

## 基于智能设备的汉语学习软件的研究与开发\*

米尔阿迪力江·麦麦提<sup>1</sup>, 吾守尔·斯拉木<sup>1,2†</sup>, 迪丽尼格尔·热夏提<sup>1</sup>, 帕尔哈提·达吾提<sup>3</sup>

(1. 新疆大学 信息科学与工程学院, 新疆 乌鲁木齐 830046; 2. 新疆大学 新疆多语种信息技术重点实验室, 新疆 乌鲁木齐 830046;  
3. 新疆大学 讯飞语音及语言联合实验室, 新疆 乌鲁木齐 830046)

**摘要:** 本文研究了汉语教学软件在Android 开发平台下的开发与实现过程, 系统主要由Android软件开发工具包和Eclipse集成开发环境进行开发, 利用Android的API函数接口, 通过自动选型规则实现了维汉英多种文字的正确显示及处理等问题, 进而给用户带来更加方便的汉语学习环境. 在存储及处理大量的相关图片和音频文件时, 利用了Android集成的小型嵌入式轻量级数据库SQLite的开发接口. 除此引用3G信息化时代中最受欢迎的语音合成技术(TTS)来实现此平台中声母韵母的标准发音, 本研究为维吾尔族民众提供随时随地都能方便的学习汉语的环境, 从而更好的满足新疆少数民族的应用需求.

**关键词:** 智能手机; 汉语学习; Android平台; Java; SQLite; 自动选型

**中图分类号:** TP311.52 **文献标识码:** A **文章编号:** 1000-2839(2014)02-0205-07

## The Research and Development of Software for Learning Chinese Based on the Smart Devices

Miradeljan Mamat<sup>1</sup>, Wushouer Silamu<sup>1,2</sup>, Dilnigar Rixat<sup>1</sup>, Parat Dawut<sup>3</sup>

(1. College of Information Science and Engineering, Xinjiang University, Urumqi, Xinjiang 830046, China; 2. Key Laboratory of Xinjiang Multilingual IT, Xinjiang University, Urumqi, Xinjiang 830046, China; 3. Joint laboratory of Iflytek & Xinjiang University for speech and language, Xinjiang University, Urumqi, Xinjiang 830046, China)

**Abstract:** Discussed the development and implementation process of Chinese teaching software on Android development platform. Using Android SDK and the Eclipse IDE to develop, using the Android API function interface, In order to achieve the multilingual hybrid data display and process problems of Chinese, English and Uighur to give users easily learn the Chinese language environment via automatic style selection. In storage and processing a large number of relative pictures and voice files, uses the Android integrated compact embedded lightweight database SQLite development interface. In addition, references voice synthesis technology(TTS) which was the most popular in 3G information age to accomplish the standard pronunciation of both consonants and vowels in this learning platform. Provide Uighur people the environment which was anytime, anywhere able to learn Chinese language conveniently. Better meet the needs of Xinjiang minorities' applications.

**Key words:** Smart phone; Chinese Learning; Android platform; Java; SQLite; Automatic style selection

## 0 引言

随着移动通信技术的飞速发展、手机应用功能的不断增加, 应用层日益成为手机软件体系中最庞杂的一层. 手机的日益普及使手机与人的关系越来越密切, 用户对于移动互联网的需求也在不断增加, 通过最基本的语言文字和最方便的形式与移动互联网另一端沟通日益成为趋势. 我国是一个多民族的国家,

\* 收稿日期: 2013-11-13

基金项目: 国家工信部电子发展项目“维哈柯文嵌入式设备支撑软件”(NO.159018).

作者简介: 米尔阿迪力江·麦麦提(1989-), 男, 新疆喀什人, 硕士生, CCF会员(E200032951G), 研究领域为嵌入式移动开发, 语音处理及自然语言处理.

† 通讯作者: 吾守尔·依斯拉木(1942-), 男, 中国工程院院士, 博士生导师, 研究领域为计算机应用与技术、自然语言处理, 嵌入式.

