

# 可定制移动数据采集系统的研究和实现<sup>①</sup>

## Research and Implementation of Customizable Mobile Data Collection System

黄志一<sup>1,2</sup> 周园春<sup>1</sup> 常青玲<sup>1,2</sup> 沈志宏<sup>1</sup> 侯元生<sup>3</sup> 阎保平<sup>1</sup>

(1 中国科学院 计算机网络信息中心 北京 100190; 2 中国科学院 研究生院 北京 100049;

3 青海省青海湖景区保护利用管理局 青海 西宁 810006)

**摘要:** 基于个人数字助理 (Personal Digital Assistant, PDA) 的数据采集系统为室外数据采集工作提供了方便快捷的工作方式。但是目前的 PDA 数据采集系统都是针对单一应用独立设计开发的。将介绍“可定制”移动数据采集系统的设计思路和实现过程。该系统的创新之处在于针对不同的应用, 进行简单的定制操作, 就可以满足不同用户的需求, 而无需开发者的多次设计和开发工作。最后通过青海湖自然保护区动植物数据调查的实践, 对该系统实施中的若干关键技术问题进行了研究试验, 取得了较为满意的效果。

**关键词:** PDA 可定制 移动数据采集 青海湖动植物数据调查

### 1 引言

室外数据采集工作在各行各业中具有相当的广泛性, 例如土地利用变更调查, 桥梁、公路养护数据采集, 电力设施数据采集等等。传统的工作方式是: 将事先设计、准备好的数据采集表格带到工作现场, 人工填写表格后, 待返回室内再将表格上的数据手动输入计算机进行后续计算和管理。这样的方式存在成本高、效率低、工作量大, 容易出错等问题。

青海湖自然保护区是青藏高原生物多样性最丰富的宝库。保护区的工作人员要定期进行环湖巡查、保护区植被样地监测、青海湖重点濒危物种监测等活动。伴随这些野外监测活动产生了大量有价值的科学数据。这些数据主要是通过工作人员在野外考察时手工填写事先印制好的各种表格来记录, 而后再输入计算机。为了减少二次录入差错, 避免由于纸质数据表格的破损、遗失而导致考察数据的丢失问题, 实现青海湖动植物数据调查工作的数字化是十分必要的。

个人数字助理 PDA 是一种数据采集的理想工具; 它功能强大, 几乎具有普通 PC 的全部功能, 同时它

小巧轻便, 可供使用者随身携带。于是, 利用 PDA 进行数据采集工作的应用系统迅速发展起来<sup>[1-3]</sup>。然而, 目前国内外此类应用系统基本上是针对单一的应用需求进行独立的开发, 这样的工作不仅开发阶段费时费力, 而且后期难以维护和扩展。

在中科院网络信息中心组织的信息化支撑服务调研中发现, 除了青海湖动植物数据调查工作迫切需要移动数据采集系统, 海洋研究所、青岛生物能源与过程研究所、沈阳应用生态研究所及东北地理与农业生态研究所都有很多科研活动是需要野外人工采集数据的。如此众多和迫切的应用需求, 成为我们设计和实现通用的可定制移动数据采集系统的强大推动力。

本文将展开可定制移动数据采集系统的研究。使用该系统, 通过简单的配置操作, 即可适用于不同应用, 满足不同的需求。这样使得开发者一劳永逸, 而不必针对不同应用, 进行重复、繁杂的开发工作; 同时可以降低后期维护的复杂度, 用户不必为一点点需求的变动, 而需要再与开发者沟通并等待其修改系统, 而是可以通过定制工具, 自己根据采集需求的变化进

① 基金项目: 中国科学院知识创新工程青年人才领域项目(O814041108); “十一五”中国科学院信息化专项项目(INFO-115-D02)

收稿时间: 2009-02-18

行调整。最后本文将该系统成功的运用在青海湖自然保护区动植物数据调查工作中，并取得了较为满意的效果。

## 2 面向单一应用的移动数据采集系统现状分析

### 2.1 相关的 PDA 技术

PDA 的操作系统目前主要有诺基亚控制的 Symbian 系统、微软的 Windows Mobile 系统、PalmOne 公司支持的 PALM 系统和摩托罗拉支持的 Linux 系统。其中微软的 Windows Mobile 系统，因为与 Windows 操作系统系出同源，和 PC 能无缝连接，所以与其他操作系统相比，更容易上手。跟 PC 上的 Windows 一样，手机上的 Windows Mobile 系统也有开始菜单、资源管理器、IE、Windows Media Player 等软件，给用户带来了极高的易用性和强大的可扩展能力<sup>[4]</sup>。本文的实验平台即选择 Windows Mobile 5.0 的系统。

