russian
journalist

5
00:00:38,286 --> 00:00:42,129
from the U.S., something the president says he has no immediate plans to do.
```

예시에서 보는 것처럼, `whisperX`가 추출한 `srt` 파일은 문장 구분이 모두 마침표 단위인 것은 아니다. 따라서 “`modifySrt`” Function을 통해서 이 구분을 전부 마침표 단위로 재구성한다.

---

## 4.10 <modifySrt>

---

### Purpose

“*whisperX*” Function의 Output인 *Video.srt* 파일의 문장 구분을 마침표 단위로 재구성하기 위함이다.

해당 Function의 Output은 추후 ChatGPT의 요약과 한 줄 단위의 코사인 유사도 분석을 진행하기 위한 Input으로 사용된다.

---

### inputs

<video\_name>.srt

`whisperX` 함수를 통해 추출한 `srt` 파일을, `pysubs2` 라이브러리를 통해 Index 단위로 가공한다.

---

### outputs

<video\_name>.srt

마침표 단위로 재구성된 `srt` 파일을 얻는다.

Ex)

```
4
00:00:33,063 --> 00:00:42,129
The journal says the White House needs to expel the Russian ambassador and Russian
journalist from the U.S., something the president says he has no immediate plans to do.

5
00:00:42,389 --> 00:00:49,453
But leaders of more than 30 news organizations around the world signed a letter to
Russia demanding his release.
```



## 4.11 <convertText>

### Purpose

영상에서 추출한 <video\_name>.srt 파일에서 자막에 해당하는 텍스트만 분리하여 *origin\_script.txt* 파일로 만들어내기 위함이다.

해당 Function의 Output은 ChatGPT에 Input으로 전달되어 ChatGPT로 하여금 영상의 스크립트를 요약한다. 또한 요약된 스크립트와 유사도 분석할 대상이 되기도 한다.

### inputs

<video\_name>.srt

마침표 단위로 재구성된 srt 파일이 Input이 된다.

### outputs

*originScript\_<video\_name>.txt*

영상의 자막만이 모아져 있는 *originScript*를 얻을 수 있다.  
*originScript\_<video\_name>.txt* 파일은 추후 “summaryScript” Function을 통해 ChatGPT에 전달된다.

Ex)

```

1 The Wall Street Journal's editor-in-chief is vowing to do everything in their power to secure the release of Evan Gershkovich.
2 The journalist who is being held in Russia on spy charges. This, as President Biden demands Gershkovich's release.
3 That's not the plan right now.
4 The journal says the White House needs to expel the Russian ambassador and Russian journalist from the U.S., something the president says
5 But leaders of more than 30 news organizations around the world signed a letter to Russia demanding his release.
6 The Kremlin says a credited foreign journalist can work safely in Russia.
7 Let's bring in Bill Browder.
8 He's the CEO of Hermitage Capital Management, and he's also the author of two books, A Red Notice and, more recently, Freezing Order, based
9 All right, so good to see you again.
10 So can you help people understand why, this time, this American journalist is targeted?
11 It's pretty straightforward that this is hostage-taking, plain and simple.
12 Putin wants to have a high-value American hostage by taking a Wall Street Journal reporter.
13 He accomplishes that.
14 Why does he want a hostage?
```

## 4.12 <summaryScript>

### Purpose

*ChatGPT*를 활용하여 영상의 자막을 요약하기 위함이다. 사용자가 선택한 옵션에 따라 다른 방법으로 요약을 처리한다.

### inputs

*originScript\_<video\_name>.txt*

### outputs

*summaryScript\_<video\_name>.txt*

ChatGPT가 요약한 텍스트 파일을 Output으로 받는다. 이후 *originScript\_<video\_name>.txt* 파일과 유사도 분석을 할 때 활용된다.

Ex)

```
1 The editor-in-chief of The Wall Street Journal and President Biden are demanding the release of journalist Evan Ge
2 The Journal is calling for the White House to expel the Russian ambassador and journalist from the U.S., but Presi
3 More than 30 news organizations signed a letter demanding Gershkovich's release.
4 Bill Browder, CEO of Hermitage Capital Management, believes that Putin wants a high-value American hostage as a tr
5 He sees Gershkovich as a more valuable hostage than Brittany Griner, who was previously taken and released.
6 Browder advises Western journalists in Russia to leave, as Putin is not playing by any set of rules.
```

## 4.13 <cosSimilarity>

### Purpose

*ChatGPT*의 요약본 *summaryScript*와 *originScript*의 코사인 유사도 분석을 진행하기 위함이다.

*originScript*의 각 문장 별로 *summaryScript*와 유사도 분석을 진행한다. 각 *originScript*의 문장마다 코사인 유사도 분석의 결과치가 제공되는데, *getOption* 함수의 Output인 *opt*에 따라 결과치에 Threshold를 다르게 적용한다. Threshold를 넘은 문장의 Index를 출력한다.

### inputs

*summaryScript\_<video\_name>.txt, originScript\_<video\_name>.txt, opt*

*opt* : *getOption* 함수로 저장된 전역변수(cosSimilarity)

### outputs

*originScript\_<video\_name>.txt*에서 *opt*에 해당하는 Threshold를 넘긴 문장의 Index

## 4.14 <clipEditor>

### Purpose

유사도 분석에서 *Threshold*를 넘긴 문장에 해당하는 *Index*에 따라 영상을 편집하기 위함이다.

원본 영상에서 *Index*에 해당하는 *Subclip*을 저장하기 위해 *MoviePy* 라이브러리를 활용한다.

### inputs

*<video\_name>.mp4*, *<video\_name>.srt*, *cosSimilarity*의 *output Index*

*<video\_name>.srt*에서 전달받은 *Index*의 타임스탬프를 가져와 원본 영상의 *Subclip*을 만든다.

### outputs

*clip\_<video\_name>.mp4(Subclip)*들, *input\_<video\_name>.txt*

*FFmpeg* 라이브러리를 이용해 *subclip*들을 하나의 영상으로 연결하기 위해서 *Subclip*의 이름이 담긴 *input\_<video\_name>.txt*를 생성한다.

## 4.15 <storageSummaryVideo>

### Purpose

여러 *Subclip*들을 편집하여 하나의 영상으로 연결하고, 이를 *public* 폴더에 저장하기 위함이다.

*Subclip*들을 연결하기 위해서 *FFmpeg* 라이브러리를 활용한다.

### inputs

*input\_<video\_name>.txt*, *clip\_<video\_name>.mp4(Subclip)*들

### outputs

*summaryVideo\_<video\_name>.mp4*

결과물인 요약 영상이 완성되면 *public* 폴더에 저장하고 사용된 *Input*을 전부 삭제한다.

Destination : /client/public

## 5. 비 기능 요구사항

---

### 5.1 성능 요구사항

동시에 여러명이 접속했을 시,

1. URL 입력에 대한 server 반응 시간 : 약 3초
2. 영상 처리 시간 : 영상 길이에 종속적임 (평균 약 3 ~ 5분)

### 5.2 소프트웨어 품질 요구사항

기능성	3 / 5
신뢰성	4 / 5
사용성	5 / 5
효율성	3 / 5
유지보수성	-
이식성	5 / 5