

# 基于 Android 平台的汉-英-泰互译有声电子词典<sup>①</sup>

刘 峰<sup>1,2</sup>, 王嘉梅<sup>1,2</sup>, 林 睿<sup>1,2</sup>, 李新运<sup>1,2</sup>, 张晨阳<sup>1,2</sup>, 石兵坤<sup>1,2</sup>, 杨莹玉<sup>1</sup>

<sup>1</sup>(云南民族大学 云南省高校少数民族语言文字信息化处理工程研究中心, 昆明 650500)

<sup>2</sup>(云南民族大学 电气信息工程学院, 昆明 650500)

通讯作者: 王嘉梅, E-mail: wangj\_2004@163.com

**摘 要:** 在中泰两国双边往来日益频繁, 以及 Android App 广泛应用的背景下, 设计并实现了在 Android 平台下的汉-英-泰互译有声电子词典软件. 该应用软件是以 Android Studio 为开发环境, 利用 Java 语言及 SQLite 数据库设计的. 以一种特别的方式由泰语语料库创建本地词库, 解决了 SQLite 可视化操作工具处理泰文会出现乱码的问题. 系统关键技术是采用 SQL 语言在创建好的本地词库中查找单词释义. 系统具有对话翻译、拍照翻译的特色, 还实现了汉-英-泰三语查询互译、泰语真人朗读等功能. 测试表明, 软件具备一定的便利性和实用性.

**关键词:** Android; SQLite 数据库; 汉英泰互译; 电子词典; Android Studio

引用格式: 刘峰, 王嘉梅, 林睿, 李新运, 张晨阳, 石兵坤, 杨莹玉. 基于 Android 平台的汉-英-泰互译有声电子词典. 计算机系统应用, 2018, 27(4): 54-62. <http://www.c-s-a.org.cn/1003-3254/6245.html>

## Audible Electronic Dictionary of Chinese-Thai-English Translation Based on Android

LIU Feng<sup>1,2</sup>, WANG Jia-Mei<sup>1,2</sup>, LIN Rui<sup>1,2</sup>, LI Xin-Yun<sup>1,2</sup>, ZHANG Chen-Yang<sup>1,2</sup>, SHI Bing-Kun<sup>1,2</sup>, YANG Ying-Yu<sup>1</sup>

<sup>1</sup>(Yunnan Province for Minority Language Information Processing Engineering Research Center, Yunnan Minzu University, Kunming 650500, China)

<sup>2</sup>(School of Electrical & Information Engineering, Yunnan Minzu University, Kunming 650500, China)

**Abstract:** In the emerging growing of bilateral exchanges between China and Thailand and the widespread use of Android apps, the study has designed and implemented Chinese-Thai-English translation audible electronic dictionary APP on the Android platform. The application is designed based on Android Studio development environment and it uses Java language and SQLite database. This study creates a local word library from the Thai language corpus in a special way, and solves the problem that SQLite visualization tool processing Thai will appear garbled. The key technology of the system is to use SQL language to inquire word meaning in the created local word library. The system has the features of dialogue translation and photo translation, and it also has the functions of trilingual query translation among Chinese, Thai, and English, and the pronunciation of Thai language by real people. The test shows that the software has certain convenience and practicability.

**Key words:** Android; SQLite database; Chinese-Thai-English translation; electronic dictionary; Android Studio

自 Google 在 2008 年正式发布 Android 手机操作系统, 并公开其所有原始码之后, Android 便成为手机第一个完整的开放开发平台<sup>[1]</sup>. 而由于其开源的特性, Android 操作系统已成为智能手机的主流. 全球众多的

开发者支持 Android 应用开发, 越来越多的软、硬件厂商, 也竞相投入 Android 相关研发计划. 可以预见的是, 未来越来越多的用户将使用 Android 智能手机, 各种功能的 Android 手机应用软件将供不应求.

① 基金项目: 国家自然科学基金地区项目 (61363085); 云南省教育厅科学研究基金重大专项 (ZD2013013)

收稿时间: 2017-06-06; 采用时间: 2017-07-12; csa 在线出版时间: 2018-03-31

与此同时,自中国国家主席习近平提出“一带一路”重大倡议以来,得到国际社会高度关注.而云南省在其中的定位是,发挥区域桥头堡的优势,打造大湄公河次区域经济合作新高地,建设成为面向南亚、东南亚的辐射中心.云南省是中国离泰国距离最短的省份,作为中国和泰国交流的排头兵,云南通过自身与泰国的合作,不断推动着中泰经贸的发展.近年来,中泰经贸关系持续稳定增长,中国是泰国最大贸易伙伴,泰国是中国在东盟国家中第四大贸易伙伴.在旅游方面,中国已成为泰国第一大入境旅游客源国.在人员交流方面,目前有2万多泰国留学生在中国学习,人数在东盟国家中是最多的,而中国则是泰国最大的留学生生源国.未来,中泰两国还将共同推进“一带一路”建设,打造亚洲命运共同体,成为国家发展睦邻友好与互利共赢的典范.

