Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.30, NO.5, Oct. 2020 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.5.895

오디오 데이터 내 개인 신상 정보 검출과 마스킹을 위한 인공지능 API의 활용 및 음성 분할 방법의 연구*

김 태 영,[1†] 홍 지 원,[1] 김 도 희,[1] 김 형 종[2‡]
[1,2]서울여자대학교 (학생, 교수)

A System of Audio Data Analysis and Masking Personal Information Using Audio Partitioning and Artificial Intelligence API*

Kim TaeYoung[1†] , Hong Ji Won,[1] Kim Do Hee,[1] Kim Hyung-Jong[2‡]
[1,2]Seoul Women's University (Undergraduate, Professor)

요 약

최근 기존 텍스트 기반 콘텐츠 외 멀티미디어 콘텐츠의 영향력이 급증함에 따라 콘텐츠 내 정보들을 처리할 수 있도록 도와주는 서비스가 콘텐츠 이용에 큰 편리함을 주고 있다. 이러한 서비스의 대표적인 기능으로는 중요 정보에 대한 검색과 마스킹이 있다. 텍스트 데이터와 이미지 데이터의 검색 및 마스킹 기술을 제공해주는 솔루션들은 활발히 보급되고 있어 쉽게 접할 수 있다. 그러나 오디오 콘텐츠의 경우, 검색 및 마스킹의 필요성은 인식되지만 기술의 난이도로 인해 범용적으로 적용되는 솔루션을 찾는 것이 쉽지 않다. 본 논문은 음성 분할을 이용하여 오디오 데이터 내 정보 검색과 마스킹 기능을 제공하는 웹 애플리케이션을 제안한다. 추가적으로, 국내외 인공지능 기반 음성 인식 API에 대한 분석을 통해 적절한 API의 선택을 진행하였으며, 정규식을 이용한 개인 신상 정보의 검출 방법을 제시하였다. 마지막으로 구현결과의 정확도를 측정하여 성능을 검증하였다. 본 논문의 기여점은 오디오 데이터 내 특정 패턴의 검출 및 마스킹 기능을 설계하고 실험을 통해 검증한 것에 있다.

ABSTRACT

With the recent increasing influence of multimedia content other than the text-based content, services that help to process information in content brings us great convenience. These services' representative features are searching and masking the sensitive data. It is not difficult to find the solutions that provide searching .and masking function for text information and image. However, even though we recognize the necessity of the technology for searching and masking a part of the audio data, it is not easy to find the solution because of the difficulty of the technology. In this study, we propose web application that provides searching and masking functions for audio data using audio partitioning method. While we are achieving the research goal, we evaluated several speech to text conversion APIs to choose a proper API for our purpose and developed regular expressions for searching sensitive information. Lastly we evaluated the accuracy of the developed searching and masking feature. The contribution of this work is in design and implementation of searching and masking a sensitive information from the audio data by the various functionality proving experiments.

Keywords: Audio Data Partition, Privacy Protection, Speech Recognition

Received(08. 20. 2020), Modified(10. 05. 2020), Accepted(10. 05. 2020)
* 본 논문은 2020학년도 서울여자대학교 교내 연구비의 지원을 받았음(2020-0312)
† 주저자, taeyoung9842@gmail.com
‡ 교신저자, hkim@swu.ac.kr(Corresponding author)

I. 서 론

스마트폰의 보급이 일반화되며 스마트폰을 처음 접하는 연령대가 낮아지게 되었고, 이러한 모바일 네이티브 세대가 인터넷상의 새로운 경제 주체로 등장하게 되었다. 넷플릭스나 유튜브 등의 영상 콘텐츠 플랫폼을 기반으로 1인 미디어, 엔터테인먼트 콘텐츠를 활발히 이용하는 모바일 네이티브 세대의 영향에 따라 영상 콘텐츠의 소비는 더욱 확대될 전망이며 관련 하드웨어와 소프트웨어 기술의 성장까지 견인하고 있다[1]. 기존의 텍스트나 이미지 기반의 콘텐츠보다 더욱 직설적으로 정보를 획득할 수 있는 영상 콘텐츠가 각광받음에 따라 오디오 콘텐츠 또한 하나의 대세로 자리 잡고 있다[2]. 최근 영상 콘텐츠만을 제공하던 플랫폼들도 1인 크리에이터의 오디오 콘텐츠 생산을 지원하고 있으며, 그 예로 아프리카 TV의 '팟 프리카', 네이버의 라이브 스트리밍 플랫폼 '나우, 오디오북 등이 있다.

영상 및 오디오 콘텐츠가 새로운 서비스 트렌드로 주목받으며 검색 시장에서도 이를 활용한 서비스가 등장했다. 해외 서비스인 'OP3Nvoice'는 시각, 음성 정보에서 텍스트를 검색할 수 있는 플랫폼으로 오디오, 비디오 등의 미디어 콘텐츠의 음성 텍스트와 대화를 검색하고 모니터링을 할 수 있도록 지원해 주는 서비스이다[3]. 한국어 버전을 지원하는 서비스로는 비디오 검색 엔진 'Panopto'가 있다. Panopto는 스마트 검색을 사용하여 화면에서 말하거나 보여주는 단어를 비디오 내부에서 검색해준다[4]. 이러한 기존 검색 서비스는 자동으로 음성을 텍스트로 변환하여 텍스트 기반의 정보를 쉽게 찾을 수 있다는 점에서 큰 편리함을 준다. 그러나 1인 미디어를 포함한 수많은 영상 및 오디오 콘텐츠의 수와 비교하면 검색 서비스의 보급률은 낮은 편이다.

영상 및 오디오 콘텐츠의 개인 간 생산과 소비가 활발히 이루어지는 가운데, 개인 정보보호의 문제가 발생하기도 한다. 개인 또는 소규모 인원이 방송 주체가 되어 콘텐츠를 제작 편집하는 1인 미디어의 특성상[5], 부주의로 개인 정보가 시각적 혹은 청각적으로 노출될 가능성이 존재한다. 이러한 영상들은 불특정 다수에게 공개되어 있어 재배포가 쉽지만 플랫폼의 정책에 따라 즉각적으로 삭제되기는 어렵다. 전 세계 인터넷 사용자 중 45%가 이용하는 영상 플랫폼 유튜브[6]의 영상 삭제 정책에 따르면, 사용자들로부터 특정 영상에 대한 부적합 신고가 다수 발생해야만 플랫폼에서 이를 검토하고 삭제 여부를 결정한다[7]. 개인 정보 노출에 대한 대처가 미미 한 것에 비해 오디오 콘텐츠를 포함한 영상 콘텐츠에서 화면과 음성으로 개인 정보가 유출될 가능성은 크다. 이처럼 1인 미디어 시대가 자리 잡음에 따라 지금보다 개인 정보 노출 사례는 더욱 증가할 것으로 보인다.

