

基于 J2ME 的维汉双语电子词典的研究与实现

张永才 吾守尔·斯拉木 (新疆大学 信息工程学院 新疆 乌鲁木齐 830046)

摘要: 在 WTK 平台上使用 J2ME 技术开发实现了维汉双语电子词典。分析设计了合理的词典的词库文件结构和良好的用户界面, 选用折半查找算法来实现快速搜索词语, 并实现了手机数字键盘和触摸屏并用的维吾尔语输入方式。

关键词: J2ME; WTK; 电子词典; 维吾尔语输入

Research and Implementation of Uigur-Chinese Dictionary Based on J2ME

ZhANG Yong-Cai, WU Shou Er· Si La-Mu

(Institute of Information Science and Engineering, Xinjiang University, Urumchi 830046, China)

Abstract: This paper Researches and implements the Uigur-Chinese electrical dictionary with J2ME technology on the WTK platform. It analyzes and designs structure of word file and well UI on the electrical dictionary, implements the rapid search word with Binary Search Method. In this paper Uigur word can be input with cellophane keyboard and touch screen.

Keywords: J2ME; WTK; electrical dictionary; Uigur word input

1 引言

随着信息技术的飞速发展, 手机及其他无线设备在给人们带来了很多的便利的同时, 自身的一些缺陷也影响了其发展: 出厂时均由硬件厂商固化程序, 程序不能增加、删除。当前市场上常见的手机、PDA 产品上多为英汉双语电子词典, 没有一种产品涉及到维文, 为了迎合维文信息处理的要求, 方便广大以维吾尔语作为母语的群体, 提高他们的学习和工作效率, 设计了一种基于 J2ME 的维汉手机词典。

J2ME 作为 Java 技术的一部分, 是专门为移动设备设计的 Java 平台, 其应用程序可以运行在不同制造商的不同设备上, 充分体现了 Sun 公司倡导的“一次编写, 随处运行”的理念。它以其不断发展的技术优势获得了手机制造商越来越广泛的支持。J2ME 平台被认为是当前移动开发领域中最杰出的手机程序开发平台之一, 它已被开发者、设备制造商、网络通信公司和消费者广泛接受。

2 J2ME体系结构

J2ME 平台由多种配置(configuration)、简表(profile)和可选包(optional package)组成。平台的实现者和应用程序的开发者可从中选择并组合出一个完整的 J2ME 运行环境来满足特定范围内的设备需求。J2ME 专家组采取这种灵活的设计结构主要是为了满足市场上不同种类的嵌入式设备的需求, 这些设备在软件和硬件特性上都存在巨大的差异, 一种规范很难将它们统一起来。

J2ME 利用 configuration 概念将所有的嵌入式设备区分为 2 种抽象类别:(1)运算能力相对较佳并在电力供应上相对比较充足的嵌入式装置 CDC(Connected Device Configuration)。(2)运算功能有限且电力供应也有限的嵌入式设备 CLDC(Connected Limited Device Configuration)。而 Profile 是架构在 Configuration 上层的规范, 规定在相同 Configuration 类别下的不同嵌入式

基金项目:新疆维吾尔自治区高技术研究与发展计划(200532112)

收稿时间:2009-11-08;收到修改稿时间:2009-12-16

设备的开发。MIDP(Mobile Information Devices Profile)是 J2ME 发展较为成熟的描述,它属于 CLDC 有限连接配置,目前已广泛应用于 JAVA 手机、双向寻呼、Palm、PDA 等设备中,成为移动数据应用的良好平台^[1]。

J2ME 的体系结构被横向地分成 3 层,纵向分成 2 个部分。配置包括一个控制配置核心类的虚拟机,具体的简表位于每个配置之上。

3 基于 J2ME 的电子词典

3.1 构建思想

随着经济的飞速发展,手机的普及程度越来越高,针对手机的应用服务也是越来越丰富,电子词典的使用只是众多服务中的一项,用户可以通过手机词典可以随时随地地查阅未知的词汇,但是各个手机所使用的系统平台都不一样,针对某一平台上设计的电子词典,在其它平台上就不能直接使用,只能按照新的系统平台所支持的语言进行软件的开发,增加了软件开发的成本和重复劳动,为此利用 java 语言的优点,开发一款电子词典可以解决上述问题。

3.2 关键技术

3.2.1 数字键盘设计

目前为止,维吾尔语手机键盘布局还没有一个统一的国家标准和行业标准,市场上所销售的手机都是厂商自己对自己生产的产品进行键盘布局,所以维吾尔语手机键盘布局是因人而异、键盘布局混乱。通过参考现有市场上手机的键盘布局、用户使用情况调查、维吾尔语字母的发音,本系统采用了图 1 所示的手机键盘布局。



图 1 手机键盘布局

将维语 32 个字母分布于手机数字键盘的第 2 到 9 号键上(0 键用于输入空格, 1 键和#键用于输入标点符号,*键用作删除键),对于维吾尔文的输入采用 22 键码^[2]输入方式:第一次按键用于选择某一个键值上的维文字母,第二次按键用于选择输入具体的维语字

母:如果要输入维语词语:ئوتلاق(牧场)则需要输入的按键序列为:21,81,31,22,71,22,61,虽然按键次数有些多,但是在本系统中安排键盘输入法是考虑到当前的一些手机不支持触摸屏方式输入,从而能够正确的输入维语,为了提高输入的效率,在下一技术难点中介绍了触摸屏方式输入维吾尔语。

