

短信息在图书查询中的应用设计

北京理工大学计算机科学与工程系(100088) 朱会东

郑州大学信息工程学院(450052) 黄艳

厦门大学电子工程系(361005) 刘祥南

摘要: 介绍了短信息服务及其在图书馆查书系统中的应用技术,论述该系统的功能及在系统中采用 MFC 多线程及动态链接库的实现方法。

关键词: 短信息服务 图书查询 多线程 动态链接库

在飞速发展的信息技术领域中,移动通信和 Internet 是两支突飞猛进的支柱产业。互联网为人们提供必要的信息资料,移动通信让人们摆脱电话线的束缚,能够随时随地进行交流。而移动通信与互联网的结合——移动互连则意味着将跨时空的计算和全方位的信息置于人们掌心。短信息服务作为移动互连技术的一种,因其不占用语音通信的信道与费用低廉等优点,目前已在许多领域得到广泛应用。

当今社会,人们对知识的需求不断增加,人们常常需要从海量的书籍中来查找自己需要的信息。网络的出现为人们通过 Web 查询书籍提供了一种方便的途径。利用短消息人们可以随时随地用手机发短消息查询书籍,既方便又快捷,同时,也可为各大书店及图书馆节省人力物力。

1 短信息服务

短信息服务 SMS(Short Message Service)是通过手机发送和接收有限长度的文本信息的功能。一条短信息可以包含 160 个英文字母(7 位编码)或 70 个非拉丁字母(16 位编码),如中文汉字或阿拉伯字等 Unicode 编码。

SMS 属于 GSM(Global System for Mobile Communications)第一阶段的标准。但目前 SMS 已经被集成到了很多网络标准中,GSM、CDMA、TDMA、PHS、PDC 等移动网络都支持 SMS,这使得 SMS 成为一项非常普及的移动数据业务。SMS 与语音传输及传真一样,都是 GSM 数字蜂窝移动通信网络提供的主要电信业务。它通过无线控制信道进行传输,经短消息业务中心完成存储和前转功能。SMS 是非对称业务,它使用 SS7 信令信道传输数据分组,即使在业务信道处于高峰期的情况下,也照样可以使用短信息顺利通信。它还可以实现全国和国际漫游。

SMS 具有收费低廉、随时随地、能够承载多种数据、可以实现推和拉二种不同的工作模式及传输安全等优点

点,目前已在资讯服务、移动商务、商务助理等领域被广泛应用。

2 系统设计

短信自动查书系统的目标就是通过手机短信息服务,实现用户的无线信息查询,快速方便地使图书查询业务通过手机方式完成。

2.1 系统结构

由于采用 CMPP 接口,移动和联通不能互发,必须将移动和联通各开通一个号码,这给管理者和读者都带来不便。考虑到书籍查询中的业务量的实际情况,采用内嵌 SIM 卡的通信器作为与 SMSC(短消息业务中心)的通信工具。通信器通过 USB 口与计算机相联,启动计算机后通信器就处于加电状态,可进行短信收发,通信器内最多可保存 10 条短信。系统的网络结构如图 1 所示。

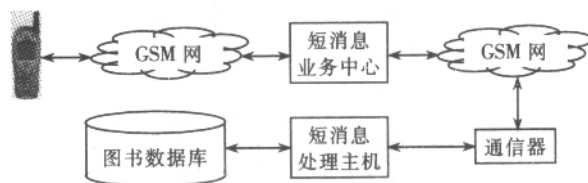


图 1 系统结构

客户通过移动电话发送短信到通信器内所嵌入的 SIM 卡的号码要求查书。该短信首先通过 GSM 网上送到短消息业务中心,然后短消息业务中心把短消息内容下传到通信器。短消息处理主机监听到有短消息到来时,立即读取通信器,然后分析读取的内容,分解出不同字段的内容,存入短消息数据库中的接收数据表中。系统根据客户所要查询的内容,启动查询线程查询图书馆内图书数据库,把查询的结果写入短消息数据库的发送数据表中。短消息处理主机把发送的内容转换成通信可识别的格式,然后写入通信器,以短消息的型式返回到客户的手机。

