



국민대학교  
전자정보통신대학  
컴퓨터공학부

# 캡스톤 디자인 I

## 종합설계 프로젝트


프로젝트 명	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스
팀 명	Summer
문서 제목	결과보고서

Version	1.4
Date	2019-05-28

팀원	김 기성 (조장)
	김 윤성
	양 성호
	정 경진
	정 예원

### CONFIDENTIALITY/SECURITY WARNING


이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인 수강 학생 중 프로젝트 “음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스”를 수행하는 팀 “Summer”의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 “Summer”의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

## 문서 정보 / 수정 내역


<b>Filename</b>	결과보고서-음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스.doc
<b>원안작성자</b>	정예원
<b>수정작업자</b>	김기성, 김윤성, 양성호, 정경진, 정예원

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2019-05-25	정예원	1.0	최초 작성	수행결과보고서 초안 작성
2019-05-25	정경진	1.1	개요 수정	수정된 목표 및 어긋난 문맥 정리
2019-05-26	김기성	1.2	개발 내용 수정	시스템 구조 및 설계도 다이어그램 추가
2019-05-27	김윤성	1.3	자기평가 수정	추가된 개발 내용 및 기술 설명 추가.
2019-05-28	양성호	1.4	문서 포맷 수정	각 문서 내용별 포맷 수정 및 통일

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

## 목 차

1	개요	4
1.1	프로젝트 개요	4
1.2	추진 배경 및 필요성	4
2	개발 내용 및 결과물	5
2.1	목표	5
2.2	연구/개발 내용 및 결과물	5
2.2.1	연구/개발 내용	5
2.2.2	시스템 기능 요구사항	6
2.2.3	시스템 비기능 요구사항	6
2.2.4	시스템 구조 및 설계도	7
2.2.5	활용 / 개발된 기술	8
2.2.6	현실적 제한 요소 및 해결 방안	11
2.2.7	결과물 목록	11
2.3	기대효과 및 활용방안	12
3	자기평가	14
4	참고 문헌	15
5	부록	16
5.1	사용자 매뉴얼	16
5.2	운영자 매뉴얼	16
5.3	테스트 케이스	18

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>	
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스
	<b>팀 명</b>	Summer
	Confidential Restricted	Version 1.4
		2019-May-28

# 1 개요

## 1.1 프로젝트 개요

TED\*와 같은 영어 강연에서 전문적이고 긴 내용을 사람들이 한번에 이해하기는 어렵다. 이러한 강연은 길이가 길어서 집중하는데 한계가 있고 중요한 부분을 놓치게 된다. 따라서 강연을 다시 보거나 음성을 녹음하는 등의 방법으로 내용을 다시 이해하려 한다. 이러한 과정에서 착안하여 음성 데이터를 텍스트로 변환한 후 중요 문장 단위로 추출하여 제공해주는 프로젝트를 계획하였다.

*\* TED는 기술, 오락, 디자인 등과 관련된 강연회이다.*


이 프로젝트의 전체적인 구조 및 개발 내용은 크게 3가지로 나뉜다.

- 음성인식을 위해 외부 API(Google Speech-to-Text API)를 사용하여 음성데이터를 텍스트로 변환한 후 그 데이터를 서버에 저장한다.
- 서버에 저장된 텍스트 데이터를 요약 알고리즘을 적용하여 핵심 문장으로만 요약되도록 데이터를 가공한다.
- 가공된 데이터를 다시 가져와서 사용자에게 제공한다.

## 1.2 추진 배경 및 필요성

### 1.2.1 개발하게 된 계기

프로젝트 개요에서 언급했듯이 긴 음성 데이터를 듣고 핵심 문장과 주제를 한번에 파악하기는 어렵다. 따라서 강연을 다시 보거나 녹화, 녹음하는 방법으로 음성 데이터를 백업하여 내용과 핵심을 파악하려고 한다. 게다가 강연을 요약하는 경우 이러한 과정을 여러 번 반복해야한다. 이번거로운 과정을 자동화해보자는 아이디어에서 본 프로젝트를 시작하였다. 본 프로젝트는 사용자에게 음성파일 또는 영상 URL을 불러오는 간단한 작업만으로 요약문을 제공하여 편리함을 더한다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>	
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스
	<b>팀 명</b>	Summer
	Confidential Restricted	Version 1.4
		2019-May-28

## 2 개발 내용 및 결과물

### 2.1 목표

본 프로젝트는 TED강연과 같은 영어 강연을 요약하여 사용자에게 제공하는 웹 서비스이다. 긴 영상 및 음성에는 불필요한 문장이 많고 내용을 이해하기 어렵기 때문에 그 데이터에 다시 접근하기 힘들다. 그래서 이러한 영상 및 음성에서 핵심 내용이 무엇인지 궁금한 사람들에게 음성 파일에서 요약본, 키워드를 뽑아내어 내용의 핵심을 쉽게 이해할 수 있도록 하는 것을 목표로 한다.

