目录

[1 引言 1](#_Toc24607)

[1.1 编写目的 2](#_Toc2595)

[1.2 背景 2](#_Toc2875)

[1.3 定义 2](#_Toc19510)

[1.4 参考文献 4](#_Toc14192)

[1.5 版本更新信息 4](#_Toc14963)

[2 总体设计 4](#_Toc13765)

[2.1 需求规定 4](#_Toc32622)

[2.2 运行环境 5](#_Toc21889)

[2.2.1设备 5](#_Toc18511)

[2.2.2支持软件 5](#_Toc9794)

[2.2.3接口 5](#_Toc26884)

[2.2.4控制 6](#_Toc19419)

[2.3 基本设计概念和处理流程 6](#_Toc7982)

[2.3.1 体系结构设计 6](#_Toc31264)

[2.3.2 基本处理流程 7](#_Toc20795)

[2.3.3 整体结构 9](#_Toc8196)

[2.5 尚未问决的问题 10](#_Toc21389)

[3 运行设计 10](#_Toc17590)

[3.1 运行模块组合 10](#_Toc19360)

[3.2 运行控制 11](#_Toc28323)

[3.3 运行时间 11](#_Toc18823)

[4系统数据结构设计 12](#_Toc27525)

[4.1 物理结构设计要点 12](#_Toc12863)

[4.3 数据库概念结构设计 13](#_Toc18916)

[4.3.1整体 E-R 图 13](#_Toc10662)

[5 系统出错处理设计 14](#_Toc6077)

[5.1 出错信息 14](#_Toc24854)

[5.2 补救措施 14](#_Toc22302)

[5.3 系统维护设计 14](#_Toc6137)

[5.4 安全保密设计 15](#_Toc24835)

# 1 引言

随着计算机技术的不断应用和提高，计算机已经深入到社会生活的各个角落。网络信息高度膨胀，导致众多的追星女孩，追星的时候抓不注重点，不能本着对偶像的憧憬和向往，端正入“饭圈”动机、提高自身思想素质、努力为入“饭圈”做好思想准备、争取早日加入“饭圈”的目的，通过对“饭圈”的性质和宗旨、“饭圈文化”的指导思想、“饭圈文化”的最终目标和现阶段的任务、“饭圈文化”组织制度、“饭圈文化”的纪律和“饭圈文化”的章程等。

本系统可以使用户认真学习“饭圈文化”，使其在“饭圈”思想上得到了升华、对“饭圈文化”的先进性理念有深刻的认识和体会、更加清楚的了解了“饭圈文化”的本质、坚定自愿为偶像进行奋斗终身的信念。

## 编写目的

追星女孩同学是个流量明星粉，特别喜欢邓超和anglebaby。她想爬取2人的微博，以txt文件或者csv文件形式保存。还要分析微博中出现的那些人，去了哪些地方，每句微博的情感，形成词云，柱状图等献给邓超和baby作为礼物。

## 背景

本项目名称：微博可视化分析系统

项目开发人员：魏振东

本产品是针对追星女孩的需求进行设计。

## 定义

在我们进行正式软件设计之前，要先了解一些概念：

（1） 开发（ develop ）：不是单纯指开发活动，还包括维护活动。

（2）项目 (project) ：该词指的是向顾客或最终用户交付一个或多个产品的收管理的相关资源的集合。 这个资源集合有着明确的始点和终点， 并且一般是按照某项计划运行。 这种计划通常会形成文件， 并且说明要交付或实现的产品、 所用的资源和经费、 要做的工作和工作进度，一个项目可能有若干项目组成。

（3）项目开发计划（ project development plan）：是一种把项目一定一过程与项目如何推进连接起来的方案。

（4）产品生命周期（ product life cycle ）：是产品从构思到不可以再使用的持续时间。

（5）模块化：模块是数据说 明、可执行语句等程序对象的集合， 模块可以单独被命名而且可通过名字来访问。

例如，过程、函数、子程序、宏等都可以作为模块；

（6） 抽象与逐步求精：软件工程过程的每一步都是对软件解法的的抽象层次的一次精化。逐步求精与抽象是紧密相关的。

（7） 模块独立性： 模块独立性是指系统软件中每个模块只涉及软的具体自功能， 而和软件系统的其他的模块接口是简单的。

（8）耦合：是对一个软件结构内各个模块之间互连程度的度量。耦合强弱取决与模块间接口的复杂程度，调用模块的方式，以及通过接口的信息。

## 1.4 参考文献

[1] 张海藩 .软件工程导论（第五版） [M]. 北京：清华大学出版社， 2008

[2] 刘乃丽 . 精通 ASP.NET2.0+SQLServer 2005 项目开发 .[ M]. 北京：人民邮电出版社 , 2007 ：100～150.

