음성 인식 서버를 이용한 모바일 사전 설계 및 구현

유재승*, 박희태*, 박미소*, 송민규**, 윤성현*
백석대학교 정보통신학부*, (주)미디어젠**

The Design and Implementation of Mobile Dictionary App based on Voice Recognition Server

Jae-Seung Yu*, Hee-Tae Park*, Mi-So Park*, Min-kyu Song**,
Sung-Hyun Yun*
Div. of Information & Communication Engineering, Baekseok University*
Mediazen, Inc**

요 약

음성 인식은 사용자의 음성을 문자로 변환하는 기술로 최근 스마트폰의 사용자 인터페이스로 사용되면서 그 활용도가 높아지고 있다. 기존의 스마트폰용 사전 프로그램은 좁은 자판으로 인하여 입력과정에 오타가 많이 발생하고, 찾고자 하는 단어의 발음은 알고 있지만 철자를 알지 못 하는 경우에 검색할 수 없다. 본 논문에서는 음성 명령을 이용한 스마트폰용 영한사전 앱을 설계 및 구현하였다. 외부 음성 인식 서버를 이용하여 음성 명령을 이해하고 클라이언트에 SQLite를 이용하여 사전 데이터 베이스를 구축한다. 클라이언트는 서버로 사용자 음성을 전달하고, 서버는 음성 정보를 텍스트로 변환하여 클라이언트에게 제공한다[1]. 클라이언트는 서버로부터 전송받은 텍스트를 이용하여 사전 데이터베이스를 검색하고, 그 결과를 사용자에게 보여준다.

1. 서론

최근 스마트폰 사용자의 급격한 증가로 사용자들의 스마트폰 앱에 대한 의존도가 높아지고 있으며 실생활에 자주 이용되는 앱에 대한 편리한 인터페이스를 요구하고 있다. 사전은 스마트폰 사용자들이 자주 사용하는 기능 중 하나로 일반적으로 스마트폰 사전을 이용하기 위해서는 영어 단어를 터치 방식의 자판으로 입력해야 한다. 스마트폰 터치 자판은 매우 좁기 때문에 텍스트 입력 시 오타가 발생할 수 있고, 길이가 긴 영어 단어를 입력할 경우에 시간이 많이 걸리게 된다. 또한, 영어 단어 발음은 알고 있지만 철자를 모르는 경우에는 사전 검색을 할 수 없는 단점이 있다.

본 논문에서는 사용자의 음성 명령을 이용한 영한사전 앱을 설계 및 구현하였다. 음성 명령을 이용하면 텍스트를 이용할 때 보다 동작 방법이 간편하고 사전에 있는 영어 단어를 검색할 때 시간적으로 수월하며 편리하게 이용할 수 있다. 스마트폰의 자판 중 'Qwerty Mode'를 선호하는 사용자들의 경우 자판 사이의 간격이 좁아 오타가 발생할 확률이 높은데 음성 명령을 이용한다면 이러한 문제점을 보완할 수 있다. 또한 음성 입력의 속도는 1급 타자 수(분당 600~800타) 정도로 일반인의 타자 입력보다 2~3배 빠른 수준이다. 게다가 사용자가 영어 단어의 철자를 모른 채 발음만을 알고 있다고 해도 그 단어의 뜻을 검색하는데 전혀 문제가 되지 않는다. 이러한 음성 명령의 장점을 이용한다면 사용자들로 하여금 앱을 사용하는데 있어서 유용성과 편리성이 증가할 것이다.

음성 인식은 사용자 음성을 텍스트로 변환해 주는 기술인데, 사람마다 인식률의 차이는 있지만 HMM(Hidden Markov Model)에 기반을 둔 최근의 음성 인식 엔진은 상용화 가능한 수준의 높은 인식률을 보여준다. 제안한 모바일 음성 사전은 외부의 음성 인식 서버를 이용하는 클라이언트-서버 방식으로 앱을 설계 및 구현하였다.

클라이언트의 음성 명령을 서버로 전송하고, 서버는 음성 인식 모듈을 이용하여 텍스트로 변환하여 클라이언트에게 전달한다. 클라이언트는 전송받은 텍스트를 Query로 미리 구축된 데이터베이스에 질의하여 해당 단어와 일치하는 단어가 있을 경우 단어의 뜻을 사용자에게 보여준다.