### 2.2 연구/개발 내용 및 결과물


#### 2.2.1 연구/개발 내용

영어 강연에서 추출된 녹음 파일 혹은 영상의 URL을 API로 처리하면, 음성 파일이 텍스트로 변환되어 Transcript 속성에 종속된 값의 형태를 가진 txt파일로 저장되는데, 이 txt파일은 음성 파일에서 추출된 모든 문장을 포함한다. 이때, API의 타임스탬프(Time Stamp)통해 각 Transcription의 출현 시간 정보를 Json 파일에 저장한다.

API의 정확도를 높이기 위한 노력으로 데이터 로깅을 이용한 고급모델을 사용하였고, 입력 음성파일에 맞는 phrase를 작성하였다. 예를들어 STT가 인식한 문장이 *transcript: "it's a swallow whole day"*. 인데 원 문장이 *transcript: "it's a shwazil hoful day"*. 라면 사용자가 phrase를 통해 hint로 ["hoful", "shwazil"] 이나 ["shwazil hopeful day"]를 제공해 원 문장으로 인식할 수 있게끔 할 수 있다.

요약 알고리즘을 적용시키기에 앞서 정확성 향상을 위해 대명사 처리(coreference resolution)를 해줘야 한다. pipeline 구조를 가진 spaCy 패키지를 이용해 영어 nlp 모델을 불러온 뒤, neural network를 이용한 머신러닝 패키지인 neural coref를 pipe에 추가해 coreference resolution을 적용시킨다.

요약하는 과정에서는 TF-IDF, TextRank 알고리즘을 사용한다. TF-IDF 알고리즘으로 문장 내에서 단어의 상대적인 가중치를 구한다. Python의 머신러닝 패키지인 Scikit-Learn을 이용해 TF-IDF 모델링을 수행한다. 그 결과 Sentence-Term Matrix가 생성된다. Matrix의 전치 행렬을

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>	
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스
	<b>팀 명</b>	Summer
	Confidential Restricted	Version 1.4
		2019-May-28

구하여 Correlation Matrix를 생성한다. 이것을 통해 문장(or단어)간 그래프로 나타낼 수 있다. 이 그래프를 입력 받아 TextRank 알고리즘을 적용한다.

TextRank 알고리즘은 Text에 관해 Google PageRank 알고리즘을 활용한 알고리즘이다. PageRank에서 중요한 웹 사이트는 다른 웹 사이트로부터 링크를 많이 받는다는 점 에서 착안해 Page의 랭킹을 매겨준다. TextRank에서는 웹 사이트를 문장(or 단어)으로 치환해 사용한다. TextRank 값이 높은 순서대로 정렬한 뒤 요약할 문장(or 단어)의 개수만큼 출력해준다.


추출된 각 문서의 중요한 키워드를 리스트의 형태로 DB에 저장한다. 중요한 키워드는 제목 옆에 키워드태그(Keyword tag)의 형태로 표시해준다. 사용자는 키워드 태그를 보고 음성 파일의 주제를 대략적으로 파악할 수 있다. 요약된 텍스트와 음성 파일의 스크립트를 웹에서 볼 수 있다.