语言作为人们沟通的工具,在中泰双边往来当中发挥着重要作用.随着汉语与泰语翻译的需求量急剧增加,如何快速且方便地翻译两国语言,对于中泰两国人民来说十分重要.传统的翻译方式大多是采用纸质词典,不能快速便捷地翻译,而聘请翻译官费用又十分昂贵,并且专业的中泰翻译人员有限,不能满足大众的需求.而移动互联网的崛起和 Android 智能手机的普及,为电子词典的发展提供了契机,正好可以解决以上的问题.设计一款基于 Android 智能手机的汉-英-泰电子词典,最大限度的满足了人们随时随地查询词典的需要,成为人们查询词典及移动学习的首选.本设计虽然是一个个案,但对于其他语种(如彝文等)信息处理的理论指导,以及泰语机器翻译系统相关领域的研究具有广泛而深远的意义.

## 1 电子词典现状分析

### 1.1 Android 平台上的电子词典

目前 Android 应用市场上的电子词典应用软件非常多,常见的电子词典涉及中、英、韩、日、德、法、西等主流语言的翻译,例如金山词霸、沪江小D词典、网易有道词典、欧路词典等应用.而像百度翻译、有道翻译官、微软必应词典等应用可选择更多语种翻译,包括爱沙尼亚语、泰语、越南语等小语种的翻译.市场上的这些应用词典基本具有查询、生词本、语音发音、句子翻译等功能,但这些客户端通常占用空间和资源较大,并且需要联网才能翻译一些小

语种语言,使用语音发音功能也需联网,使用网络流量较大,对用户的网络流量和网络传输速率要求较高.

### 1.2 汉-英-泰电子词典发展现状

当前存在一些涉及泰语翻译的 Android 词典软件,如微软必应词典、百度翻译等都可以在众多语种中选择泰语翻译,但这些应用只能汉译泰或英译泰的双向互译,并不能汉-英-泰三语同时互译,有的应用不能翻译泰语,有的可以翻译包括泰语的众多语种,但专门翻译汉-英-泰的软件少之又少.这些涉及泰语的电子词典使用其相关功能时需要网络流量,而且几乎没有能从图片中识别泰文的应用.而本文设计的汉-英-泰互译有声电子词典是一款专门针对汉-英-泰互译的词典 APP,可实现汉-英-泰三语同时互译功能,能准确翻译泰语,具有原汁原味泰语真人发音功能,有丰富的本地词库,实现真正意义上的随时随地汉-英-泰查询互译,还具有对话翻译和拍照翻译功能,填补了泰语电子词典应用市场的空白.

## 2 系统相关技术介绍

### 2.1 Android 系统简介

Android 是 Google 于 2007 年 11 月推出的一款基于 Linux 平台的开源手机操作系统.它是一种基于 Linux 的自由及开放源代码的操作系统,主要用于移动设备.如今 Android 已成为市场占有率最高的手机操作系统<sup>[2]</sup>. Android 系统架构粗略地可分为以下 5 个部分: Linux 内核层、系统库、Dalvik/Art、应用框架层、应用程序层<sup>[3]</sup>, Android 系统架构如图 1 所示.在 Android Studio 开发环境下开发 Android 应用,官方推荐采用视图和逻辑分离的原则,也就是使用 XML 文件来控制界面布局,然后通过 Java 代码来控制这些布局和组件的功能<sup>[4]</sup>.

### 2.2 Android 开发环境搭建

文章设计的汉-英-泰电子词典是在 Android Studio 开发环境下完成实现的.搭建 Android Studio 开发环境,首先要下载并配置 Java JDK,在 Oracle 官方网站下载最新的 JDK 后,将其安装至电脑上,然后配置计算机的 JAVA\_HOME、CLASSPATH、Path 三个环境变量.其次在 Android 开发者官网,下载并安装 Android Studio,至此,Android 开发环境便搭建好了. Android Studio 集成开发工具界面如图 2 所示.