본 논문은 영상 및 오디오 콘텐츠의 활용성과 콘텐츠 내 개인 신상 정보의 기밀성 보장을 위해 오디오 데이터 내 정보 검색 및 마스킹 기능을 제공하는 웹 애플리케이션 ADAM(Audio Data Analysis and Masking)을 제안한다. ADAM은 음성 인식 API를 통해 음성 파일을 텍스트 데이터로 변환하여 사용자가 찾고자 하는 정보가 오디오 데이터 내의 어느 위치에서 재생되는지 출력해준다. 개인 정보와 욕설과 같이 프라이버시 보호가 필요한 영역들은 마스킹을 위해 정규식으로 탐지하는데, 이 때 탐지 정확도를 높이기 위해 음성을 분할하여 음성 파일의 조각들을 활용한다. 음성 파일 조각들에서 탐지하고자 하는 정보의 위치는 Wav 형식 파일의 raw 데이터가 진폭에 대한 정보를 포함한다는 점[8]을 이용하여 비율적으로 계산한다. 이러한 기술의 구현을 위해 본 연구에서는 다양한 국내외 음성인식 API의 분석, 정규식 기반 개인 신상 정보 검색 및 구현결과의 정확도 분석을 진행하였다.

본 논문의 2장은 개인 신상 정보 마스킹에 관련한 기존 연구를 제시한다. 3장은 본 연구에서 제안하는 시스템의 구조 및 기능과 음성 분할 기법을 사용하여 서비스의 정확도를 증가시킨 방법에 대해 설명한다. 4장에서는 3장의 설계를 바탕으로 시스템을 구현한 결과를 나타낸다. 5장은 시스템 구현 방식과 결과의 유효성을 위한 테스트 결과를 제시한다. 마지막 6장은 본 연구의 결론 및 후속 연구에 대해 다룬다.

II. 기존 연구

박석천(2017)의 이미지 인식 기반 개인정보 식별 및 마스킹 시스템에 대한 연구에서는 개인정보가 포함된 이미지를 SNS나 블로그에 업로드하여 일어나는 유출 문제를 방지하기 위하여 마스킹 시스템을 설계 및 구현하였다[9]. 이미지 내의 문자를 식별하기 위해 광학 문자 인식 오픈소스인 TesseractOCR를 사용하였고 CNN 딥러닝 알고리즘으로 특징 벡터를 추출하여 인식 정확도를 높였다. 이미지 내의 문자가 텍스트로 변환된 후에는 정규식을 통해 개인정보 패

턴을 탐지한다. 개인정보 패턴 중, 주민등록번호는 '-'를 기준으로 앞 6자리와 뒤 1~7자리 숫자로 이루어져 있고, 연락처는 첫번째 그룹 마지막 숫자가 0, 1, 6, 7, 8, 9 중의 하나이며 두번째 그룹은 3자리 혹은 4자리이고 마지막 그룹은 4자리로 구성된다.

류재철, 조은선, 원유재(2018)의 연구는 활자 및 필기 텍스트로 이루어진 숫자 형태의 개인정보를 유사도 분석과 문자 수정을 사용하여 기존보다 효율적이고 정확하게 검출하는 방법을 제시하였다[10]. 활자를 텍스트로 변환할 때에는 인식률을 고려하여 주민등록번호, 전화번호 등을 표현한 정규식에 임의로 설정한 임계값 이상의 유사성을 보이면 개인정보로 가정하였다. 그 예로, 주민등록번호는 첫 번째 그룹이 6자리이고 두 번째 그룹이 7자리인 정규식으로 표현되어 활자 텍스트에서 총 11자리의 숫자만 인식이 되었다면 활자 텍스트와 정규식의 인식률은 85.7%인 것이다. 임계값을 설정함으로써 문자학습에 불필요한 작업을 줄이고 효율적으로 인식률을 향상시킬 수 있다. 이는 기존 TesseractOCR 보다 35% 높은 인식률로 개인정보를 탐지하였다.

이미지 인식 기반의 두 연구 모두 필기 텍스트를 인식하여 텍스트 데이터로 변환한 후, 정규식을 통해 개인정보 패턴을 탐지하였다. 이 때 필기된 텍스트는 대화체로 작성되지 않고 문서 등에서 조건에 따라 기입된 형태이다. 따라서, 정규식을 정립할 때 각 패턴 중간에 불순 문자가 들어갈 가능성을 고려하지 않고 있다. 이는 대화체 혹은 자유 형식으로 작성된 문서에서는 정규식을 통해 개인정보를 탐지하지 못한다는 점에서 한계점을 갖는다.

음성 데이터 기반의 연구로는 오종훈(2019)의 음성 키워드 분석을 통한 보이스 피싱 방지 기법에 대한 연구가 있다[11]. 지능화된 보이스피싱 범죄를 예방하기 위해 전화 음성을 텍스트 파일로 변환한 후, 텍스트 매칭 알고리즘을 통해 보이스피싱에 주로 사용되는 단어를 탐지하였다. 음성 인식을 위해 구글 Cloud Speech-To-Text API를 사용하였으며 텍스트 매칭은 KMP 알고리즘을 사용하였다. 전화를 하는 상황에서 동작하는 시스템이므로 실시간으로 탐지가 이루어진다. KMP 알고리즘을 사용한다면 약 30초 분량의 전화음성에서 약 40초 이내에 보이스피싱 여부를 탐지할 수 있다. 텍스트 매칭은 '서울중앙지검', '대출' 등의 의심 단어들을 키워드 기반으로 검색하여 이루어지는데, 이는 키워드로 등록되지 않은 단어들은 보이스피싱으로 탐지되지 않는다는 한계점을 갖는다. 또한, 음성 인식 결과값을 기반으로 한 키워드 검색에 의존하므로 음성 인식 정확도에 시스템 전체의 성능이 좌우된다.

텍스트가 아닌 이미지나 음성 기반의 데이터 내에서 특정 키워드 및 패턴을 검출하기 위해서는 우선적으로 텍스트 데이터로의 변환이 필요하다. 키워드 및 패턴 검출은 이 텍스트 데이터 내에서 이루어지므로 변환 방식에 따라 전체 시스템의 성능이 좌우된다. 음성 데이터의 변환에는 Speech to Text(음성 인식) 즉, 음성 언어를 입력받아 문자 데이터로 처리하는 기술이 사용되는데 최근 검색 DB를 많이 보유한 회사들이 음성 인식 API를 공개함으로써 이를 이용하여 프로그램을 용이하게 개발할 수 있게 되었다[12]. 현재 다양한 기업들은 음성 인식기술을 오픈 API로 제공하고 있으며[13], 대표적인 오픈 API로는 Google의 Cloud Speech-to-Text, IBM의 Watson Speech to Text, Microsoft의 Azure Speech to Text, Amazon의 Transcribe, Kakao의 Speech-to-Text System, Naver의 Clova Speech Recognition 등이 있다. 본 연구에서는 여러 음성 인식 API를 본 연구의 목적에 따라 평가하여 적합한 API를 선정한다. 그리고 서비스의 유형에 따라 서비스 정확도를 증가시킬 수 있도록 음성 파일을 분할한 후 음성 파일 조각들의 음성 인식 결과를 분석한다. 패턴 검출에 있어서는 정규식을 이용하되 음성 인식 결과 값의 부정확성과 구어체의 비정형성을 고려하여 정규식을 다양화하였다. 이를 통해 발화 시에 탐지하고자 하는 특정 패턴 사이에 관련 없는 단어들이 포함되어도 효과적으로 패턴을 탐지해낼 수 있다.

