

# News All-in-one

## 新闻聚合网站

**Date : 2017-4-25**

**B/S 体系软件设计报告-V1.0**

**张昌琳作品**

专业：软件工程 学号：3140103541 联系方式：17816862765

指导老师：胡晓军老师

# 目录

1.引言 .....	4
1.1 编写目的 .....	4
1.2 背景 .....	4
1.3 项目概述 .....	4
2.总体设计 .....	5
2.1 需求规定 .....	5
2.1.1 系统功能 .....	5
2.1.2 系统性能 .....	5
2.1.3 输入输出要求 .....	5
2.1.4 数据管理能力要求 .....	6
2.2 技术选型和处理流程 .....	6
2.3 内部接口 .....	6
3 模块设计说明 .....	7
3.1 信息抓取模块设计说明 .....	7
3.1.1 模块描述 .....	7
3.1.2 功能 .....	7
3.1.3 性能 .....	8
3.1.4 输入项 .....	8
3.1.5.输出项 .....	8
3.1.6 接口： .....	8
3.2 聚合展示模块设计说明 .....	9
3.2.1 模块描述 .....	9
3.2.2 功能 .....	9
3.2.3 性能 .....	9
3.2.4 输入项 .....	9
3.2.5.输出项 .....	10
3.2.6 接口： .....	10

3.3 账户管理模块设计说明 .....	10
3.3.1 模块描述 .....	10
3.3.2 功能 .....	11
3.3.3 性能 .....	11
3.3.4 输入项 .....	11
3.3.5.输出项 .....	12
3.3.6 接口: .....	12
3.4 推荐调整模块设计说明 .....	12
3.4.1 模块描述 .....	12
3.4.2 功能 .....	13
3.4.3 性能 .....	13
3.4.4 输入项 .....	13
3.4.5.输出项 .....	13
3.4.6 接口: .....	14
4.系统数据结构设计 .....	14
4.1 账户 User.....	14
4.2 新闻信息 News .....	14
6.运行设计 .....	15
6.1 运行模块组合 .....	15
6.2 运行时间 .....	15
7.系统出错处理设计 .....	15
7.1 出错信息 .....	15
7.2 补救措施 .....	17
a. 后备技术.....	17
b. 降效技术.....	17
c. 恢复及再启动技术.....	17
7.3 系统维护设计 .....	17

# 1.引言

## 1.1 编写目的

本设计说明书的编写目的，是从总体设计的角度，明确 News All-in-one 的总体架构，流程，数据结构，数据库设计。

目的在于：

- 为编码人员提供依据。
- 为修改维护提供条件。
- 明确各模块外部接口，内部接口，用户接口

## 1.2 背景

- 软件系统名称  
News All-in-one
- 用户  
系统管理员、注册用户、游客
- 实现该软件的计算机网络  
由若干台 PC 机组成的局域网
- 该软件系统同其他系统或其他机构的基本的相互来往关系  
面向的应该是互联网全体用户。

## 1.3 项目概述

New All-in-one 将实现以下功能模块：

- 信息抓取模块  
实现定时到多个主流新闻门户网站抓取即时信息。
- 聚合展示模块  
将抓取到的信息整理在自身系统的类目下，对用户展示。
- 账户管理模块

实现个人账号管理相关逻辑。开发个人账号管理界面，包括用户信息注册修改，用户登录。

➤ 推荐调整模块

1、用户可以设置感兴趣的类目，对其显示会据此有相应的调整。

2、根据收集用户使用情况进行自身学习，调整展示内容。

另：在移动端展示的方式还包括要能实现推送。

此次 B/S 系统设计和实现出于实验性目的，需要能够充分利用现有资源高效地开发出符合需求的系统。在设计上更进一步还需要刺激用户的参与度，比如转发。在功能上争做“新闻界的网易云”，以及“看新闻一个就够，News All-in-one”。除了功能模块以外，性能和安全也需要关注。因为新闻的特殊性，要避免通过恶意攻击得以借由本系统发布信息，以及要保证良好的并发访问量以避免出现新闻爆点时系统崩溃。

## 2. 总体设计

### 2.1 需求规定

#### 2.1.1 系统功能

信息抓取、聚合展示、账户管理、推荐调整等功能。

#### 2.1.2 系统性能

本系统的性能将由服务器端数据库，网络数据传输延时，以及并发访问该系统的用户数量决定。客户端方面将实现 AJAX 技术，减轻服务器压力，增加带宽利用率，并为用户提供良好的界面体验。