PDA 上的数据采集系统，需要将数据先保存在 PDA 的存储卡上，再通过同步软件同步到后台数据库中。于是我们需要能够运行在 PDA 上的数据库软件帮助存储和管理数据，嵌入式数据库正好满足了我们的需求。嵌入式数据库系统是指支持移动计算或某种特定计算模式的数据库管理系统，它通常与操作系统和具体应用集成在一起，运行在智能型嵌入式设备或移动设备上。目前国内外市场上有多种嵌入式数据库产品，每种数据库产品在技术上都有不同的侧重点，以满足嵌入式应用不同层次的需求。国外嵌入式数据库有 Pocket Access、SQL Server CE、Oracle 9i Lite、Berkley DB 和 SQLite 等，国内则有人大的小精灵、东北大学的 OpenBase Mini 等<sup>[5]</sup>。SQLite 是用一个小型 C 库实现的强有力的嵌入式关系数据库管理系统。它提供了对 SQL92 的大多数支持，支持多表、索引、事务、视图、触发和一系列的用户接口，实现了一个完备的、可嵌入的、零配置的 SQL 数据库引擎<sup>[6]</sup>。本文在 PDA 上的数据存取模块即采用 SQLite 数据库。

### 2.2 单一应用系统现状分析

面向单一应用的移动数据采集系统<sup>[1-3]</sup>，一般包括数据采集和数据导出两个模块。数据采集模块主要用来完成 PDA 上数据的录入、修改、保存工作。如果是带有 GPS 功能的 PDA，数据采集模块还可以自动获取

GPS 数据。数据导出模块主要用来将 PDA 上保存好的数据导出到 PC、经过有效性验证后导入后台数据库中。数据导出到后台数据库系统后，就可以供基础数据管理系统调用、浏览和检索。系统结构图如图 1。

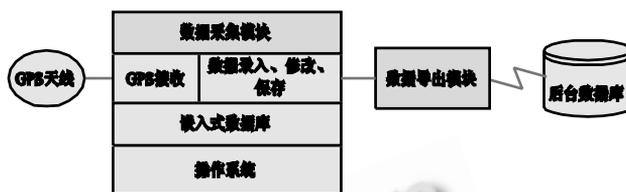


图 1 面向单一应用的移动数据采集系统结构图

数据采集模块的设计，是根据应用需求设计数据采集所需要的表格，根据录入字段的不同类型设计录入界面，并为每个表格进行数据存取的编码，以完成录入、修改、保存的功能。另外带 GPS 功能的 PDA，实时获取室外数据采集的经纬度、海拔信息，这样可以减少工作人员输入 GPS 数据的烦恼，同时也为将来采集路径的可视化展现提供位置信息。

数据导出模块的设计是根据应用需求将 PDA 上保存的数据(包括录入数据和 GPS 数据)导出到 PC 上，并根据具体应用规则进行有效性验证，或者是提供用户浏览修改的界面，最后将通过验证的数据插入到后台数据库中。

## 3 可定制移动数据采集系统的设计方案

面对中科院多个研究所对移动数据采集系统的迫切需求，通过对这些不同数据采集需求的研究分析，发现它们存在很大的共性，他们都是基于表格的数据录入，都需要一些字典数据，获取的数据都需要导出到后台数据库以进行下一步的利用。基于这些共性，本文的研究目标是实现一个通用化的移动数据采集系统，通过简单的定制操作，就可以满足不同用户的需求，而无需开发者的多次设计和开发工作。

本文设计的可定制移动数据采集系统的使用流程如下：首先根据应用的具体需求进行定制操作(包括定制 PDA 上的数据采集菜单，每个数据采集页面的采集内容，每个采集内容的数据类型、显示风格等)，系统将用户的定制信息保存成固定格式的 XML 文件，然后用户将 XML 文件复制到 PDA 上，运行 PDA 上的数据采集程序，该程序根据 XML 中的定制信息动态生成 PDA 上的采集菜单，以及对应的采集

页面。用户录入的信息保存在 PDA 上的 SQLite 数据库文件中，最后将存储在 SQLite 数据文件中的数据与后台数据库同步(适用于 Oracle、SQL server、MYSQL 等各种数据库)。