III. ADAM 시스템의 구조와 기능

본 장에서는 음성 분할 기법을 이용한 오디오 데이터 내 개인 신상 정보 검출 및 마스킹 시스템의 구조 및 기능에 대해 설명한다.

3.1 시스템 구조도

ADAM는 Fig.1.과 같이 jsp 파일과 java 파일이 연동되어 동작한다. 사용자는 검색 서비스와 마스킹 서비스 중 하나를 선택하고 음성 파일을 업로드한다. 검색 서비스를 선택한 사용자는 SearchForm.jsp로 이동하여 검색 키워드를 입력

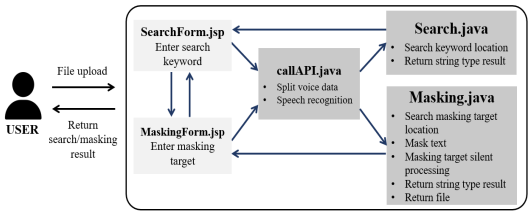

Fig. 1. Composition of ADAM System

하게 되고, 마스킹 서비스를 선택한 사용자는 MaskingForm.jsp로 이동하여 마스킹 타깃을 입력하게 된다.

검색과 마스킹 서비스 모두 callAPI.java에서 음성 데이터 분할과 음성 인식을 거친 후 검색 서비스는 Search.java로, 마스킹 서비스는 Masking.java로 이동한다. 그 후, ADAM이 사용자에게 서비스에 따른 결과를 반환한다. 사용자는 서비스 이용 도중 자신이 업로드한 음성 파일을 유지한 상태로 검색 화면과 마스킹 화면을 자유롭게 전환하며 편리하게 이용할 수 있다.

3.2 음성 내의 키워드 검색기능 설계

검색 기능의 대략적인 플로우 차트는 Fig.2.와 같다. 사용자가 음성 파일과 검색 키워드를 입력하면, 키워드의 위치를 용이하게 찾기 위해 음성 파일을 분

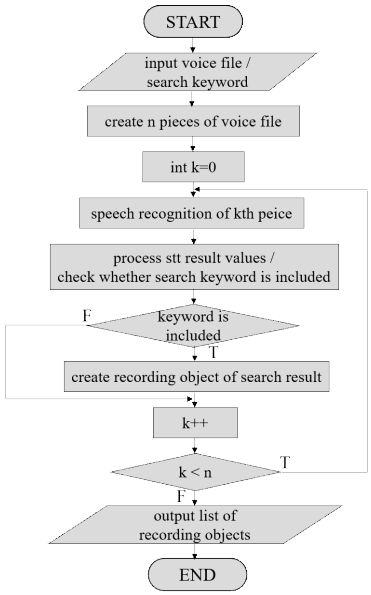

Fig. 2. Flowchart of Searching keywords

Table 1. Elements of Search Result Recording Object

element	description
startTime	kth voice file piece's playback start position
endTime	kth voice file piece's playback end position
originalSentence	original result of speech recognition API response
highlighted -Sentence	result of highlighting search keywords in originalSentence

할하여 n개의 조각을 생성한다. 음성 파일 분할의 단위와 목적은 3장에서 설명된다. n개의 음성 파일 조각들 중, k번째 조각을 음성 인식 API의 요청에 포함시켜 문자열로 이루어진 음성 인식 결과 값을 받는다. 음성 인식 결과 값은 원활한 키워드 탐지를 위해 띄어쓰기와 특수 문자를 제거한 후 검색 키워드가 포함되었는지를 확인한다. 키워드가 포함되어 있는 조각이라면, Table 1.의 요소들을 포함하는 검색 결과 기록 객체를 생성한다. n개의 음성 파일 조각에 대한 키워드 검색이 끝났다면 기록 객체 리스트를 출력하여 검색 결과를 반환하게 된다.

3.3 개인 신상 정보의 마스킹 기능 설계

마스킹 기능의 대략적인 플로우 차트는 Fig.3.와 같다. 원본 음성 파일에서 n개의 조각을 생성한 후, 이를 순서대로 음성 인식 API의 요청에 포함시켜 음성 인식 결과 값을 받아오는 과정은 검색 기능과 동일하다. 그러나 사용자가 입력한 마스킹 타깃이 포함되어 있는지 확인하는 부분에서 차이점이 존재한다. 사용자는 마스킹 할 개인 정보를 체크박스를 통해 입력하거나 직접 마스킹 타깃을 키워드 형태로 입력할 수 있다. 키워드 형태로 입력한 타깃은 검색 기능과 동일하게 음성 인식 결과 문자열에서 포함 여부를 확인하지만, 체크 박스로 입력한 개인 정보는 사전에 정의해 놓은 정규식을 통해 패턴에 부합하는 지를 검사한다. 사전 정의한 정규식은 5장에서 확인할 수 있다.

이러한 두 가지 방식을 통해 각 음성 인식 결과 내의 마스킹 타깃 포함 여부를 확인하고, 타깃이 포함되었다면 타깃을 발화하는 위치의 재생 시간을 추

정하여 마스킹 범위를 지정한다. 이 때 재생 시간은 전체 음성 파일을 바이트배열로 변환한 파일을 분석하여 추정하며 이에 대한 실험 결과는 5장에서 확인 가능하다. 마스킹은 특정 인덱스의 바이트 값들을 초기화하여 수행한다. 각 음성 인식 결과에서 마스킹 타깃이 탐지될 때마다 Table 2.의 요소들을 포함한 마스킹 결과 기록 객체가 생성된다. 마스킹 결과 기록 객체는 마스킹된 음성 파일과 함께 마스킹 기능의 결과로써 사용자에게 반환된다.

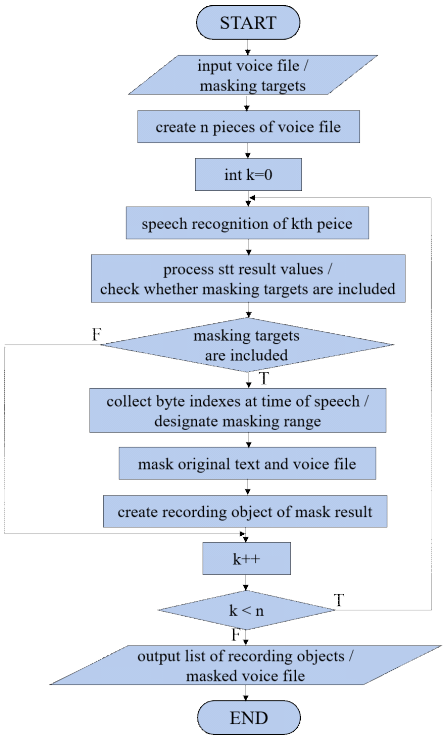

Fig. 3. Flowchart of Masking sensitive information

Table 2. Elements of Masking Result Recording Object

element	description
startTime	kth voice file piece's playback start position
endTime	kth voice file piece's playback end position
originalSentence	original result of speech recognition API response
maskedSentence	result of masking target masked with '*' symbol in originalSentence

3.4 음성 파일의 분할 기법

음성 인식 API는 요청에 포함하는 음성 파일의 내용과 품질에 따라 상이한 결과의 인식 결과를 출력한다. 본 연구에서는 5장에 제시된 실험을 통해 적절한 음성 파일의 분할이 특정 패턴의 정확한 탐지에 도움이 되는 것을 파악하였다. 여기서는 본 연구에서 제시한 음성의 분할 기법의 설계를 제시한다.