3.2.2 触摸屏设计

支持触摸屏技术的手机在我们的生活中的应用已经是越来越常见了,相比与传统的按键操作,触摸屏技术可以使用户进行快速直观的功能选择,这样就可以省去了频繁按键的麻烦,只需轻触屏幕就能完成操作。本系统同时也提供了触摸屏方式输入维吾尔文,键盘布局是将 32 个维吾尔语字母布局到屏幕当中,用户通过触摸维吾尔语字母相应的区域,即可将相应的维吾尔语字母输入到文本框中,对于这种方法来说,难点在于触摸屏技术与维文输入时自动选型结合。维吾尔语的每个字母根据在单词中的位置不同有单独形,词首形式,词中形式,词尾形式等 4 种表现形式。将字母的每一种表现形式看作一个字符,32 个字母共有 120 多个字符,在输入时根据该字母在词中的位置自动选择出合适的字型。

3.2.3 词库文件的建立

决定电子词典优劣的一个重要的指标就是词库设计^[3],只有一个丰富的词库才能够满足用户的需求,本系统中的词汇来自于拥有 46750 维汉词条的基础语料库,从而在数量上保证了用户的常用需求,为了保证查询结果的实时性和准确性,词库文件结构设计的是否合理对电子词典系统的实现显得尤为关键。

本系统中词库文件包括两个方面:汉维词库文件和维汉词库文件。它们的文件结构和设计过程是一致的,即汉语词条和维语词条在词典文件中的位置构建成一一对应的索引关系,只要能够查找出词语在词典文件中的位置,就能够迅速找到相应的词语解释的位置,将查询结果返回并输出到手机屏幕供用户查阅。

在汉维词库的文件设计中,首先,按照词语中汉字的个数进行分块,分成独字,二字词语,三字词语,四字词语等,其次以汉语词语为关键词进行排序,这样做是为了后续的词语的查询。然后,将汉语词语和维语解释写入到两个文件当中。目的是为了在用户进行汉——维词语查询时,通过查找到汉语词语在汉语词库文件中的位置,将此位置的索引号返回给系统,通过该索引号在维语词库中进行查找,并将查找到的维语解释输出到手机屏幕,这就是利用了这两个词库

文件中词条的一一对应关系。

维汉词库的文件设计过程与汉维词库的文件设计过程基本一致,但也存在细微差别,不同之处在于:首先以维吾尔词语为主关键词进行排序,然后按照维吾尔词语的首字母将词典文件分成 31 块(其中没有以 ئ字母开头的维吾尔词语)。

3.2.4 查找算法的设计

经过多种查找算法的比较,结合词典文件中的词语全部为有序排列;而折半查找法适用于不经常变动而查找频繁的有序列表,即满足顺序存储结构和词语有序排列。因此本系统采用折半查找算法对词典文件进行匹配查找,即可最大程度节约系统查找时间,提高查找速度。

在获得用户的词语输入情况下,查词的步骤如下:

(1) 查找时,首先获取用户输入的词语并判断输入词语的长度,根据词语的长度(或者维文词语的首字母)定位到某一小块字典词库文件中,转(2);如果不符合要求,转(4);

(2) 使用折半查找算法在字典词库文件中进行词语查找,匹配成功转(3);如果匹配失败,转(4);

(3) 将查找到的词语解释输出到手机屏幕上,转(4);

(4) 结束本次查词,输出提示信息。

4 实验分析

根据上述对该系统的总体结构设计,在 Eclipse3.3.2+ wtk2.5.2 环境下,采用 J2ME 技术实现了维汉双语电子词典,在 wtk2.5.2 模拟器中成功地进行了测试。

打开电子词典后,用户界面以菜单的形式显示,用户可以根据自己的查询需求来选择相应的菜单项,选择“汉语—维语”或“维语—汉语”查询方式。在文本框中输入想要查找的汉语或维语词语,点击查询选项后系统将进行匹配查找,若匹配成功,系统将词语对应的解释显示在手机模拟器中;若匹配失败,系统返回“无此解释”提示信息。

下图以“维语—汉语”查询方式为例,系统的默认维语输入方式为触摸屏输入,用户通过点击维文字母的相应区域就可以将维语字母输入到文本框中(图 2),在输入过程中维语字母按照在词语中出现的位置自动进行变形,输入完成所要查找的维语词语后,点击 سۇرۇش(查询) 菜单项,对应的汉语解释如图 3 所示,对于不支持触摸屏输入的手机用户,可通过选择

“touchkey”菜单项(图 2)切换到数字键盘输入方式,采用 22 键码进行维语输入。图 4 为“汉语—维语”查询方式的结果界面。



图 2 维-汉查询操作界面



图 3 维-汉查询结果



图 4 汉-维查询结果

5 结语

测试结果证明了电子词典的词库文件结构设计的合理性,在查找解释的过程中速度快而准确,该电子词典的所有代码均采用 J2ME 技术,用户只需在界面上稍加修改即可应用到 PDA、学习机等多种小型移动设备上。其词库的设计方式可以用于其它语种(哈萨克语,柯尔克孜语,阿拉伯语)的电子词典^[4]。

该电子词典的功能还不够完善,词库以常用词汇为主,不能查到所有的词汇,为此可以考虑建立词典服务器,通过网络方式查阅词库中没有的词汇,从而充分利用 J2ME 在无限开发领域的优势。

参考文献

- 詹建飞.J2ME 开发精解.北京:电子工业出版社,2006.1 - 10.
- 张淑萍,袁保社.一种手机数字键盘的维吾尔文字母输入法.电脑知识与技术,2005,26:66 - 68.
- 商陆民,周然,朱丹群,邢海蓉.电子词典的数据结构设计.电子器件,1995,18(1):50 - 54.
- 董治江,吴健,钟义信.在 ICU 中实现少数民族文字的处理.中文信息学报,2003,18(2):67 - 73.