新疆是个多民族地区之一,有900多万人使用维吾尔语进行交流和沟通。然而针对基于Android平台的汉语学习软件一直以来都是市场上的空缺。因此,在通讯技术飞速发展的今天,开发一套嵌入式汉语学习平台对少数民族用户来说是必要的。以前用维文键盘映射和键盘事件识别来处理维、汉、英多文输入,但是在本系统中没有此概念,因此无法嵌入到本嵌入式汉语学习平台。本文根据维吾尔文的特征和手机显示屏幕的物理特征,设计了维吾尔文的手机软键盘布局。除此,在改善传统教学的教育方式、教学效率和教学质量等方面也将起到重要作用。本软件的特点是随时随地、不需连网就可利用该平台提供有效的资源,从而为学者带来更加方便的学习环境,而且系统实用性强、简单、易用。

## 1 平台的开发技术和应用

### 1.1 平台的系统架构

在本项研究中,使用的平台是目前最流行的Android平台。该平台的系统架构由四层结构<sup>[1]</sup>组成的,从上层到下层分别为应用程序层、应用程序框架层、系统运行库层以及Linux内核层。分别介绍如下:

1) 应用程序层。Android平台不仅仅是一个操作系统,也包含了许多应用程序,诸如SMS短信客户端程序、电话拨号程序、图片浏览器、Web浏览器等应用程序<sup>[2,3]</sup>。

2) 应用程序框架层。是我们从事Android开发的基础,很多核心应用程序也是通过这一层来实现其核心功能。

3) 系统运行库层。Android包含一些C/C++库,这些库被Android系统中不同的组件使用<sup>[4]</sup>。它们通过Android应用程序框架为开发者提供服务。

4) Linux内核层。Android是基于Linux2.6内核,其核心系统服务如安全性、内存管理、进程管理、网络协议以及驱动模型都依赖于Linux内核<sup>[5]</sup>。

### 1.2 搭建开发平台

本应用软件是典型的第三方应用开发软件,因此需要搭建Android开发环境。先来看看Android开发的系统与软件需求:

操作系统:Windows XP、Vista或是Win7、Mac OS X 10.4.8及更高版本(仅限x86)、Linux(在Ubuntu Jaunty Jackalope上测试通过)。

开发环境:Eclipse IDE,其他开发环境或IDE。

具体操作步骤如下:安装JDK→安装Eclipse→安装Android SDK→安装ADT(Android Development Tools);安装ADT成功后,在File->New选项中会出现Android Project项。完成以上4步,就可以进行Android程序的开发了<sup>[6,7]</sup>。

## 2 系统总体架构

### 2.1 软件层次结构

通过对用户的需求进行一系列地分析,最终明确了本系统的使用对象及其功能。即该软件的主要任务是在最正确最有效的原则上为学者提供最方便的学习环境。该系统的功能划分如图1所示。

此软件根据用户的习惯,为用户提供方便的切换使用,从而把系统分为课程学习和综合学习等两部分。课程学习部分针对学者的水平由初、中、高等三个模块组成。刚开始学汉语者可以从初级学习开始,具有基础汉语水平者可从中级开始,而高级模式是为有一定汉语水平的学者而提供;综合学习部分是整个流程的容纳形式。从图1可以看出,综合学习部分包含各类学习模式的同时,按照其内容进行了系统的分类,如:语音、生词、课文、汉子、句型、练习等等。此外,该软件另一个优点就是灵巧地把少儿拼音学习部分添加到课程学习的初级部分,而且把少儿拼音的标准发音用语音合成技术来实现出来了。

在处理及显示维文时,移动设备不支持从右向左的文字输入和不支持系统输入法,这为维吾尔文输入及处理带来一定的困难。维文字母与汉英文的书写规则不同,比如书写方向相反。汉字和西文的书写方向是从左到右,而维文的字符从右到左,行向从上到下;维吾尔文字母根据在单词中的位置不同会有四种变形,即首写、中写、尾写形和独立形,在文字输入时要根据字母在文字中的位置来确定使用何种形式。维吾尔文在设备中主要是用自动选型处理和字母序列转换函数来处理维吾尔文的正常显示和处理,如图2。

图2中设*i*为当前需要选形的字母, *i-1* 是当前字母前面的字符, *i+1*是当前字母后面的字符, 设定*i-1*、*i*、*i+1*字符值为0(特殊字母或非维文字母)或1(普通字母), 当*i-1*、*i*、*i+1*字符值为不同的值时, *i*字符选形也不同.