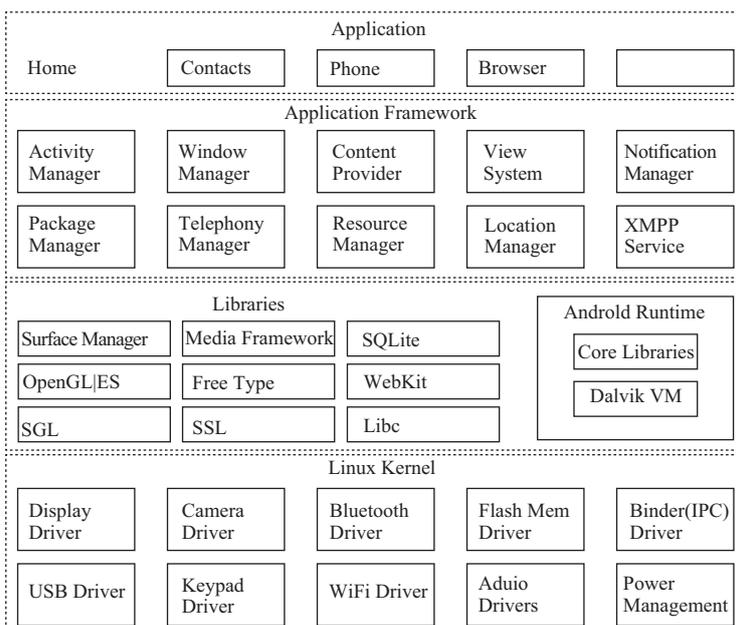


图1 Android 系统架构

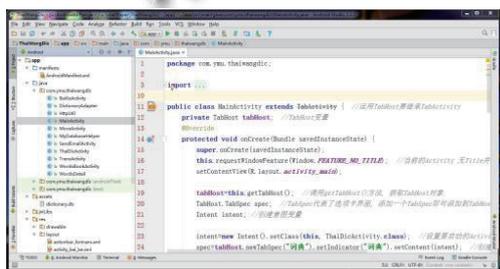


图2 Android Studio 集成开发工具界面

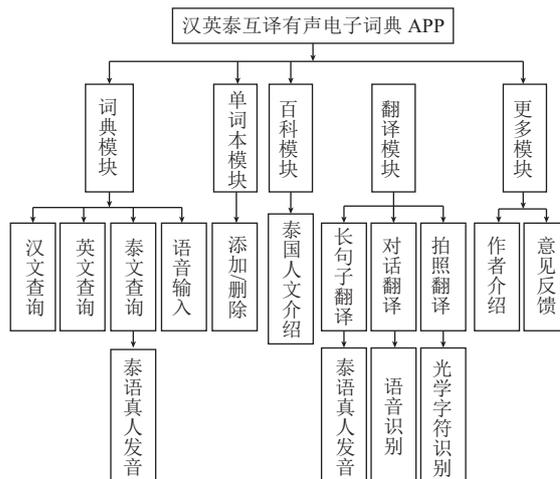


图3 系统总体功能模块图

### 2.3 Android 数据持久化技术

数据持久化技术就是将数据保存至存储设备中, 保存在存储设备中的数据是处于持久状态的. Android 系统内置了 SQLite 数据库, SQLite 是一款轻量级的关系型数据库, 它的运算速度非常快, 占用资源很少. SQLite 不仅支持标准的 SQL 语法, 还遵循了数据库的 ACID 事务, 文章设计的基于 Android 平台的汉-英-泰电子词典 APP 的本地词库, 就是通过 SQLite 数据库实现的.

## 3 电子词典客户端研究概括

### 3.1 系统功能结构总体设计

本系统主要可分为五大功能模块: 词典模块、单词本模块、百科模块、翻译模块和更多模块. 本系统的各功能模块的组成如图3所示.

#### (1) 词典模块

此模块为本系统的关键模块, 主要是实现汉-英-泰互相查询的功能, 用户根据实际需要, 输入汉英泰任何一种语言的单词或词组, 点击查询按钮后, 便可查询检索出其对应的其他两种语言的释义, 并且可实现对应的泰语的发音, 还可将生词添加至单词本中.

#### (2) 单词本模块

主要是方便个人学习的, 在此模块可以查看生词, 添加或删除生词, 以达到复习生词的目的.