2.2 系统功能模块

在系统需求及功能的分析上,把系统划分为如图 2 所示的子系统 and 功能模块。

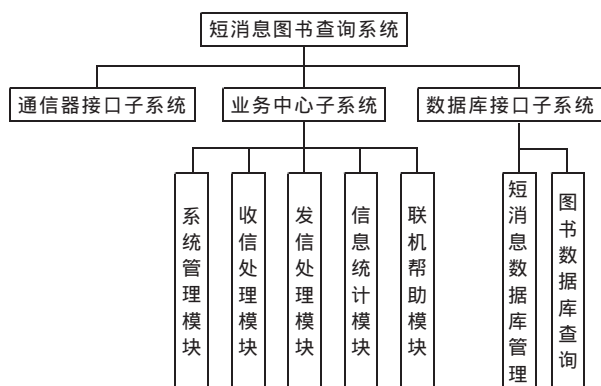


图 2 系统功能模块划分

2.2.1 通信接口子系统

该子系统中硬件的控制较为复杂,微小的变动都有可能导导致系统不能正常运行。由于对硬件的访问需求比较固定,所以采用动态链接库技术(DLL)来实现对硬件的控制。完成的功能主要包括实现硬件上的短消息的自动接收和发送、信号量的获取、通信器工作状态读取等。

2.2.2 业务中心子系统

该子系统主要完成整个系统的管理及各种数据的处理。主要实现以下基本功能:

(1)系统管理。完成操作员的管理,包括操作员的增加、注销、权限的设置、密码的设置及修改;系统参数的设置,包括设置短消息中心号码以及图书数据库的用户名、密码以及短消息的格式。

(2)收信处理。系统启动后创建短消息接收线程,线程不断查询通信器。如果有短信接收,通信器立即读取内容,然后分析读取的内容,分解出短消息的内容和手机号码,存入收信数据表中。接收短消息参数字符串的格式为:手机号|短消息格式|接收日期|接收时间。

(3)发送处理。系统定时查询发送表,如果发送表有内容,就创建短消息发送线程,再把发送表中的数据合成通信器可识别的格式发送出去,然后删除发送表中已发送出去的内容。发送短消息参数字符串的格式为:手机号|短消息内容|日期|时间|加密标志|发送等级。

(4)信息统计。以短消息数据库里的内容为基础,对短消息查询的类别、内容、读者类别、返回结果等的查询、汇总统计、分析,以图表的形式显示出来,并可生成用户指定格式的报表。

2.2.3 数据库接口子系统

实现通过局域网来访问图书信息数据库。首先根据短消息数据库收信表中的内容分析要查询的内容和查询条件,然后生成 SQL 查询语句的形式,对图书数据库进行查询,把查询结果放到发件箱。

3 系统实现中的关键技术

短消息查书系统采用 VC 作为开发工具,短消息数据库采用 SQL Server2000,整个系统在 Windows 2000 Server 操作系统下运行。

3.1 动态链接库技术

动态链接库是用作共享函数库的可执行文件。一般情况下,DLL 是一个库中所有函数的集合,并通过包含在可执行文件中的信息进行检索来使用。在收发短信息时需与通信器交互,这里是通过动态链接库实现的。DLL 中的函数有导出函数和内部函数二种,导出函数可以被外部程序调用,内部函数只能在动态链接库内部使用。

对通信器控制的动态链接库提供的导出函数包括短消息的发送、接收等

```

#ifndef _Public
#define _Public
extern "C" _declspec(dllexport) int InfoDataComSend
(char * Message); //发送短信息
extern "C" _declspec(dllexport) char * GetAllRecvMes-
sage(int device); //接收短信息
.....
#endif

```

下面是对动态链接库导出函数的引用:

```

#include "Public.h" //包含导出函数的头文件
.....
Cstring revmsg;
strcpy(revmsg,GetAllRecvMessage(0)); //读取通信器
//分解出不同的字段;
.....
Cstring sendmsg;
//合成发送短消息各字段到字符串 sendmsg;
retu=InfoDataComSend(sendmsg); //发送到通信器;
.....