#### 2.1.3 输入输出要求

客户端通过网页展现给用户一个友好的界面，用户可以通过提交表单或者点击超链接向服务器提供数据与命令。

### 2.1.4 数据管理能力要求

**安全：**服务器将予以数据库最高等级的保护，防止黑客从后台下载数据库，防止通过网页 SQL 注入等方式从数据库中获取信息或者破坏数据库。

**性能：**对于频繁访问数据库的操作，后台需要建立持久的数据库连接，以避免重复连接数据库耗费资源。

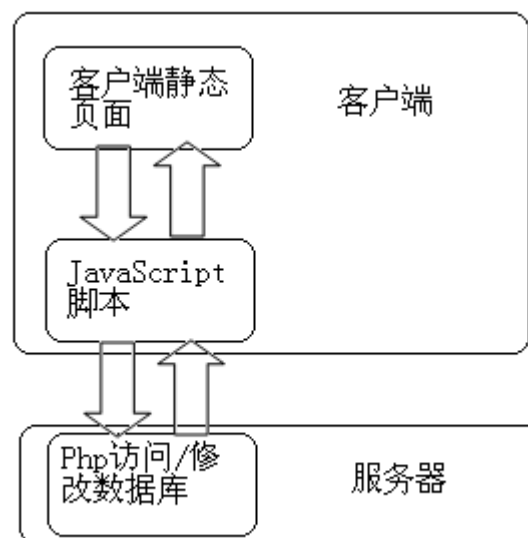
## 2.2 技术选型和处理流程

本系统分为客户端和服务端：

**服务器：**以 nginx 为 web 服务器，服务器端 python 脚本语言编写。数据库使用 mongodb。

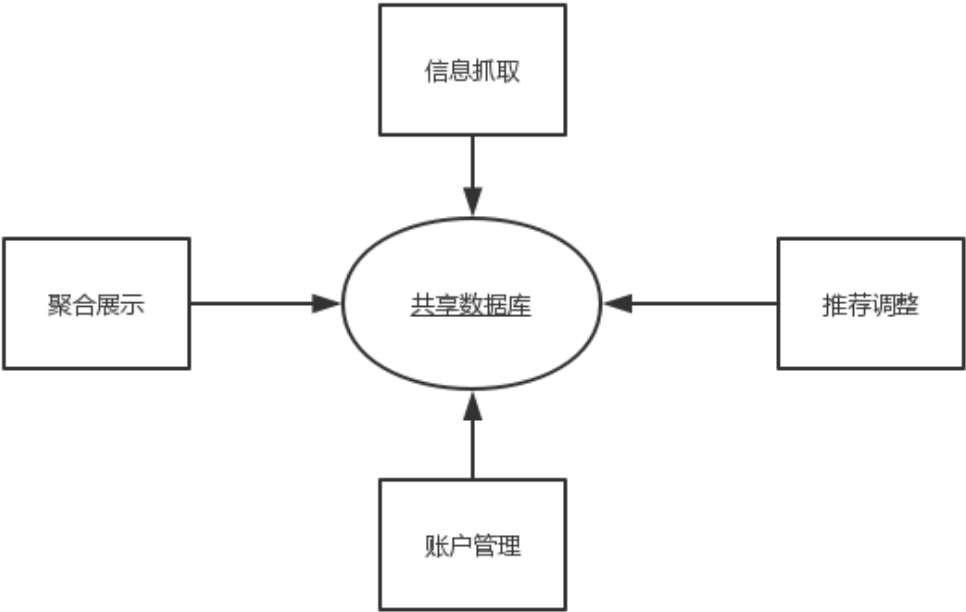
**客户端：**使用 JavaScript MVC 框架 React。

处理流程图如下：



## 2.3 内部接口

本系统内不同模块通过共享数据库完成通信。



News All-in-one 各个模块共享数据库图

### 3 模块设计说明

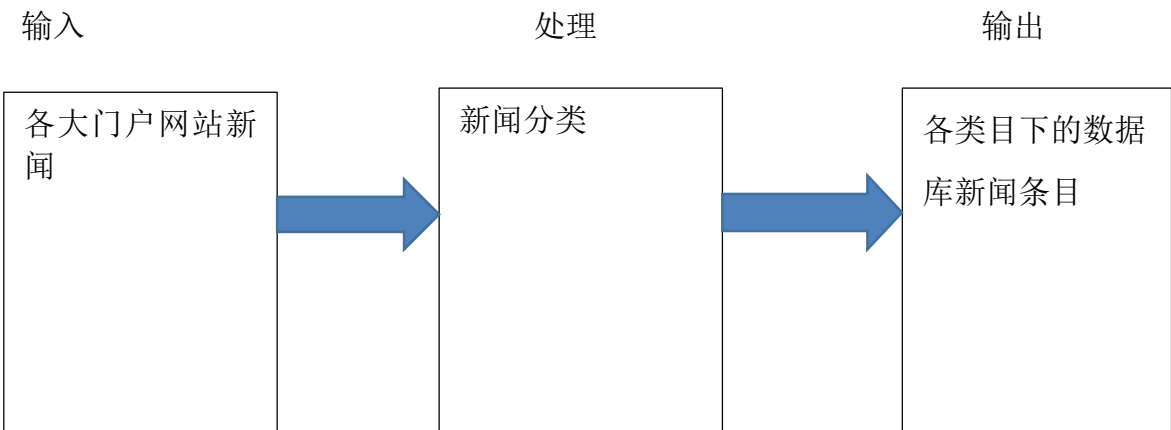
#### 3.1 信息抓取模块设计说明

##### 3.1.1 模块描述

此模块在特定时间点（比如每天的偶数整点）抓取各大门户网站新闻，将内容分门别类后存入数据库。（是一个和用户不直接交互的模块）

