

基于 Android 平台和云计算的电子词典^①

张卫芬, 管文强, 刘培梅, 王 娜

(解放军 72850 部队, 济南 250031)

摘 要: 主要探讨了基于 Android 平台和云计算的电子词典设计与实现过程, 系统本地客户端使用 Android 软件开发工具包以及 Eclipse 集成开发环境进行开发, 云服务器端则综合运用负载均衡、网络存储、虚拟化、结点管理等技术进行搭建. 系统采用二级索引方式存储词汇数据, 大大提高了查询检索词汇的速度.

关键词: 安卓; 云计算; 电子词典; 云服务器; 二级索引

Electronic Dictionary Based on Android Platform and Cloud Computing

ZHANG Wei-Fen, GUAN Wen-Qiang, LIU Pei-Mei, WANG Na

(School of Software, Fujian Normal University, Fuzhou 350108, China)

Abstract: This paper discussed the design and implementation process of electronic dictionary based on Android platform and cloud computing. The local-client used Android SDK and the Eclipse IDE to develop. The cloud server was constructed by load balance, network storage technologies, virtualization, node management and other technologies. The system used two-level index method to store data, greatly improved searching speed.

Key words: Android; cloud computing; electronic dictionary; cloud server; two-lever index

电子词典作为一种机器翻译工具已被人们广泛使用, 尤其是随着国内外经济贸易、文化教育等交流日益增多, 其作用也日益凸显. 移动电子产品较之传统的电脑具有体积小、便于携带等优势, 因而当前移动平台电子词典开发领域前景较为广阔. Android 系统相比于其他操作系统具有开源性、可移植性及可定制性等优势, 成为电子词典设计与开发所使用的主流操作系统.

本文提出了基于 Google Android 和云计算的电子词典的设计与开发, 借助网络将云计算平台上强大、快速、准确的翻译服务提供给移动终端使用. 系统使用 Android 平台推出的 SDK 和 Android 内置的 SQLite 小型嵌入式数据库开发接口, 综合运用负载均衡^[1]、结点管理、数据存储等技术构建云服务器, 在不影响用户查询性能的前提下有效缩小了客户端的体积.

1 Android系统及云计算技术

1.1 Android 系统架构

Android 系统架构从整体上划分为四层结构, 从上层到下层分别是应用程序层、应用程序框架层、系统

运行库层以及 Linux 内核层^[2]. 电子词典的 Android 应用程序开发是基于框架和组件的, Android 本身已在框架中提供了许多组件供应用程序使用, 我们除利用 Android 平台提供的一些组件外, 自己开发了新的组件并放入应用框架中, 供应用程序的使用; Android 开发库和运行环境包含了一组核心库, 提供了 Java 语言核心库内的大部分功能和 Dalvik 虚拟机; Linux 内核层则用来提供系统的底层服务.

从 Android 的系统架构来看, Android 和其他平台相比, 有着自身的一些特点, 如集成了 Dalvik 虚拟机等模块. 我们充分利用 Android 提供的模块接口, 开发出具有特色的应用程序, 使用了 Google 提供的 Android SDK 中的工具集. 它包括编译器、资源管理器、调试器和模拟器, 并使用 Eclipse 加上 Android Developer Tools 插件作为集成开发环境.

1.2 云计算

2006 年 8 月 9 日, Google 首席执行官埃里克·施密特(Eric Schmidt)在搜索引擎大会首次提出“云计算”的概念, 短短的几年时间里得到了广泛关注及快速发展.

^① 收稿时间:2014-05-16;收到修改稿时间:2014-07-07

它是由分布式计算、并行计算、效用计算、网络存储等多种传统计算机和网络技术发展融合的产物,其目的在于通过网络把相关计算资源整合成一个可定制的共享池,再借助商业模式将其强大的资源优势分布到终端用户手中。云计算平台从底层提供几乎无限的计算、存储和网络资源,其上的应用程序就像一个个具有人工智能的独立单元。从用户使用服务的角度来说,云计算的出现使得大量的计算工作从终端设备逐渐迁移到云计算平台上,进而减少终端的处理负担,并能根据用户自身需求选择合适的终端来使用云计算服务,享受云端强大的计算处理能力。

本文设计并实现了基于 Google Android 平台的云客户端,既具有本地识别功能,也可以通过网络接入云服务器,利用云计算技术极大地提高资源利用率,有效降低能源消耗。确保大量用户访问时不会造成终端运行速度降低,有效提高资源利用率并降低成本。