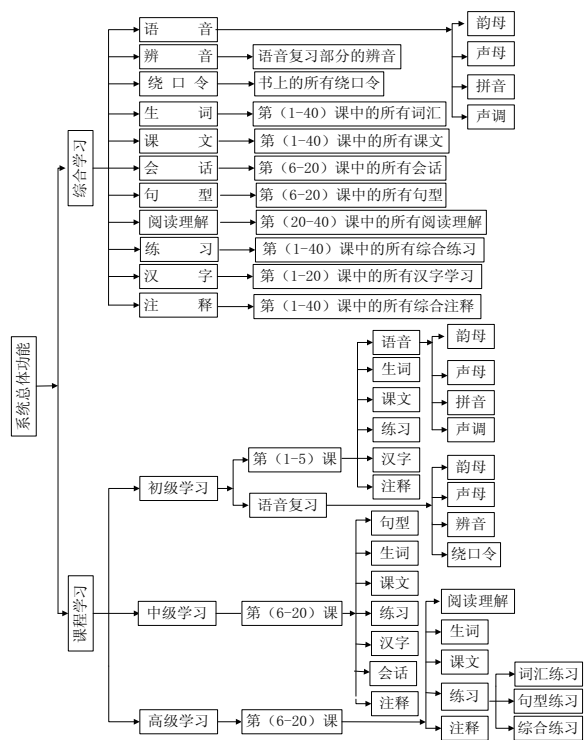


图 1 软件总体结构功能图

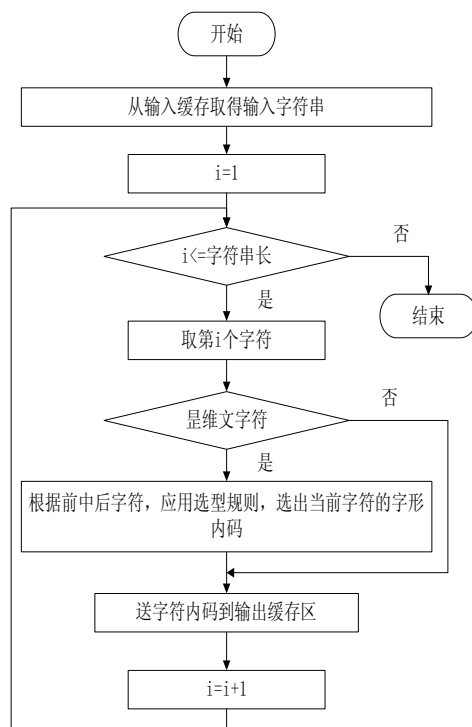


图 2 输入处理模块流程图

### 2.2 移动数据库选型

由于SQLite是轻量及嵌入式的关系型数据库的缘故, 只占非常小的内存空间, 可以避免产生数据冗余的情况. 它不仅支持SQL 语句, 而且在不影响整个系统的实时性能的前提下, 完全可以保证在嵌入式智能手机实时操作系统上正常运行. 目前在嵌入式平台中使用的数据库<sup>[8]</sup>为PostgreSQL、MySQL、mSQL、Berkeley DB、XML和SQLite等几种. 尽管都是嵌入式数据库, 但是各有各的优缺点<sup>[9]</sup>. mSQL是SQL的简化版, 方便开发, 但有使用期限限制; PostgreSQL有较大体积的缺点; 用Berkeley DB的开发成本较高; 对于应用的中小型数据库, MySQL为首选, 不过消耗太大存储空间等问题尚未得到解决. XML技术在处理少量数据时灵活性高, 但存储效率低, 因而适用于处理少量数据. 统计、处理实验数据速度如表1所示. 因此, XML不作为本系统选择, 而选择SQLite嵌入式数据库<sup>[10]</sup>. 在创建数据库时, 首先打开SQLiteExpertProfessional来建立并且命名为Chineselearning\_data.db3的数据库, 然后通过表设计器分别创建软件相关的声母表、韵母表、词汇表、课文、对话、练习、句型表等, 如表2所示的词汇表.

