

1 引言	2
1.1 编写目的.....	2
1.2 背景.....	2
1.3 定义.....	2
1.4 参考资料.....	2
2 任务概述	3
2.1 目标.....	3
2.2 用户的特点.....	3
2.3 假定和约束.....	4
3 需求规定	4
3.1 对功能的规定.....	4
3.2 对性能的规定.....	5
3.2.1 精度.....	5
3.2.2 时间特性要求.....	5
3.2.3 灵活性.....	6
3.3 输入输出要求.....	6
3.4 数据管理能力要求.....	6
3.5 故障处理要求.....	7
3.6 其他专门要求.....	7
4 运行环境规定	7
4.1 设备.....	7
4.2 支持软件.....	8
4.3 接口.....	8
4.4 控制.....	8

软件需求说明书

1 引言

1.1 编写目的

本要求规格说明书对垃圾邮件过滤系统的需求进行了全面的分析,明确所要开发的软件应具有的功能、性能与界面,使小组开发人员能清楚地了解用户的需求,并在此基础上进一步提出概要设计说明书和完成后续设计与开发工作。为明确软件需求、安排项目规划与进度、组织软件开发与测试,撰写本文档。

本文档仅供小组开发人员参考。

1.2 背景

本项目全称为《Clean Mailbox》,由张森、严灏共同提出设计,并由张森、严灏协同开发的一款基于机器学习算法的垃圾邮件过滤系统。本软件需要在 windows XP 及以上系统,连接互联网的环境下运行,并可以完成在互联网环境下收发邮件、标记删除垃圾邮件等功能。

本软件需要通过邮箱账户获取来自电子邮件服务商的账户数据,并通过对服务商发送请求完成电子邮件的相关操作。

1.3 定义

- **系统:** 若未特别指明,统指本邮箱软件;
- **QT Creator:** Qt Creator 是跨平台的 Qt IDE, Qt Creator 是 Qt 被 Nokia 收购后推出的一款新的轻量级集成开发环境 (IDE)。此 IDE 能够跨平台运行,支持的系统包括 Linux (32 位及 64 位)、Mac OS X 以及 Windows;
- **Anaconda:** 一个开源的 Python 发行版本,包含了 conda、Python 等 180 多个科学包及其依赖项;
- **分类器:** 在已有数据的基础上学习的一个分类函数或构造出的一个分类模型;
- **SQA:** 软件质量保证 (SQA-Software Quality Assurance) 是建立一套有计划,有系统的方法,来向管理层保证拟定出的标准、步骤、实践和方法能够正确地所有项目所采用。

1.4 参考资料

《项目开发计划 2.0》见数据仓库 <https://github.com/airzs/EmailBox/>

《软件工程——实践者的研究方法》 Roger S. Pressman, Bruce R. Maxim 编著，机械工业出版社，2015

《数据库系统概念》 Abraham Silberschatz, Henry F. Korth, S. Sudarshan 编著，机械工业出版社，2012

《统计学习方法》 李航编著，清华大学出版社，2012

《机器学习》 周志华编著，清华大学出版社，2016

《SQL Server 实用简明教程》 闪四清，清华大学出版社，2003

2 任务概述

2.1 目标

数据显示，近年来电子邮件用户收到的垃圾邮件所占比重已经超过总邮件数的一半，且呈缓慢增长的趋势。在工作生活高度信息化的今天，大量的垃圾邮件不仅严重影响了电邮用户的体验，更加降低了用户的工作效率。如何有效的标记、删除垃圾邮件便成了一个亟待解决的问题。而我们的项目则旨在为用户更好的标记垃圾邮件，为用户删除垃圾邮件提供便利，提高用户的工作效率。

团队计划在本学期内完成此垃圾邮件过滤系统项目，为用户提供邮件收发，垃圾邮件标记分类等功能，结合软件工程方法加强小组成员之间的团队协作，做到框架的清晰、简洁。代码整体做到低聚合性和高耦合性，重用性高，减少代码编写过程中的错误。核心分类算法到达较理想的期望值，前端界面用户友好，为用户提供尽量好的使用体验。让工作生活中经常使用电子邮件的用户更方便地标记垃圾邮件，为用户删除垃圾邮件提供便利，提高用户的工作效率。

本软件依赖于电子邮件服务商的数据库，需要通过邮箱账户获取相关的账户数据，并通过对服务商发送邮件请求完成电子邮件的收发、删除等操作。

2.2 用户的特点

- 系统的操作人员需要具有一定的计算机和电子邮件的基础操作知识；
- 开发维护人员需要有 QT Creator 的开发经验以及相关程序设计经验，有电子邮件

协议和邮件文本分类相关的算法设计的基本知识。

2.3 假定和约束

设计开发用 Windows PC 两台，完成时间限制在 2017 年 5 月 24 日之前，无资金供给，无资金限制。

3 需求规定

3.1 对功能的规定

编号	名称	简述	输入	处理(算法)	输出
1	登录	用户输入账号密码登录邮箱	账号、密码	验证账号密码正确性并返回结果	若账号密码正确，进入邮箱主界面，否则提示用户输入错误
2	保存账号密码	对用户账号密码进行保存	选择“保存账号密码”功能	对所输入账号密码进行储存	在登录界面显示已保存账号信息
3	写信	用户进行邮件输入	收件人、主题、正文	对所输入内容进行储存	保存用户输入内容并在界面显示所有内容
4	发送邮件	用户通过点击发送进行邮件发送	点击“发送”功能	对邮件进行数据结构转换并通过协议发送邮件	若发送成功，提示用户操作成功，否则提示用户发送失败
5	查看邮件	用户通过收件箱查看相应邮件	点击收件箱并选择相应邮件	根据用户选择提取相应的邮件并显示全部内容	显示用户所选择邮件的所有内容
6	回复邮件	用户对当前查看邮件进行回复	点击“回复”功能	提取当前邮件信息并进入写信界面，提示用户输入回复内容	进入写信界面并且已标识收件人