#### (3) 百科模块

主要包含了一些泰国人文信息, 例如泰国旅游实

用口语、泰国名言警句、泰语基础知识等,都可在此模块查看.

#### (4) 翻译模块

针对长语句翻译,在此模块可自动检测输入的语言类型,也可手动选择翻译类型,包括对话翻译和拍照翻译功能.

#### (5) 更多模块

主要是一些其它的功能,包括作者介绍和意见反馈等.

### 3.2 语料库建设

本课题组在工程中心机构的支持下,结合我校东南亚语言学科的优势,发挥少数民族语言文字、自然语言处理、电子信息等文理交叉、多学科交融综合研究的特色,建立了泰语语料库.语料库的建设是在我校国际交流学院留学生,民族文化学院研究生等协助下,通过收集高等学校泰语专业教材等方法,在对话料纠正、翻译和标注后完成的<sup>[5]</sup>.最终的泰语语料库是以 Excel 表形式呈现的,语料库共包含 29945 条泰语词汇及其汉语和英语的释义,这将为后续泰语词汇数据库建设提供帮助.部分泰文词汇表如图 4 所示.

num	wod	s	eng	chi
6-00019	ตา	n	eye(s); mater	眼睛; 外祖父;
6-00020	ทวด	n	great-grandp	曾祖父母; 外
6-00021	ทวดน้อย	n	great-grandu	大师-伯叔祖;
6-00004	บรรดาศักดิ์	n	the conferre	贵族的等级被
6-00022	บรรพบุรุษ	n	eleg.ancesto	祖宗; 祖先
6-00023	บุพการี	n	parent; ances	父母; 祖先
6-00005	ประยูร	n	family; clai	家族
6-00024	ปู่	n	grandfather;	祖父
6-00006	พี่	n	gentleman; la	绅士; 淑女; 一
6-00007	พงศ์	n	[pertaining	后裔; 家族
6-00002	มองโกลเลีย	n	Mongolia	蒙古人; 蒙古
6-00003	มองโกลเลียน	n	Mongolian	蒙古人
6-00001	มองโกล	n	Mongolian	蒙古族
6-00008	มอญ	n	Mon; Pegu; Peg	孟族
6-00009	มูเซอ	n	N; Musser (La	慕瑟; 一个山
6-00010	ลอร์ด	n	noble; nobi	贵族
6-00011	ลื้อ	n	Lue; a Thai e	你; 泰仂族; 来
6-00012	ศากย	n	Sakya; an anc	一个印度的远

图 4 泰-汉英语料库部分展示

### 3.3 词汇数据库建设

本设计的目的是在 Android 智能手机的环境下,实现具有汉-英-泰词汇相互查询功能的电子词典,其核心是关于汉-英-泰词汇查询检索的功能,本设计决定采用本地词库的方式实现上述功能,因此建立汉-英-泰词汇数据库是必要的.此外,生词本的建立以及显示历史查询记录等功能,也需要将数据储存至数据库表中.以上

提到的功能,是依托建立相关的 SQLite 数据库来实现的.

#### (1) 建立汉-英-泰词汇数据库

经仔细分析,对于在 Android Studio 开发环境下创建的项目,当项目中创建一个 SQLite 数据库时,SQLite 数据库就是一个后缀为“.db”的文件,它是保存在 data/data/包名/databases 目录下的,其中包名是指当前已经创建数据库的项目的包名.操作 SQLite 数据库是通过 Java 语句实现的,每当在数据库表中添加一个词汇就要编写一条 Java 语句,如果有很多词汇的话,创建数据库时就要编写大量的 Java 语句,这是不现实的也是十分耗费项目编译时间的.因此需要在另一个项目中创建好所需的 SQLite 数据库,然后将此 SQLite 数据库导入所需的项目中.

由此可知,把汉-英-泰词汇数据库做成“.db”文件,再将其导入至项目的 data/data/包名/databases 目录下,项目就可以不用创建数据库,而直接使用 SQLite 数据库了,本设计就是通过项目中导入数据库实现的.