基于以上设计目标，采用模块化设计方式，系统由三大模块组成：定制模块、数据采集模块、数据同步模块。系统结构图如图 2。

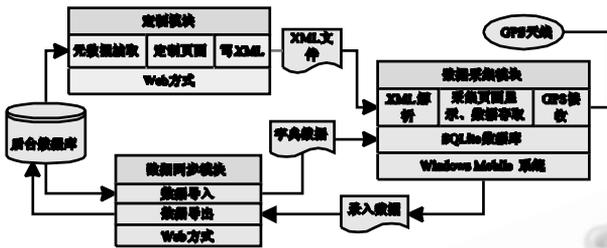


图 2 可定制移动数据采集系统结构图

数据定制模块以 Web 方式提供用户定制界面，用户配置好与后台数据库的链接后，定制模块抽取数据库的元数据信息；基于元数据信息定制模块为用户提供每个数据库表的定制页面，最后将用户关于表、字段相应的定制信息保存到 XML 文件中。

数据采集模块又由 3 个子模块构成，分别是 XML 解析模块、采集页面的显示与数据存取模块、GPS 接收模块。XML 解析模块抽取 XML 文件中的定制信息，供采集页面显示和数据保存模块使用。采集页面的显示需要采用动态显示的技术，动态添加每个输入控件以及按钮，而非像面向单一应用的设计那样，在系统开发的阶段就静态确定表单的每个输入控件。

数据同步模块采用 Web 方式实现，由数据导入子模块和数据导出子模块构成。由于每个应用系统都可能有一些静态的字典数据，所以数据导入子模块提供一个用户选择界面，从后台数据库中选出作为字典数据的关系表，抽取数据形成 SQLite 数据库文件，供 PDA 上的数据采集模块使用，以减少用户需要录入的字段。数据导出子模块的功能与单一应用系统一致，但是因为面向不同应用，所以关于从 SQLite 数据库文件中读取记录插入到后台数据库的 SQL 语句编码需要动态编码。

#### 4 可定制移动数据采集系统主要功能的实现

基于 Web 的定制模块和数据同步模块采用 jsp 语言，开发环境：eclipse，resin 服务器；PDA 上的数

据采集模块采用 C#语言开发，开发环境为：Visual Studio 2005、Windows Mobile 5.0 Pocket PC SDK、.Net Compact Framework 2.0 (VS2005 自带)、SQLite3 和 ActiveSync 4.5。

#### 4.1 定制模块

定制模块以 Web 方式提供用户定制界面。由于移动数据采集应用往往是在已有数据库应用系统基础上，或者即使是新的应用系统，也都需要将数据存于数据库中供以后做进一步的数据应用。所以本文所设计的定制功能是以设计并建立好的数据库表结构为基础的，并且该定制模块可适用于 oracle、SQL Server、MYSQL 等各种类型的数据库。

定制模块提供一个关于数据库链接的配置页面，用户将所使用的数据库产品类型、版本号、数据库名、IP 地址、端口、用户名、密码填写正确后，定制模块将链接数据库并抽取数据库中的表结构、表字段等元数据信息。以青海湖数据库为例，图 3 显示了青海湖数据库中部分的关系表。如图 4，针对每一张表，定制模块还将抽取出所有字段名，类型，大小等信息，并提供一个关于每个字段的定制页面(如图 5)。用户需要为每个字段定义字段标题，对应 PDA 上该字段的名称；还可以定义提示文字(当用户输入该字段给与的提示)，字段类型，显示风格(单行、多行、下拉框、复选框等)、等信息。

标题	实例url	表名	注册时间	操作
鸟类表	cn.cob.qinghai.birds	birds	2008-12-28 16:00:41	修改 删除
环湖调查表	cn.cob.qinghai.circlelake	circlelake	2008-12-28 16:15:18	修改 删除
鸟类调查统计表	cn.cob.qinghai.circlelakePerbird	circlelake_perbird	2008-12-28 16:18:22	修改 删除
环湖调查点表	cn.cob.qinghai.circlelakePerplace	circlelake_perplace	2008-01-04 10:38:22	修改 删除
路线调查表	cn.cob.qinghai.sampleRoute	sample_route	2008-01-11 13:08:04	修改 删除
路线调查表	cn.cob.qinghai.sampleRouteSurvey	sample_route_survey	2008-01-11 13:08:22	修改 删除
路线调查表	cn.cob.qinghai.sampleRouteSeytdl	sample_route_seytdl	2008-01-11 13:08:37	修改 删除
placet	TABLE			注册 登录

图 3 抽取出的数据库表

定制模块还有关于表的主键、外键的设置，根据外键设置表与表之间的父子关系。最后用户可以选择需要编辑的表格，以及对于每个表格需要编辑的字段，即用户可以仅选取部分表格，或者部分字段放到 PDA 上录入数据。最后，定制模块将用户关于数据库、表、