```

3.2 多线程技术

Windows 是一个多任务操作系统。在 Win32 中,每一个进程可以同时执行多个线程,这意味着一个程序可以同时完成多个任务。图书查询系统需要对大型的数据库进行访问,系统既要进行耗时的工作,又要保证用户输入响应的及时性,尤其当检索的数据量很大、数据处理任务很重时,如果不采用多线程,检索时的漫长的等待是难以忍受的。在采用多线程时,关键的问题是要采取措施来保持线程间的同步。

短消息查书系统主要包括消息发送线程、短消息接收线程、图书数据库查询线程。在对收信箱和发信箱操作时,使用的是临界区(CCriticalSection)同步对象,对图书数据库的查询采用信号量同步机制,信号量初始化时资源计数设置为 3,这样最多三个查书线程可同时运行,避免因查询数据库的线程太多而造成堵塞。

临界区的使用:

```
CCriticalSection send_cs, rev_cs;
send_cs.Lock(); //锁定发信箱临界区资源
对发信箱进行操作;
send_cs.Unlock(); //释放对发信箱临界区资源的锁定
rev_cs.Lock(); //锁定收信箱临界区资源
对收信箱进行操作;
rev_cs.Unlock(); //释放对收信箱临界区资源的锁定
信号量的使用:
HANDLE hsem=CreateSemaphore(NULL,3,3,NULL);
//建立信号量对象
.....
WaitForSingleObject(hsem,INFINITE); //引用信号量
AfxBeginThread(QueryProc,GetSafeHwnd(),
```

```
THREAD_PRIORITY_NORMAL); //创建图书数据库
//查询线程
.....
ReleaseSemaphore(hsem,1,NULL); //释放信号量
```

3.3 数据库

数据库是系统的核心和基础,短消息数据库结构设计的合理性及优劣性将直接影响整个系统的性能。对图书数据库的访问方式也同样影响访问数据库的执行效率。在对短消息数据库的设计中遵循数据独立性原则,按规范化设计数据库,可减少数据间的依赖和数据冗余。

在对图书数据库访问时采用 ADO 访问数据库技术。ADO 是基于 COM 接口的技术,使用这种技术可以直接对数据库的驱动程序进行访问,从而大大提高访问的速度。

(下转第 52 页)

(上接第 25 页)

发送号码子程序流程图如图 5 所示。无线发射时,要取出所播对方的号码与本机号码共 8 个数字逐一发出。接收端接收时,需计数。计数满 8 个后,再将前 4 个数字与本机号码比较,通过后调用发声、来电显示子程序。发送号码子程序执行前,必须先关中断,否则会由于单片机的中断而引起发送失败。