## 2.2.2 시스템 기능 요구사항

- 로컬에 있는 음성파일을 업로드 한다. (변경)  
⇒ YouTube 영상의 URL을 입력으로 받아 음성 파일을 업로드 할 수 있는 기능을 추가하였다.
- 핵심내용을 요약할때 생성 요약을 한다. (변경)  
⇒ 핵심 내용을 요약할때 생성 요약이 아닌 추출 요약으로 변경하였다.
- 타임스탬프 기능 - 타임스탬프 정보를 이용하여 사용자는 원하는 단어를 검색하여, 그 단어가 포함된 파일과 출현시간을 알아낼 수 있다. (변경)  
⇒ 원하는 단어를 검색하는 대신 전체 overall view에서 Transcript 마다 타임스탬프로 명시한다.
- 키워드 태그(keyword tag) 기능 - 음성파일에서 추출된 중요한 키워드 3개를 의미하며, 사용자는 이 키워드를 확인한다. (완료)

## 2.2.3 시스템 비기능(품질) 요구사항

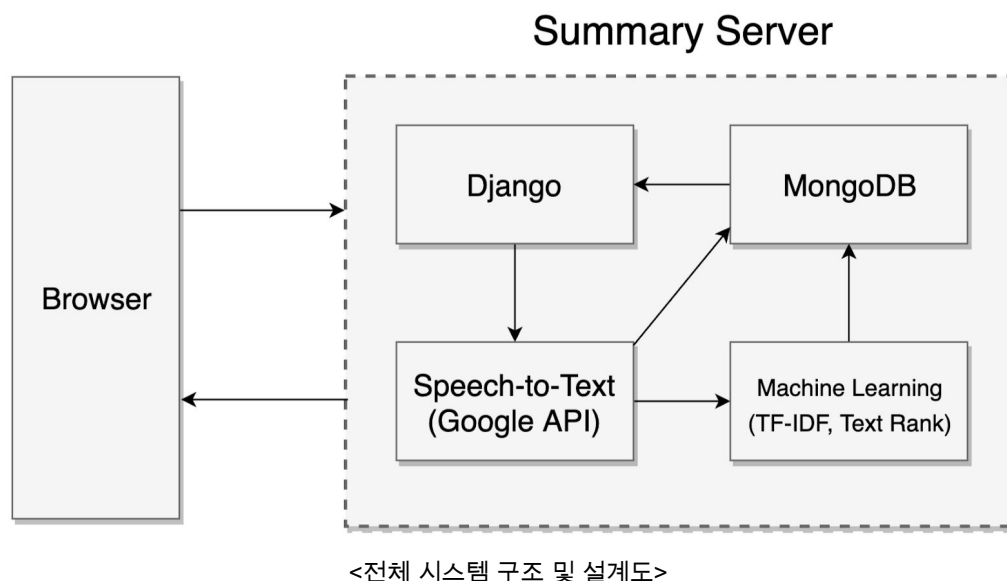
- 음성파일을 입력 받고 핵심내용추출까지 걸리는 시간이 음성 파일 길이의 30% 이내여야 한다. (미달성)  
: 실행시간보다 인식률이 더 중요하다고 판단하여 인식률을 실행시간보다 높은 우선순위에

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28


두었다. 결과물의 품질을 높이기 위해 데이터 로깅 사용을 설정한 후에 음성인식을 위한 특수 모델조합을 사용하였다.

- 입력으로 들어오는 음성 파일의 길이가 480분 이내여야 한다. (미달성)  
: 사용자가 480분 이상의 음성파일을 불러올 경우 파일이 업로드되지 못하도록 하고 경고문을 띄운다.
- DB에 저장된 키워드를 불러와서 글꼴 및 테두리를 강조표시한다. (달성)
- 음성파일의 확장자를 제한없이 입력받는다. (미달성)  
: Speech-to-Text API가 입력받을 수 있는 음성파일의 확장자가 제한적이며 저음질에서 고음질로 변환하는데에 기술적인 한계가 있었다. flac 파일의 음성 품질이 가장 우수해서 flac 확장자의 음성파일만 입력받는 것으로 수정되었다.
- 한국어 강의에 대한 요약을 제공한다. (미달성)  
: Speech-to-Text API가 인식하는 한국어의 정확도가 미흡하여 영어에 대한 요약만을 제공하는 것으로 프로젝트가 수정되었다.