本课题的泰语语料库是 Excel 表格形式的,表格中有 29945 条泰语词汇及其对应的英汉释义,将 Excel 表格转化成后缀为.db 的 SQLite 数据库文件,通常的做法是利用可视化的 SQLite 操作工具(如 SQLite Expert Professor、Navicat 等),但在实际操作过程中,Excel 转换为 CSV 格式后,无论是编码方式改为 UTF-8 还是用记事本打开文档,泰文都出现了乱码的情况,故不采用上述的方法.我们创造性地想出了一个新的方法,此方法的流程如图 5 所示,其基本思路是:在 Android Studio 中另外建一个项目用于构建汉-英-泰词汇的 SQLite 数据库,首先在这个新建的项目 app/libs 文件夹中导入操作 Excel 表格的 jar 包;其次,将泰语词库的 Excel 表格复制至项目的 assets 文件夹中;再次,在项目的 Java 代码中编写循环语句,使得项目可读取 Excel 表格每一行的内容然后添加至 SQLite 数据库表中;最后,在 File Explorer 中将创建好的 SQLite 数据库(“.db”文件)拷贝至电脑上.之后便可将这个已经创建好的汉-英-泰词汇 SQLite 数据库,导入至所需要的项目中,就可以在项目中使词汇数据库了.

#### (2) 其它数据库表的建立

本设计除了需要汉-英-泰词汇 SQLite 数据库之外,为了实现生词本及历史查询记录的功能,还需要在 SQLite 数据库中建立两张表.在项目中新建 MyDatabaseHelper 类继承自 SQLiteOpenHelper,在这

个类中建立 words 表和 history 表,用于存放生词和历史查询记录的数据,他们的结构如表 1 和表 2 所示。

径,然后通过路径将音频文件取出,然后交给媒体播放器播放,就实现了软件中朗读泰文的目的。



图 5 构造词汇数据库过程示意图

表 1 words 表结构

字段名	数据类型	备注
_id	Integer	主键
Thai	Text	泰文
Chinese	Text	中文
English	Text	英文
Number	Integer	录音编号

表 2 history 表结构

字段名	数据类型	备注
_id	Integer	主键
First	Text	所查的词
Second	Text	第一释义
Third	Text	第二释义
Number	Integer	录音编号

### 3.4 语音库建设

我们利用构建好的泰语语料库,以泰语文字和标注序号作为录音的依据,采用前期工作开发的原音录音软件,在安静的实验室里完成了整个泰语录音过程。录音人员是我校汉语国际教育专业的泰国留学生,能够保证泰语录音的原汁原味。最终我们得到了以 WAV 为后缀的泰语录音文件,录音成果如图 6 所示。至此,课题组建立了泰语语音库,这将为后续汉-英-泰电子词典应用的泰语发音功能提供支撑。

数据库中我们只存储声音音频文件的路径,具体的声音文件则是存储于手机 SD 卡中,当需要播放某泰语录音文件时,只要通过数据库读取其录音文件的路

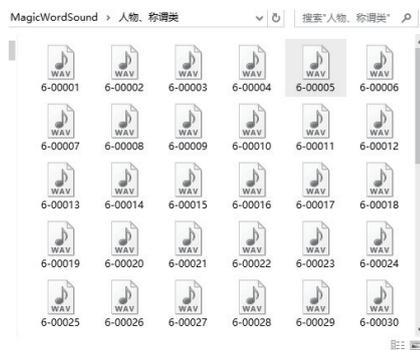


图 6 泰语音频录制成果展示

## 4 电子词典客户端功能实现

### 4.1 项目结构

在 Android Studio 中创建一个名为 ThaiWangDic 的项目,以此项目来开发汉-英-泰电子词典,Android 模式下的项目目录结构如图 7 所示。其中 asstes 文件夹中存放的是已经创建好的汉-英-泰词汇 SQLite 数据库文件,当安装此应用时,程序会将此词库导入至相关目录中。而 res/layout 文件夹中存放的是应用各种界面布局的 XML 文件<sup>[6]</sup>。Java/com.ymu.thaiwangdic 中存放的则是包含各种界面逻辑功能的 Java 代码<sup>[7]</sup>。Android Manifest.xml 文件则是整个 Android 项目的配置文件,Activity 等组件需要在这个文件里注册,另外还有一些权限申明要在其中添加<sup>[8]</sup>。

### 4.2 本地词库的实现

将之前创建好的汉-英-泰词汇 SQLite 数据库(.db 文件)导入项目中,在安装应用程序时词库会加载至对应路径上,这就实现了词库的本地化。在安装应用时,程序会检查是否存在数据库文件的目录,若不存在则创建目录,若存在则将项目 assets 文件夹中的.db 文件加载到该目录中。

### 4.3 系统主界面

系统主界面采用的是底部导航栏的形式,实现界面的切换。底部的导航栏存放了多个选项卡,包括:词典、单词本、百科、翻译和更多五个部分,点击选项卡可切换至相关的界面。系统主界面默认一开始是停留在词典界面的,主界面如图 8 所示。

