

# 캡스톤 디자인 I

## 종합설계 프로젝트

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

Version	1.2
Date	2019-04-17

팀원	김기성 (조장)
	김윤성
	양성호
	정경진
	정예원
지도교수	박수현 교수




### 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.

## 문서 정보 / 수정 내역

Filename	17조-중간보고서(1.2).doc
원안작성자	김기성, 김윤성, 양성호, 정경진, 정예원
수정작업자	김기성, 김윤성, 양성호, 정경진, 정예원

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2019-04-17	김기성	1.0	최초 작성	
2019-04-17	정경진	1.1	내용 수정	수정된 내용 추가
2019-04-17	정예원	1.2	내용 수정	수정된 내용 추가

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

## 목 차

1	프로젝트 목표 .....	4
2	수행 내용 및 중간결과 .....	5
2.1	계획서 상의 연구내용 .....	5
2.2	계획서 외 연구내용 .....	6
2.3	수행내용 .....	7
3	수정된 연구내용 및 추진 방향 .....	10
3.1	수정사항 .....	10
4	향후 추진계획 .....	11
4.1	향후 계획의 세부 내용 .....	11
5	고충 및 건의사항 .....	13

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

## 1 프로젝트 목표

현재까지 음성인식 후 요약된 필기를 제공해주는 소프트웨어가 개발되어 있지 않다. 사용자입장에서는 녹음파일의 데이터를 가시적으로 확인하고 싶고 더 나아가서 중요하거나 핵심으로 여겨지는 부분만을 접근하고 싶어하는 것이 대개 그렇다. 그래서 개발할 시스템으로 인해 전체 문장을 요약하여 핵심적인 내용만을 제공받을 수 있게 하는 것이 필요하다 생각하였다.

예시로 공식적인 발표나 연설에서 음성 녹음을 하고 마친 뒤에 그 음성 파일을 처음부터 끝까지 듣고 텍스트 변환하고 요약하는 번거로운 작업을 거쳐야만 한다. 이에 대해 사용자가 음성 파일만 있다면 텍스트 변환하여 음성 파일에 대한 모든 데이터를 확인할 수 있고 더 나아가 중요한 핵심문장들을 요약하여 확인할 수 있게 된다면 매우 편할 것이다.

이 소프트웨어를 사용하여 장문의 음성 데이터(연설이나 수업, 미팅 등)에서 핵심적인 내용이나 중요한 부분을 놓쳐도 다시 정보를 얻을 수 있는 유용함을 얻을 수 있다. 요약하자면 미사여구가 많고 내용을 이해하기 어려우며 긴 음성데이터나 동영상을 본다는 건 굉장히 큰 저항감을 일으킨다. 그러한 영상 들에서 핵심 내용이 무엇인지 궁금한 사람들에게 음성 파일에서 요약본, 키워드를 뽑아내어 쉽게 내용의 핵심을 이해할 수 있도록 하는 것을 목표로 한다.

 <div> <p>국민대학교</p> <p>컴퓨터공학부</p> <p>캡스톤 디자인 I</p> </div>	중간보고서		
	프로젝트 명	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	팀 명	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

## 2 수행 내용 및 중간결과

### 2.1 계획서 상의 연구내용

긴 녹음 파일을 API로 처리하면, 음성 파일이 텍스트로 변환되어 Transcript 속성에 종속된 값의 형태를 가진 Json 파일로 저장된다. Json 파일은 DB에 저장한다. 이때, API의 타임스탬프(Time Stamp) 마커 기능을 통해 각 단어가 파일의 어느 시간대에 출현하는지 나타내는 정보를 Json 파일에 함께 저장한다. 이 타임스탬프 정보를 이용하여 사용자는 원하는 단어를 검색하여, 그 단어가 포함된 파일과 출현 시간을 알아낼 수 있다.

API의 출력 결과인 Json 파일에서 Transcription만 추출, 저장하여 텍스트 파일(이해를 위해 스크립트라 부름)을 생성한다. 이 스크립트 파일에는 음성 파일에서 추출된 모든 문장이 포함되어 있다. 요약하는 과정에서 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 전치행렬을 구하여 Correlation Matrix를 생성한다. 이것을 통해 문장(or 단어)간 그래프로 나타낼 수 있다.

이 그래프를 입력 받아 TextRank 알고리즘을 적용한다.