3.4.1 음성 파일의 분할

본 연구의 5장의 실험에서 제시된 바와 같이 긴 시간의 음성파일을 활용하여 키워드 검색 및 마스킹을 수행하게 될 때 발화지점 탐색에서 생성된 오차의 누적이 음성파일의 길이(재생시간)에 비례하여 나타난다. 이러한 문제를 해결하기 위해 음성 인식 API에 요청을 보내기 전, 전체 음성 파일은 Fig.4.와 같이 n개의 음성 파일 조각으로 분할된다.

이 때 음성 파일 분할 단위에 따라 키워드 혹은 패턴의 탐지가 불가능할 수 있으므로 이를 고려하여 적절한 분할 단위를 결정해야 한다. 예를 들면, 전체 음성 파일을 5초 단위로 분할한다고 가정하였을 때, 사용자가 입력한 검색 키워드나 마스킹 타깃이 5초 이내에 담기지 않을 가능성이 있다. 특히 마스킹 타깃의 경우, 주소와 같은 개인 정보를 5초 이내에 발화하는 경우는 드물다. 따라서 분할 단위는 키워드나 타깃을 충분히 담을 수 있어야 한다. 그러나 이것은 분할 단위를 음성 파일의 전체 재생 시간을 초과할 정도의 큰 단위로 설정해야 함을 의미하지는 않는다. 20초 이상의 단위로 분할하는 경우, 키워드 및 타깃의 실제 재생 위치와 알고리즘을 통해 추정한 위치에 차이가 발생할 수 있다.

본 연구에서의 추정 위치는 분할 단위가 커질수록 부정확해지는 한계점이 있으므로 마스킹을 통해 개인 정보를 노출시키지 않는 목적에 바람직하지 않다. 또

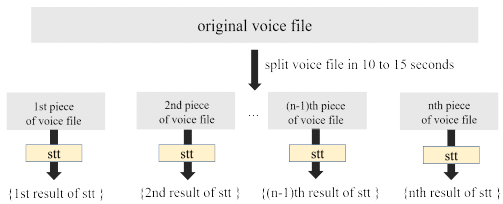

Fig. 4. Partitioning voice file to maintain accuracy of pattern detection

한, 20초 미만의 단위로 분할함으로써 사용자에게 더욱 구체적인 키워드 및 타깃의 재생 위치를 반환할 수 있다. 따라서 전체 음성 파일은 10~15초 단위로 분할하여 음성 인식 API 요청에 포함시켰으며, 구체적인 분할 단위는 서비스의 종류 혹은 음성 파일의 전체 재생 시간에 따라 가변적으로 설정할 수 있다.

3.4.2 음성 파일 중복 분할

음성 파일을 분할한 목적은 특정 키워드 및 타깃의 탐색 정확도를 높이기 위함이다. 그러나 바이트 배열이 나누어짐에 따라 분할 경계에 위치한 바이트는 이전 혹은 이후 바이트들과의 관계성을 잃게 되어 음성 인식 결과 값 자체가 불확실해지는 문제점이 발생한다. 본 연구에서는 이를 해결하기 위해 분할 범위를 중복시켜 분할 경계의 음성 인식 부정확성을 보완하였다. 예를 들어, 음성 파일의 분할 단위가 10초일 때, 5초씩 중복시킨 모습은 Fig.5.와 같다. 첫 번째 조각에서 분할 경계에 위치했던 바이트는 두 번째 조각에서 전후 바이트들과 연결되며 경계에서 벗어나게 된다. 이와 같은 분할 단위 중복을 통해 음성 인식 결과 값을 온전하게 활용할 수 있다. 이러한 설계에 대한 검증은 5장에서 실험을 통해 제시하였다.

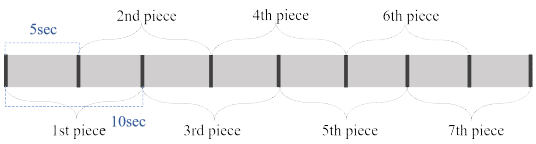

Fig. 5. Overlapping pieces in 5 seconds when partitioning in 10 Seconds

IV. 연구 결과의 구현

본 장에서는 이전 장에서의 설계를 기반으로 실제 웹 애플리케이션을 개발한 결과를 제시한다.

4.1 개발 환경

ADAM은 JSP 기반 웹 애플리케이션으로 서버 구축을 위해 AWS 클라우드 컴퓨팅 서비스를 사용하였다. Apache Tomcat을 활용한 WS(Web Server)는 사용자로부터 요청을 받아 WAS(Web Application Server)에 전달한다.

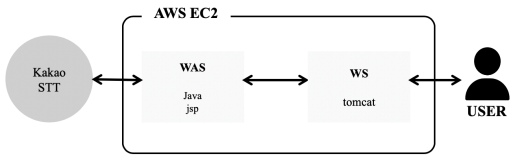

Fig. 6. Development Environment of ADAM System

WAS에서는 요청에 포함된 음성 파일을 카카오 음성 인식 API에 포함하여 요청을 전송하고 음성 인식 결과에 대한 응답을 받아온다. 여러 음성 인식 API의 테스트 결과 및 선정 조건을 기반으로 카카오 음성 인식 API를 선정한 과정은 5장에서 확인할 수 있다. 이후 WAS에서 WS를 통해 전달된 사용자 요청에 대한 결과 값을 도출하여 WS를 통해 다시 사용자에게 반환하게 된다. Table 3.는 AWS 클라우드 컴퓨팅을 통해 사용한 인스턴스의 사양이다.

Table 3. Instance Specification

Instance Specification	
Ubuntu Server	version(18.04 64bit)
	cpu(1core)
	memory(1GB)
	hdd(8GB)
Java	1.8.0
Tomcat	8

4.2 개발 결과 및 UI

ADAM은 오디오 데이터 내의 개인 신상 정보를 포함한 다양한 패턴의 검색 서비스과 마스킹 서비스를 제공하는 웹 애플리케이션이다. 보다 편리한 멀티미디어 콘텐츠의 사용을 위한 검색 서비스는 사용자가 입력한 파일과 키워드를 기반으로 이에 대한 음성 데이터 분할 및 인식을 거쳐 키워드의 재생 위치를 찾아 사용자에게 반환한다. 개인정보보호에 초점을 둔 마스킹 서비스는 음성 데이터 내에 포함된 개인 신상 정보를 무음 처리함으로써 보다 안전한 콘텐츠 생산에 기여한다. 마스킹 기능은 음성 파일에서 마스킹 타깃이 포함된 재생 위치와 마스킹된 음성 파일을 반환한다.

