基于 Symbian OS短信智能过滤设计与实现

冯 松,欧阳鑫,黄青松
(昆明理工大学 信息工程与自动化学院,云南 昆明 650051)

摘要: 分析了目前短信过滤 2种方式的局限性,论述了对已知号码运用黑白名单过滤和对未知号码运用有害信息特征库扫描短信内容的方法,采用面向对象和组件化的思想,设计和实现了一种在基于 Symbian OS手机的短信智能过滤系统. 论文从系统设计目标出发描述了系统总体结构,介绍了用户界面和过滤引擎组件. 对过滤引擎的关键类进行了详细描述.

关键词: 手机操作系统; 垃圾短信; 消息服务器; 过滤引擎
中图分类号: TP319　　**文献标识码**: A　　**文章编号**: 1007 - 855X(2007)04 - 0043 - 04

Design and Implementation of Intelligent Short Message Filtering Based on Symbian OS

FENG Song, OUYANG Xin, HUANG Qing-song
(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650051, China)

Abstract: The limitations of two current methods of short message filtering are analyzed. And the methods, filtering black and white list for the known telephone numbers and scanning the short messages contents for the unknown numbers by adopting the feature library of junk information, are also discussed. By adopting the oriented object and component, an intelligent short message filtering system based on Symbian OS cell phones is designed and implemented. The general system architecture from the goal of system is described and the components of user interface and filtering engine, whose key classes are described in detail, are introduced.

Key words: mobile phone operating system; junk short message; message server; filtering engine

0 引 言

我国手机短信业务大幅增长始于 2000年,当年的发送总量为 10亿余条,2005年已经达到 3 046亿条. 然而同时,在短信发送中,有 3成短信属于不健康的、垃圾和带有诈骗性的信息,使短信成为垃圾和不健康信息传播的温床,对这类短信进行过滤和拦截已是迫在眉睫. 短信发送与接收是由通讯网络中的短信服务中心来完成的. 它依靠短信服务中心的存储和转发机制. 发送方发送的短信由短信服务中心对其进行存储转发,短信发送到短信服务中心后,如果对方处于关机或不在服务区,信息在短信中心储存 24 h. 从短信传输过程看,对信息过滤可在两处进行: ①在短信服务中心:当对实时接收的短信息进行存储转发的时候,对有害信息实施监控. 这种监控方式是根据短信发送频次监督模式,如对短信发送量超过 100条 /h和发送内容重复进行监控[4]. 但目前通信公司对短信过滤还缺少足够的法律依据,所以大部分公司对有害信息仅采取监控措施; ②在手机终端:在接收信息时,根据预先在手机中设定的黑名单来识别是否对信息拦截和过滤,即"短信防火墙". 目前这种功能仅出现在一些少量高端手机中,普通的手机不具备此功能. 另外这种方法也存在局限性:其一有害信息的发送者经常使用多部号码发送短信,使得采用预先设定黑名单的监控方式失效;其二我们经常仅需要过滤掉含有某些关键字的信息,而不是将所有的信息封堵.

Symbian操作系统是一个专用于手机等移动设备的操作系统. 目前采用 Symbian操作系统的手机占据

收稿日期: 2007 - 01 - 05. **基金项目**:云南省教育厅科学研究基金项目 (项目编号: 6Y00790);昆明理工大学科研启动基金项目 (项目编号: 2006 - 53).

第一作者简介:冯松 (1971 -),男,硕士,讲师. 主要研究方向:中文信息检索、信息安全. **E - mail:** fengsong@public.km.yn.cn

了智能手机操作系统市场70%以上的份额.本文论述基于Symbian操作系统采用在手机端的方式对垃圾短信进行智能过滤拦截的设计及其实现.

1 系统设计目标

系统目标是设计嵌入到手机中的具有较高精度的短信过滤平台,应具备如下特点:①提供强大的用户自定义功能:黑白名单规则管理,垃圾信息规则特征库的管理;②多种过滤技术相结合:根据黑白名单过滤和根据短信内容过滤;③智能的模糊过滤:对伪装短信进行快速识别,使用多规则对短信文本进行过滤拦截;④高效率和资源占用少:目前手机作为一种受限资源,它的硬件性能和价格还无法同现在的计算机相比较,系统的运行必须具有较好的时间和空间复杂度.