7	删除邮件	用户可选择删除无用邮件	选中相应邮件并选择“删除”功能	对用户所选取的邮件进行删除并删除其储存	刷新邮箱并提示用户成功删除邮件
8	标识垃圾邮件	用户选择垃圾邮件并标识移入垃圾箱	选中相应邮件并选择“移入垃圾箱”功能	将用户所选取的邮件移入垃圾箱	将邮件从收件箱移入垃圾箱并刷新界面
9	邮件存草稿	用户将当前输入邮件进行存草稿	点击“存草稿”功能	对用户当前操作邮件进行存草稿移入草稿箱	在草稿箱生成当前邮件副本
10	垃圾邮件过滤	对已收邮件进行垃圾邮件过滤,将垃圾邮件移入垃圾箱	所有邮件内容	通过学习算法对邮件中出现的关键词进行解析过滤,标识垃圾邮件并进行移动处理	将垃圾邮件移入垃圾箱,普通邮件放入收件箱
11	退出邮箱	用户选择安全退出邮箱	点击“退出”功能	根据用户点击退出邮箱系统,返回登录主界面	返回登录主界面

3.2 对性能的规定

3.2.1 精度

- 输入数据精度的要求：输入数据为文本，无精度要求
- 输出数据精度的要求：小数点后保留 2 位有效数字
- 传输过程中的精度：小数点后保留 2 位有效数字

3.2.2 时间特性要求

- 响应时间：不超过 3 秒；
- 更新处理时间：软件界面更新响应为即时刷新；
- 数据的转换和传送时间：数据传输处理为即时处理
- 解题时间：分类器训练时间不超过 1mins，分类器分类时间不超过 3 秒；

3.2.3 灵活性

- 操作方式上的变化：该系统适用于 Windows XP 及以上的操作系统
- 运行环境的变化：在 Windows PC 上通用
- 精度和有效时限的变化：无
- 计划的变化或改进：软件升级后不影响用户原有数据

3.3 输入输出要求

输入输出要求如下表所示：

	类型	格式	数值范围	精度
输入	账户、密码	字符串	无	无
	邮件收件人、标题、正文	字符串	无	无
	联系人备注、账号	字符串	无	无
输出	当前信息	字符串	无	无
	邮件发件人、标题、正文	字符串	无	无
	提示信息	字符串	无	无

3.4 数据管理能力要求

- 需要管理的文卷：存放登录用户信息的数据库，包括用户信息表，邮件信息表，联系人表，草稿表，系统日志表等。
- 基本数据设定：基本数据设定要符合实际情况，既要保证数据的保密性，又要保证数据的安全性，将用户利益放在首位。
- 表和文卷的大小规模：表中的数据要根据登录用户信息的改变及邮件信息的更新而不断的增加，数据管理能力要快、要有序。

- 按可预见的增长对数据及其分量的存储要求作出估算：根据邮件数目和联系人数目的增多，表中的相应字段也会逐渐增多，记录会逐渐增大，要求有较多的储存空间。
- 用户信息的存储：将系统所用级别的用户的登录验证信息准确的存储到数据库中，还包括对数据的增、删、改等操作。

3.5 故障处理要求

系统可能出现的故障以及故障处理的要求如下表所示：

故障类型	故障后果	故障处理要求
PC 断电	用户信息丢失	程序定时保存用户工作内容，保证用户数据不会完全丢失
互联网络断开	用户信息丢失	程序定时保存用户工作内容，保证用户数据不会完全丢失
PC 端数据文件丢失	软件运行出错	尽量利用电邮服务商提供的数据，减少本地数据依赖
内存耗尽	软件崩溃	定时保存用户工作内容，保证用户数据不会完全丢失

3.6 其他专门要求

无其他专门要求

4 运行环境规定

4.1 设备

- Windows PC 一台；
- 酷睿 i3 以上的处理芯片一枚及以上；
- 1GB 显存及以上的兼容显卡；
- 500M 的存储空间及以上

- Internet 互联网接入
- 联机 VGA 彩显一台

4.2 支持软件

- 操作系统：Windows XP 及以上操作系统
- 编程开发软件：Visual Studio 2015, QT Creator 5.7.0, Anaconda 3.4.0

4.3 接口

本系统采用的数据通信协议主要有：邮局协议（Post Office Protocol，也即 POP 协议），传输控制协议/因特网互联协议（Transmission Control Protocol/Internet Protocol，也即 TCP/IP 协议）。软件各功能模块化，各模块过程之间采用函数调用、参数传递、返回值的方式进行消息传递。接口传递的信息将是数据以数据结构封装了的数据，以参数传递或返回值的形式在模块之间传递。