ADAM의 메인 화면에서 검색 혹은 마스킹 서비스를 선택하면 Fig.7.과 Fig.8.과 같이 서비스가 진행된다.

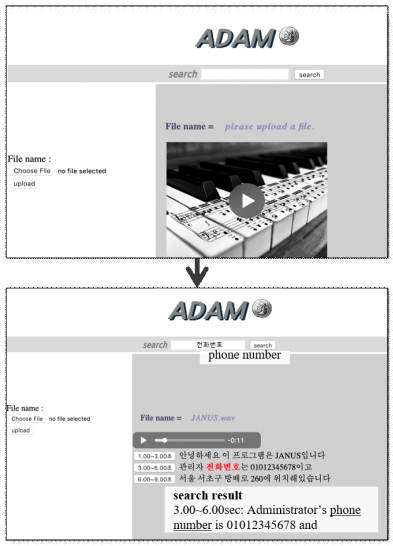

Fig. 7. Search Page of ADAM System

Fig.7.은 검색 화면으로 상단의 그림은 음성 파일 업로드 전의 모습을 나타낸다. 음성 파일이 업로드 되면 파일의 형식에 따라 영상 파일은 비디오 플레이어가 노출되고 음성 파일은 오디오 플레이어가 노출된다. 하단의 그림은 음성 파일을 업로드 한 후, 검색 키워드를 입력하여 검색 결과가 출력된 화면이다. 오디오 플레이어 하단에 음성 파일에서 입력한 검색 키워드가 재생되는 위치가 리스트로 출력된다.

Fig.8.는 마스킹 화면으로 상단 그림을 음성 파일

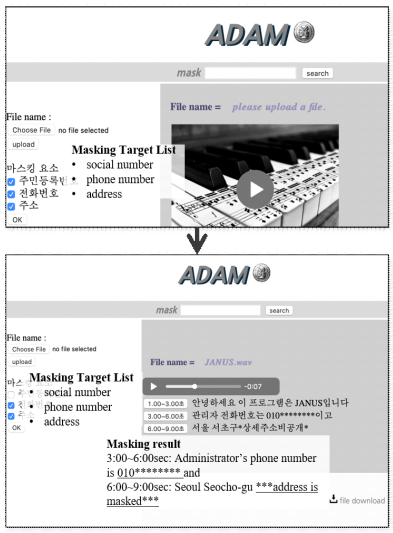

Fig. 8. Masking Page of ADAM System

업로드 전의 모습을 나타낸다. 검색 키워드를 문자열 형태로만 입력하였던 검색 서비스와 다르게 마스킹 서비스는 주민등록번호, 전화번호, 상세주소와 관련된 마스킹 타깃을 체크 박스를 통해 선택할 수 있다. 문자열 형식의 마스킹 타깃도 사용자가 직접 입력하여 추가할 수 있다. 하단의 그림은 음성 파일을 업로드 한 후, 마스킹 타깃을 입력하여 마스킹 결과가 출력된 화면이다. 출력 화면에 있어 마스킹 서비스와 검색 서비스의 차이점은 마스킹 서비스에서는 화면 우측 하단에 다운로드 할 수 있는 마스킹 된 파일이 표시된다는 것이다.

V. 음성 분할 기법의 개인 신상 정보 검출 및 마스킹을 위한 테스트 및 검증

본 장에서는 ADAM 시스템을 개발하기 위해 진행된 기술 검토 및 실험 내용을 통해 본 연구의 유효성을 제시한다.

5.1 기존 음성 인식 API 분석 및 검토

본 논문에서는 음성 인식 API인 Google의 Cloud Speech-to-Text, IBM의 Watson Speech to Text, Microsoft의 Azure Speech to Text, Amazon의 Transcribe, Kakao의 Speech-to-Text System, Naver의 Clova Speech Recognition 중 ADAM에 적합한 음성 인식 API를 선정하기 다음과 같은 전제 조건을 설정하였다.

첫째, 한국어 지원이 될 것, 둘째, REST API로 제공될 것, 셋째, 일정 사용량 이하로는 과금하지 않을 것, 넷째, 한국어 명사 및 고유명사 인식률이 어느 수준 이상일 것, 다섯 번째 한글/영어와 숫자가 번갈아 나타날 시 숫자 인식률이 어느 수준 이상일 것이다. 이와 같이 총 5가지를 음성 인식 API 선정 조건으로 설정하였다. 아래 Table 4.은 선정 조건에 따라 Google, Watson, Naver, Kakao의 음성 인식 API를 비교한 결과이다.

4가지 API 모두 한국어가 지원되고, 한국어 명사 및 고유명사의 인식률도 높았다. 또한 RESTful하게 서비스를 제공하여 HTTP 프로토콜을 따른다는 점에서 활용도가 높았다. 비용 관점에서는 Kakao만이 일정 사용량 이하로는 API 사용료를 과금하지 않았다. 각 API의 한국어 명사 및 고유명사와 숫자

Table 4. Speech recognition API analysis results

	Google	Watson	Naver	Kakao
condition 1	O	O	O	O
condition 2	O	O	O	O
condition 3	X	X	X	O
condition 4	O	X	O	O
condition 5	O	X	O	O

의 인식률은 '이름', '상세주소', '주민등록번호', '전화번호'와 같은 개인 신상 정보 형태의 데이터를 포함한 문장을 녹음한 실제 음성 파일을 이용하여 테스트해보았다. Table 5.와 동일한 문장을 발화한 녹음 파일을 각 API당 최소 5번 테스트하였으며 출력 결과들 중 각 API의 대표적인 성능을 보여주는 결과를 사용하였다.

Table 6.는 발화문장에 대한 고유명사, 숫자, 전치사의 오차 개수를 Google, Watson, Naver, Kakao API 순서로 수치화하여 표현한 내용이다.

전체적인 API의 음성인식 결과를 분석해보았을 때, 정확도 측면에서 Google의 성능이 가장 높았다. 반면, Watson은 오차 개수가 총 7개로 API 중 가장 낮은 인식률을 보였다. Watson은 문자와 숫자를 구별하여 출력하지 않았으며 고유명사의 인식률도 현저히 낮았다. Naver는 숫자 인식에 있어 오차가 적은 편이지만, 고유명사 인식에는 오차 개수가 3개로 Watson 다음으로 인식률이 낮았다. Google과 Kakao는 총 오차 개수가 2개로, 문자를 숫자화하여

Table 5. Speech Recognition test sentences

Sentence	안녕하세요 제 이름은 홍길동입니다. 저는 서울여자대학교 정보보호학과에 재학 중입니다. 저희 학교는 서울시 노원구 화랑로 621에 위치해 있고, 수업은 주로 제2과학관 ***호에서 합니다. 제 전화번호는 010 1234 5678이고, 주민등록번호는 980323 *******입니다. 지금은 음성인식 API를 비교하고 있습니다.
	Hello my name is Gil-dong Hong, I am attending the Department of Information Security at Seoul Women's University. Our shool is located at 621 Hwarang-ro, Nowon-gu, Seoul, and classes are mainly held at 2nd Science Hall ***. My phone number is 010 1234 5678, and my social security number is 980323 *******. We are now comparing the voice recognition API.