## 3 软件功能实现

### 3.1 实现各个模块

#### 3.1.1 课程学习模块

实现此模块时, 首先需要将课程的目录列出. 由于需要一端是汉文, 另一端是维吾尔文, 因此利用Android 提供的ListActivity类来实现列出这些目录<sup>[11]</sup>. 因为课程学习部分的初级、中级及高级部分的内容复杂、没有规律的安排内容, 而且中级学习模块中的词汇学习、课文学习和高级学习模块中的词汇学和课文学习子模块的结构根本不一样, 前者课文是带图片、带汉字、带拼音, 但是后者没有带词汇拼音, 这给Android屏幕中每个汉字和所对应的拼音对齐显示带来一定的麻烦. 通过从Activity类来衍生出来的, 而且默认自带ListView 控件的ListActivity类来创建View加载layout\_courselist01.xml、layout\_courselist02.xml、layout\_courselist03.xml和为每个列表项设计的Item 不一样, 所以在以上的三个xml之外还设计list\_

item01.xml、list\_item02.xml及list\_item03.xml等三个xml达到目的.

表 1 100和1000条记录数据库操作时间结果表

操作	SQLite(s)	MSXML/JAVA(s)
Insert 100 records	<2	16
Insert 1000 records	3~4	17
Delete 100 records	<2	18
Delete 1000 records	3~4	21
Update 100 records	<2	19
Update 1000 records	2~3	20
Select 100 records	<2	19
Select 1000 records	<2	21

表 2 词汇表

字段名称	类型	长度	主键	是否空	备注
ID	Integer	4	是	否	编号
Word_ch	Text		否	否	汉文
Word_sort	Text	4000	否	否	类型
Word_ch_pin	Text	8000	否	否	拼音
Word_uy	Text	8000	否	否	维文
Course_id	Integer	4	否	否	课程 编号
Picture_list	Text	255	否	是	路径

由于直接调用ListView<sup>[12]</sup>的onItemClickListener方法不能实现其维汉对应的效果,并且ListActivity提供的ListView控件没有TypeFace属性,因此需要重新编写一个实现Filterable接口继承BaseAdapter类的MyListAdapter类和List<map<string, string>>类型的getData()方法.总之,通过调用自己编写好的MyListAdapter类的setViewText(Textview v, String text)方法和BindView(int position, View view)方法来解决此问题.这样使那些从未接触过我们国语的汉语学习者可以更加方便地学到其需要的知识.以上难点得到解决以后,初级学习模块的语音学习(如图3所示)、辨音、绕口令等子模块也都将按以上的方法来处理,而语音学习子模块中的韵母(单、复及鼻韵母等3种韵母),声母(唇音、舌尖后音和舌尖前音等6种)都是通过语音合成技术(Text To Speech, TTS)来解决声母、韵母的标准发音.假如用Android的多媒体开发技术,音频处理<sup>[13]</sup>实现Music接口的AndroidMusic类的Play()方法来解决的话,需要大量预先录音出来的音频文件,这个会导致系统运行速度变慢,和大量消耗本来就小的手机存储空间及内存.

中级学习模块中的每篇课文会话都有3个对话.作者选择了Android提供的TabHost控件来解决同一个界面中显示符合用户选择条件的课程的不同的3种会话.本界面设计中选择此控件的原因是,该控件的各个标签中能够显示这篇课文所包含的对应的对话内容.通过创建继承TabActivity类的dialog类来实现对话标签界面,使用TabHost类创建一个mytabHost对象,然后mytabHost对象被TabHost类的getTabHost()构造方法来初始化.LayoutInflater类的inflate方法主要负责把course\_dialog.xml布局设计和mytabHost.getTabContentView()方法作为形参,能够实现三种对话的切换.然后调用mytabHost的addTab方法来实现加载每篇课文的3个不同的对话,具体地说mytabHost.addTab(tabHost.newTabSpec("xxx").setIndicator("", getResources().getDrawable(R.drawable.yyy)).setContent(R.id.zzz))形式来添加标签页.高级学习模块中的练习部分是比较复杂且没有制定的规则,因此还是在SQLite的帮助下处理了大量的图片文件.