4.4 控制

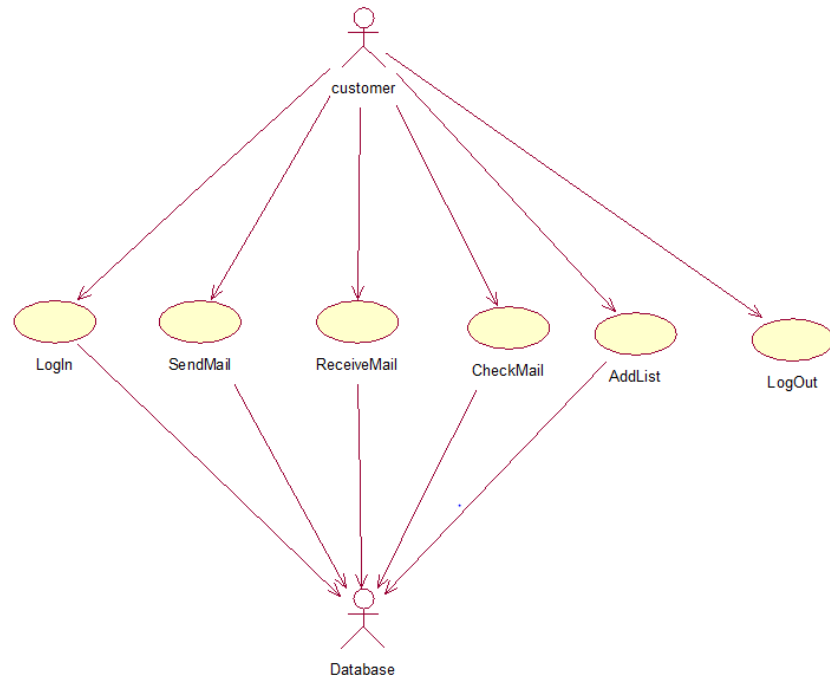
本系统有完善的用户登录及安全机制，通过登录提供用户名和口令后才能访问系统。用户暂时离开时，可以轻松注销，防止其他人员进入，保证用户数据的安全性。

控制信号

- 时钟：用来同步各种操作
- 总线请求：说明这些控制信号的来源
- 总线允许：表示需要获得总线使用权的部件已获得了使用权
- 存储器写：将数据总线上的数据写至存储器的指定地址单元内
- 存储器读：将指定存储单元中的数据读到数据总线上
- I/O 读：从指定的 I/O 端口将数据读到数据总线上
- I/O 写：将数据总线上的数据输出到指定的 I/O 端口内
- 传输响应：表示数据已被接收，或已将数据送至数据总线上