2 系统总体设计

2.1 系统总体功能结构

根据电子词典的现实需求及特点,本系统的总体功能划分如图 1 所示。

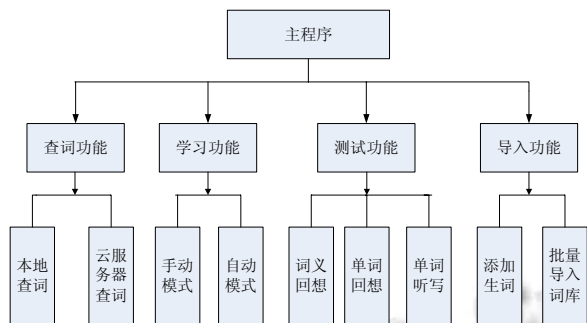


图 1 电子词典功能结构图

(1)查词功能

实现单词或词组的查词功能,提供中英文释义、英文发音、成语、汉语内容、反义词等。依据用户实际需求,既可以选择本地查词也可以使用云服务器实现查询。查询时支持模糊查询,由用户输入与查找词汇相关的内容,如其中的某些字等,系统查找出相应的一个或多个词条,全部显示在词条列表中,用户可直接翻阅全部词条。此外,词典还支持语音查询,系统调用语音库与用户口述内容进行比对,语音录入准确率达到 90% 以上。

(2)学习功能

为方便用户学习设计了手动和自动两种学习模式。手动模式需要用户自己根据实际需求保存生词到生词本、添加重点生词进行系统化学习。自动模式则是由系统根据历史记录自动筛选重难点词汇,用于系统化学习。

(3)测试功能

分为词义回想、单词回想和单词听写三种方式对用户进行测试。

(4)导入功能

为方便用户的个性化需求,我们开发了词库导入更新功能,分为单个词语导入和批量导入两种模式。其中单个词语名称及解释的导入根据个人需求自行添加词条,分别输入相应的词条解释,保存录入系统数据库中;批量内容导入则可根据词库种类将词库文件导入制定的数据库中。

2.2 云服务器架构图

云服务器是云计算服务的重要组成部分,是一个具有可弹性伸缩的云平台,该平台整合了传统意义上的互联网应用三大核心要素:计算、存储、网络,面向用户提供公有化的互联网基础设施服务。平台的规模可以根据应用实际灵活增减,从而保证良好的服务性能^[2];同时,云平台也具有较强的可靠性,当某个节点发生故障,可以在不影响其他节点运行基础上通过热迁移技术保证服务的正常运行。对于用户来说,云服务器的架构完全是透明的,用户与其交互就如同与一个单独的高性能服务器交互一样便捷,总体而言云服务器具有以下几大特点:

(1)横向扩展:云计算的第一个特点就是“超大规模”,因而云服务器的横向扩展能力就尤为重要。利用 LVS 集群技术搭建云服务器,使得云服务器可以灵活增减服务器集群中的物理设备,在完全不影响用户使用过程的前提下保证云数据中心的灵活性。

(2)高可靠性^[3]:虚拟机可以突破单个物理机的限制,动态的资源调整与分配消除服务器及存储设备的单点故障,实现服务的高可靠性。当一个计算节点的主机需要维护时,可以将其上运行的虚拟机通过热迁移技术在不影响其他节点运行的情况下迁移至其他空闲节点,用户毫无感觉。在计算节点物理损坏的情况下,也可以在 3 分钟左右将其业务迁移至其他节点运行。

(3)共享数据存储: 存储服务器是若干具有超大容量、配置了 SAS 硬盘+Raid5 阵列的服务器, 通过集群文件系统组成一个统一的存储池, 为节点内的虚拟机提供逻辑磁盘存储、非结构数据存储以及整合备份服务。

云服务器的系统架构如图 2 所示. 其中云服务器由一个负载均衡结点(LB)、一个磁盘阵列(RAID)以及多个物理服务器结点(RSX)和虚拟服务器结点(VMX)组成. 当用户的访问信息接入云服务器中, 首先由负载均衡结点 LB 将访问请求根据各个服务器结点(RSX 和 VMX)的权重和负载情况进行分析再将用户请求分发给它们处理, 然后服务器结点将处理结果返回给用户。