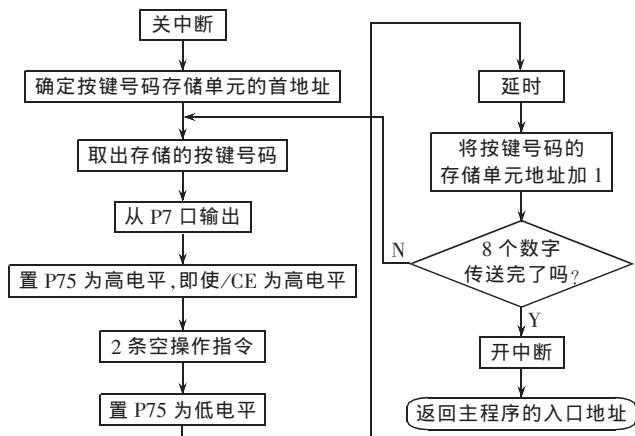


图5 发送号码子程序流程图

中断服务子程序流程图如图 6 所示。其中,/INT 是外

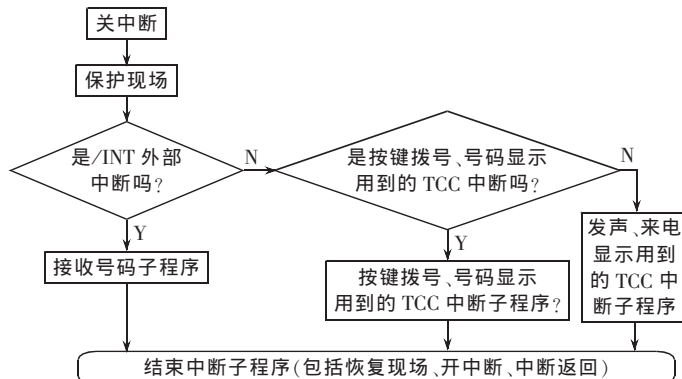


图6 中断服务子程序流程图

部中断。接收号码子程序流程图如图 7 所示。按键拨号、号码显示子程序和发声、来电显示子程序所用到的 TCC 中断子程序可参考文献[1]。

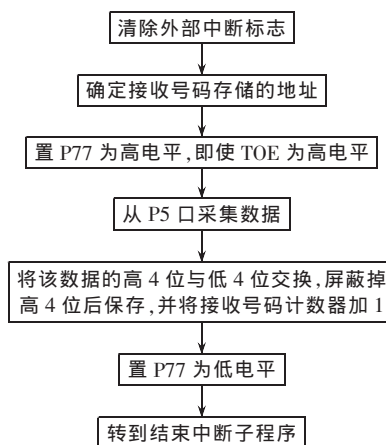


图7 接收号码子程序流程图

5 结束语

该系统充分利用软件实现了较多的功能,因此所用元器件极少。虽然目前市场上通信设备很多,但这种小巧、功能较强、成本低廉的有线传呼系统在某些场合也有它的生存空间。此外,还可以将本系统的对讲时信息加密等功能进一步完善,并可以根据国家信息产业部无线电管理委员会公布的公众对讲机的技术规范做功能更强、使用更方便的无线传呼对讲机,以满足人们郊外旅游,特别是对一些移动通信信号达不到的地区的通信需求,满足一些需要频繁通信的如大超市、大酒店、运动会等的通信需求。若要做成无线传呼对讲机,可以考虑在这个系统的基础上再解决无线发射、接收等问题。

参考文献

1 邓锦城.EM78447 入门与实操.台北:台湾宏友图书开发股份有限公司,2000

(收稿日期:2003-10-18)

其中 m 为特征向量维数, α_k 为二个文档对应特征词条的四位码字的十进制数值差的绝对值。由于这种相似性的计算使用的是整数, 所以计算速度和精度得到一定的提高。

可以利用简单的示例验证公式(5)的合理性。当二个文档完全相似时, $\text{sim}(d_i, d_j)$ 的值等于 1, 而二个文档完全不同时它的值为 0。这种方法不仅反应了文档之间的差异, 而且定量地描述了这种差异性, 从而为文档的聚类提供了依据。下面通过对具体的 Web 文档进行实验并进一步地验证。

3 实验

实验用的文档是从搜狐的中文网站上获取的娱乐类文档, 选用其中的 1 500 篇。对这 1 500 篇文档进行手工分类, 如表 1 所示共分为 10 类。

表 1 选取的 Web 文档手工分类结果

分类	游戏	旅游	时尚	音乐	影视	休闲	人物	车类	收集	其他
文档数	211	127	86	350	307	43	134	98	62	82

衡量信息检索性能的召回率和精度也是衡量分类算法效果的常用指标。然而聚类过程中并不存在自动分类类别与手工分类类别确定的一一对应关系, 因此无法像分类一样直接以精度和召回率作为评价标准。为此本文选择了平均准确率作为评价的标准。平均准确率通过考察任意两篇文章之间类属关系是否一致来评价聚类的效果。