### 3.1.2 综合学习模块

此模块(如图4所示)中最难实现的是该模块中的课程难度、结构、先后关系不一样等缘故导致的用户界面的不同,而在这样结构不一样的课程内容、课文,特别是练习部分必须在一个容器中显示出来,因此还需要利用课程学习模块中的ListActivity类来解决综合学习模块中的各个课程的正常显示<sup>[14]</sup>.同时实现做完练习后调用Compare()方法来比较从数据库中的Answer表读取的相应答案和用户的答案,然后调用Mark()方法来打分用户做的各种层次的练习.

汉字学习也一样,在课程学习时只能学习用户选择的课程中的汉字(如图5所示),但综合学习时需要显示整个课程中的所有汉字.为了实现这个功能,如果重新创建Xml会影响系统的运行速度,因此需要在同一个界面上显示两个结构完全不同的内容,此时调用Android提供的Intent组件<sup>[15]</sup>(Android的四大组件之一)的setFlags()方法来处理此难题.除汉字和练习子模块以外,语音、辨音、绕口令、声调、课文、会话、句型、阅读理解以及注释等几个子模块没有创建Activity和Xml,而是调用<sup>[16]</sup>setFlags()方法来解决节省存储空间,缩短运行时间等问题.



图3 声母学习模块



图4 综合学习模块



图5 汉字学习模块

### 3.2 维吾尔文输入及显示模块

在该学习平台的输入和显示模块,有嵌入式移动智能设备不支持文字方向从右向左的情况及系统中没有自配系统输入法等两大难题,因此,必须做维吾尔文连续正常显示和文字处理系统的开发.维吾尔文跟其他语言文字有明显的区别,主要表现在如下两种情况:

1) 跟汉文和英文比较的话,文字方向不同.既中文与英文的书写方向从左向右,而维吾尔文具有从右向左的特征;

2) 维吾尔文字母具有前连后连,前不连后不连,前不连后连和前连后不连等4种显示形式.在这4种情况下字母的形状是不同,也就是说维吾尔文字母根据在单词中的位置不同会有首写、中写、尾写形和独立形等4种书写法.在文字输入时用何种形式取决于字母在文字中实际的位置.因此,输入及显示时需要进行前后判断,下面给出了维文软键盘国家标准Unicode编码对应表(表3)和维文标准软键盘布局图(如图6).

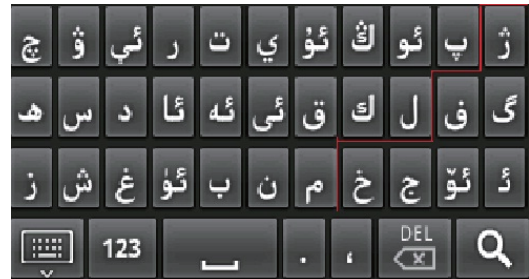


图6 维文标准软键盘布局图

表3 维文软键盘国家标准Unicode编码对应表

序号	字母	编码	序号	字母	编码	序号	字母	编码	序号	字母	编码
1.	چ	0686	10.	پ	067E	19.	ك	0643	28.	ن	0646
2.	ق	06CB	11.	ز	0698	20.	ل	0644	29.	م	0646
3.	ق	06D0	12.	ح	06BE	21.	ف	0641	30.	خ	062E
4.	ر	0631	13.	س	0633	22.	گ	06AF	31.	ج	062C
5.	ت	062A	14.	د	062F	23.	ز	0632	32.	ئ	06C6
6.	ي	064A	15.	ڤا	0627	24.	ش	0634	33.	ۋ	0626
7.	ئ	06C7	16.	ڤه	06D5	25.	غ	063A	34.	空格	0020
8.	ڤك	06AD	17.	ڤى	0649	26.	ئ	06C8	35.	٠	002E
9.	ئو	0648	18.	ق	0642	27.	ب	0628	36.	٠	060C