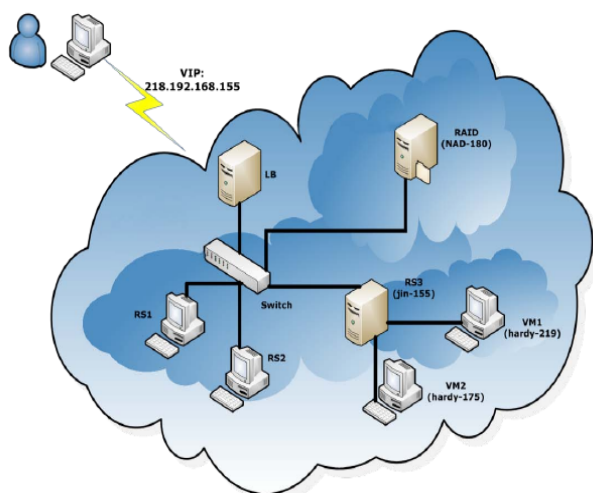


图 2 云服务器架构图

3 系统关键技术及实现

3.1 Gzip 网络数据压缩

Gzip 是一种流行的文件压缩算法, 现在的应用十分广泛, 尤其是在 Linux 平台^[5]. 手机作为一种便捷的移动设备, 其使用的网络具有自身的特殊性, 即流量、网速有一定限制. 因而基于其上的网络应用程序就需要对网络数据传输采用压缩技术, 本客户端使用了 GZip 技术. HTTP 协议上的 GZip 编码是一种用来改进 WEB 应用程序性能的技术, 大流量的 WEB 站点常常使用 GZip 压缩技术加快数据在网络中的传输速度. 当用户访问服务器时, 服务器运用该技术将网页内容压缩后传输到访问的设备中显示出来, 通常对纯文本内容可减少 70%以上的文件大小, 这取决于文件中的内容. 经过压缩后实际上降低了网络传输的字节数, 最

明显的好处就是可以加快网页加载的速度, 从而达到节省流量以及改善用户浏览体验的目的。

具体应用步骤为: 首先服务器端安装此功能模块, 并且以 GZip 的格式来返回内容. 在设置请求消息头 Header 中声明 Accept-Encoding 为 gzip, 这样你将会得到一个响应, 根据消息头 Content-Encoding 为 gzip 你可以知道, 传输过来的数据是经过 gzip 压缩的. 接收到服务器的返回时, 要根据 Response 的 Header 中 Content-Encoding 的值来决定是否需要解压 GZip 内容, Android 的 SDK 中内置了 GZIPInputStream 可以帮助我们解压 GZip 流得到一个正常的输入流。

3.2 二级索引存储

电子词典的查询效率在很大程度上取决于后台词库结构的设计, 由于词典的单词量较大, 若采用普通的线性存储结构, 每次查找时都需遍历数据表, 显然速度很慢. 因此, 本文采用二级索引结构^[6]来存储单词, 结构如图 3 所示。

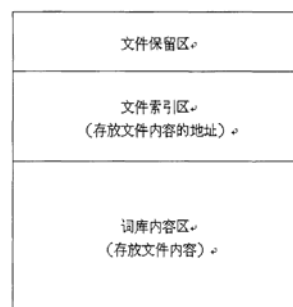


图 3 词典库存储结构图

基于二级索引存储结构的词典设计, 在查询单词时的基本流程是: 首先用户输入需要查询的生词, 系统根据该生词计算出相应的地址; 然后根据该地址在文件索引区读取其中的内容, 该内容则为存放单词的地址; 因此最后根据读取的地址在词库内容区找到具体单词及相应解释. 由此可见, 该方法不同于传统方法, 不用遍历词库, 而是通过文件索引方式间接读取查询的内容, 有效提高了查询单词的速度。

3.3 云端查询及通信流程

用户使用该电子词典实现翻译功能既可以采用本地服务也可以通过云服务器实现. 本文设计了一套通信算法使本地翻译功能接入云服务器, 算法的详细流程图如图 4 所示. 每一个云服务器都有多个结点, 每个结点都有相应的 IP 地址及端口用于联接云客户端。

在用户输入单词过程中对云服务器进行轮询，直到找到空闲服务结点。确认查询后，客户端将需要查询的词汇信息发送给空闲服务结点，再由云服务器进行相应处理，最后返回识别结果。由此可见，云客户端的作用就是负责接收、传输信息，而识别处理工作则都打包交给云服务器处理。同时我们也考虑到移动终端设备是基于无线网络，因而信号连接不顺畅或中断的情况时有发生，此时云客户端则可以选择调用本地翻译引擎进行识别，确保云客户端的适应性、可靠性。

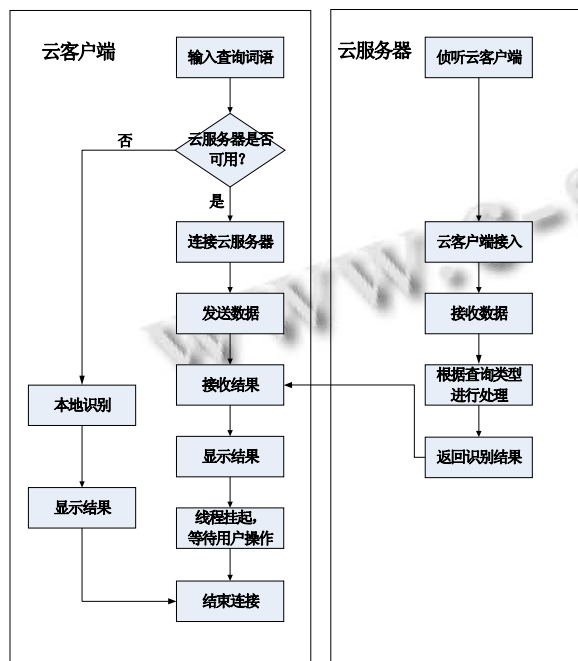


图 4 通信算法流程图

3.4 使用 xml 格式批量导入

项目中具有单词录入和批量录入功能，针对新的单词，可以使用单词录入功能，用户将单词写入到程序的录入界面当中，使用 AHibernate 框架，将单词录入到数据库当中，但对于大数据量的录入就会有一定的问题，因此，开发中使用比较流行的 xml 文件格式^[6]，将每个单词拼装成指定的格式，然后将文件放到手持终端目录下，就可以选中文件进行批量导入。

文件格式如下：

```
<root>
<obj>
<word></word>
<f_trans> </f_trans>
<s_trans></s_trans>
<t_trans></t_trans>
.....
```

```
</obj>
</root>
```

Root 为根节点每个 obj 为一个单词，word 属性值为单词，f_trans,s_trans,t_trans 依次为单词的第一种，第二种，第三种解释。

程序将文件内容读入，根据指定的 xml 解析方案，会将每一个单词解释出来，并进行入库。这样既能提高效率，也能保证录入时的稳定性。

4 系统测试

客户端的开发基于 Android SDK 2.2,Java JDK 1.6, 开发环境为 Eclipse3.6. 通过在一台服务器上虚拟多台虚拟机构成云服务器，对客户端的词典系统进行了功能模块和系统性能测试，均达到了设计要求，各个功能模块运行正常。实验通过用户数量以阶梯式方式不断增长到峰值对平均响应时间和相关性能进行测试分析。测试结果显示，100 个并发用户的响应时间仅为普通服务器的 1/3 左右，300 个并发用户的平均响应时间为 0.08 秒，500 个并发用户平均响应时间为 0.1 秒，且吞吐量非常稳定。

5 结语

本文设计并实现了一种基于 Android 平台的电子词典，同时针对海量用户，引入云计算技术。将众多的计算资源通过虚拟化技术组建成一个可动态伸缩的资源池，既具有本地识别功能，也可以通过网络接入云服务器，大大减少用户端的处理负担，并能根据个人需求来享受其强大的计算处理能力。

参考文献

- 1 周晓.上海轨道交通清分系统架构中应用云计算技术的探讨.地下工程与隧道,2012(2):53-56,69.
- 2 张仕成.基于 Google Android 平台的应用程序开发与研究.电脑知识与技术,2009,28.
- 3 冯登国,张敏,张妍,等.云计算安全研究.软件学报,2011, 22(1):71-83.
- 4 朱变,朱海.云计算中基于 TPM 的多银行电子现金系统.计算机应用研究,2014,31(2):551-554.
- 5 刘荣.质检专业电子词典的设计与研究[硕士学位论文].南京:南京理工大学,2011.
- 6 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(2):228-232.
- 7 何鸿君,王明昕,史晓东,郑若忠.一种简单、高效的电子词典组织策略.计算机科学,1996,2.