Table 6. Speech recognition API analysis results

API provider	proper noun	number	preposi-tion	sum
Google	1	0	1	2
Watson	3	4	0	7
Naver	3	1	0	4
Kakao	1	1	0	2

표현하는 정확도가 높았으며 고유명사의 인식률 또한 탁월한 것을 확인할 수 있었다.

Google과 Naver는 기본 요금이 존재하며 API 사용량이 증가함에 따라 과금량 또한 증가하는 반면, Kakao는 Kakao 측에서 설정한 기본 쿼터에 대해서는 API 사용이 무료로 제공된다. 또한 Kakao API는 이미 카카오 미니, 다음 앱, 카카오 내비, 카카오 맵, 현대자동차에 적용되는 기술로 실제 산업에 적용될 수준의 정확도를 갖추고 있으며 현재 API 업데이트가 활발히 진행되는 중이다. 이에 따라 ADAM의 서비스를 위한 음성 인식 API로 Kakao를 선정하게 되었다.

5.2 정규식 설계 및 정확도 테스트

ADAM은 개인 정보를 탐지하기 위해 정규식을 활용하였다. 개인을 식별할 수 있는 주민등록번호, 전화번호, 상세주소에 대한 정규식을 정립하였고, 정규식에 의해 패턴이 탐지된다면 해당 부분을 마스킹 처리하여 추가 정보 없이는 개인을 식별할 수 없도록 하였다. Fig.9.는 ADAM의 마스킹 타깃인 전화번호, 주민등록번호, 상세주소의 정규식이다.

전화번호의 첫 번째 그룹은 010, 011, 016~019로 지정되어있으므로 정규식에서 앞 3자리 숫자의 범위를 지정하였다. 두 번째 그룹과 세 번째 그룹은 특정 패턴이 존재하지 않고, 음성 인식 결과에서 일부 숫자가 부정확하게 출력될 가능성이 있으므로 총 7자리나 8자리의 숫자가 탐지되면 전화번호로 인식한다.

주민등록번호의 기본 정규식의 생년월일 패턴에 맞춰 작성한 앞 6자리와 나머지 7자리의 숫자이다. 그러나 주민등록번호 또한 13자리의 숫자 모두 음성 인식 결과가 정확하게 출력되지는 않을 수 있으므로 이를 고려하여 총 10자리에서 16자리까지는 주민등록번호로 인식하도록 작성하였다.

상세주소의 경우 도로명 주소에 자주 나오는 번,

1. Phone Number

```
MOBILE_PATTERN = "^(\\S*)((01)(0|1|[6-9])(\\D*)(\\d)(\\D*)(\\d)(\\D*)(\\d{1,2})(\\D*))((\\d)(\\D*)(\\d)(\\D*)(\\d{1,2}))(\\S*)(.*)$";
```

2. Social Number

```
SOCIAL_PATTERN = "^(\\S*)[0-9]{0,4}(\\S*)(0[1-9]|1[0-2])(\\S*)(0[1-9]|[12][0-9]|[3][01])(\\S*)[1-4](\\S*)[0-9](\\S*)[0-9](\\S*)[0-9](\\S*)[0-9](\\S*)[0-9]{0,2}(\\S*)[0-9]{0,2}(\\S*)$";
```

3. Address

```
ADD_PATTERN =
"^(\\S*)(\\D{2,4})((\\uB85C)|(\\uBC88\\uAE38)|(\\uAE38))(\\S*)((\\d{1,5})|(\\d{1,5})(\\uB3D9)|(\\d{1,5})(\\uCE35)|(\\d{1,5})(\\uD638))(\\S*)$";

ADD_PATTERN1 =
"^(\\S*)(((\\uC544\\uD30C\\uD2B8)|(\\uBE4C\\uB77C)|(\\uBE4C)|(\\uC8FC\\uC0C1\\uBCF5\\uD569)|(\\uC624\\uD53C\\uC2A4\\uD154)|(\\S*)(\\d{1,5})|((\\uD638)|(\\uCE35))|(\\D*)(\\S{1,4})(\\d{1,5}\\uB3D9)(\\S*)(\\d{1,5\\uD638))(\\S*)$";
```

Fig. 9. Regrex of Masking target

길, 로 등의 특정 문자들을 주소 형식에 맞게 작성하였다. 고유 명사인 건물명은 정규식을 통해 탐지할 수 없다는 한계점을 보완하기 위해 아파트, 빌라 등 건물명과 함께 사용될 수 있는 특정 단어들과 동, 호, 층 등 세부 주소를 표현하는 단어를 포함하는 정규식을 추가적으로 작성하였다.

각 정규식은 음성을 텍스트로 변환한 데이터 내에서 개인 정보를 인식하는 데에 사용되므로 개인 정보 패턴 사이에 관련 없는 단어들이 섞여 있을 가능성을 고려해야 한다. 따라서 개인 정보 패턴에 해당하는 정규식 그룹 사이에 공백 문자를 제외한 모든 문자들을 의미하는 '\S'를 포함시켜 탐지 정확도를 높였다.

Table 7.는 전화번호, 주민등록번호, 상세주소 형태의 데이터를 각 1000개씩 생성하여 정규식의 정확도를 측정한 결과이다. 총 데이터는 개인 정보 데이터를 입력하면 데이터의 문자 사이에 한글 단어, 숫자를 무작위로 삽입하는 방식으로 생성하였다. 각 개인 정보 유형별로 실제 개인 정보 패턴에 해당하는 데이터뿐만 아니라 패턴은 유사하지만 개인 정보가 아닌 데이터 또한 생성하였다.

실험 결과로는 3개의 마스킹 타깃 모두 실제 개인 정보 패턴의 데이터에 한해 100%에 가까운 탐지율을 보였다. 그러나 이는 개인 정보가 아닌 데이터를 개인 정보로 탐지하는 오탐도 발생시켰다. 통신 장비에 적용되는 서비스일 경우에는 오탐이 빈번할수록 성능에 문제가 생길 수 있지만, ADAM은 웹 애플리케이션으로 사용자의 부주의로 노출된 개인 신상 정보를 탐지하는 역할을 한다. 따라서 본 연구의 정규식은 미탐을 줄이는 데에 초점을 맞추어 개인 정보와 밀접한 패턴을 가진 데이터는 개인 정보로 탐지하도록 하였다.

Table 7. Regrex detection rate results

type of info.	input (examples)	No. of Pass	No. of Fail	rate (%)
phone number	010 123 456	188	812	18.8
	010 1234 567	1000	0	100
	010 1234 5678	1000	0	100
social number	9803 ******	117	883	11.7
	98032 ****	118	882	11.8
	980323 *****	1000	0	100
	980323 ******	1000	0	100
	980323 *******	1000	0	100
address	Apgujeong-ro, Seoul Garosu-gil Apple Store	115	885	11.5
	Hyundai***** ***_**** in Matan 3-dong, Yeongtong-gu, Suwon	1000	0	100
	Hyundai***** ***_**** on Gwonkwang-ro 260 36 Yeongtong-gu, Suwon	1000	0	100
	Seoul Women's University *** in Hwarang-ro 621, Nowon-gu, Seoul	1000	0	100

5.3 개인 신상 정보 발화 위치 탐색 실험

본 연구에서는 음성 파일 내에서 발화하는 부분과 발화하지 않는 부분을 구분하기 위해 두 경우의 Wav 형식 음성 파일을 raw 데이터인 바이트 배열로 출력하여 비교하였다. 비교 결과, 바이트 값이 진폭을 나타냄에 따라 발화하는 음성 파일의 바이트 값은 Fig.10.과 같이 절댓값이 큰 수로 이루어져 있지만 발화하지 않는 음성 파일의 바이트 값은 Fig.11.과와 같이 대부분 0에 가까운 수이다.