##### 3.1.2 功能

IPO 图如下：



信息抓取模块 IPO 图

3.1.3 性能

对于收集到的新闻要能够快速去重。

3.1.4 输入项

登录模块输入项表

名称	标识	类型和格式	输入方式
新闻 ID	ID	string	抓取
新闻标题	Title	string	抓取
新闻内容	Content	string	抓取
新闻来源	Source	string	输入源
新闻分类	Type	string	抓取

3.1.5.输出项

登录模块输出项表

名称	标识	类型和格式	输出方式
新闻 ID	ID	string	数据库条目
新闻标题	Title	string	数据库条目
新闻内容	Content	string	数据库条目
新闻来源	Source	string	数据库条目
新闻分类	Type	string	数据库条目

3.1.6 接口：

a) 内部接口：

前端通过不同种类元素种类与后端衔接，从数据库导出数据到网页页面。



后端通过 py 框架输出数据库信息。

b) 用户接口:

无。

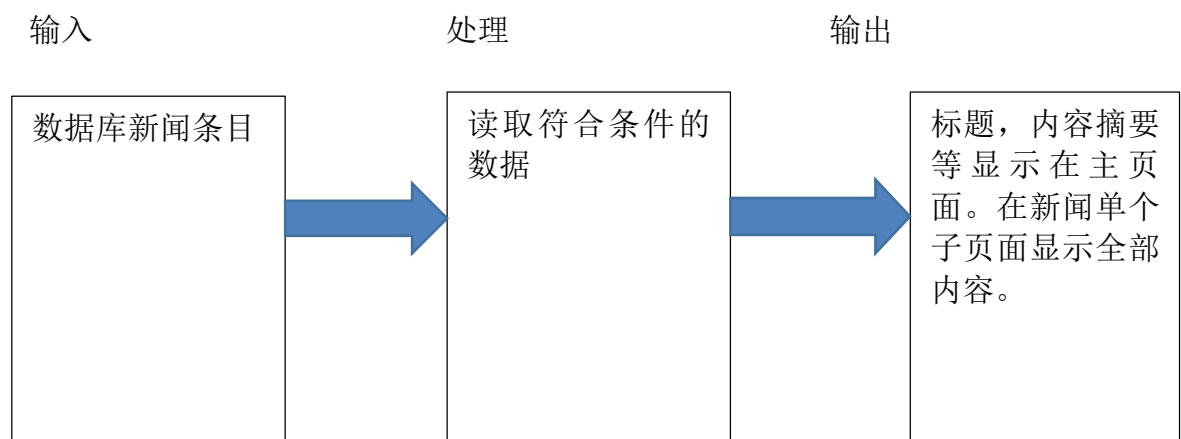
### 3.2 聚合展示模块设计说明

### 3.2.1 模块描述

此模块包括了政治、经济、社会、娱乐等类目，可以从导航栏分别进入，注册用户和游客均可以点击查看新闻内容。

### 3.2.2 功能

IP0 图如下:



聚合展示模块 IPO 图

### 3.2.3 性能

展示页面要能够吸引用户。

### 3.2.4 输入项

### 聚合展示模块输入项表

名称	标识	类型和格式	输入方式
----	----	-------	------

新闻 ID	ID	string	抓取
新闻标题	Title	string	抓取
新闻内容	Content	string	抓取
新闻来源	Source	string	输入源
新闻分类	Type	string	抓取

### 3.2.5.输出项

聚合展示模块输出项表

名称	标识	类型和格式	输出方式
新闻 ID	ID	string	数据库条目
新闻标题	Title	string	数据库条目
新闻内容	Content	string	数据库条目
新闻来源	Source	string	数据库条目
新闻分类	Type	string	数据库条目

### 3.2.6 接口：

a) 内部接口：

后端通过在数据库查找对应的新闻信息显示到前端页面上。

b) 用户接口：

点击导航栏即可跳转到对应的信息页面，点击新闻标题即可跳转到新闻子页面。

## 3.3 账户管理模块设计说明

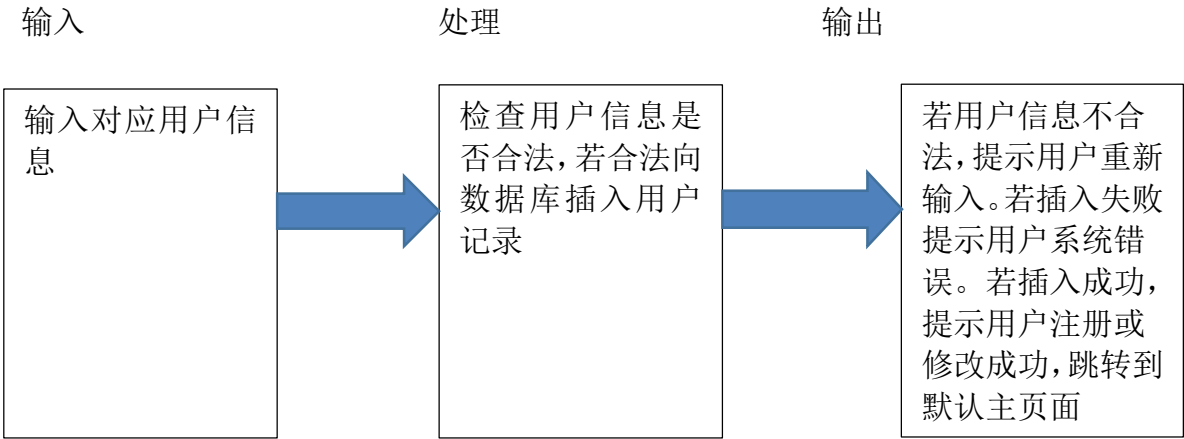
### 3.3.1 模块描述

此模块是给用户注册、登录、注册信息修改等操作的。

3.3.2 功能

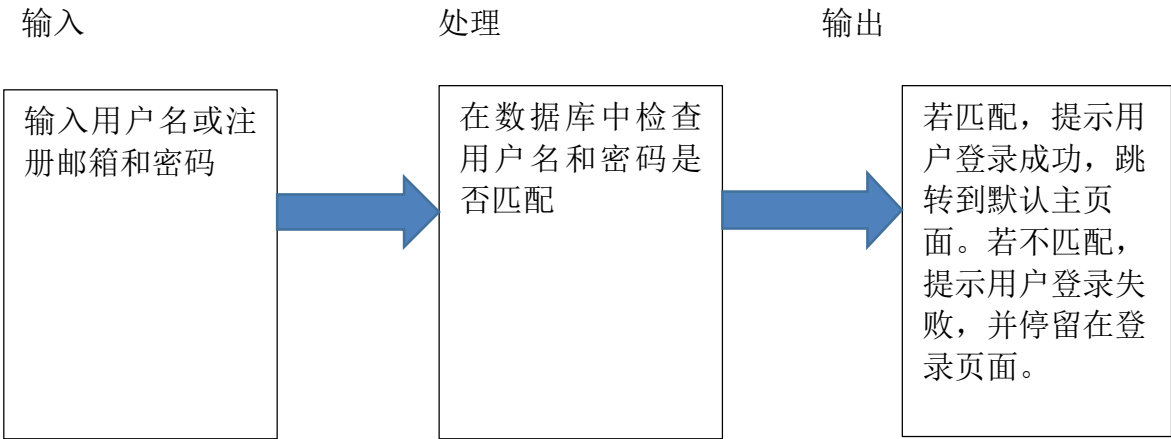
IPO 图如下：

注册或注册信息修改：



账户管理模块注册 IPO 图

登录：



账户管理模块登录 IPO 图

3.3.3 性能

用户手机端和网页端可以同时登录，修改信息以后提交覆盖前提交。

3.3.4 输入项

账户管理模块输入项表

名称	标识	类型和格式	输入方式
----	----	-------	------