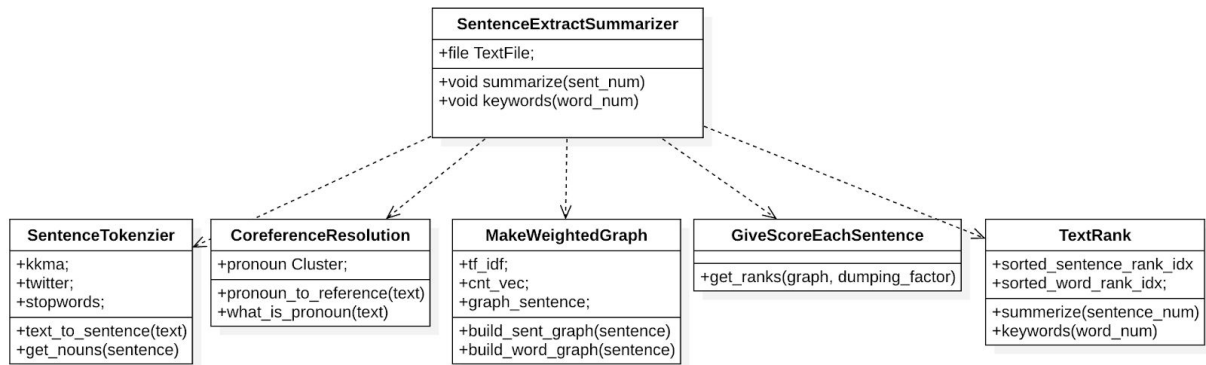
## 2.2.4 시스템 구조 및 설계도



- 브라우저에서 음성 데이터를 업로드하면 Server에서 Google Speech-to-Text API를 호출한다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

- API를 통해 변환된 텍스트에 요약 알고리즘을 적용한다.
- 요약된 텍스트를 Browser에 출력한다.



<요약 알고리즘 오브젝트 다이어그램>

## 2.2.5 활용/개발된 기술


### 1. Google Speech-to-Text API

Google Speech-to-Text는 머신러닝 기반 음성 텍스트 변환 기술이다. 오디오를 텍스트로 변환하는 작업을 할 수 있다.

Google은 데이터 로깅을 통해 수집된 데이터를 기반으로 이러한 고급 모델을 만들고 개선한다. 데이터 로깅을 선택하면 음성 인식을 위한 특수 머신러닝 모델 조합을 사용할 수 있다. 사용자는 오디오 파일 스크립트 작성 시 Speech-to-Text가 고품질 결과를 만들 수 있는 고급 모델을 사용하도록 요청할 수 있다.

### 2. Text-Rank 알고리즘



 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	팀 명	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

$$TR(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} TR(V_j)$$

- $TR(V_i)$ : 문장 또는 단어( $V$ ) $i$ 에 대한 TextRank 값
- $w_{ij}$ : 문장 또는 단어  $i$  와  $j$  사이의 가중치
- $d$ : *damping factor*, PageRank에서 웹 서핑을 하는 사람이 해당 페이지를 만족하지 못하고 다른 페이지로 이동하는 확률로써, TextRank에서도 그 값을 그대로 사용(0.85로 설정)

TextRank는 Text에 대해 Google의 PageRank를 활용한 알고리즘이다. PageRank에서 중요한 웹 사이트는 다른 웹 사이트로부터 링크를 많이 받는다는 점에서 착안해 Page의 랭킹을 매겨준다. TextRank에선 웹 사이트를 문장(or 단어)으로 치환해 사용한다. TextRank값이 높은 순서대로 정렬한 뒤 요약할 문장(or 단어)의 개수만큼 출력해 준다.

### 3. TF-IDF


$$tfidf(t, d, D) = tf(t, d) \times idf(t, D)$$

TF-IDF(Term Frequency - Inverse Document Frequency)는 정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치이다. 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정하거나, 문서들 사이의 비슷한 정도를 구하는 등의 용도로 사용할 수 있다.

- 불린 빈도:  $tf(t, d) = 1$  if  $d$ 에 한 번이라도 나타나면 1, 아니면 0;
- 로그 스케일 빈도:  $tf(t, d) = \log(f(t, d) + 1)$ ;
- 증가 빈도: 최빈 단어를 분모로 target 단어의 TF를 나눈 값으로, 일반적으로는 문서의 길이가 상대적으로 길 경우, 단어 빈도 값을 조절하기 위해 사용한다.

$$tf(t, d) = 0.5 + \frac{0.5 \times f(t, d)}{\max\{f(w, d) : w \in d\}}$$

TF(단어 빈도, Term Frequency)는 특정한 단어가 문서 내에 얼마나 자주 등장하는지를 나타내는 값으로, 이 값이 높을수록 문서에서 중요하다고 생각할 수 있다. 하지만 단어 자체가 문서군 내에서 자주 사용되는 경우, 이것은 그 단어가 흔하게 등장한다는 것을 의미한다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

이것을 DF(문서 빈도, document frequency)라고 하며, 이 값의 역수를 IDF(역문서 빈도, inverse document frequency)라고 한다. TF-IDF는 TF와 IDF를 곱한 값이다.