문자열 형식의 음성 인식 결과와 음성 파일의 데이터 바이트 배열을 통해 특정 패턴의 발화 시점을 탐색하기 위해서는 발화하지 않는 부분에 대한 고려

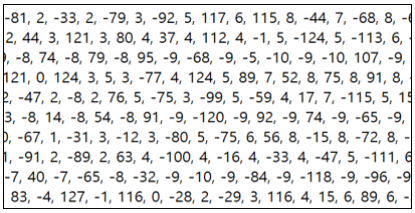

Fig. 10. Byte values of speaking part

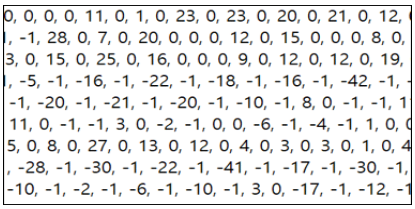

Fig. 11. Byte values of unspeaking part

가 필요하다. 문자열 형태의 음성 인식 결과는 발화 시의 공백을 고려하지 않으므로 문자열 상에서 2/4~3/4 지점에 검색 키워드 및 마스킹 타깃이 등장한다고 가정했을 때, 전체 데이터 바이트 배열 또한 2/4~3/4 지점에서 찾고자 하는 키워드 및 타깃이 나타난다고 추정한다면 실제 위치와 추정한 위치가 일치하지 않을 수 있다.

이에 따라 본 연구에서는 Fig.12.와 같이 전체 데이터 바이트 배열에서 0보다 큰 값을 가지는 바이트의 인덱스들을 모아 발화하는 부분만을 가리키는 새로운 배열을 생성하였다. 새로운 배열의 2/4~3/4 범위 값들이 찾고자 하는 검색 키워드 및 마스킹 타깃의 발화 위치가 된다. 이와 같은 방식을 통해 발화하지 않는 부분은 검색 키워드 및 마스킹 타깃의 재생 위치 탐색에 영향을 미치지 않게 되어 보다 정확하게 재생 위치를 지정할 수 있게 된다. 마스킹 기능에서는 마스킹 타깃의 위치를 탐색한 후, 이를 무음

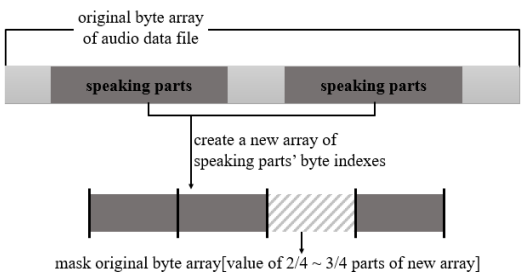

Fig. 12. Mask audio file considering speaking parts

처리 해야 하므로 위의 방식으로 찾은 마스킹 타깃의 인덱스에 해당하는 바이트 값을 모두 0으로 초기화 하였다.

5.4 개인 신상 정보의 마스킹 기능 구현 및 테스트

개인 신상 정보 마스킹 기능 구현 결과의 테스트를 위해 Table 8.와 같이 각각 이름, 주소, 전화번호, 주민등록번호를 담은 문장을 발화한 음성 데이터를 사용하였다. 이름, 전화번호 및 주민등록번호, 주소에 해당하는 음성 파일의 재생 시간은 각각 4.68초, 10.1543초, 10.847초이다.

각 음성 데이터에서 실제로 개인 신상 정보를 발화하는 위치와 알고리즘을 통해 탐지하여 마스킹 처리한 위치를 비교하여 서비스 정확도를 검증하였다. 검증 결과는 Table 9.와 같다.

실제 음성 파일을 재생하였을 때, 마스킹 타깃이 재생되는 위치와 알고리즘을 통해 마스킹 된 위치는 소수점 둘째 자리의 초 단위로 표현하였다. Table 9.의 마스킹 바이트 인덱스 범위는 알고리즘을 통해 음성 파일에서 마스킹 처리 된 실제 바이트 인덱스 범위이다. 알고리즘의 오차를 수치적으로 나타냈을 때, 개인 신상 정보 시작 지점의 표준편차는 약 0.26초이고 종료 지점의 표준 편차는 약 0.07초로 나타났다.

Fig.13.은 Table 9.의 초 단위의 실제 마스킹 타깃 범위와 알고리즘을 통한 마스킹 범위를 비교하여 나타낸 그래프이다. 이름, 전화번호, 주민등록번

Table 8. Data used to verify service accuracy

Field	Data
Name	안녕하세요 제 이름은 홍길동입니다 Hello, my name is Hong Gil Dong.
Address	저희 학교는 서울시 노원구 화랑로 621 에 위치해 있고 수업은 주로 제 2 과학관 ***호에서 합니다 My school is located in 621 Hwarang-ro, Nowon-gu Seoul and Classes are mainly held in second science building
Phone Number	제 전화번호는 010 1234 5678 이고 주민등록번호는 980323 *******입니다 My phone number is 010 1234 5678 and my social numer is 980323 *******
Social Number	

Table 9. Result of service accuracy experiment

Information Type	Actual Pattern Appearance Range	Masking Range
Name	from ≈3.5sec to ≈3.95sec	from ≈2.98sec to ≈4.03sec
Address	from ≈4.1sec to ≈9.8sec	from ≈3.44sec to ≈9.71sec
Phone Number	from ≈2.26sec to ≈4.25sec	from ≈1.76sec to ≈4.37sec
Social Number	from ≈6.6sec to ≈9.14sec	from ≈6.53sec to ≈9.37sec
Information Type	**Index Range of Masked Bytes**	**Standard Deviation**
Name	133935 ~ 181178	Start Point: 0.26 End Point: 0.07
Address	153158 ~ 431824	
Phone Number	78391 ~ 194942	
Social Number	416987 ~ 451902	

호의 경우, 알고리즘을 통한 마스킹 범위가 실제 마스킹 타깃 범위를 포괄하여 타깃이 완전히 마스킹 되었음을 볼 수 있다. 주소의 경우에는 알고리즘을 통한 마스킹의 종료 지점이 실제 마스킹 타깃의 종료 지점보다 약 0.09초가 빨라 0.09초 분량의 마스킹 타깃이 노출되는 것으로 나타났다.