已获取的数据库元数据信息基础上, 提供用户选择字典表的界面, 将这些字典表的数据导出成 SQLite 数据库文件, 如图 7 所示。同样采用 Web 方式, 用户将从 PDA 上拷贝出来的 SQLite 数据库文件上传到网站上, 点击导出按钮即将 SQLite 数据文件中的记录一条条插入到后台数据库中(其中, 要主键外键的更新一致问题)。



图 7 导出字典表

## 5 青海湖动植物数据调查中的应用

作为中国最大的内陆盐湖和国际重要湿地, 青海湖是青藏高原生物多样性最丰富的宝库。除了普氏原羚、雪豹、藏野驴、黑颈鹤等国家一、二类保护动物外, 青海湖保护区还是候鸟重要栖息地, 每年在青海湖停息的候鸟有 189 种、12 万多只。

青海湖自然保护区的野外科研考察活动产生了大量有价值的科学数据和资料。为了更好地开展该地区的生态保护、科学研究工作, 中科院计算机网络信息中心联合相关研究所和保护区管理局共同开发“青海湖国家级自然保护区联合科研基地数据基础平台”, 建立相关的数据库, 长期保存积累野外科研考察采集到的数据, 以满足保护工作和科学研究的需求。本文的移动数据采集系统即是在此背景下展开研发与应用的。

若采用传统工作方式, 青海湖保护区工作者的动植物数据调查工作相当繁杂而且效率低下。以环湖调查为例, 一次调查绕湖一周, 需要近 10 天, 在 20 多个地点进行鸟类、样地植被、普氏原羚调查, 需要带各类表格近 100 张, 最后还要输入到计算机中。如果采用移动数据采集系统, 动植物数据调查工作的效率将大大提高。

下面将本文实现的定制移动数据采集系统运

用于青海湖鸟类资源调查和样地植被调查。其中鸟类资源调查包含环湖调查、样线调查、繁殖地定点调查。环湖调查涉及到三个表: 环湖调查表、地点调查表、鸟类调查细表。环湖调查表包含调查开始时间、终止时间、调查人员、调查计划等信息, 地点调查表、鸟类调查细表在第 4.2 节中已详述。因为一次环湖调查将在多个地点进行, 所以环湖调查表作为地点调查表的父表。点击菜单项“环湖调查表”, PDA 上显示以上 3 个表对应的编辑页面(如图 6)。样地植被调查包括样地植被调查表、样地样方表、样地洋房细表。系统需要的字典表为: 鸟类信息表、地点信息表、样线信息表、样地信息表。青海湖保护区工作者仅需要按系统的定制规则将以上数据表格的相关信息配置恰当, 并将获得的 XML 文件和字典数据库文件拷贝到 PDA 上, 即可获得简单实用的 PDA 数据采集工具, 如图 8 所示。

如果数据采集表格发生变化, 可能需要新添字段、或者删除字段、或者新增表格, 只要稍微进行定制操作, 重新拷贝 XML 文件就可以获得更新后的 PDA 数据采集页面, 而无需涉及程序的修改, 使得整个应用系统具有很好的扩展性。



图 8 PDA 采集页面

## 6 小结与展望

移动数据采集是室外调查工作的关键环节, 传统的手工采集方式已难以适应新形势下信息化发展的需要, 调查内容的日益完善给数据采集的精度和效率提出了更高要求。可定制的移动数据采集系统为室外数据采集工作提供了一种方便灵活、快捷高效的全新手段, 它不仅实现了 PDA 采集数据与后台数据库系统自

(下转第 32 页)

动对接和数据自动入库的“傻瓜式”工作模式，使科考人员更加方便、有效地解决了数据录入和入库的负担，而且通过定制的方式，使得该系统可以适用与众多同类型的应用，不仅减少开发人员的工作量，更使得系统可配置，易扩展。该系统正在青海湖科考中试用，将进一步完善并推广到海洋研究所、植物研究所、农业生态研究所使用。

### 参考文献

1 曾维军. GPS-PDA 在土地利用变更调查中的应用研究. 科技情报开发与经济, 2007,17(28):169-171.

2 邹国平,黄铮,郝国昌.基于 Pda 的高速公路养护数据采集系统设计与实现. 交通标准化, 2007,(9):97-101.

3 蔡华,刘剑欣.基于 Pda 的变电站自动化系统数据采集装置. 电力系统通信, 2007,28(8):58-60.

4 Wigley A, Moth D, Foot P. 张大威,译. Microsoft Mobile 移动应用开发宝典. 北京:清华大学出版社, 2008.

5 周丽芬,杨春金.嵌入式系统中移动数据库的研究. 通讯和计算机, 2007,2:58-60.

6 庄学山.嵌入式数据库 SQLite 在水情信息监测系统中的应用. 工业控制计算机, 2007,20(8):48-50.