**역문서 빈도**는 한 단어가 문서 집합 전체에서 얼마나 공통적으로 나타나는지를 나타내는 값이다. 전체 문서의 수를 해당 단어를 포함한 문서의 수로 나눈 뒤 로그를 취하여 얻을 수 있다.

$$\text{idf}(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$$

- $|D|$ : 문서 집합  $D$ 의 크기, 또는 전체 문서의 수
- $|\{d \in D : t \in d\}|$ : 단어  $t$ 가 포함된 문서의 수.(즉,  $\text{tf}(t, d) \neq 0$ ). 단어가 전체 **말뭉치** 안에 존재하지 않을 경우 이는 분모가 0이 되는 결과를 가져온다. 이를 방지하기 위해  $1 + |\{d \in D : t \in d\}|$ 로 쓰는 것이 일반적이다.

IDF 값은 문서군의 성격에 따라 결정된다. 예를 들어 '원자'라는 낱말은 일반적인 문서들 사이에서는 잘 나오지 않기 때문에 IDF 값이 높아지고 문서의 핵심어가 될 수 있지만, 원자에 대한 문서를 모아놓은 문서군의 경우 이 낱말은 상투어가 되어 각 문서들을 세분화하여 구분할 수 있는 다른 낱말들이 높은 가중치를 얻게 된다.

#### 4. Coreference resolution


언어학에서 'coreference'는 한 텍스트에서 두 개 이상의 표현이 같은 사람이나 사물을 지칭할때 발생한다. 'coreference resolution'은 텍스트에서 동일한 실체를 가리키는 대상을 클러스터링하는 작업이다. 또한 텍스트에서 올바른 해석을 도출하거나 다양하게 언급된 주제에서 상대적 중요성을 평가하기 위해, 대명사 및 기타 참조 표현이 올바른 대상과 언급되어야 한다. 핵심 문제를 해결하기위한 알고리즘은 참조 표현식과 호환되는 가장 가까운 선행 개체를 먼저 찾는다.

예를 들면, 아래 문장에서 "I", "my", "she"는 같은 클러스터에 속하며, "Obama"와 "he"는 같은 클러스터에 속한다.

```

+-----+
|               |
"I voted for Obama because he was most aligned with my values", she said.
|               |               |
+-----+-----+-----+

```

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	결과보고서		
	프로젝트명	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	팀명	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

## 2.2.6 현실적 제한 요소 및 그 해결 방안

- 샘플링 레이트가 낮으면 정확도가 떨어질 수 있습니다.  
 : 16,000Hz 이상의 샘플링 레이트로 오디오를 사용한다.
- 여러 사람이 동시에 또는 다양한 음량으로 대화하는 경우, 대화는 백그라운드 노이즈로 해석되어 무시될 수 있다.  
 : 발화자가 두 명 이상이어서 음성이 겹치는 경우, API를 통해 텍스트로 변환하기 어렵다. 따라서 본 프로젝트에서는 TED 강연과 같은 발화자가 한 명인 음성에 대해서 요약하는 것으로 제한하였다.
- 불분명한 대명사를 처리하는 것에 대해 무엇을 가리키는지 정확히 파악하기 어렵다.  
 : coreference resolution을 처리할 수 있는 파이썬 패키지인 neuralcoref 를 사용하여 대명사처리를 한다.
- 화자의 발음이나 억양이 좋지 않는 경우 정확한 결과가 출력되지 않는다.  
 : 비교적 발음이 정확한 영어 영상에서 음성을 추출하는 것으로 제한한다.