[3] ( 美) 沃尔特，本 - 甘，萨卡 . Microsoft SQL Server 2005技术内幕 -T-SQL 程序设计[ M]. 北京：电子工业出版社， 2007:50 ～80.

## 1.5 版本更新信息

本文档的更新记录如表1－１所示。

表1-1 版本更新记录

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **修改编号** | **修改日期** | **修改后版本** | **修改位置** | **修改内容概述** |
|  |  |  |  |  |
|  |  |  |  |  |

# 2 总体设计

## 2.1 需求规定

追星女孩同学是个流量明星粉，特别喜欢邓超和anglebaby。她想爬取2人的微博，以txt文件或者csv文件形式保存。还要分析微博中出现的那些人，去了哪些地方，每句微博的情感，形成词云，柱状图等。

## 2.2 运行环境

### 2.2.1设备

操作系统：Windows 7及以上

CPU：Intel i7酷睿双核及以上

内存：16G及以上

硬盘空间：100G以上

软件配置如下：

开发工具：MyEclipse，pycharm

数据库：Mysql等

Web服务器：Tomcat

### 2.2.2支持软件

先进可靠安全性高，可扩展性且性价比高，支持 Java Web/J2EE规范。

操作系统： Microsoft Windows 2000 Advanced Server 支持环境： IIS 5.0 ;

数据库系统： Mysql;

### 2.2.3接口

硬件接口：本软件不需要特定的硬件或硬件接口进行支撑， 486 以上 PC 机均可运行此软件，对机器要求不太严格． 因为可能涉及数据的备份应该保持打印机和光盘刻录机的接口。

软件接口： 运行于 Windows7 及更高版本具有 WIN64 API 的操作系统之上。 因为可能

涉及一些文档、报表的处理应该保持与常用软件的办公软件的接口。

局域网内部接口： 为微博可视化分析系统交换信息， 为相关部门或主管提供参考数据和决策支持数据，采取中间数据库或表的方式并遵循相应的数据交换协议。

外部系统的接口：与外界网站建立信息交换，与国家管理机构相关系统的数据交换，遵 循 TCP/IP 网络传输与 RPC 远程调用数据通信协议。

### 2.2.4控制

本系统初步决定采用 B/S 架构，用户通过浏览器访问微博可视化分析系统， 在权限范围内对其所属信息和附件可增删改。 管理员可以通过浏览器方式管理和维护微博可视化分析系统， 或者远程控制软件对后台系统进行管理和维护

## 2.3 基本设计概念和处理流程

### 2.3.1 体系结构设计

C/S体系结构的数据库应用，由两部分组成：客户应用程序和数据库服务器 程序。二者可分别称为前台程序与后台程序。 运行数据库服务器程序的机器， 称

为应用服务器，一旦服务器程序被启动，就随时等待响应客户程序发来的请求； 客户程序运行在用户自己的电脑上， 对应于服务器电脑， 可称为客户电脑。 当需 要对数据库中的数据进行任何操作时， 客户程序就自动地寻找服务器程序， 并向其发出请求， 服务器程序根据预定的规则做出应答， 送回结果。 从原理和经验上看，对本案例或类似的应用， C/S结构是目前技术条件下，能较好适应不确定和变化的需求环境的比较现实的方案。 它可以以较低的投入， 实现将易变与稳定的要素分离，快速地增添和替换 “瘦小”而互相独立的前台应用，保持数据的连续性和继承性。