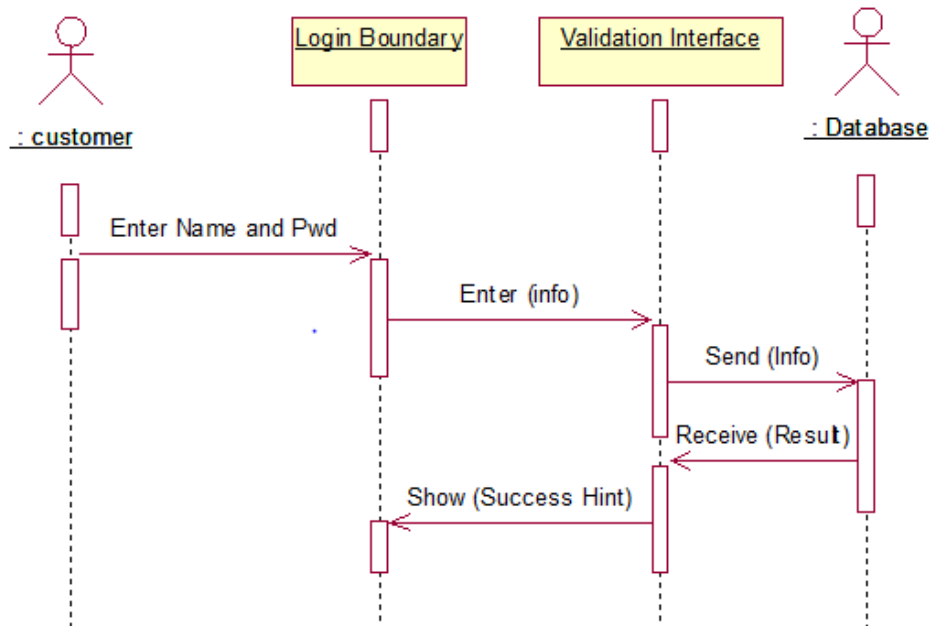
5 附录

5.1 用例图：

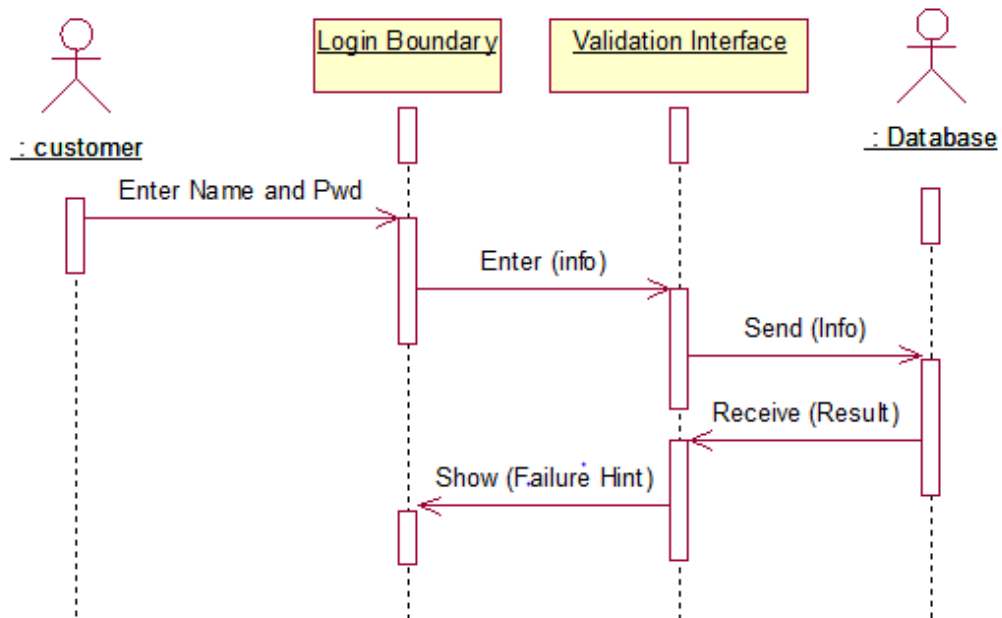


5.2 时序图：

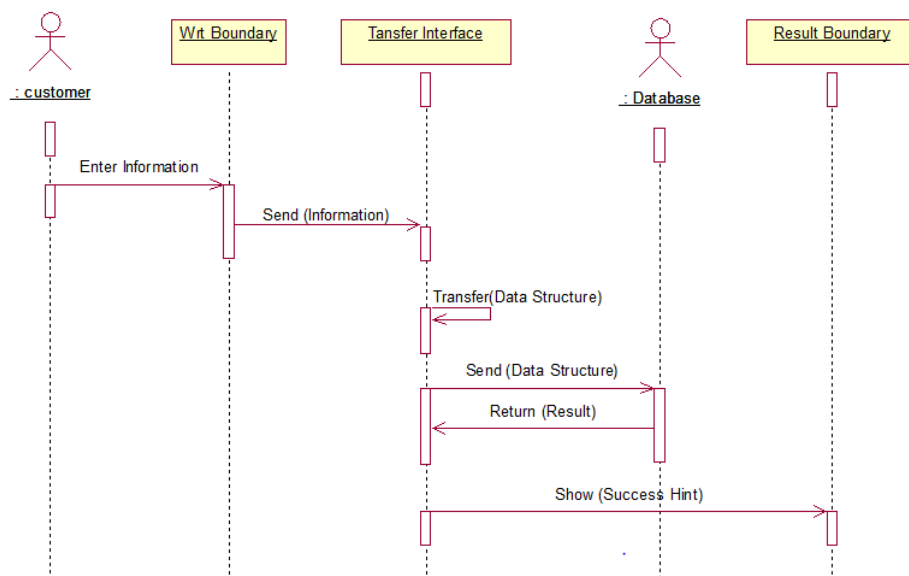
login(success):



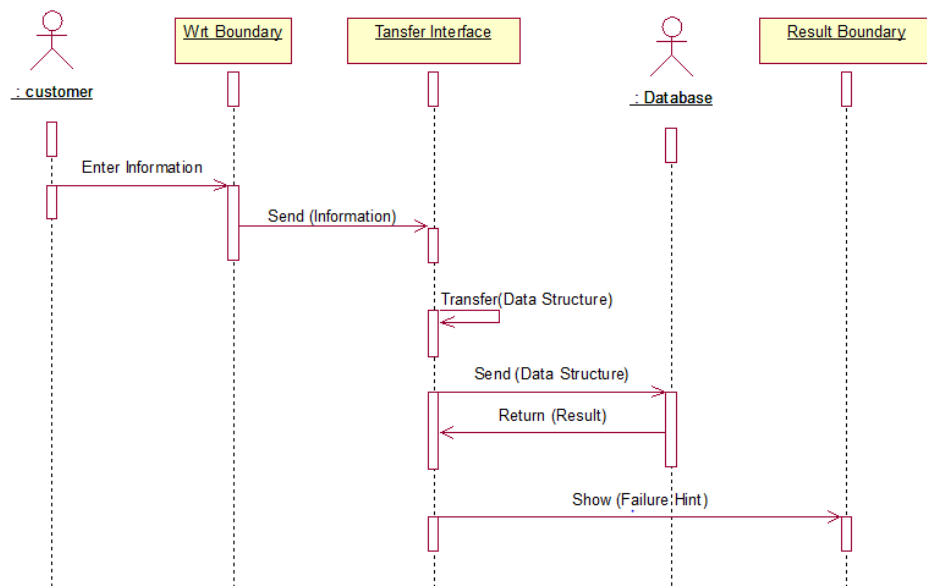
login(fail):



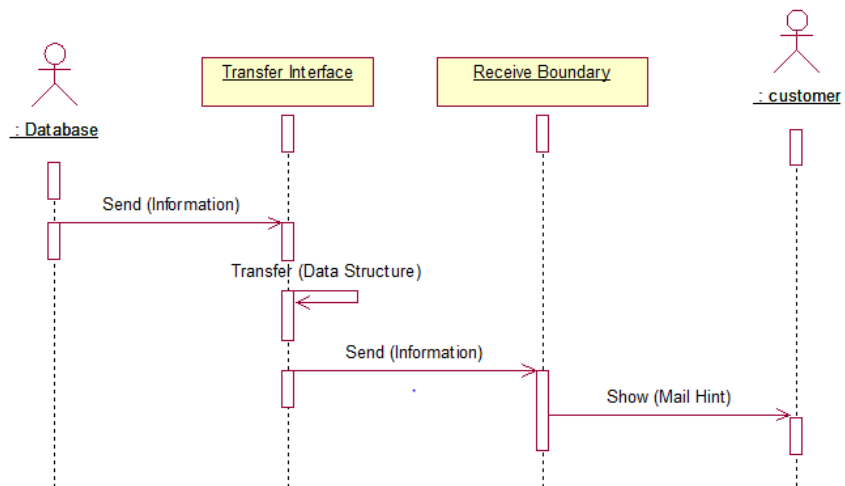
sendMail(success):