2 장에서는 서버 기반의 음성 인식 기법에 대해서 알아본다. 3 장에서는 제안한 모바일 음성 사전 VoiDic을 설계하고 4 장에서 기능 및 구현 결과를 설명한다.

2. 서버 기반 음성 인식 기법

음성 인식 기법은 컴퓨터가 음향학적 신호를 텍스트로 매핑하는 과정으로, 일반적으로 마이크나 전화를 통하여 얻어진 음향학적 신호를 단어나 단어 집합 또는 문장으로 변환하는 과정이다.

<그림 1> 음성 인식 프로세스

그림 1은 음성 인식 프로세스를 나타낸다. 전처리부는 입력된 음성의 질(Quality)을 개선하는 단계로 사용자 음성 데이터에 섞여있는 잡음을 제거하는 과정이다. 인식부는 음성 데이터로부터 사용자 주파수를 구하고 음성 모델 DB에 등록되어 있는 음성 사전을 기반으로 텍스트를 추출해 내는 과정이다[3].

최근의 음성 인식 인터페이스는 음성 인식 엔진을 서버에 두는 클라이언트-서버 방식이 주요 추세이다. 아이폰의 시리(SiRi), 구글, 다음, 네이버 등 대형 포털에서 제공하는 음성 인식 인터페이스가 여기에 해당한다[2]. 음성 인식 엔진이 클라이언트에 있으면 인식에 걸리는 지연 시간이 짧아서 빠른 응답을 제공하지만, 사투리 등 다양한 사용자들의 발음을 모두 고려할 수 없어서 수요자가 제한될 수 있다. 서버 기반 음성 인식은 네트워크를 이용하기 때문에 응답 시간이 가변적인 단점이 있지만, HMM 엔진을 사용하여 서버에 접속하는 사용자들의 발음에 대한 학습을 함으로써 다양한 사용자에게 질 높은 서비스를 제공할 수 있다.

HMM 기법은 음운, 단어와 같은 음성 단위를 통계적으로 모델화 한 것으로써 음성 신호의 변동을 확률적으로 취급하기 때문에 입력 음성의 흔들림 등을 잘 표현할 수 있고 대량의 음성 데이터로부터 음성 모델을 구축하는데 적격이다.

클라이언트-서버 기반 음성 인식에서 사용자는 클라이언트 앱(프로그램)에 음성을 입력하고, 음성 데이터를 서버로 전송한다. 서버는 음성 데이터를 가공, 처리할 수 있는 음성 인식 모듈을 통해서 잡음을 제거하고, 특징 값들을 추출한다. 이 값들은 단어가 저장되어 있는 음성 모델 데이터베이스를 이용하여 가장 발음이 유사한 단어를 찾아내고 유사도가 높은 단어들을 배열 형태로 구성하여 클라이언트에게 전달해 준다.

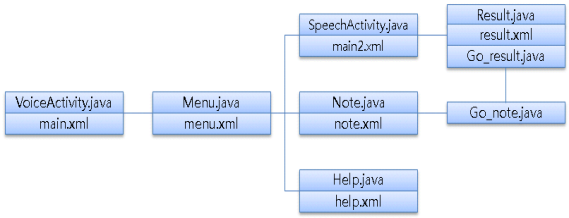
3. 모바일 음성 사전 VoiDic 설계 및 구현

제안한 모바일 음성 사전 VoiDic의 전반적인 구조와 음성 인식 인터페이스, 내장 데이터베이스 구조에 대해서 살펴본다.

3.1 프로그램 구조

그림 2는 VoiDic 앱의 전반적인 구조를 보여준다. 사용자가 음성 명령을 내리면 음성 데이터는 음성 인식 서버로 전송되고 서버에서 음성 모델 데이터베이스를 이용하여 발음이 유사한 단어의 리스트를 사용자에게 다시 보내준다. 이후 사용자는 리스트에서 원하는 영어 단어를 선택하여 개인의 단말기에 저장되어 있는 데이터베이스에 접근하여 전송받은 텍스트로 질의하게 된다. 구축되어 있는 사전 데이터베이스는 입력받은 단어와 동일한 단어를 찾아 해당 단어의 뜻을 사용자에게 보여준다.