主要的实现代码如下:

```

.....
for (int i = 0; i < count; i++) {
final View v = view.findViewById(to[i]);
if (v != null) {
final Object data = dataSet.get(from[i]);

```

```

String text = data == null ? "" : data.toString();
if (text == null) {
text = ""; }
boolean bound = false;
if (binder != null) { bound = binder.setViewValue(v, data, text); }
if (!bound) {
if (v instanceof Checkable) {
if (data instanceof Boolean) {
((Checkable) v).setChecked((Boolean) data);}
else if (v instanceof TextView) {setViewText((TextView) v, text);}
else {throw new IllegalStateException(v.getClass().getName() + " should be bound to a Boolean, not a " +
(data == null ? "<unknown type>" : data.getClass()));}}}} }
.....

```

#### 4 测试环境

在Eclipse IDE 开发环境下,使用Java 语言和Android的有些类库,及时调用自己编写的几个核心类来设计和实现了基于Android 平台的汉语教学的应用程序,并在模拟器上测试了各项功能,均能输出预期的结果.采用的软硬件测试环境是API 水平分别为为6、8、9、10的Android 2.0、2.2、2.3、2.3.3或以上等不同版的软件环境,ROM256, RAM128, CPU单核等配置的硬件环境.

由于Android是基于JAVA的面向对象的程序设计语言、它的代码部分由类、方法、属性、接口、包等组成,因此用单元测试.但单元测试过程中我们只能知道每个类和方法(也包含类方法)是否完成各自的任务和功能,而无法检查整个模块的结果.单元测试成功后,对整个模块进行集成测试,主要目标是发现与接口有关的问题.这样此软件主要的两种测试,在不同版本的Android操作系统的设备(平板电脑、嵌入式设备和Android智能手机)上都能够正常的运行起来,如下给出其中之一,如图7.



图 7 HTC hero上的测试结果

#### 5 结论

本文针对目前发展较快的Android操作系统,结合维吾尔语的语言特点及少数民族当中汉语学习者的需求,首次实现了Android的维汉英混合输入处理和屏幕显示等文本处理功能.其次,实现了韵母、声母、辨音和绕口令等初级学习模块及引用语音合成技术(TTS)来实现声母韵母的标准发音.再次,实现Android系统上的学习汉语的平台.此学习平台具有合理的数据库设计方法和程序设计方法,是属于嵌入式应用系统,其开发重点包括后台数据库的建立及前台应用程序的开发等两个方面.用户体验无明显停滞感、读词准确,实现了基本的教学功能.本软件作为维吾尔族信息技术中首次创建的手机汉语教学软件,将填补少数民族信息化发展中的一个空白.

基于移动智能设备具有较小的存储空间的缘故,数据处理速度及效率确实比较低,而且大部分桌面应用程序不能完全地嵌入到手机来.此外,移动设备操作系统的多样性和核心技术的不可重用性也是阻碍手机教学软件发展的一个重要原因.因此,根据目前教学软件的体系结构,特别是汉语教学软件的发展程度,汉语教学下一步将在用户体验上做深入研究,进一步提高汉语教学软件的总体功能.最终,为少数民族提供一个更好的汉语学习软件平台,实现他们学习汉语的目的.

#### 参考文献:

- [1] 郝玉龙.Android程序设计基础[M].北京:清华大学出版社,2011,461-465.