试验中对使用公式(3)和(5)的改进 k-means 算法和原 k-means 算法的平均准确度进行了比较, 实验结果如表 2 所示。

实验结果表明, 改进后的 k-means 算法与原 k-means 算法在运行速度上基本相同甚至略快, 平均准确度则比原算法有了普遍提高, 尤其在正确指定聚类数 k 时, 平均准确度提高了近 7%, 说明此算法具有较高的准确性。由于实

(上接第 49 页)

ADO 的主体接口对象有三个: ConnectionPtr、CommandPtr 和 RecordsetPtr。ConnectionPtr 返回一个记录集或空指针, 通常用来创建一个数据连接; CommandPtr 接口返回一个记录集, 它提供一种简单的方法来执行返回记录集的存储过程和 SQL 语句; RecordsetPtr 记录集对象, 它能够提供更多的控制功能。

其程序实现为:

```

:: CoInitialize(NULL);
m_connection.CreateInstance(__uuidof(Connection));
m_recordSet.CreateInstance(__uuidof(Recordset));
pConn.CreateInstance(__uuidof(Connection));
pConn->Open("Provider=SQLOLEDB;SERVER=server;
DATABASE=dbname;UID=user;pwd=password", "", "", -1);
pRs.Open("select * from cmk where name=msg",
pConn, adOpenDynamic, adLockOptimistic,
adCmdUnknown);
    
```

表 2 实验结果

聚类数	8	9	10	11	12
未改进	38.60%	41.05%	53.18%	43.54%	36.44%
改进后	43.74%	45.21%	60.04%	44.16%	40.35%

验中使用的文档集很小, 所以改进的算法优势不很明显。

4 结束语

本文对 k-means 算法进行了改进。根据不同位置的特征词条对文档内容的不同决定程度, 提出一种新的文档特征词条的权重评价函数, 并在此基础上提出一种文档相似性的度量方法。实验表明改进后的算法不仅保留了原 k-means 算法效率高的优点, 而且在平均准确度方面比原算法有了较大提高。实验还表明, k-means 算法要依赖原始聚类数 k 的选择。如何为初始文档集选择合适的聚类数 k 以及进一步提高平均准确度是今后改进 k-means 算法的主要研究方向。

参考文献

- 1 Fraley C, Raftery A E. How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. Department of Statistics University of Washington Technical Report, 1998
- 2 Xu L. How Many Clustering?: A YING-YANG Machine Based Theory For A Classical Open Problem In Pattern Recognition. IEEE Trans, Neural Networks, 1996; 3(10)
- 3 Jiang M F, Tseng S S, Su C M. Two-phase clustering process for outliers detection. Pattern Recognition Letters, 2001; (22) (6-7)
- 4 Michaud P. Clustering techniques. Future Generation Computer System, 1997; 13(6)
- 5 Pedrycz W, Bagiela A. Granular Clustering: A Granular Signature of Data. IEEE Trans, Neural Networks, 2002; 32(2)

(收稿日期: 2003-10-20)

```

.....
:: CoUninitialize();
    
```

4 结束语

目前移动电话的普及已使得 SMS 的应用极为广泛。本文应用 SMS 技术通过使用 MFC 多线程编程及动态链接库实现的自动查书系统, 已经在图书馆及书店中得到很好的使用。系统不但方便了图书的管理, 也给人们的学习生活带来了极大的便利。

参考文献

- 1 刘刀桂, 孟繁晶. Visual C++ 实践与提高, 数据库篇. 北京: 中国铁道出版社, 2001
- 2 周学泳, 石丹. 短消息(SMS)与 WAP 的开发及应用. 北京: 电子工业出版社, 2002
- 3 Petzold C. Windows 程序设计. 北京: 北京大学出版社, 1999
- 4 侯俊杰. 深入浅出 MFC. 武汉: 华中科技大学出版社, 2001

(收稿日期: 2003-10-18)