## 2.2.7 결과물 목록

### 1. 음성 데이터를 텍스트로 변환한 전체 텍스트

instrument\_benefits -  
Overall View

Download

Time	Script
00:13.05	Did you know that every time musicians pick up their instruments there are fireworks going off all over their brain on the outside. They may look calm and focused reading the music and making the precise and practice movements required but inside their brains, there's a party going on. How do we know this? Well in the last few decades neuroscientists have made enormous breakthroughs in understanding how our brains work by monitoring them in real-time with instruments like fmri and pet scanners when people are hooked up to these machines tasks such as reading or doing math problems, each have corresponding areas of the brain where activities can be observed but when researchers got the participants to listen to music, they saw fireworks multiple areas of their brains were lighting up at once as they process the sound took it apart to understand elements like Melody and Rhythm and then put it all back together into unified musical experience and our brains do all this work in the split-second between one
01:13.05	First hear the music and when are foot starts to tap a long but when scientists learn from observing the brains of Music listeners to those of musicians, the little backyard fireworks became a jubilee. It turns out that while listening to music in Gage's the brain in some pretty interesting activities playing music is the brains equivalent of a full body workout the neuroscientist some multiple areas of the brain light up simultaneously processing different information in intricate interrelated and

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

## 2. 1번을 요약한 텍스트

instrument\_benefits -  
Summary View

Overall View

Download

### Script

They may look calm and focused reading the music and making the precise and practice movements required but inside their brains, there's a party going on.

How do we know this?

Well in the last few decades neuroscientists have made enormous breakthroughs in understanding how our brains work by monitoring them in real-time with instruments like fmri and pet scanners when people are hooked up to these machines tasks such as reading or doing math problems, each have corresponding areas of the brain where activities can be observed but when researchers got the participants to listen to music, they saw fireworks multiple areas of their brains were lighting up at once as they process the sound took it apart to understand elements like Melody and Rhythm and then put it all back together into unified musical experience and our brains do all this work in the split-second between one

First hear the music and when are foot starts to tap a long but when scientists learn from observing the brains of Music listeners to those of musicians, the little backyard fireworks became a jubilee.

It also combines the linguistic and mathematical Precision in which the left hemisphere is more involved with the novel and creative content that the right excels in for these reasons playing music has been found to increase the volume and activity in the brain is corpus callosum the bridge between the two hemispheres allowing messages to get across the brain faster and through more diverse routes.


## 3. 음성 데이터에서 추출된 핵심 키워드

files	duration	upload date	keyWords		
INSTRUMENT_BENEFITS_QFYQ4PW	00:04:44	28 May 2019 12:10	ALREADY	EXHIBIT	FASTER

## 2.3 기대효과 및 활용방안

### 1. 요약 작업 과정에서 편리함 제공


글을 잘 이해하고 있는지 알아보는 가장 좋은 방법은 요약을 해보는 것이다. 하지만 글을 요약하는 일은 여간 쉬운 일이 아니다. 짧은 신문 기사를 요약하는 작업에도 어려움을 느끼는데, 잘 이해되지 않는 전공에 관한 글을 요약하는 작업은 더 어려울 것이다. 이러한 요약 작업 과정에서 이 프로그램이 편리함을 제공할 것이다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

## 2. 중요한 내용을 파악하는 과정에서 편리함 제공


글에서 중요한 내용을 구분해내는 것은 매우 어려운 일이다. 예를 들어, 강연에서 강연자의 농담이 강의 주제와 전혀 맞지 않는다면 불필요한 내용이다. 청취자의 입장에서선 이러한 불필요한 내용을 빼고 필요한 부분만을 얻고 싶어한다. 이 과정에서 이 프로그램이 제공하는 요약본이 도움을 줄 수 있을 것이다.

## 3. 영상 내용의 키워드를 추출해주는 기능을 이용해서 영상의 내용을 대략적으로 파악할 수 있다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>	
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스
	<b>팀 명</b>	Summer
	Confidential Restricted	Version 1.4
		2019-May-28

### 3 자기평가


1. 음성 및 영상파일을 업로드하기만 하면 요약본을 추출할 수 있어서 간편하게 핵심내용을 파악할 수 있다.
2. 음성 파일 업로드 뿐만 아니라, 영상 URL 입력을 통해 업로드가 가능하므로 사용자의 편리성을 높였다.
3. 음성인식 후 텍스트로 변환하는데 시간이 오래 소요되므로 사용하기에 불편할 수 있다.
4. 요약본의 정확성에 대해서 coreference resolution을 적용하여 전처리하였으므로 결과물을 신뢰할 만하다.
5. DB를 사용하였지만 로그인 기능 없어서 사용자를 구분하지 못한다. 사용자의 개인 데이터에 다른 사용자가 접근할 수 있으므로 프라이버시를 보장하지 못한다.
6. STT에서 고급 모델을 사용하였고, phrase를 추가해주었기 때문에 인식의 정확도가 높아졌다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

## 4 참고 문헌

번호	종류	제목	출처	발행년도	저자	기타
1	서적	TextRank 알고리즘을 이용한 한국어 중요 문장 추출	누리미디어	2009	홍진표	
2	웹	TextRank를 이용한 문서요약	<a href="https://excelsior-cjh.tistory.com/93">https://excelsior-cjh.tistory.com/93</a>			
3	웹	Cloud Speech-to-Text 문서	<a href="https://cloud.google.com/speech-to-text/docs/">https://cloud.google.com/speech-to-text/docs/</a>			
4	웹	Django 문서	<a href="https://docs.djangoproject.com/ko/2.1/">https://docs.djangoproject.com/ko/2.1/</a>			
5	서적	(Django로 쉽게 배우는 배프의 오지랖) 파이썬 웹프로그래밍	배프	2019	배프	
6	강의	파이썬으로 장고(Django) 공략하기: 입문	인프런(inflearn)			
6	웹	The MongoDB 4.0 Manual	<a href="https://docs.mongodb.com/manual/?_ga=2.238564611.283686187.1552306197-1599669186.1552306197">https://docs.mongodb.com/manual/?_ga=2.238564611.283686187.1552306197-1599669186.1552306197</a>			
7	웹	Coreference Resolution in spaCy with neural network	<a href="https://github.com/huggingface/neuralcoref">https://github.com/huggingface/neuralcoref</a>			
8	웹	hsaudiotag3k - Read audio file metadata	<a href="https://github.com/hsoft/hsaudiotag">https://github.com/hsoft/hsaudiotag</a>			
9	웹	YouTube-Video-Converter	<a href="https://github.com/aksalcido/YouTube-Video-Converter">https://github.com/aksalcido/YouTube-Video-Converter</a>			

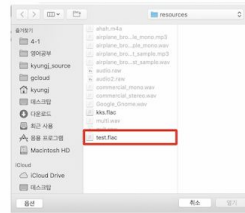
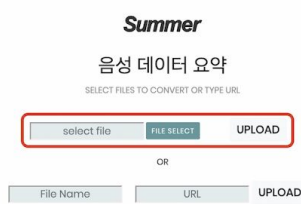


 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>	
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스
	<b>팀명</b>	Summer
Confidential Restricted	Version 1.4	2019-May-28

## 5 부록

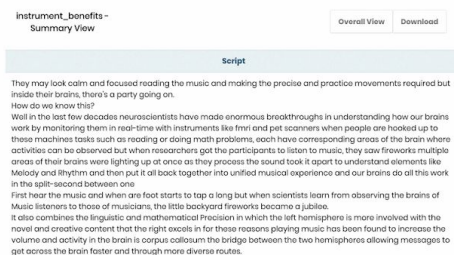
### 5.1 사용자 매뉴얼

#### 사용 방법



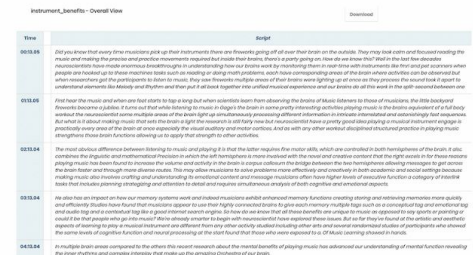
1. 강연 음성 데이터를 불러온 후, UPLOAD 버튼을 클릭한다.

2. 요약될 텍스트의 길이를 선택한 후 SUMMARY 버튼을 클릭하면 요약이 진행된다. 30%는 전체 텍스트의 30% 길이만큼 요약한다는 의미이다.