- [2] 张海藩.软件工程导论[M].北京:清华大学出版社,2006,112-114.
- [3] 韩万江.软件工程案例教程[M].北京:机械工业出版社,2007,45-48.
- [4] 韩超,梁泉.Android系统原理及开发要点详解[M].北京:电子工业出版社,2010,340-343.
- [5] Katysovas T.A first look at Google Android.[s.l.]:Free University of Bolzano, 2008,12-26
- [6] DIMARZIO J F. Android: a programmer's guide[M].New York: McGraw-Hill, 2008.
- [7] RICK Rogers, BLAKE Meike, ZIGUARD Medniek, *et al.* Android应用开发[M].李耀亮,译.北京:人民邮电出版社,2010.
- [8] 田俊静,张波,黄湘情.Android基础教程[M].北京:人民邮电出版社,2011,53-57.
- [9] 吴亚峰,索依娜.Android核心技术与实例详解[M].北京:电子工业出版社,2010,248-250.
- [10] 林城.Android 2.3应用开发实战[M].北京:机械工业出版社,2011,160-180.
- [11] 李金霖,赖超,龙曦等.基于Android平台的手机日程管理系统[J].计算机与数字工程:2011,3(39):67-68.
- [12] 汪永松.Android平台开发之旅[M].北京:机械工业出版社,2011,186-189.
- [13] 巢文涵.Android多媒体开发高级编程[M].北京:清华大学出版社,2012,125-129.
- [14] 宋小倩,周东升.基于Android平台的应用开发研究[J].软件导刊,2011,10(2):104-106.
- [15] 赵亮,张维.基于Android技术的界面设计与研究[J].电脑知识与技术,2009,29(5):184-185.
- [16] 蓝坤,张跃.Android在远程医疗信息系统中的应用[J].计算机应用,2013,33(6):1791-1792.

责任编辑:闫新云

(上接第204页)

### Acknowledgements

The authors thank the Physics and Chemistry Detect Center of the Xinjiang University for the XRD analyses and the IR experiments.

### 参考文献:

- [1] Hellsten M. Drag-reducing surfactants [J]. J Surfactants Deterg, 2001, 4(1):65.
- [2] Li G P, Yang R, Wang K H. The new progress of drag reducer in research and production at home and abroad[J]. Oil & Gas Storage and Transportation, 2000, 19(1): 3.
- [3] Hu T N. Optimization strategy on batch transportation of oil product[J]. Oil & Gas Storage and Transportation, 1997, 16(6): 11-14.
- [4] Boschat J, Sabathier J. Ecuador plans expanded crude-oil line[J]. Oil & Gas Journal, 1995, 1: 37.
- [5] Sellin R H J, Hoyt J W, Scrivener O. The effect of drag-reducing additives on fluid flows and their industrial applications part 1: basic aspects[J]. J Hydraulic Res, 1982, 20(1): 29.
- [6] Brostow W. Drag reduction in flow: Review of applications, mechanism and prediction[J]. J Ind Eng Chem, 2008, 14(4): 409-416.
- [7] Moussa T, Tiu C, Sridhar T. Effect of solvent on polymer degradation in turbulent flow[J]. J Non-Newton Fluid, 1993, 48(3): 261.
- [8] Kanwal F, Liggat J J, Pethrick R A. Ultrasonics degradation of polystyrene solutions[J]. Polym Degrad Stabil, 2000, 68: 445.
- [9] Nurulla I, Tanimoto A, Shiraishi K, *et al.* Preparation of [pi]-conjugated polymers consisting of 2-decylbenzimidazole and thiophene units and chemical properties of the polymers[J]. Polymer, 2002, 43:1287.
- [10] Mi H Y, Wang J D, Li H P, *et al.* Study on oil-soluble drag reducing agent with ultra high molecular weight made by bulk polymerization[J]. Journal of Functional Polymers, 2005, 18(3): 499-503.
- [11] Boor J Jr. Ziegler-Natta Catalysts and Polymerization[M]. New York: Academic Press Inc, 1979.
- [12] Huang B T, Wang Z Q. The Progress of Olefins & Diolefins Coordination Polymerization[M]. Beijing: Science Press, 1998.
- [13] Zhao H K, Li H P, Zhou Q X. Effect of dimethoxydiphenylsilane on active centers of  $\alpha$ -olefin polymerization[J]. Petro & Chemical Technology, 2005, 34(8): 744-748.
- [14] Paul D R, Barlow J W. A binary interaction model for miscibility of copolymers in blends[J]. Polym Rev, 1984, 25(4): 487-497.

责任编辑:周蓉