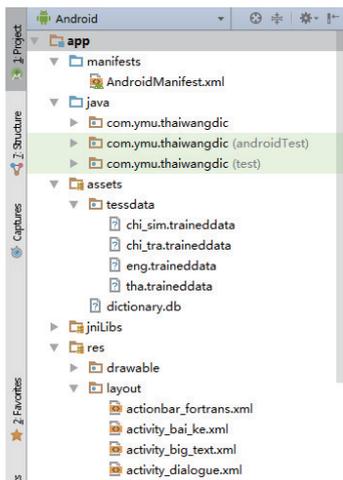


图7 Android 模式下的项目结构



图9 词典界面

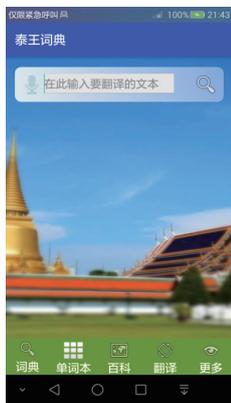


图8 系统主界面

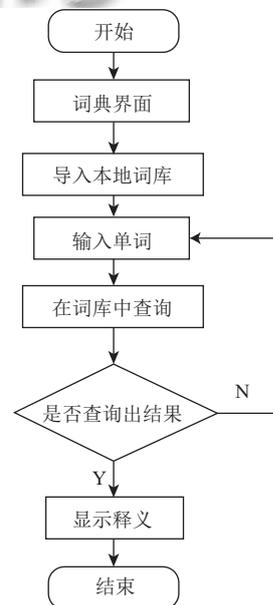


图10 单词查询流程图

#### 4.4 词典界面

词典界面主要是实现汉-英-泰相互查询的功能, 这个功能也是应用软件的核心功能. 当安装应用程序时, 该软件会使用 `openDatabase()` 的方法在词典界面的 `onCreate` 方法中导入汉-英-泰词汇 SQLite 数据库, 代码如下: `db=openDatabase()`; 其中 `db` 是定义的 `SQLiteDatabase` 类型变量. 当用户点击界面上方的文本输入框, 输入汉泰英任意一种语言的单词或词组, 点击查询按钮后, 在查询按钮的单击事件中, 会调用 `db` 的 `query` 方法在数据库中查找相应的单词释义, 并将查询结果显示出来. 运行界面如图9所示. 该界面的文本输入框由 `AutoCompleteTextView` 组件实现, 当用户输入两个字符后, 将进行模糊查询, 检索到与前两个字符相同的单词, 以下拉菜单的形式显示出来, 实现模糊查询. 其整个流程如图10所示.

部分关键 Java 代码如下 (以查询英文为例):

```

Cursor cursor = db.query("chinathai", null,
"english=?", new String[]{word.getText().toString()},
null, null, null); //调用 query() 方法查找单词释义
if(cursor.getCount() > 0) { //若查询有结果
cursor.moveToFirst(); //数据的指针移到第一行
...
result =cursor.getString
(cursor.getColumnIndex("taiwen")); //取出泰语释义
...
cursor.close(); //关闭数据的指针
txv_word.setText(result); } //显示释义
else{txv_word.setText("抱歉, 没有本地释义"); }
    
```

基于 Android 平台的电子词典除实用性与便捷性之外,与传统纸质词典最大的区别在于可以实现单词语音发音.在词典界面中,当查询结果显示出来后,界面的右上角有一个像喇叭一样的按钮,点击此按钮,将会播放此单词对应的泰语真人发音.还有一个类似五角星的按钮,点击此按钮可将当前查询到的单词及其相关的释义保存至单词本中,以便日后查看和复习使用.

#### (1) 泰语真人发音模块

此软件使用 Android 中的 MediaPlayer 类来实现音频的播放,将所有词汇的泰语真人发音资料保存于手机的 SD 卡当中,项目中首先创建一个 MediaPlayer 对象,当查询检索单词有结果时,调用 setDataSource() 方法来设置对应词的泰语音频文件的路径,再调用 prepare() 方法使 MediaPlayer 进入到准备状态,当播放按钮按下时,便会调用 start() 方法播放相关音频了.最后当播放完音频后需调用 reset() 方法,这样下次播放才会正常.还要注意的,要在项目中加入访问 SD 卡的权限,才能成功读取 SD 卡中的音频文件并播放.