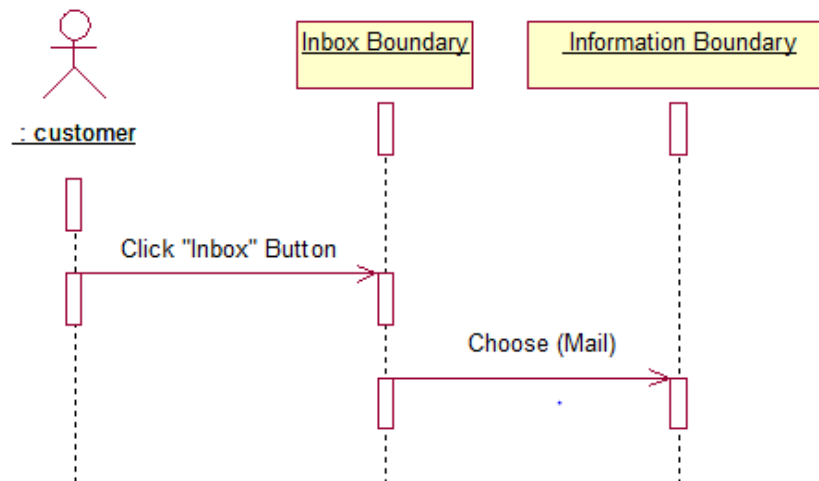
sendMail(fail):



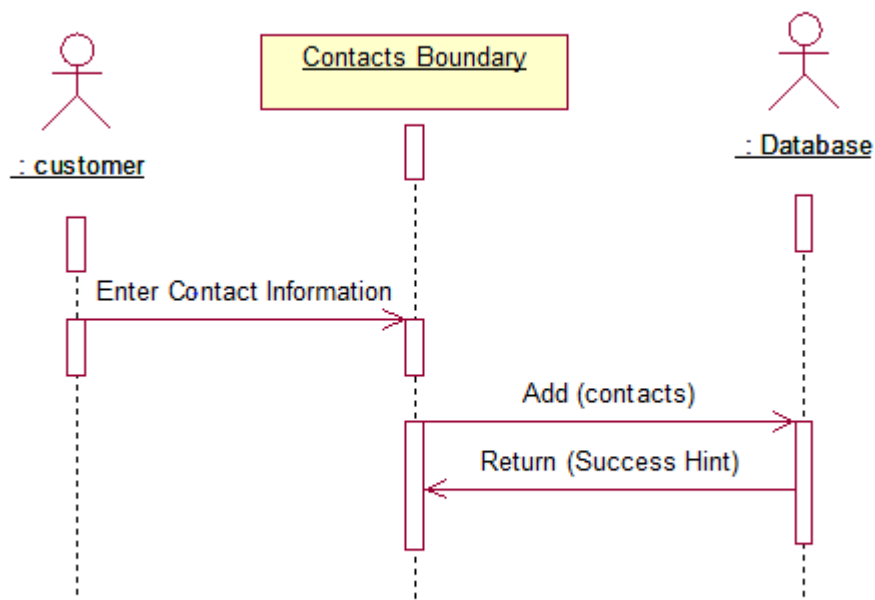
receiveMail:



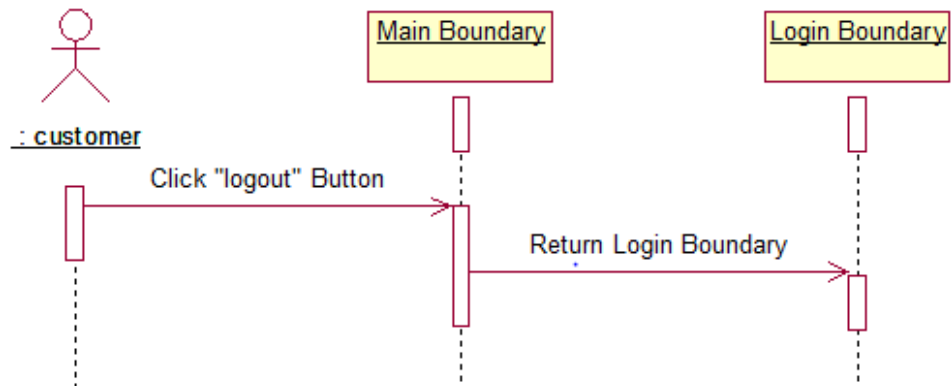
checkMail:



addList:

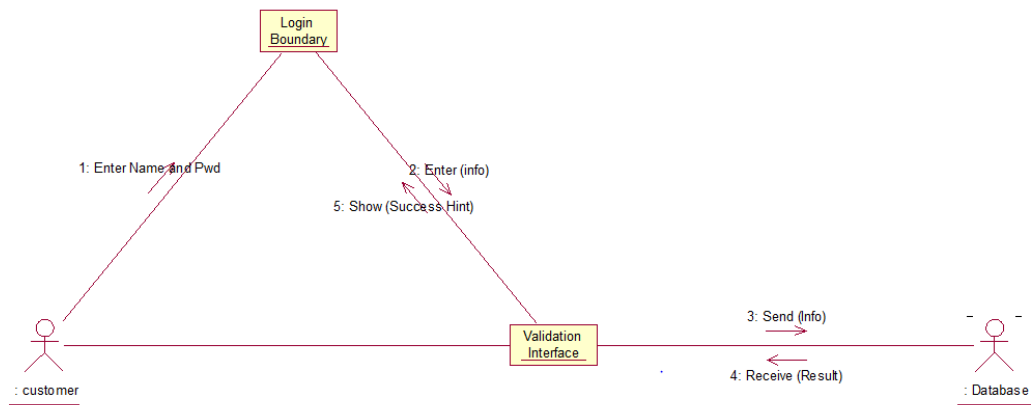


logout:

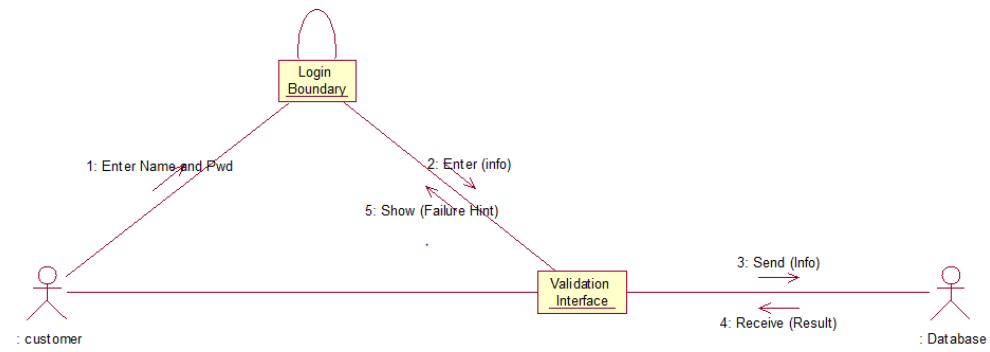


5.3 协作图:

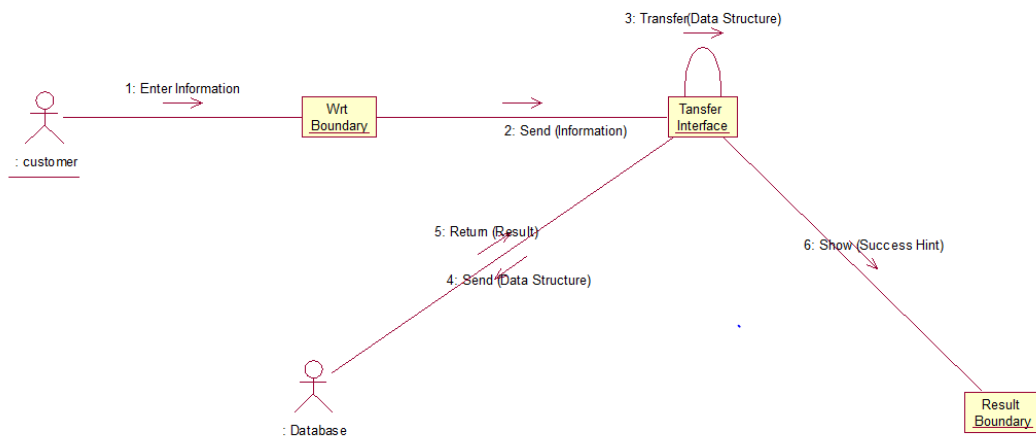
login(success):



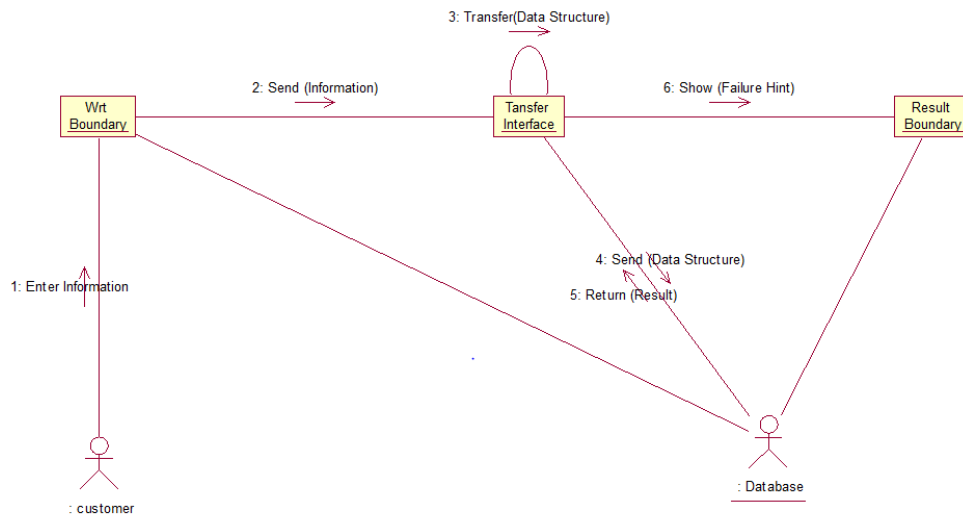
login(fail):



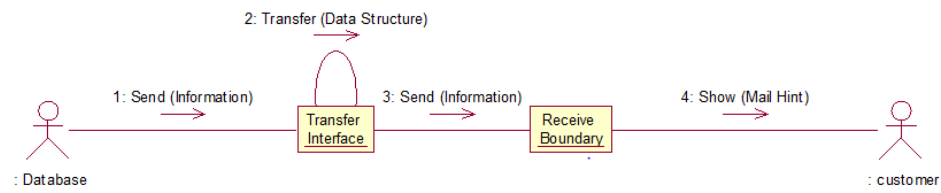
sendMail(success):



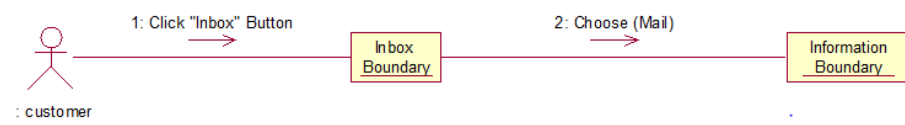
sendMail(fail):