<그림 2> 제안한 모바일 음성 사전 구조



<그림 3> 프로그램 구조도

그림 3은 모바일 음성 사전의 프로그램 구조를 보여준다. 처음 앱을 실행시키면 VoiceActivity.java 파일에 의해 main.xml 화면이 보여 진다. 적정시간(2sec)후 Menu.java 파일이 호출되며 menu.xml 화면이 나타난다. 사용자가 선택하는 메뉴에 따라 각 java 파일과 xml 화면이 나타나게 되며 검색 결과를 단어장에 추가할 경우에 Go_note.java 파일이 호출되며 그 과정에서 Note.db와 note테이블이 생성된다. 검색 과정에서는 Example.db와 recent_search 테이블이 생성되어 사전 데이터베이스가 사용자의 앱에 구축 된다.

3.2 음성 인식 인터페이스 구현

안드로이드 SDK에서 제공하는 음성 인식기(Voice Recognizer)를 사용하기 위해서는 음성 인식 액티비티를 이용해야 한다. 액티비티란 안드로이드 기반의 앱에서 사용자에게 보여지는 화면을 의미한다. 한 액티비티는 소스코드와 레이아웃 파일로 이루어지며 레이아웃 파일로 액티비티의 화면을 구성하고 소스코드로 동작을 처리한다.

음성 인식 액티비티는 사용자의 음성을 인식하여 파형을 분석하고, 액티비티 종료 후에 음성 인식 결과인 문자열의 리스트를 반환한다. 액티비티를 실행하기 위해서는 음성 인식 행위를 의미하는 시스템 상수 값을 이용하여 인텐트 객체를 생성해야 하며, 이 값은 RecognizeIntent 클래스에 정의되어 있다. 인텐트는 앱 컴포넌트가 무엇을 할 것인지를 담는 메시지 객체이며 메시지 자체일 수 있다. 인텐트를 사용하는 목적은 서로 다른 액티비티를 실행하는 것과 그 사이에서 데이터를 주고받기 위함이다.

제안한 모바일 음성 사전 앱은 RecognizeIntent 클래스

에 정의되어 있는 상수 값 중 음성을 인식하기 위한 액티비티를 시작하도록 전달하는 값인 ACTOIN_RECOGNIZE_SPEECH를 매개변수로 하는 인텐트 객체를 사용하였다.

3.3 음성 사전 데이터베이스 설계

모바일 사전 데이터베이스는 안드로이드 SDK에서 제공하는 SQLite를 이용하여 설계 및 구축하였다. 사전 데이터를 클라이언트에서 관리함으로써 사용자가 직접 단어를 추가, 삭제 및 조회할 수 있다.

사용자가 음성 인식 서버로부터 전송받은 텍스트를 이용하여 SQL 문으로 질의를 하면 데이터베이스에 저장되어 있는 단어와 비교한 후에 일치하는 단어의 뜻을 반환해 주는 방식이다. 표 1은 사전의 단어가 저장되어 있는 테이블, 표 2는 검색 후 단어장에 저장하기 위한 테이블을 보여준다.

[표 1] 사전 단어-의미 테이블

recent_search	Column	Type
	word	Text
	mean	Text

[표 2] 단어장 검색 테이블

note	Column	Type
	word	Text

사용자가 음성 인식 서버로부터 전송받은 텍스트를 이용하여 사전 데이터베이스에 질의하면 표 1의 테이블에 있는 단어들과 비교하게 된다. 또한 검색 결과 화면에서 원하는 단어를 단어장에 저장하기 위해 '단어장에 추가하기' 버튼을 선택하면 표 2의 테이블이 생성되어 해당 단어가 단어장에 저장된다.

4. 구현 결과

본 논문에서 구현한 음성 명령을 이용한 모바일 사전 VoiDic의 개발 환경, 구성 및 구현 결과에 대해서 기술한다.

4.1 개발 환경 및 메뉴 구성

VoiDic 앱은 클라이언트 앱과 사전 데이터베이스로 구성된다. 클라이언트 앱은 Android 2.3.3 기반으로 eclipse 통합 개발 툴을 이용하여 Java, XML로 구현하였고 사전 데이터베이스는 안드로이드 내장 데이터베이스인 SQLite를 이용하였다. 음성 인식 알고리즘은 구글 음성 인식 API를 사용하여 구현하였다.