#### (2) 添加生词模块

点击添加生词按钮,系统便会将查询到的单词添加至之前创建好的数据库的 words 表中,以便在单词本界面中查看.

### 4.5 单词本界面

单词本界面主要是显示用户添加的生词,方便用户浏览,点击一个生词会显示其详细的释义.在此界面,用户可以添加或删除生词,管理生词十分方便,通过单词本界面,用户可以及时复习未掌握的生词.在此界面,系统运用的是 ListView 组件显示生词的,当界面启动时,系统用 cur = db2.rawQuery("SELECT \* FROM words", null);语句查询数据库 words 表中的所有词汇数据,然后通过 SimpleCursorAdapter 配置器将生词显示于 ListView 组件中.此界面如图 11 所示.

### 4.6 百科界面

在百科界面,介绍了一些泰国实用信息,如泰国旅游实用口语、泰国名言警句、泰语基础知识等.界面如图 12 所示.此界面运用了 ScrollView 组件嵌套一个 TextView 组件来显示泰国的实用信息,之所以实用 ScrollView 组件是因为它很方便,它可以允许用户通过滚动的方式查看屏幕以外的内容.



图 11 单词本界面

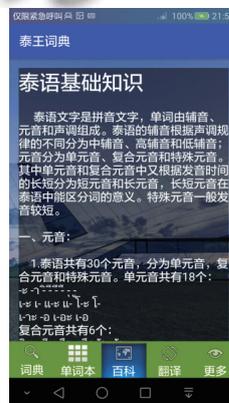


图 12 百科界面

### 4.7 翻译界面

翻译界面的功能和词典界面的类似,只是在此界面我们可以翻译较长的语句,界面顶部有一个下拉选项框,可以选择翻译的类型,默认是语言自动检测,即自动识别输入的语言类型,然后将其翻译出来.右上角有对话翻译和拍照翻译按钮.界面如图 13 所示.



图 13 翻译界面

#### 4.8 更多界面

更多界面是为了实现一些其它更多的功能, 以更好地迎合用户的需要. 此界面包含了设置、意见反馈、检测新版本、关于等模块. 其中关于模块是对软件的作者介绍及申明版权所属的. 此界面如图 14 所示.



图 14 更多界面

### 5 系统特色

#### 5.1 特色一: 调用百度语音技术实现对话翻译

对话翻译是指用户对着应用说话(中文或英文)后, 应用立即翻译出对应的释义并播放相应的泰语真人发音, 实现即听、即译、即说的效果. 而利用百度语音识别技术可将用户的语音转化为文字, 运用百度语音识别技术, 首先要在项目的 `AndroidManifest.xml` 文件中添加使用麦克风录音等权限; 其次在百度语音开放平台上注册申请 API Key 和 Secret Key, 并下载相关 jar 包和 so 库; 然后在项目的 `app/libs` 中添加 jar 包, 在 `armeabi` 和 `mips` 中添加 so 库; 最后在对话翻译 `dialogueActivity.java` 的 `onCreate` 方法中设置 API Key 和 Secret Key 等相关参数, 并实例化百度语音识别对话框. 在百度语音识别回调接口中重写 `onResults` 方法, 使其用识别的文字查询 SQLite 词汇数据库, 将释义显示出来并由 `MediaPlayer` 播放泰语发音, 便可实现对话翻译功能. 对话翻译界面的 `activity_dialogue` 布局主要运用了强大的滚动控件 `RecyclerView`, 它定义在 `support` 库当中, 在项目的 `build.gradle` 中添加相应的依赖库便可使用, 添加后需点击 `Sync Now` 来进行同步. 对话翻译界面如图 15 所示, 百度语音识别对话框如图 16 所示.

#### 5.2 特色二: 运用 Tesseract-OCR 引擎实现拍照翻译

Tesseract 是开源的 OCR (Optical Character Recognition) 引擎, 可识别多种格式的图像并将其转换成文本, 目前已支持 60 多种语言. 本系统利用 Tesseract-OCR 引擎识别图片中的文字, 然后翻译实现拍照翻译

功能. 由于 Tesseract 是以 C++ 实现的, 在 Android 上运用需要封装 Java API, 而 `tess-two` 是 Tesseract Tools for Android 的一个 Git 分支, 在 Android 项目的 `build.gradle` 中添加 `tess-two` 的依赖库便可运用 Tesseract 技术.