注册邮箱（主键）	emailID	string	用户输入
用户名	displayname	string	用户输入
兴趣类目	favorite	String array	用户选择
用户行为记录	behavior	string	用户行为

### 3.3.5.输出项

账户管理模块输出项表

名称	标识	类型和格式	输出方式
操作结果	opeState	String	提示框
跳转结果	pageState	html	网页

### 3.3.6 接口：

a) 内部接口：

前端注册登录按钮都会触对话框，提交后，修改后端数据库的信息。

b) 用户接口：

前端的按钮以及按钮里的信息输入，都是用户的直观接口。

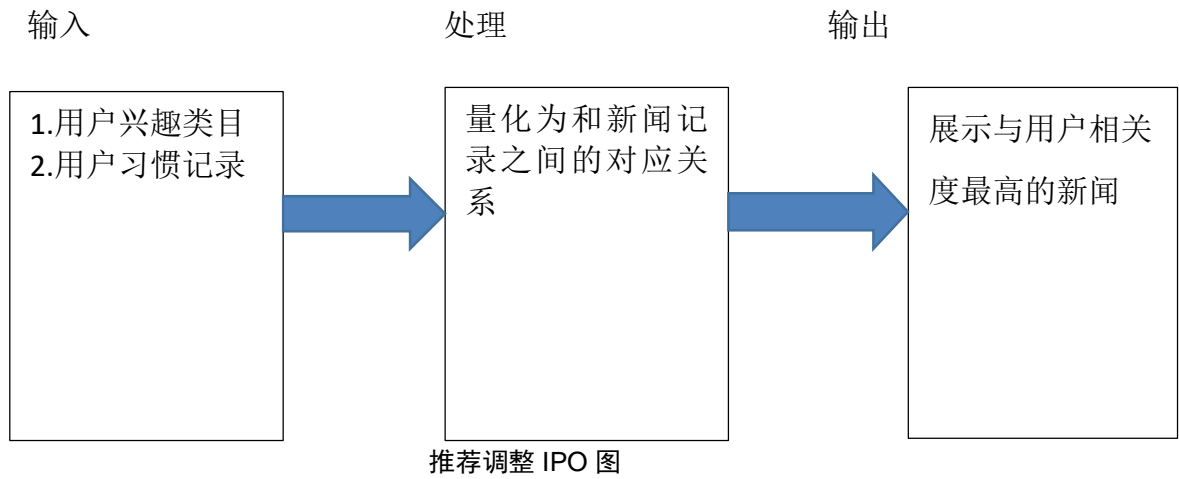
## 3.4 推荐调整模块设计说明

### 3.4.1 模块描述

此模块是改善用户体验、个性化调整展示模块，也是这个作品质量和用户吸引力的成败之处。推荐调整收到两方面的影响。一方面，用户可以设置感兴趣的类目，对其显示会据此有相应的调整。另一方面，根据收集用户使用情况进行自身学习，调整展示内容。另外，因为客户端平台特点不同，在移动端展示的方式还包括要能实现推送。

### 3.4.2 功能

IPO 图如下：



### 3.4.3 性能

需要快速计算相关度以及进行倒排搜索查询。

### 3.4.4 输入项

推荐调整模块输入项表

名称	标识	类型和格式	输入方式
兴趣类目	favorite	String array	用户选择
用户行为记录	behavior	string	用户行为

### 3.4.5.输出项

推荐调整模块输出项表

名称	标识	类型和格式	输出方式
新闻 ID	ID	string	数据库条目
新闻标题	Title	string	数据库条目
新闻内容	Content	string	数据库条目