마스킹 기능에서는 발화하는 부분과 발화하지 않는 부분을 고려하여 마스킹 범위를 보정하였지만 마스킹 범위와 실제 개인 신상 정보 발화 범위에 약간의 차이가 발생할 수 있다. 마스킹 범위의 오차 수준은 음성 파일의 품질, 음성 인식 결과의 부정확성 등의 알고리즘과 무관한 요인들에 영향을 받기도 한다. 예시로 나타난 마스킹 기능 구현 테스트의 오차 수준은 마스킹 범위가 크게 벗어난 경우가 아니므로 출력된 음성 파일 내의 개인 신상 정보는 모두 무음처리 되어 유출될 시에도 개인 신상 정보의 내용을 유추할 수 없는 상태이다.

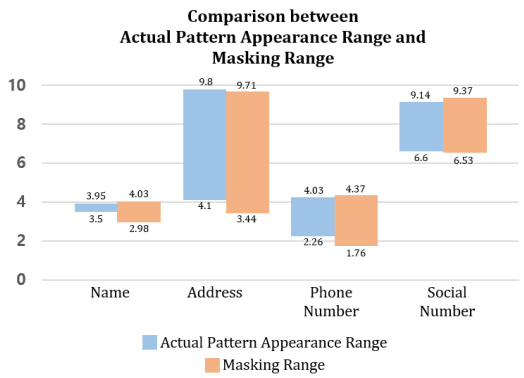

Fig. 13. Floating Column Chart of Actual Pattern Appearance Range and Making Range

VI. 결 론

본 논문은 인공지능 음성인식 API와 음성 분할 기법을 이용하여 오디오 데이터 내 정보검색과 마스킹 기능 갖는 애플리케이션 ADAM을 개발하였다. ADAM은 미디어 플레이어를 통한 재생 기능과 동시에 데이터 내의 검색 및 키워드 위치 반환을 제공한다. 마스킹 기능을 이용해 새롭게 생성된 파일을 통해 사용자의 개인정보와 원치 않는 개인 신상 정보의 노출을 최소화 할 수 있다.

구체적이고 정확한 키워드의 재생 위치와 정확한 음성 인식을 위하여 음성 파일 분할 및 분할 범위의 중복 기법을 제안하고 이를 구현하였다. 음성 파일의 분할 단위의 크기가 증가할수록 추정 위치의 오차가 증가하는 것을 파악하고 적정 수준의 분할 지점을 확인하고 음성 파일의 분할 지점을 중첩해 음성인식의 결과 값을 유지할 수 있도록 하였다. 또한 wav 형식의 음성 파일에서 바이트의 크기가 음량에 비례한 것을 이용하여 실제 발화지점을 찾아 탐색의 정확도를 높였다. 정규식은 발화 상황에 따라 개인 정보 패턴 사이에 관련 없는 말이 섞이는 상황이 존재하기 때문에 각 정규식에 공백 문자를 제외한 모든 문자를 의미하는 \S를 정규식에 포함해 작성하였다. 정규식 성능 확인을 위해 전화번호, 주민등록번호, 상세주소가 포함된 각 1,000개의 데이터에 대해 탐색하였고 실제 음성 파일에 대한 마스킹을 진행한 결과 유의미한 수준의 성능을 보였다.

향후 연구로서, 본 연구에서 제시된 ADAM 시스템은 추가적인 정규식의 설계 및 구현을 통해 여권번호, 개인통관 고유부호 등의 추가적인 개인정보에 대한 검색 및 마스킹이 가능하도록 하는 것이 필요하다. 또한, 검색 및 마스킹 대상이 되는 정보를 개인정보와 같은 개인 신상 정보에 한정하지 않고 필요정보로 확장할 경우 좀 더 다양한 응용환경에서 활용할

수 있을 것으로 보인다.

References

[1] Siwon Kim and Jaehee Chung, "Proposal of home multimedia device reflecting the behavior of video content consumption trend for binge-watchers of generation z," Journal of Industrial Design Studies, 14(1), pp. 13-24, Mar. 2020

[2] Digital Content Next, "Audio audiences are on the rise," https://digitalcontentnext.org/blog/2019/03/12/audio-audiences-on-the-rise/, Aug. 2020

[3] Silicon Hills News, "Techstars london startup OP3Nvoice moves to austin," http://siliconhillsnews.com/2013/12/16/techstars-london-startup-op3nvoice-moves-to-austin/, Aug. 2020

[4] A Smarter Video Search Engine, "Panopto," https://www.panopto.com/features/video-search/, Aug. 2020

[5] Min Ho Kim and Byung Soo So, "Personal media and fake news regulations," Korean Law Association Journal, 68(5), pp. 7-41, Oct. 2019

[6] Business of Apps, "YouTube revenue and usage statistics (2020)," https://www.businessofapps.com/data/youtube-statistics/#2, Aug. 2020

[7] Chron.com, "How to have YouTube remove someone's video," https://smallbusiness.chron.com/youtube-remove-someones-video-68262.html Aug. 2020

[8] Z. Mingguang and L. Zhitang, "A wav-audio steganography algorithm based on amplitude modifying," 2014 Tenth International Conference on Computational Intelligence and Security, pp. 489-493, Nov. 2014.

[9] Seok-Cheon Park, "Design and implementation of personal information identification and masking system based on image recognition," The Journal of The Institute of Internet, Broadcasting and Communication, 17(5), pp. 1-8, Oct. 2017

[10] YoungKyung Lee, "A method of detecting personal information in type and handwritten text using improved data learning," Master's Thesis, Chungnam National University, Aug 2018.

[11] Oh, Jong-Hoon, "A study on the prevention of the voice phishing by keyword matching," Master's Thesis, Graduate School of Soongsil University, June 2019.

[12] Hee-Kyung Roh and Kang-Hee Lee, "A basic performance evaluation of the speech recognition API of standard language and dialect using google, naver, and daum kakao APIs," Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, 7(12), pp. 819-829, Dec. 2017

[13] Seung Joo Choi and Jong-Bae Kim, "Comparison analysis of speech recognition open APIs' accuracy," Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, 7(8), pp. 411-418, Aug. 2017

〈저 자 소 개〉

김 태 영 (Kim TaeYoung) 학생회원
2017년 3월~현재: 서울여자대학교 정보보호학과 학사과정
〈관심분야〉 정보보호, 블록체인, 클라우드 보안

홍 지 원 (Hong Ji Won) 학생회원
2017년 3월~현재: 서울여자대학교 정보보호학과 학사과정
〈관심분야〉 정보보호, 인공지능

김 도 희 (Kim Do Hee) 학생회원
2016년 3월~현재: 서울여자대학교 정보보호학과 학사과정
〈관심분야〉 정보보호, 네트워크 보안, 인공지능

김 형 종 (Hyung-Jong Kim) 종신회원
1996년: 성균관대학교 정보공학과 공학사
1998년: 성균관대학교 정보공학과 공학석사
2001년: 성균관대학교 전기전자 및 컴퓨터공학과 공학박사
2001년~2007년: 한국정보보호진흥원 수석연구원
2004년~2006년: 미국 Carnegie Mellon University, CyLab 국제공동연구원
2013년~2014년: 미국 Carnegie Mellon University, ECE, Visiting Professor
2007년~현재: 서울여자대학교 정보보호학과 정교수
〈관심분야〉 안드로이드 환경의 IoT서비스 개인정보보호, 블록체인 서비스 성능평가, 클라우드 서비스 보안 모델