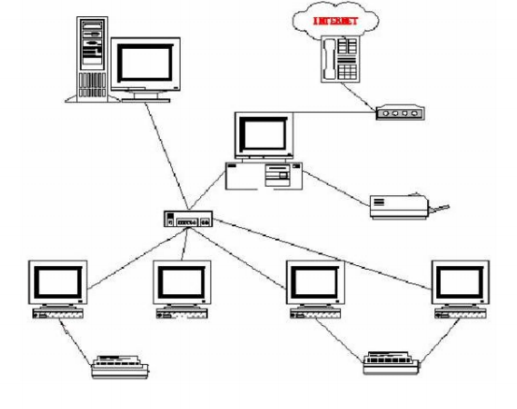


图2-1 体系结构

### 2.3.2 基本处理流程

（1）登录：

用户模拟打开浏览器，然后找到输入框和密码框，输入密码和账号，然后点击登录。如图2-2所示

1. 可视化：

用户爬取数据，然后进行存储，用户对存储之后的数据进行清洗，然后分析可视化。如图2-2所示

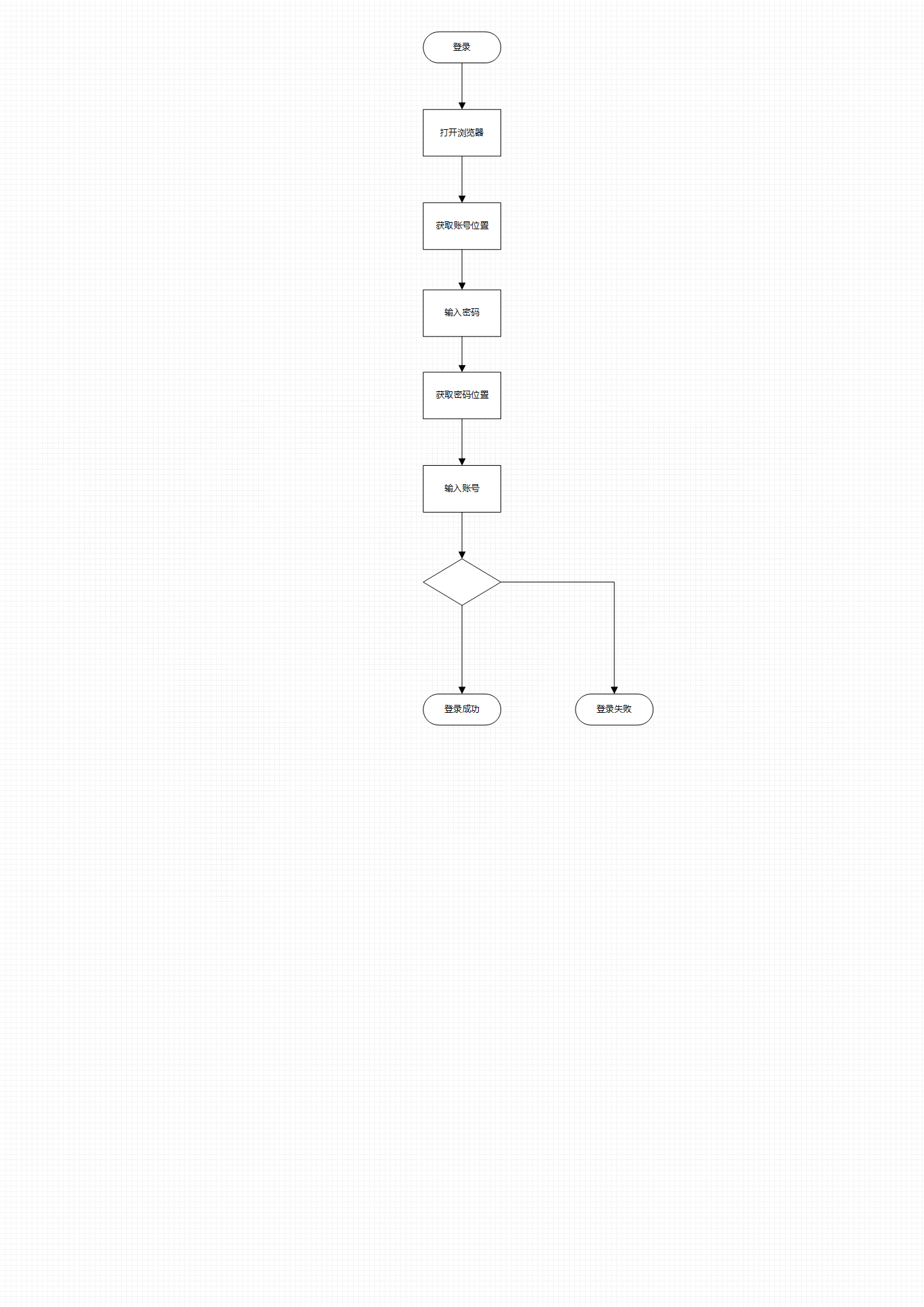
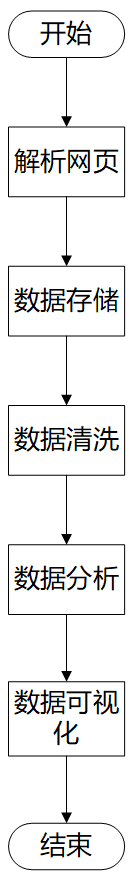


图2-2程序流程图

### 2.3.3 整体结构

根据微博可视化分析系统的实际需求，可以将微博可视化分析系统划分为登录，数据爬取，数据存储，数据清洗，数据分析，数据可视化5个部分各个部分的具体功能的系统功能结构图如图 3-1 所示

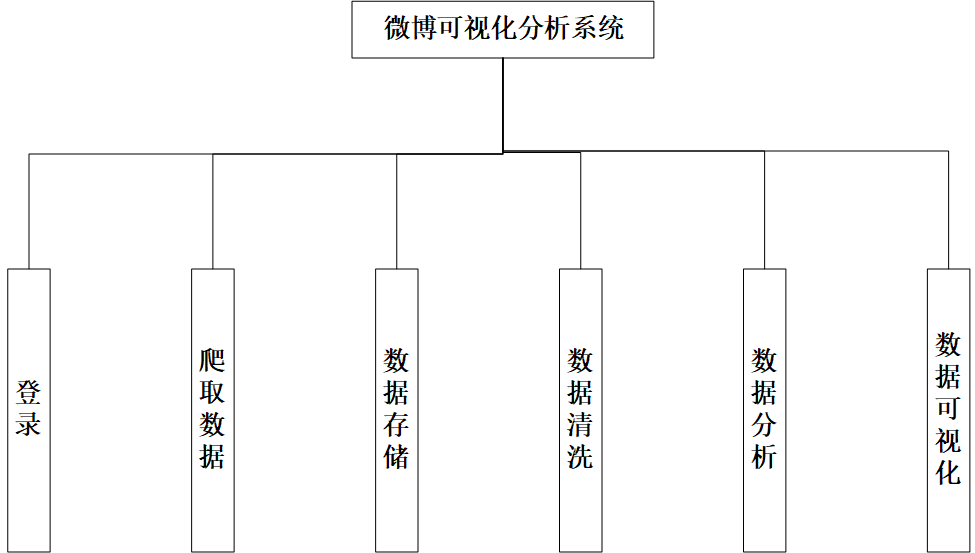


图 2-3 功能模块图

## 2.5 尚未问决的问题

（1）只对系统的功能进行了描述， 没有阐述各个模块之间是如何协调合作；

（2）没有描述模块的实现方法，以及在这个过程具体的流程。

# 3 运行设计

## 3.1 运行模块组合

客户机程序在有输入时启动接收数据模块， 通过各模块之间的调用， 读入并对输入进行格式化。在接收数据模块得到充分的数据时，将调用网络传输模块， 将数据通过网络送到服务器， 并等待接收服务器返回的信息。 接收到返回信息后随即调用数据输出模块，对信息进行处理，产生相应的输出。

服务器程序的接收网络数据模块必须始终处于活动状态。接收到数据后， 调用数据处理查询模块对数据库进行访问，完成后调用网络发送模块，将信息返回客户机

施加不同的外界运行控制时所引起的各种不同的运行模块组合如表4-1所示：

## 3.2 运行控制

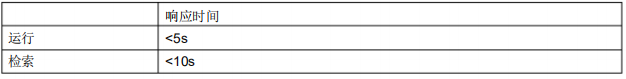
运行控制将严格按照各模块间函数调用关系来实现。在各事务中心模块中，需对运行控制进行正确的判断，选择正确的运行控制路径。

在网络传输方面， 客户机在发送数据后， 将等待服务器的确认收到信号， 收到后，再次等待服务器发送回答数据， 然后对数据进行确认。 服务器在接到数据后发送确认信号，在对数据处理、访问数据库后，将返回信息送回客户机，并等待确认。

## 3.3 运行时间

在软体的需求分析中，对运行时间的要求为必须对做出的操作有较快的反应。网络硬件对运行时间有最大的影响， 当网络负载量大时， 对操作反应将受到 很大的影响。所以将采用高速 ATM 网络，实现客户机与服务器之间的连接，以 减少网络传输上的开销。 其次是服务器的性能， 这将影响对数据库访问时间即操作时间的长短，影响加大客户机操作的等待时间， 所以必须使用高性能的服务器，建议使用 Pentium III 处理器。硬件对本系统的速度影响将会大于软件的影响。

表3-1运行时间



# 4系统数据结构设计

5.

## 4.1 物理结构设计要点

表4.1 文章信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段 | 类型 | 允许空 | 备注 | 是否为主键 |
| id | int | NOT NULL | id | 是 |
| Name | varchar | NOT NULL | 文章名称 | 否 |
| Content | longtext | NOT NULL | 内容 | 否 |

表4.2 停用词表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段 | 类型 | 允许空 | 备注 | 是否为主键 |
| id | int | NOT NULL | id | 是 |
| name | varchar | NOT NULL | 停用词 | 否 |
| a\_id\_id | int | NOT NULL | 与文章表关联 | 否 |

表4.2 替换词表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段 | 类型 | 允许空 | 备注 | 是否为主键 |
| id | int | NOT NULL | id | 是 |
| oldname | varchar | NOT NULL | 旧词 | 否 |
| newname | varchar | NOT NULL | 新词 | 否 |
| a\_id\_id | int | NOT NULL | 与文章表关联 | 否 |

## 4.3 数据库概念结构设计

### 4.3.1整体 E-R 图

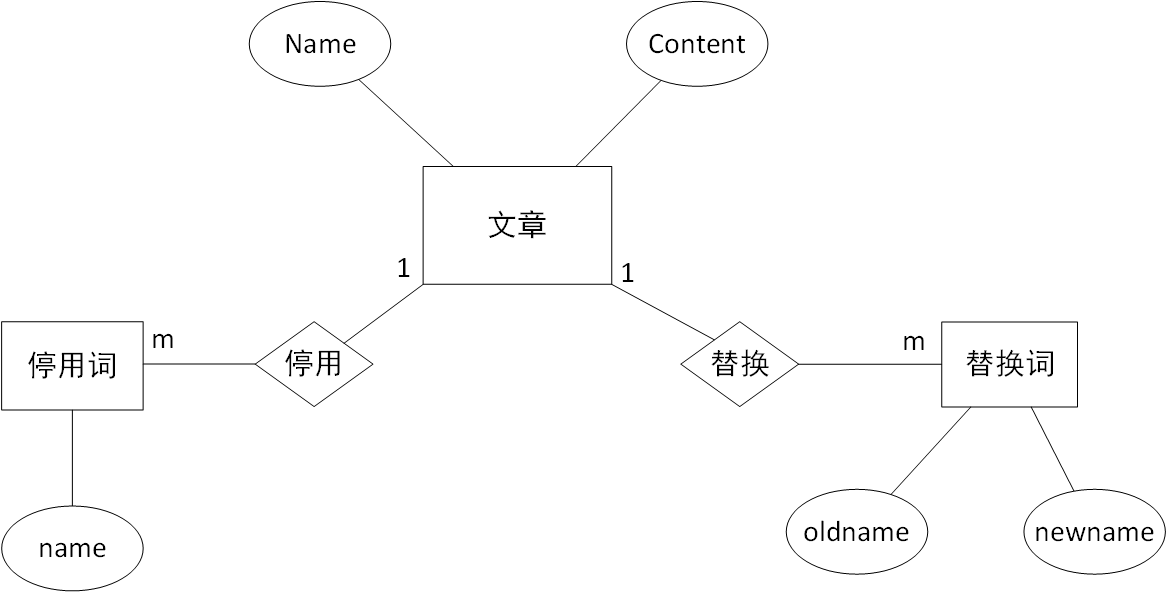


图4-1 e-r图

# 5 系统出错处理设计

## 5.1 出错信息

（1）验证用户的身份若读者的密码、借阅证号有误则发出“验证错误”信息；

## 5.2 补救措施

（1）后备技术说明准备采用的后备技术，当原始系统数据万一丢失时启用的副本的建立和启动的技术， 例如周期性地把磁盘信息记录到磁带上去就是对于磁盘媒体的一种后备技术；

(2 ) 降效技术说明准备采用的后备技术， 使用另一个效率稍低的系统或方法来求得所需结果的某些部分， 例如一个自动系统的降效技术可以是手工操作和数

据的人工记录；

(3 ) 恢复及再启动技术说明将使用的恢复再启动技术，使软件从故障点恢复执行或使软件从头开始重新运行的方法。

## 5.3 系统维护设计

及时清理数据库，提高数据库运行能力； 同时对更新的数据及时保存，以防止数据丢失。

## 5.4 安全保密设计

（1）对用户输入的所有内容都应进行安全性和合法性检查， 避免恶意输入；

（2）为了确保数据库的安全，数据库采用冗余备份、定时备份和即时备份相结合，详细记录日志文件，以备灾难恢复。

（3）系统对用户注册密码要进行安全性提示，确保密码不易丢失或失窃，同时为用户提供密码找回功能，确保安全。

（4）系统对用户的所有操作和相关信息都记录在案。