新闻来源	Source	string	数据库条目
新闻分类	Type	string	数据库条目

### 3.4.6 接口：

#### a) 内部接口：

通过当前用户用户名修改和读取后端数据库用户兴趣类目和用户行为记录，填入前端页面。

#### b) 用户接口：

无。

## 4.系统数据结构设计

### 4.1 账户 User

账户表物理结构设计

数据名称	数据类型	Nullable	IS_PK
emailID	Varchar(30)	No	Yes
displayname	Varchar(10)	Yes	No
favorite	int	Yes	No
behavior	Varchar(50)	Yes	No

### 4.2 新闻信息 News

新闻信息表物理结构设计

数据名称	数据类型	Nullable	IS_PK
ID	Varchar(10)	No	Yes
Title	Varchar(20)	No	No
Content	Text	No	No
Source	Varchar(10)	No	No

Type	Int	Yes	No
------	-----	-----	----

## 6.运行设计

### 6.1 运行模块组合

本系统按功能划分模块，每个模块又按流程划分为客户端界面，客户端脚本，服务器后台程序。功能模块之间相互不会共享界面（主界面除外），后台程序只共享建立数据库连接的方法。

### 6.2 运行时间

由于采用 Ajax 技术，实现客户端静态更新信息，静态分页等等，因此，服务器的带宽利用率大大提高，唯一会影响服务器的运行时间的，应该只是隔一段时间需要建立销毁数据库连接所带来的时间开销（可以放在访问量低峰时期做）。

## 7.系统出错处理设计

### 7.1 出错信息

表格 7-1：系统出错信息

系统输出信息的形式	含义	处理办法
数据库连接不上	由于并发操作的用户数量很大，导致数据库访问读写效率降低。或者数据库配置不对，导致数据库连接失败。	修改数据库配置，并限制同一个用户两次访问数据库之间的时间间断限制。

Sql 语句执行错误	由于一些用户恶意在表单中插入 sql 语句，企图破坏数据库，导致访问数据库出错。	对表单中的数据在服务器端进行过滤，检测。
网站信息丢失	由于一些用户恶意不通过登陆界面而直接访问网站或者后台，导致网站信息破坏或被窃取。	禁止直接访问一些需要登陆的网页。
服务器崩溃	服务器由于长时间连续运作，导致错误积累过多，负荷过大，最终崩溃	定期对服务器进行维护
账户被盗	由于客户端被黑客植入木马或者其他方式引起用户账户被盗，密码被修改。	建议用户经常对电脑进行杀毒，不要在陌生的地点用其他人的电脑登陆客户端。数据库中账户信息泄露黑客通过后台下载服务器中的数据库，盗取所有用户的账户信息。对账户中密码进行加密，如 3 重 DES 加密算法，MD5 加密算法等等。
磁盘损坏	由于需要频繁地操作数据库，导致磁盘损坏，数据丢失。	使用磁带对磁盘进行周期性地备份。
数据库读取乱码	在操作数据库过程中，由于写入出错或者读出出错导致出现乱码	一般出现乱码的错误为数据库编码与写入数据编码不一致缘故，此时需要统一编码图表-系统出错一览表



## 7.2 补救措施

### a. 后备技术

周期，频繁备份数据库，并且将其储存在更加稳定的介质上，比如磁带。

将系统所需的不同数据库部署到不同的计算机上，减小因硬件问题而导致数据全部丢失的可能性。

### b. 降效技术

紧急情况下，启用人工干预手段。

### c. 恢复及再启动技术

系统崩溃后，通过系统运行日志记录恢复数据。

## 7.3 系统维护设计

连接数据库方面，需要在创建数据库连接，销毁数据库连接，执行 sql 语句的模块使用 try catch 语句捕获异常。于是发生此类错误便能很快得知。

网络方面需要检测是否有特定的 IP 地址频繁访问系统，不能排除其攻击服务器的可能性。

中央管理系统中，需要检测撮合进程是否进入死锁，一些未被处理的指令永久处于饥饿状态。

内部人员如（开户人员，系统维护人员）操作留下操作痕迹，使用权管理层可以定期或不定期地稽核系统。