图 15 对话翻译界面



图 16 百度语音识别对话框

为了运用 Tesseract-OCR 引擎识别图片中的文字, 还需将 Tesseract 识别文字的字库 (中英泰字库) 放入项目的 `assets/tessdata` 文件夹中, 应用安装时会把字库复制至 SD 卡中, `AndroidManifest.xml` 文件也要加入调用相机等权限. 拍照界面如图 17 所示, 其布局 `activity_take_photo` 中添加了自定义的相机 `CameraPreview` 并在 `TakePhotoActivity.java` 的 `onCreate` 方法中实例化自定义相机、拍照按钮和顶部 `Spinner` 组件, `Spinner` 组件用于选择文字识别类型, 当用户点击拍照按钮或从相册中选择一副图后, 系统会设置 `Visibility` 属性值使拍照界面消失, 而将截图界面显示出来, 如图 18 所示. 在截图界面确定按钮的监听器中, 新建 `Intent` 对象的 `startActivity` 方法会开启识别界面, 如图 19 所示. 在识别界面 `ShowCropperedActivity.java` 中, 为了提高识别的准确性, 图片先进行灰度化处理, 再由新建

TessBaseAPI 对象 baseApi 的 init 方法设置字库 SD 卡路径、语言类型等相关参数后, 在子线程中执行 getUTF8Text 方法即可以 UTF-8 编码类型识别灰度化图像的文本, 翻译此文本即可实现拍照翻译功能。



图 17 拍照界面



图 18 截图界面



图 19 识别界面

## 6 系统测试

测试系统, 可以检测其整体性能和运行效率是否符合设计需求。软件测试需要一个运行的载体, 可以是

一部 Android 手机, 也可以是 Android 模拟器 (Android Virtual Device, AVD)。在本软件开发完成后, 我们在 Android Studio 上创建的 Android 模拟器 (Android 5.1, API Level 为 22) 和一台华为 KIW-TLOOH 型号的 Android 手机 (Android 5.1.1) 上, 分别进行了软件测试。测试结果为: 软件各项性能, 均达到预期效果, 软件的界面和功能实现比较理想, 系统运行稳定流畅, 易于操作, 具备一定的便利性和实用性。

## 7 结束语

目前, Android 在移动领域已经得到广泛应用, Android 手机应用软件发展迅猛, 几乎各行各业的企业和公司都推出了与其服务相关的 Android 版移动客户端<sup>[9]</sup>。本文利用 Android Studio 开发平台, 使用 Java 语言及 SQLite 数据库设计并开发了在 Android 平台下的汉-英-泰互译有声电子词典应用软件, 实现了关于汉-英-泰的单词查询、生词本录入、语音输入、泰语真人朗读、句子翻译、对话翻译、拍照翻译、百科知识等功能。系统界面美观大方, 贴近用户, 运行稳定流畅, 易于操作, 具备一定的便利性和实用性。

在中泰两国经贸交流越发频繁的背景下, 软件可提供给赴泰的游客、泰国留学生以及其他的泰文学习者使用, 使用者可以随时随地学习泰语并且能及时应用泰语与他人进行交流, 相信此软件将会对他们的泰语和汉语学习起到较好的促进作用。

## 参考文献

- 1 郭霖. 第一行代码: 2 版. 北京: 人民邮电出版社, 2016.
- 2 施威铭. Android APP 开发入门. 北京: 机械工业出版社, 2016.
- 3 郭金尚. Android 经典项目案例开发实战宝典. 北京: 清华大学出版社, 2013.
- 4 黄艺锋, 闫巧. 基于 Android 平台电子词典的设计与实现. 计算机应用, 2011, 31(S2): 228-232.
- 5 胡刚, 王嘉梅, 李炳泽, 等. 汉英-泰互译有声语料的数据库研究. 计算机系统应用, 2016, 25(9): 223-229. [doi: 10.15888/j.cnki.csa.005242]
- 6 李肖. 基于 Android 的电子词典系统设计与实现[硕士学位论文]. 长春: 吉林大学, 2013.
- 7 王佳琪. 基于 Android 的电子词典手机客户端的设计与实现[硕士学位论文]. 北京: 北京交通大学, 2012.
- 8 周宇, 尹生强, 王冬青, 等. 基于 Eclipse 和 Android 系统的 App 开发平台搭建研究. 青岛大学学报, 2016, 31(3): 49-53.
- 9 赵志成. 基于 Android 平台的维汉多媒体电子词典的设计与实现[硕士学位论文]. 乌鲁木齐: 新疆大学, 2013.