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

- $TR(V_i)$ : 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)

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

 <div> <b>국민대학교</b>  <b>컴퓨터공학부</b>  <b>캡스톤 디자인 I</b> </div>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

만큼 출력해준다.


추출된 각 문서의 중요한 키워드를 리스트의 형태로 DB에 저장한다. 중요한 키워드는 제목 옆에 해시태그(Hash tag)의 형태로 표시해준다. 사용자는 해시태그만 보고 음성 파일의 주제를 간단하게 파악할 수 있다. 또한, 음성 파일의 요약된 텍스트와 음성 파일의 스크립트를 웹에서 볼 수 있다.

## 2.2 계획서 외 연구내용

기존의 알고리즘에선 대명사(it, that, he, she) 들은 불용어 처리를 한 후에 TF-IDF와 TextRank 알고리즘을 적용시킨다. 하지만 문장에 포함된 대명사가 어떤 의미를 가지고 있는지에 따라서 중요한 문장이 될 수도 있다. 그래서 대명사를 추측하는 방법을 찾아보았다.

- **Deterministic** : 빠른 rule-based의 coreference resolution. (영어와 중국어 지원)
- **Statistical** : 머신러닝 기반의 coreference resolution. 다른 시스템들과는 다르게 dependency parses가 필요하다. (영어 지원)
- **Neural** : 가장 정확함. 하지만 neural-network-based라 느리다.(영어와 중국어 지원)

‘Coreference Resolution’ 키워드로 찾아보니 Stanford CoreNLP의 CorefAnnotator를 이용해 대명사의 의미를 유추하는 방법을 알아냈다. CorefAnnotator엔 3가지 coreference system이 있었다.

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17


## 속도 측정의 결과

SYSTEM	LANGUAGE	PREPROCESSING TIME	COREF TIME	TOTAL TIME	F1 SCORE
Deterministic	English	3.87s	0.11s	3.98s	49.5
Statistical	English	0.48s	1.23s	1.71s	56.2
Neural	English	3.22s	4.96s	8.18s	60.0

## 2.3 수행내용

음성파일(flac 포맷)을 클라이언트가 업로드하게 되면 서버로 음성파일 데이터가 저장되고 여기서 Google Speech Api를 이용해서 음성 데이터를 텍스트 데이터로 변환하는 작업이 시작된다. Transcript 속성에 종속된 값의 형태를 가진 Json 파일로 만들어진 것을 MongoDB에 저장하게 되는데 이때, 단순 음성데이터에 대한 텍스트로 변환된 transcript만 나오는 것이 아니라 여기에 더해 API의 타임스탬프 마커 기능을 통해 각 단어가 어느 파일의 시간에 출현하는 지에 대한 정보 또한 포함한다. 이 정보는 후에 만들어지는 기능에 사용이 되는데, 사용자가 원하는 단어를 검색하여, 그 단어가 포함된 파일과 출현 시간을 알아낼 수 있게 한다. 그 후에 서버에서는 DB에 담겨 있는 텍스트 데이터를 가지고 와서 어떠한 필터링 과정을 거치게 되는데 그 필터링 과정이 바로 요약할 알고리즘 적용이다.


요약 알고리즘에서 사용되는 메인 알고리즘은 TF-IDF, TextRank 이 두 가지이다. 파이썬 패키지인 scikit-learn을 이용하여 TF-IDF 알고리즘에 대한 모델링을 하게 되는데 TF-IDF가 문장내에서 단어의 상대적 가중치를 구하는 개념으로 scikit-learn 패키지를 사용하여 상대적 가중치에 대한 값이 Sentence-Term Matrix 형태로 나타나게 되는데 이것을 TextRank 알고리즘에 적용시킨다. TextRank는 문장(or 단어)들을 전 처리하였던 sentence-term에 대한 상대적 가중치들을 Ranking해주기 위한 알고리즘으로 이 작업까지 거치면 요약할 문장 개 수만큼 나오게 된다. 이 문장들과 중요하다고 매겨지는 키워드들이 리스트의 형태로 MongoDB에 저장된다. 여기서 중요 키워드는 후에 추가 기능에서 해시태그의 형태로 표시해주기 위해 데이터를 저장한 것이다. 마지막으로 웹UI를 통해 요약된 텍스트와 음성 파일의 스크립트를 선택적으로 모두 보여줄 수 있게 구현한다.

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

**<계획상 진도와 수행하고 있는 진도 비교 분석>**

마일스톤	개요	시작일	종료일
중간자문평가	핵심 키워드 추출 및 요약본, 웹서버 구축 <b>산출물 :</b> <ol style="list-style-type: none"> <li>1. 프로젝트 1 차 중간보고서</li> <li>2. 프로젝트 진도 점검표</li> <li>3. 1 차분 구현소스 코드</li> </ol>	2019-03-24	2019-04-19
구현완료	시스템 구현 완료 <b>산출물 :</b> <ol style="list-style-type: none"> <li>1. 최종 소스 코드</li> </ol>	2019-04-22	2019-05-20
테스트	시스템 통합 테스트 <b>산출물:</b> <ol style="list-style-type: none"> <li>1. 완성된 웹 사이트</li> </ol>	2019-04-24	2019-05-23
최종보고서	최종 보고 <b>산출물 :</b> <ol style="list-style-type: none"> <li>1. 전시용 자료(포스터 및 소개 책자)</li> <li>2. 온라인 평가용 자료</li> <li>3. 최종결과보고서</li> </ol>	2019-05-25	2019-05-31



 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

**<실제상 진도와 수행하고 있는 진도 비교 분석>**

마일스톤	개요	시작일	종료일
중간자문평가	핵심 키워드 추출 및 요약본, 웹서버 구축 <b>산출물 :</b> <ol style="list-style-type: none"> <li>1. 프로젝트 1 차 중간보고서</li> <li>2. 프로젝트 진도 점검표</li> <li>3. 1 차분 구현소스 코드(API 호출 및 요약 알고리즘)</li> </ol>	2019-03-24	2019-04-19
구현완료	시스템 구현 완료 <b>산출물 :</b> <ol style="list-style-type: none"> <li>1. 최종 소스 코드</li> <li>2. 대명사 처리에 대한 클래스 구현</li> </ol>	2019-04-22	2019-05-20
테스트	시스템 통합 테스트 <b>산출물:</b> <ol style="list-style-type: none"> <li>1. 완성된 웹 사이트</li> <li>2. 결과물에 대한 피드백</li> <li>3. 버그 수정</li> </ol>	2019-04-24	2019-05-23
최종보고서	최종 보고 <b>산출물 :</b> <ol style="list-style-type: none"> <li>1. 전시용 자료(포스터 및 소개 책자)</li> <li>2. 온라인 평가용 자료</li> <li>3. 최종결과보고서</li> </ol>	2019-05-25	2019-05-31

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

### 3 수정된 연구내용 및 추진 방향

#### 3.1 수정사항

기존의 제안서의 내용은 사용자가 강의 내용을 녹음해서 요약 프로그램을 사용하는 것이었지만, 사용자가 녹음을 한 파일이 아닌 유튜브 등에서 쉽게 구할 수 있는 음성파일을 입력 파일로 받는 것으로 수정되었다.

강의를 녹음하는데 생각보다 교수님의 목소리가 잘 들리지 않고, 주변의 잡음이 음성 인식을 방해해 Speech-to-Text에서 에러율이 너무 높았다. 아무래도 녹음기가 전문 장비가 아니고, 강의실의 환경을 생각하지 못해서 생긴 일이었다. 텍스트를 요약하는 데 있어 Speech-to-Text가 제대로 되지 않는다면 요약본의 결과가 엉뚱한 문장으로 나오는 것은 당연해서 일단 Speech-to-Text의 결과가 잘 나오는 음성 파일로 test를 진행하기로 하였다.

계획서 발표 후 김준호 교수님께서 문장에서 나오는 대명사 처리를 어떻게 할 것인가에 대해 질문을 하셨다. 그 후에 담당 교수님과 강승식 교수님과의 미팅 결과 coreference resolution이라는 키워드로 이 문제를 해결할 계획이다.