#### <요약된 텍스트>

3. Overall View 버튼을 클릭하면 요약 전 텍스트를 볼 수 있다.



#### <요약 전 텍스트>

4. Download 버튼을 클릭하면 요약된 텍스트와 요약 전 텍스트를 로컬에 저장할 수 있다.

### 5.2 운영자 매뉴얼

#### 5.2.1 설치 및 설정 가이드

1. Google Cloud Storage bucket 생성, API key 얻기
2. AWS EC2 instance Ubuntu Server 18.04 LTS(Port 80, Elastic IP)
3. Package(Django, Summary, STT, URL to FLAC, MongoDB)
 

```
sudo apt-get update
```

```
sudo apt-get install python3-pip, nodejs, npm, ffmpeg
```

```
pip3 install newspaper3k, scikit-learn, nltk, neuralcoref --no-binary neuralcoref
```



결과보고서	
프로젝트 명	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스
팀 명	Summer
Confidential Restricted	Version 1.4
	2019-May-28



국민대학교  
컴퓨터공학부  
캡스톤 디자인 I


```

sudo su -> pip3 install youtube-dl, pymongo, django -> exit
sudo su -> python3 -m nltk -d /usr/local/share/nltk_data all, -m spacy download en -> exit
npm install --save @google-cloud/speech, sf를 설치 후
nano /node_modules/@grpc/grpc-js/build/src/index.js
(45)const supportedNodeVersions = "; // 코드수정
sudo apt install gsutil, sox, mongodb-server-core, mongodb-clients
sudo snap install google-cloud-sdk --classic
cd / -> sudo mkdir data -> cd data -> sudo mkdir db

```

## 5.2.2 운영 가이드

1. AWS EC2 pem키를 이용해 ssh로 접속한다.(터미널 2개 이용)
2. Git에서 manage.py가 있는 디렉토리로 이동한다
3. sudo python3 manage.py runserver 0.0.0.0:80 (one)
4. sudo mongod (another)
5. AWS에서 EC2에 연결한 Elastic IP로 서비스를 배포한다.

 <b>국민대학교 컴퓨터공학부 캡스톤 디자인 I</b>	<b>결과보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.4	2019-May-28

### 5.3 테스트 케이스

대분류	소분류	기능	테스트 방법	기대 결과	테스트 결과
파일	파일 업로드	사용자가 웹에 파일을 업로드한다.	상단 메뉴바에서 'FILE SELECT' 버튼을 클릭하면 1) 불러올 파일을 선택할 수 있는 파일 리스트 창을 열어준다. 2) 파일을 선택해서 '열기'를 클릭하면 파일명을 볼 수 있다. 3) 'UPLOAD' 버튼을 클릭한다.	웹 중단부에 올리게 될 파일명이 출력된다.	성공
파일	파일 업로드	음성 파일의 길이가 480분 이상이면 파일이 업로드되지 못하도록 한다.	480분 이상의 음성 파일을 업로드한다.	웹에 파일을 업로드할 수 없으며, 길이가 너무 길다는 문구를 띄워준다.	실패
파일	파일 가공	업로드된 사용자의 음성파일을 요약한다.	업로드가 완료된 텍스트 이름이 화면에 띄워지면 1) 요약될 텍스트의 길이를 %단위로 설정한다. 2) 'SUMMARY' 버튼을 클릭한다.	웹 하단부에 리스트 형태로 파일의 이름, 파일의 길이, 날짜, 키워드가 추가된다.	성공
파일	파일 view	요약문을 보여준다.	웹 하단부 리스트에서 파일이름을 클릭한다.	modal view에서 요약문을 볼 수 있다.	성공
파일	파일 view	전체문을 보여준다.	요약문 modal view 우측 상단에 'Overall View' 버튼을 클릭한다.	새로운 modal view 형태로 전체 원본 텍스트가 이전 modal view 위에 출력된다.	성공
파일	요약본 및 전체본 저장	텍스트 파일을 다운로드 할 수 있다.	modal view의 오른쪽 상단에 'Download' 버튼을 클릭한다.	화면 아래에서 텍스트 파일의 다운로드 성공여부를 확인할 수 있다.	성공
요약 알고리즘	정확도	요약된 텍스트의 신뢰성을 확인한다.	불필요한 문장을 삽입한다.	임의로 삽입한 불필요한 문장이 걸러진다면 요약된 텍스트의 신뢰성이 보장된다.	성공
요약 알고리즘	정확도	Coreference resolution (대응어 처리)	이전 버전과 Coreference resolution을 추가한 버전의 ROUGE 값을 비교한다.	추가한 버전의 ROUGE값이 높음.	성공
STT	정확도	고급 모델 사용, phrase hint 추가	사용한 버전과 사용하지 않은 버전의 인식 정확도를 비교한다.	추가한 버전의 단어들이 구문에 알맞는 단어로 인식.	성공