4.2 구현 결과

<그림 4> 모바일 음성 사전 메뉴

VoiDic 앱은 그림 4와 같이 영한사전, 단어장, 도움말로 구성된다. 영한사전을 선택하면 단어를 검색할 수 있고 검색한 단어를 단어장에 저장할 수 있다. 또한 정확한 발음을 듣고자 하는 사용자들을 위해 TTS(Text To Speech)를 이용한 단어 발음 기능을 제공한다. 단어장 메뉴는 검색 후 저장한 단어를 리스트 형태로 볼 수 있으며, 단어를 클릭하면 뜻을 볼 수 있고, 길게 클릭하면 해당 단어 목록을 단어장에서 삭제할 수 있다.

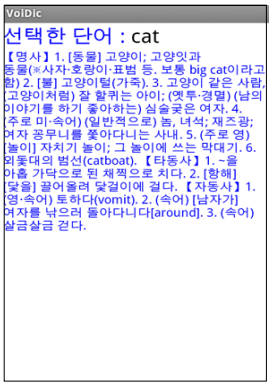
<그림 5> 인식 화면

<그림 6> 인식 결과

그림 5는 사용자의 음성을 입력받을 준비가 된 상태를 보여준다. 그림 5의 화면에서 마이크 버튼을 클릭하면 나타나는 화면이다. 이 상태에서 검색하고자 하는 단어를 말하면 사용자의 음성 데이터가 서버로 전송되고 서버 측에서 가공 후 텍스트로 변환하여 사용자에게 다시 보내준다.

구글의 음성 인식 API는 사용자의 음성 데이터와 발음이 가장 유사한 단어들의 리스트를 보내준다. Dialog를 구성하여 이 리스트를 Return 받고, 그림 6과 같이 사용자에게 인식 결과를 보여준다. 이 리스트는 사용자의 음성 데이터와 발음이 유사한 정도를 기준으로 가장 유사한 단어를 리스트의 첫 번째에 기록한다.

그림 7은 영어 단어의 검색 결과 화면을 보여준다. 해당 단어로 'cat'을 검색하였고, 단어에 대한 뜻이 화면에 출력되었다. 이 화면에서 메뉴 버튼을 누르면 단어장에 저장할 수 있는 기능과 발음을 들을 수 있는 기능을 제공한다. '단어장에 추가하기' 버튼을 누르면 Note.db 데이터베이스가 생성되고 note 테이블이 생성된다. 해당 테이블에 추가한 단어가 입력된다.



<그림 7> 검색 결과

<그림 8> 단어장

그림 8은 단어장 화면을 보여준다. 검색 결과 화면에서 추가하고 싶은 단어를 선택한 경우에 단어장에 추가된다. 상단에 보이는 버튼을 누르면 리스트가 초기화되어 모든 단어 목록이 삭제되고 단어를 한번 클릭하면 다시 해당 단어의 결과 화면으로 돌아간다. 만일 Long-Click 이벤트가 발생하면 해당 단어를 삭제할 수 있는 대화상자가 나타난다.

5. 결론

본 논문에서는 사용자의 음성 데이터를 이용하여 원하는 단어를 검색할 수 있는 스마트폰용 사전 앱을 설계 및 구현하였다. 외부 음성 인식 서버를 이용하여 음성 명령을 처리하고 클라이언트에 SQLite를 이용하여 사전 데이터베이스를 구축하였다. 클라이언트는 서버에게 사용자 음성을 전달하고, 서버는 음성 정보를 텍스트로 변환하여 클라이언트에게 변환된 텍스트를 제공한다. 클라이언트는 수신한 텍스트를 이용하여 사전 데이터베이스를 검색하고 그 결과를 사용자에게 보여준다.

Acknowledgement

* 이 논문은 2012년도 정보통신산업진흥원 IT멘토링 팀프로젝트 지원사업으로 수행된 연구임

참고문헌

[1] Eric Thelen, Stefan Besling, US6487534, 2002.

[2] 이윤근, “음성인터페이스 기술 개요 및 스마트폰 환경에서의 서비스 동향,” 한국통신학회지29(4), pp. 3-9, 2012.

[3] Johan Schalkwyk, Doug Beeferman, Francoise Beaufays, Bill Byrne, Ciprian Chelba, Mike Cohen, Maryam Garret, Brian Strope, “Google Search by Voice: A case study,“ Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics, Springer, pp. 61-90, 2010.

[4] Bou-Ghazale, S.E, “HMM-based stressed speech modeling with application to improved synthesis and recognition of isolated speech under stress,” IEEE Transactions on Speech and Audio Processing, pp. 201 - 216, 1998.