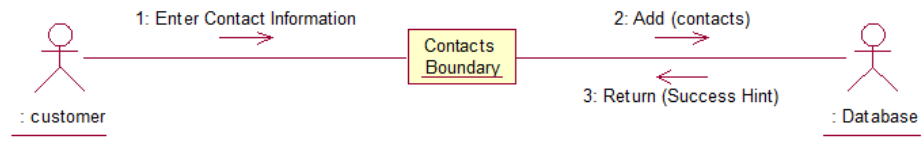
receiveMail:



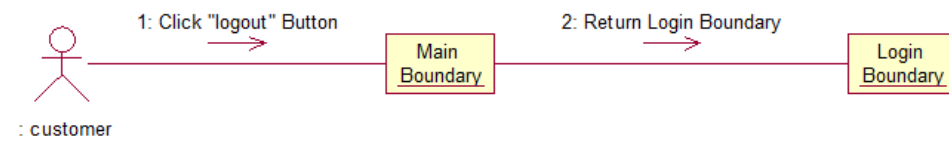
checkMail:



addList:

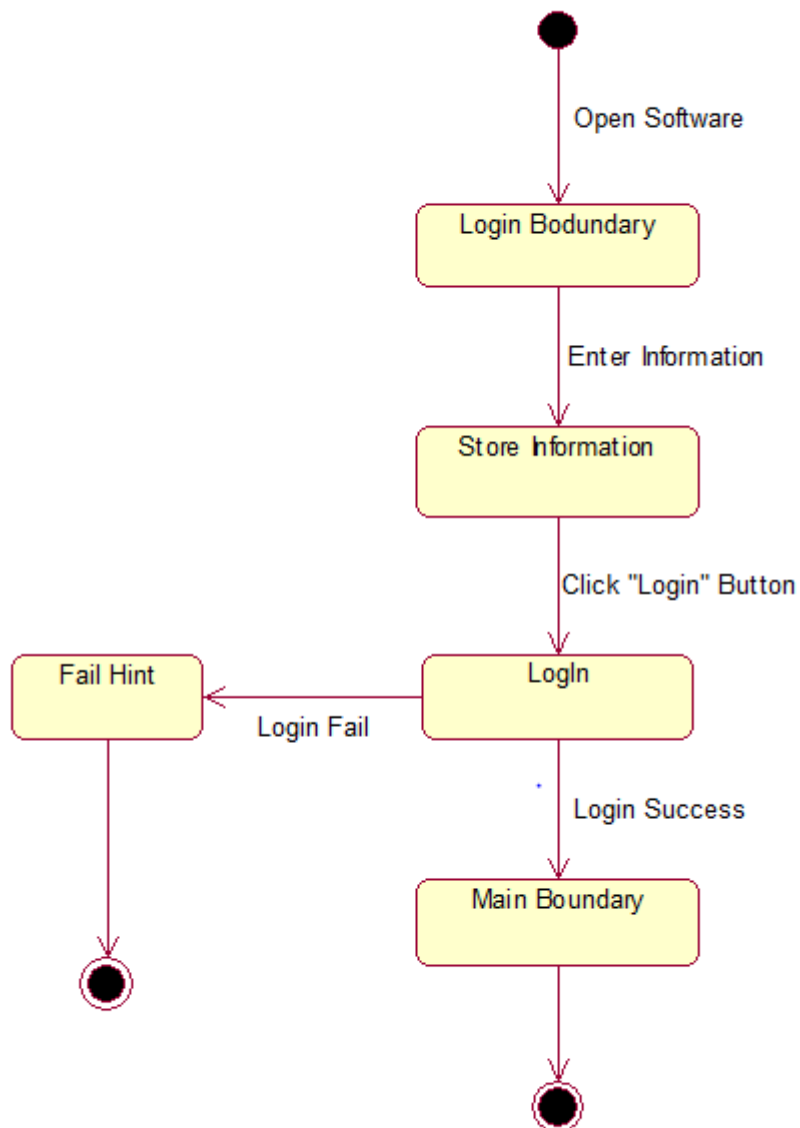


logout:

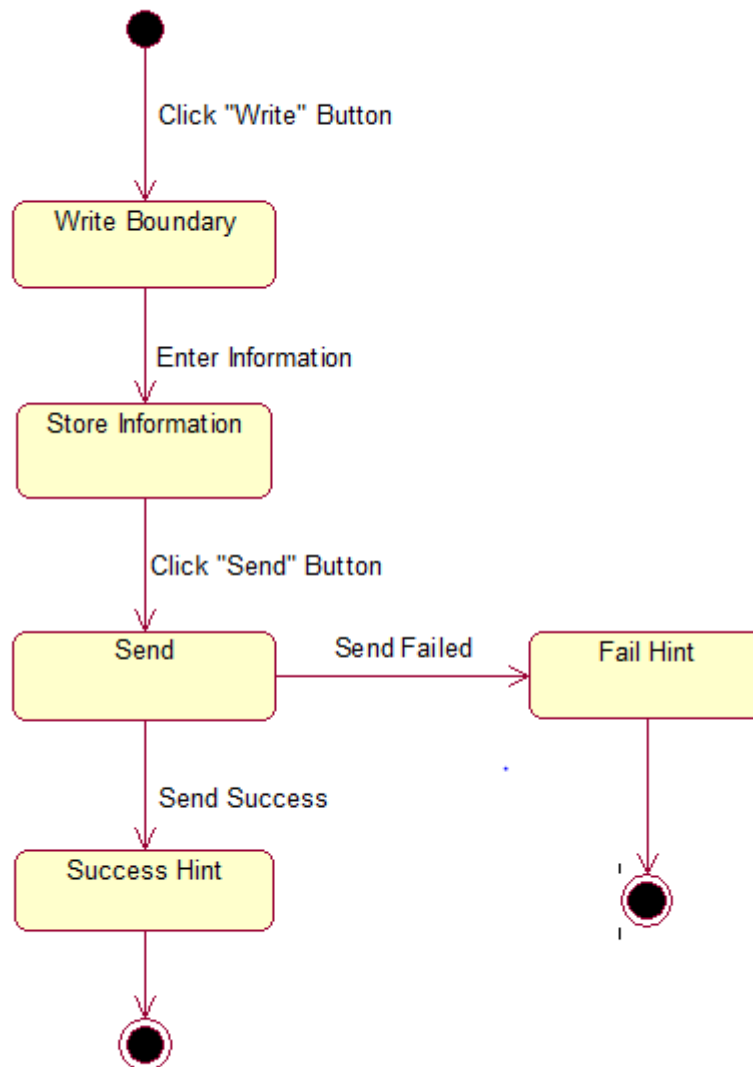


5.4 状态图：

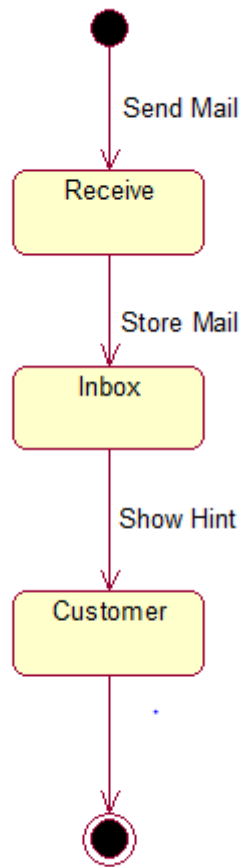
LogIn:



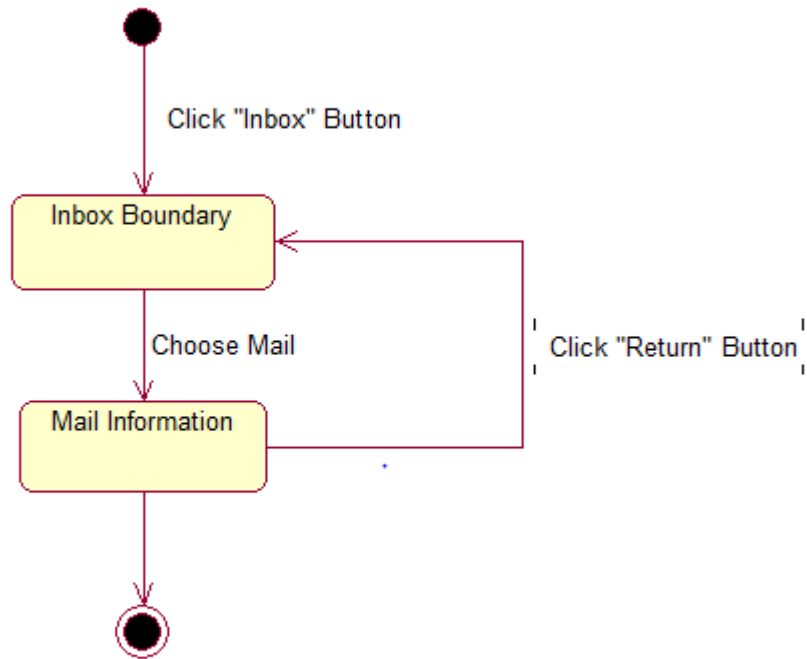
SendMail:



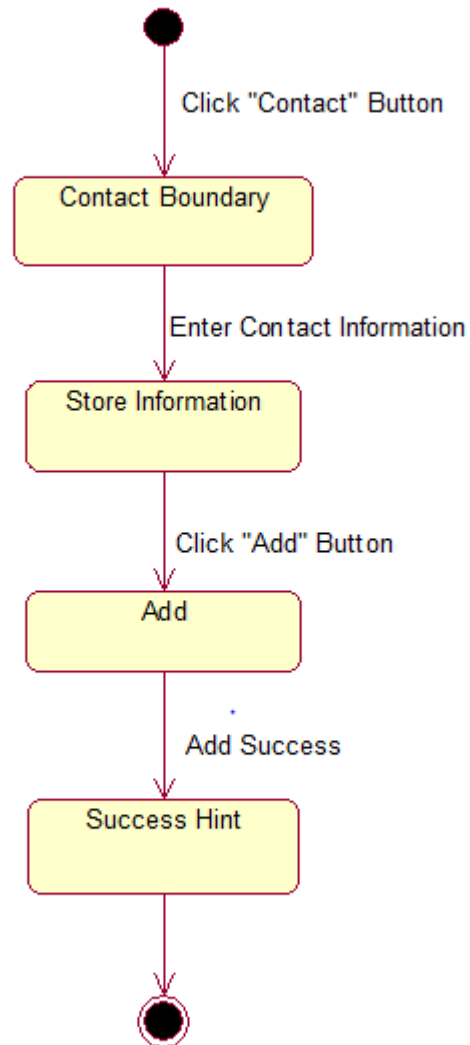
ReceiveMail:



CheckMail:



AddList:



LogOut:

