

고령층의 디지털 소외 방지를 위한 ASR(Automatic Speech Recognition, 음성 인식 기술) 기반 복지 정보 검색 모델 연구

하장원¹, 임화랑², 정동규³, 이해원⁴, 김영종⁵
^{1,3,4}승실대학교 소프트웨어학부 학부생
²승실대학교 글로벌미디어학부 학부생
⁵승실대학교 소프트웨어학부 교수

hos05020@gmail.com, imhwarang4@gmail.com, 1013aq@gmail.com,
ifpffk@gmail.com, youngjong@ssu.ac.kr

ASR (Automatic Speech Recognition)-based welfare information search model to prevent digital alienation of the elderly

Jang-Won Ha¹, Hwa-Rang Im², Dong-Gue Jung³, Hye-won Lee⁴
Youngjong Kim⁵
^{1,3,4}School of Software, Soongsil University
²School of Global Media, Soongsil University
⁵School of Software, Soongsil University

요 약

복지 정보와 인터넷 사용에 대한 이해도가 낮은 고령층의 디지털 소외 문제를 해결하고자, 고령층 친화 UI/UX 및 음성 인식 기술 등의 기술을 활용한 <고령층의 디지털 소외 방지를 위한 ASR 기반 복지 정보 검색 모델>의 개발을 제안한다.

1. 서론

ICT 기술이 급속도로 발전하는 오늘날, 다양한 디지털 정보 서비스가 제공되고 있지만 이를 이용하기 위해서는 일정 수준의 웹/앱 활용 능력이 기본적으로 요구된다. 그러나 이러한 서비스들은 디지털 기기의 사용에 익숙한 청년층을 대상으로 설계 및 개발되는 경우가 대부분이기 때문에 노년층이 필요한 정보를 쉽게 습득하기는 어려운 것으로 보인다[1].

본 논문에서는 디지털 정보 서비스에 대한 지식이 부족한 고령층의 사람들이 필요한 정보를 웹/앱에서 쉽게 습득하지 못하는 것을 문제로 정의했고 이를 해결하기 위해 음성 인식 기술(Automatic Speech Recognition, ASR)을 이용한 복지 정보 검색 모델을 연구하고자 한다.

이를 통해 본 연구는 고령층의 디지털 소외 방지를 해결하고 사용자들의 정보 접근성을 높이는 데 기여할 것으로 기대한다.

2. 고령층의 디지털 소외 방지를 위한 오픈 API 기반 복지서비스의 요구 사항

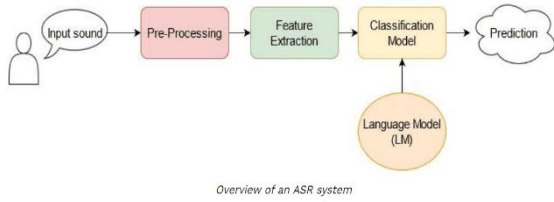
2.1 관련 연구

본 연구에서 기술적인 해결 방안을 제시하기 위하여 선행되어야 할 기술 분야들, 즉 고령층에게 친화적인 UI/UX, 음성 인식 기술(ASR)에 대한 선행연구에 관해서 서술한다.

2.1.1 고령층에게 친화적인 UI/UX

고령층이 원활하게 웹 애플리케이션을 사용하기 위해서는 적절한 텍스트, 크기 쉬운 용어, 가시적 시스템 상태 등을 잘 고려하여 UI/UX 배치를 구성해야 한다[2].

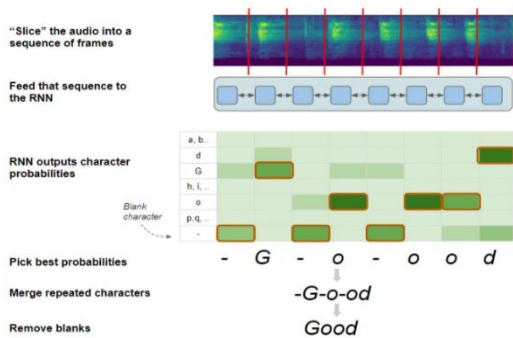
2.1.2 음성 인식 기술 (ASR)



[그림 1] ASR 처리 흐름[3]

인지능력과 운동 반응이 감소한 고령층에게 키보드의 입력 에러[4]로 인한 잘못된 복지 정보 검색은 흔히 볼 수 있는 일이다. 이러한 상황 속보다 쉽게 복지 정보를 검색하기 위해서 음성 인식 기반 복지 정보 검색 서비스가 필요하다. 이에 따라 ASR(Automatic Speech Recognition) 기술을 사용한다. 사람이 음성을 입력하면 전처리 후 특징을 추출하여 분류 모델이 이에 대한 출력 텍스트를 생성한다. 이후 언어 모델에서 분류 모델의 출력 텍스트를 인식 및 수정 작업을 수행한다.

2.1.3 Connectionist Temporal Classification



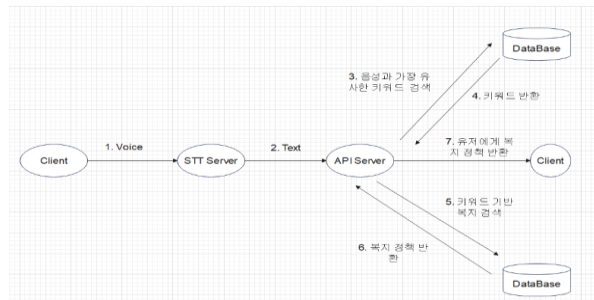
[그림 2] CTC 알고리즘[5]

음성 인식 기술(ASR)의 문제점 중 하나는 각 프레임의 경계가 정확히 어디에 있는지 알 수 없는 경우 음성이 문자로 정확히 변환되지 않는다는 점이다. 따라서 이 문제를 해결하기 위해 CTC(Connectionist Temporal Classification) 기술을 사용한다. CTC는 입력이 연속적이고 출력이 이산적이며 입력을 출력 시퀀스의 요소에 매핑하는 데 사용할 수 있는 명확한 요소 경계가 없을 때 입력 및 출력 시퀀스를 정렬하는데 사용한다. 각 프레임이 순환 신경망에 입력되면 순환 신경망은 어휘에서 각 문자에 대한 확률을 예측하고, CTC 알고리즘은 이러한 문자 확률을 가져와 올바른 문자 시퀀스를 도출한다.

2.2 요구 사항

위의 선행 연구를 바탕으로 애플리케이션 설계 단계에서 갖춰야 할 구체적인 요구사항을 제시한다. 연구에서 제안하는 해결책은 음성 인식 기술을 이용하여 음성을 텍스트로 변환하고, 변환된 텍스트로 사용자에게 복지 정보 검색을 도와주는 것을 핵심으로 하고 있다.

3. 고령층의 디지털 소외 방지를 위한 오픈 API 기반 복지서비스의 제안사항



[그림 3] 데이터 처리 흐름

본 시스템은 사용자가 찾고자 하는 복지 정보 관련된 단어를 음성으로 입력하면, ASR에서 이를 텍스트로 변환하여 서버로 전달한다. 서버에서는 이러한 텍스트를 가지고 DB 내의 복지 키워드와 비교해 텍스트와 가장 유사한 복지 키워드를 찾는다. 이후 찾은 복지 키워드와 관련 있는 복지 정보를 클라이언트에게 나타낸다.

4. 결론

본 논문에서는 고령층에게 친화적인 UI/UX 설계와 음성 인식 기반(ASR) 복지 정보 검색 모델을 제안하였다.

참고 문헌

[1]이진천.(2022). 디지털 기기의 보급과 디지털 소외 계층. 설비저널,51(9),92-93.
 [2]서울디지털재단, 고령층 친화 디지털 접근성 표준 모바일웹-앱 1.0, 2021-03-29, 10대 지침
 [3] Ilias Papastatis, "Speech Recognition: a review of the different deep learning approaches",2021-07-14,
 [4] Trewin, S. & Pain, H. 1999. Keyboard and mouse errors due to motor disabilities. Int. J. Human-Computer Studies., 50, 109-144
 [5] Ketan Doshi, "Audio Deep Learning Made Simple: Automatic Speech Recognition (ASR), How it Works", 2021-03-26,