2 系统总体结构模型

依据系统的设计目标,手机短信智能过滤系统由短信监控、短信提取、黑白名单检索、内容分析、垃圾短信处理、信息维护模块和拦截短信日志库、垃圾短信特征库和黑白名单数据库组成.系统结构模型如图1所示.

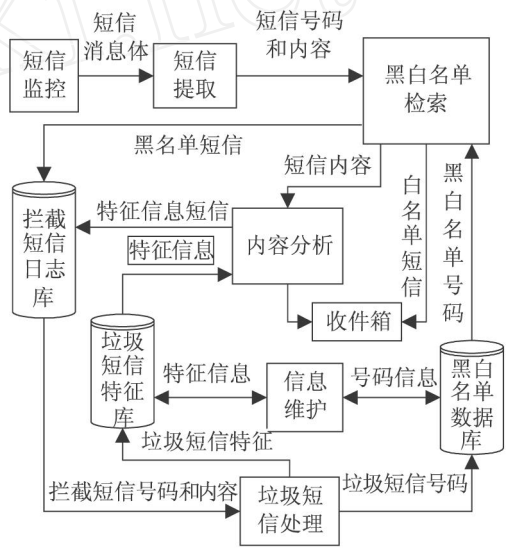

图1 系统总体结构模型

Fig.1 General system architecture

1) 短信监控模块:该模块主要对发送到用户的短信息进行实时捕获.它实时监控手机短信接收端口,当短信接收端口接收到短信息后,将消息体送给短信提取模块.

2) 短信提取模块:该模块的作用就是把短信解码,从短信的消息体中提取发送方的号码和短信内容,并将其送给黑白名单检索模块.

3) 黑白名单检索模块:"白名单"即是不阻拦名单,从白名单发来的任何信息均不会被阻挡."黑名单"是绝对阻拦名单,从黑名单发来的信息全部都会被阻挡.该模块就是根据预先设定的号码与由短信号码进行匹配,来决定是否允许通过.如果发现黑白名单中都不存在此号码,就将短信内容送入内容分析模块.

4) 内容分析模块:内容分析模块是短信过滤系统的核心,该模块主要对来自于未知号码的短信进行文本检测.根据手机用户预先设定的垃圾信息特征库运用模式匹配算法对短信内容进行扫描检测.如果短信内容中包含了用户预先设置的特征信息,将接收的短信移到拦截短信日志库中,否则允许通过.

5) 垃圾短信处理模块:在短信过滤的实现中,可能出现将有用短信误判为垃圾短信的情况,同时对真正的垃圾短信还需要对它进行特征分析,提取特征信息,对垃圾短信特征库更新.

6) 系统维护:主要提供用户对黑白名单数据库和垃圾信息特征库进行设置和维护.

3 系统实现

用于开发测试的智能手机是诺基亚6670,该手机采用嵌入式Symbian操作系统7.0.开发平台选择VC6,并运用Nokia公司提供的Series60_2ndSDK包.

3.1 系统软件结构

为了实现软件的可扩展性、高效性和灵活性,系统采用组件技术和面向对象软件设计思想,将系统分为两部分:过滤引擎和用户界面(GUI),并采用了两者分离的原则.过滤引擎是一个驻留手机内存的EXE程序,它包含了短信获取和解析,黑白名单检索和内容分析等模块,用于对短信进行实时处理.用户界面是一个APP应用程序,用于实现系统和用户交互,包括系统的设置、黑白名单和垃圾特征库维护,对拦截信息处理等功能组成.系统主要组件类图如图2所示.

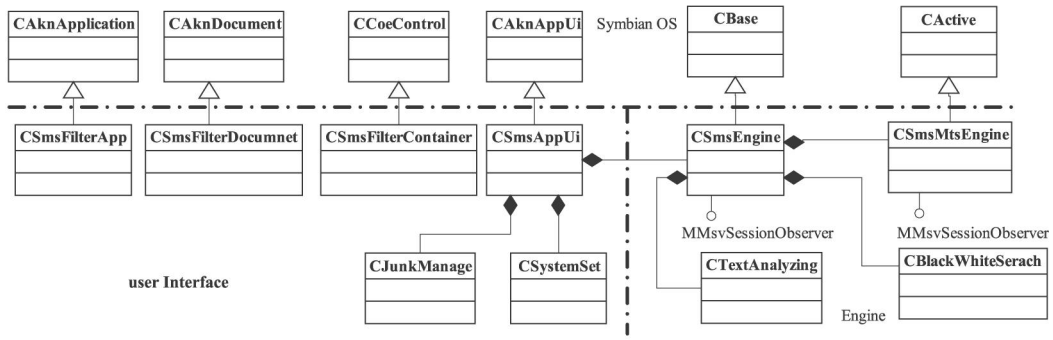

图2 软件组件类图

Fig.2 Class diagram

系统组件类图主要由 Symbian OS, User Inter face, Engine3个部分构成, Symbian OS部分是系统实现所继承和使用的系统基类. User Interface部分的组件设计采用了基于 Symbian OS控件的应用程序架构,系统定义并实现了 2个类: CJunkManage类, CSystemSet类. CJunkManage类用于用户对拦截短信日志库进行管理; CSystemSet类用于对黑白名单和垃圾短信特征库进行设置和管理. Engine部分是系统的核心,下面将重点描述过滤引擎组件的实现.

4 过滤引擎实现的关键技术

Symbian操作系统把对手机中的 SMS、MMS、Email操作称为消息传送架构. 它是基于 Client/Server机制, Symbian系统提供消息服务器,它负责管理手机上的所有消息资源. 为获取新接收的短信息,过滤系统与消息服务器建立通讯通道,结构如图 3所示.

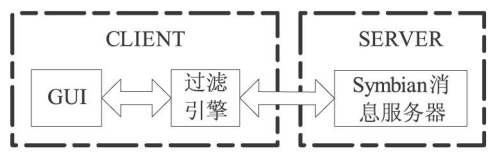

图3 Client/Server结构

Fig.3 Client/Server architecture

这个过程需要 2步: ①要使用系统类 CMsvSession的 OpenAsyncL()方法创建一个异步 Session接口,同消息服务器的会话建立连接; ②通过实现混合接口类 MMsvSessionObserver的 HandleSessionEventL()方法并实例化. 通过接收该实例 HandleSessionEventL()方法中的 EMsvServerReady事件来获得会话连接成功的通知. 短信过滤系统将过滤引擎作为 SymbianOS的消息服务器的 Client端,通过 Session来获取实时接收的短信.

4.1 CSmsEngine类

它是过滤引擎主体类,实现了系统混合接口类 MmsvSessionObserve. 该类有 3个功能: ①打开与 Symbian OS消息服务器的通讯端口,并建立会话; ②监听消息服务器,获得新短信并将其传送给 CsmsMtsEngine类; ③负责调度 CBlackWhiteSearch类和 CTextAnalyzing类对有害信息进行处理. 类中的主要数据结构如下:

```
class CSmsEngine : public CBase{
……
private:
CMsvSession * iMsvSession;
CMsvEntry * iMsvEntry;
TMsvId iNewMessageId;
HandleSessionEventL();
}
```

类中的最主要的数据成员有 MsvSession, MsvEntry, iNewMessageId. SymbianOS中的各种消息都是以数据项(Entry)形式供程序操作. 消息服务器为每个数据项分配一个惟一的数值做为标识,它的数据类型是 TmsvId. 跟接收短信有关的数据项是收件箱,它的 ID是 KmsvGlobalInBoxIndexEntryId. iNewMessageId就是通过获得收件箱的 ID来获得新短信的信息. MsvSession是 CmsvSession类的一个实例,通过它与消息服务器建立会话,并获得 CMsvEntry上下文对象. MsvEntry是数据项的句柄,它提供了操作数据项(Entry)的

各种接口,并根据指定 ID通过该接口获得包含数据的数据项.在这里主要是新短信所在数据项的句柄.

在 CsmsEngine类中最重要的方法是实现接口混合类 MmsvSessionObserver的 HandleSessionEventL()方法. MmsvSessionObserver类提供了 EMsvEntriesCreated, EMsvEntriesChanged, EMsvEntriesDeleted, EMsvEntriesMoved, EmsvServerReady等与短信有关的事件.

在 Symbian OS中,新短信到来时消息传送服务器将先创建一个类型为"消息"的对象,用于存储新短信,同时产生一个是 EMsvEntriesCreated事件,当消息服务器存储消息 Entry结束时触发一个 EMsvEntriesChanged事件.获取新短信要先将这两个事件相结合进行处理.

EMsvEntriesDeleted和 EMsvEntriesMoved事件用于调度 CBlackWhiteSearch类和 CTextAnalyzing类对短信号码和内容进行扫描检测.

4.2 CTextAnalyzing类

内容分析模块是短信过滤系统的核心,因为大量的垃圾和不良欺诈短信往往来自不可预知的号码.由于发送垃圾短信者经常对所发的短信息中进行伪装,以便穿透短信过滤系统的拦截和封堵,所以在进行内容分析前要对短信内容进行预处理.然后根据手机用户预先设定的垃圾信息特征库对短信内容进行扫描检测. CtextAnalyzing类主要功能列举如下.

4.2.1 短信文本预处理

系统主要对以下 2种情况进行处理:

1)短信文本预处理[3]:对夹杂在短信中的一些数字, @, +等噪声符号进行处理.如"现有走@私套+牌车销售".处理后变为"现有走私套牌车销售".

2)组合关键字预处理:运用关键字对短信进行扫描时,各有害关键字之间是"或"的关系,即在短信文本中只要存在一个有害关键字即可判定该信息有害.但很多时候一个在短信中找到一个关键词,并不能判断该短信是否有害,往往需要结合其它的关键词.例如:只看到"销售",并不能认定该短信有害,如果在同一条短信中出现了"走私",就可以认定该短信有害.对这类关键字采用静态链表方式建立关联关系.在对文本扫描检测时,如果发现一个关键字,还应判断是否为组合关键字,如果是,还要与这个关键字相关联的关键字进行匹配.

4.2.2 内容分析

目前经典的多模式精确匹配算法有 Aho_Corasick算法、AC_BM算法和 Wu_Mander算法等.这些算法大多采用有限自动机和 HASH的思想.内容分析的实现运用了 WM算法思想.根据短信文本长度不超过70个字符的特点和手机的软硬件的资源受限的条件,结合中文字符串匹配中多数情况下为首字符匹配失败的情况,设计了 2个表,一个是 HASH,表中的索引值即为有害特征信息的第一个汉字的 Unicode编码值;另外一个是 PATTERN表,存放所有的有害信息.

假设有害信息特征库为"购物,彩玲,二手车".构建 Hash表和 PATTERN表过程为:取每一个关键词的第一个汉字的 Unicode编码作为 HASH表的索引值,根据这个值对所有关键词进行排序,然后将其存放在 PATTERN表中.对于 PATTERN[i] (第 i输入模式)的 Hash值即为该模式第一个汉字的 Unicode编码值,设为 H,那么就将 HASH表的第 H项设置为 i,即 HASH[H] = i例如,彩铃的编码是 5F69 73B2,它在 PATTERN中的索引为 1(即为 i),则将 i填入 HASH表中 5F69位置(即为 H,"彩"的 Unicode编码),将 HASH表中的 5F70(即 H+1)位置填入 i+1(即哈希值为 5F69的模式只有 1个).

扫描匹配过程如下:

1) 从待扫描短信文本编码的左端开始,依次取 2个字节(即 1个汉字 Unicode编码)得到的结果为 H.

2) 检查 HASH[H]和 HASH[H+1],若两者相等,即在模式集合中不存在此字符,转到(1)继续;若两者不等,即有可能与某个模式发生了匹配,转到(3).

3) 设当前 HASH[H]≤index<HASH[H+1]的每一个 index值,取出 PATTERN[index]值(即模式匹配串),与短信文本当前位置的编码值进行比较,如果不等,转到(1);如果相等,则发生了匹配,完成了内容的分析.

(下转第 56页)

究"[7],其研究结果为图7所示的一条曲线,这条曲线表明:人的偏好与产品的视觉复杂程度没有直接的关系,人们偏爱的是有一定的视觉复杂度、能给人留下深刻印象的产品.具象产品在这方面具有得天独厚的优势.这是因为:

(1)具有吸引力的主要决定者不是该产品的复杂性,而是使用者观察到的视觉复杂度.因此,一个实际上很复杂的产品可能给人的感觉是简洁的.

(2)具象化产品对消费者有抽象设计无法比拟的熟悉度,另一方面,具象产品设计的趣味性强,它能很快抓住观察者的注意力,增加消费者的熟悉度;熟悉度是产品设计吸引力的又一因素,因为具象的物品的熟悉度高,记忆性强,所以增加了产品的吸引力,这也是具象产品设计的一个潜在优势.例如图6中的花篮式座椅设计本身是复杂的,但消费者对花篮有很高的熟悉度,所以整体看上去给人一种简洁的感觉.总之,在设计过程中,设计者要选用适当的具体形象,使产品的功能和形象更好的结合,来体现设计的主题.

具象设计作为产品设计的重要方式之一,是应该得到重视和研究的.具象思维是设计师在设计过程中,因设计的需要而产生的设计思路,它能将设计师的思路拓宽,使设计师能够从更多更深的层次思考问题.具象设计可以帮助设计师充分发挥想象力,对产品的研究开发有积极作用,它将会为产品设计注入新的生机与活力.

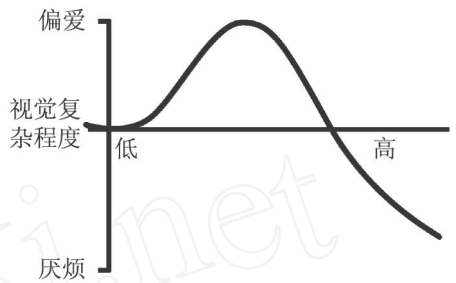

图7 视觉吸引力与外形复杂度的关系

Fig.7 Relationship between visual attraction and formal complexity in design

参考文献:

[1] 李妮.产品的趣味化设计方法研究[J].工程图学学报. 2006,(5):117-120.
[2] 杨君顺,马镭,杨刚.产品的非物质设计及其可持续发展[J].包装工程,2006,27(3):153-155.
[3] 诸葛铠.图案设计原理[M].南京:江苏美术出版社,1991:53.
[4] 柳冠中.历史——怎样告诉未来[C].装饰艺术文萃.北京:北京工艺美术出版社,1991:318.
[5] 刘国余.产品设计[M].上海:上海交通大学出版社,2000:42-44.
[6] Mike Baxter. Product Design: A Practical Guide to Systematic Methods of New Product Development[M]. Cheltenham: Stanley Thornes Ltd, 1995.
[7] 郑仁华,干静.设计中的三段式[J].包装工程,2006,27(5):243-245.

(上接第46页)

在过滤引擎中涉及的类还有CBlackWhiteSearch类和CSmsMtsEngine类.CBlackWhiteSearch类用于对接收的短信进行黑白名单的检测,主要功能有:白名单检测、黑名单检测等操作.CSmsMtsEngine类用于对手机上的短信文件夹(如收件箱,草稿箱,发件箱等)进行操作,主要功能有短信复制,移动和删除,获取短信号码及内容等操作.这两个类被CSmsEngine类调用.

5 结束语

本文介绍了基于SymbianOS智能手机的短信智能过滤系统的设计过程,并重点介绍了软件的组件类以及过滤引擎的实现.论文所论述的短信智能过滤系统对于智能移动通讯设备的信息安全的研究具有重要的现实意义.

参考文献:

[1] Leigh Edwards Richard Barker. Developing series60 Applications[M]. 北京:人民邮电出版社,2005.
[2] Wu S,Manber U. A Fast Algorithm for Multi-Pattern Searching[J]. Technical Report TR-94-17, University of Arizona, 1994.
[3] 陈儒,张宇,刘挺.面向中文特定信息变异的过滤技术研究[J].高技术通讯,2005,15(9):10-15.
[4] 王春晖.关于治理垃圾短信的若干意见[EB/OL]. http://www.law-lib.com/lw/lwml.asp. 2006-05.