마지막으로 default 언어는 한글에서 영어만 지원하는 것으로 바뀌었습니다. 현재 Google Speech API에서 한글에 대한 음성 인식률이 제한된 조건에서만 이해할 수준의 인식을 보였다. (실제로 강의를 녹음하여 Google Speech API를 통해 음성 인식 후 텍스트로 변환한 결과 이해하기 힘든 수준으로 변환되었다.) 하지만 영어에 대해서는 한글 인식률에 비해 많이 개선된 음성 인식률을 보였다. (실제로 원어 강의를 녹음한 샘플링 작업을 시도하지는 못했지만, 원어 영상 강의인 TED 강의나 유튜브 영상에서 음성을 추출한 샘플링 작업을 시도했을 때는 한글에 비해 정확도가 높았다.) 따라서 default 언어를 영어로 바꾸었고 한글은 추가 지원하는 식으로 개발할 예정이다.

## 4 향후 추진계획

### 4.1 향후 계획의 세부 내용

현재 핵심 키워드 추출 및 요약본 생성, 웹 서버가 구축되었다. 따라서 이를 바탕으로 프로그램 개발을 진행할 예정이다.


- 알고리즘 정확도 향상

기존의 코드는 파이썬 언어로 구현을 했는데 Stanford CoreNLP는 Java Library이다. 그래서 Stanford CoreNLP Server를 이용한 패키지를 다운 받아서 파이썬 내에서 Stanford CoreNLP의 기능을 사용할 예정이다.

- 해시 태그 구현

NAME	DURATION	UPLOADED	KEYWORD
<input type="checkbox"/> How to keep human bias out of AI	0:20:35	08 Mar 2019 17:17	AI machine recognition
<input type="checkbox"/> Creative ways to get kids to thrive in school	0:31:13	12 Apr 2019 12:32	keyword1 key2 key3
<input type="checkbox"/> A simple way to break a bad habit   Judson Brewer	0:18:42	15 Apr 2019 13:52	keyword1 key2 key3

추출된 중요한 키워드를 위 그림과 같은 형태로 사용자에게 보여준다. 사용자는 키워드를 보고 이 파일이 어떤 내용의 음성 파일인지 직관적으로 파악할 수 있다. 음성 파일명이 의미 없는 단어일 경우 키워드를 보고 주제를 파악할 수 있다.

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

- 타임 스탬프 구현

0:08:32

Artificial intelligence can be classified into three different types of systems analytical, human-inspired, and humanized artificial intelligence.

0:09:08

Analytical artificial intelligence has only characteristics consistent with cognitive intelligence generating a cognitive representation of the world and using learning based on past experience to inform future decisions. Human-inspired AI has elements from cognitive and emotional intelligence.

Understanding human emotions, in addition to cognitive elements, and considering them in their decision making. Humanized AI shows characteristics of all types of competencies, for example cognitive, emotional, and social intelligence, is able to be self-conscious and is self-aware in interactions with others.

위 그림처럼 전체 스크립트 뷰에서 각 문장 앞에 해당 문장의 발화 시간을 보여 준다.

- Mongo DB 구현

Django 서버에서 사용되는 파이썬 언어로 인해 MongoDB를 구축하기 위해서 커넥터 모듈인 파이몽고를 파이썬 path의 패키지 라이브러리에 설치하여 Mongo DB와 연동시킨다. 연동된 DB를 통해 전체 텍스트 데이터, 각 단어에 대한 키워드 텍스트 데이터, 각 단어에 대한 키워드 타임 스탬프를 저장한다.

 <b>국민대학교</b> <b>컴퓨터공학부</b> <b>캡스톤 디자인 I</b>	<b>중간보고서</b>		
	<b>프로젝트 명</b>	음성인식 머신러닝에 기반한 핵심 요약본 추출 웹 서비스	
	<b>팀 명</b>	Summer	
	Confidential Restricted	Version 1.2	2019-04-17

## 5 고충 및 건의사항

프로젝트 팀원들 모두 각자 수강 스케줄이 존재해서 수업이 모두 끝나는 시간인 오후 6시 이후에 회의를 진행하여야 하는데 7호관의 대학원 랩실이나 동아리 방을 제외하고는 모든 강의실이 오후 9시 이후에는 이용이 불가능 합니다. 작년까지만 해도 철야 신청을 할 수 있었으나 다음 날 강의가 있는 경우 강의실 관리가 안된다는 이유로 철야 신청을 막았습니다. 오후 9시 이후 작업을 하는 경우, 따로 외부 장소를 잡아서 회의를 진행해야 합니다. 장소를 지원해 주셨으면 합니다.