高考志愿填报建议系统

概要设计说明书

**版本 1.2**

**修订历史记录**

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 版本 | 说明 | 作者 |
| 02/11/2020 | 1.0 | 创建，首次编辑 | 王雪飞、朱宇博、夏宇祺、吴杰乐、潘雨非、李斯文 |
| 08/11/2020 | 1.1 | 添加算法、数据处理、类图、补充接口设计 | 王雪飞、朱宇博、夏宇祺、吴杰乐、潘雨非 |
| 13/11/2020 | 1.2 | 添加详细操作流程，修改详细用户接口 | 王雪飞、李斯文 |
|  |  |  |  |

**目录**

[1．引言 2](#_Toc17692)

[1.1 编写目的 2](#_Toc3643)

[1.2背景 2](#_Toc7613)

[2．总体设计 2](#_Toc20616)

[2.1需求规定 2](#_Toc28259)

[2.1.1系统功能 2](#_Toc9100)

[2.1.2系统性能 3](#_Toc923)

[2.1.3输入输出要求 4](#_Toc8637)

[2.1.4故障处理要求 4](#_Toc5525)

[2.1.5其他专门要求 5](#_Toc29636)

[2.2运行环境 5](#_Toc32455)

[2.2.1设备 5](#_Toc13562)

[2.2.2支持软件 5](#_Toc29917)

[2.2.3接口 5](#_Toc5427)

[2.3基本设计概念和处理流程 6](#_Toc15844)

[2.3.1 运行流程 6](#_Toc27046)

[2.3.2数据处理设计 7](#_Toc27666)

[2.3.3数据库设计 10](#_Toc11812)

[2.3.4推荐算法设计 12](#_Toc15521)

[2.4结构 15](#_Toc16640)

[2.4.1 模块图 15](#_Toc23180)

[2.4.2 类图 16](#_Toc14175)

[2.5功能需求与系统模块的关系 16](#_Toc3123)

[2.6人工处理过程 17](#_Toc21694)

[3．时序设计 17](#_Toc19254)

[3.1登录时序图 17](#_Toc18418)

[3.2用户功能时序图 18](#_Toc3)

[3.3管理端功能时序图 18](#_Toc8436)

[4．接口和操作 19](#_Toc9957)

[4.1客户端 19](#_Toc20017)

[4.2管理端 22](#_Toc4771)

[5．系统出错处理设计 24](#_Toc9528)

[5.1出错信息 24](#_Toc24808)

[5.2补救措施 25](#_Toc23036)

[5.3系统维护设计 26](#_Toc5895)

# 1．引言

1.1 编写目的

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。方便后续对该项目的开发，明确该系统各功能的定义、输入与输出，对系统本身性能的规定，输入输出的要求，数据管理能力的要求，故障处理要求，除了软件本身，还对其运行环境进行了相应的规定。以此文档作为后续软件开发设计的重要参考文档。该说明在与甲方确定过后，若没特殊情况，不做修改，后续开发都以此为准。本文档的预期读者包括：程序开发经理、项目组所有人员、测试组人员、SQA人员。

1.2背景

本项目的委托单位是北京理工大学，开发单位是07111705-4组，组长王雪飞。

# **2．总体设计**

## **2.1需求规定**

### **2.1.1系统功能**

总的概括来说，高考志愿填报建议系统旨在根据对于往年录取情况的分析，为广大参加高考的考生提供个性化的、正确的志愿填报建议。本系统可以根据用户提供的考试省份，高考分数以及考生类型，给出在当前分数下较为合适的选择，为考生规划出冲刺院校、适中院校、保底院校。用户可以对各个高校进行查询，获取高校的一些简介和往年的录取情况，也可以对各个专业进行查询，获取该专业的高校排名。

细分来看，该软件会先要求使用者拥有一个属于自己的账号，注册可以自己设定用户名、密码，还需要提供手机号码，在登录中，可以通过用户名或手机号码登录。登录过后，便进入主界面，会要求用户完善个人基础信息，其中包括了姓名、学校、身份证号、高考年份、高考省份。接着便进入了对学校的要求填写，可以查看各学校各专业的高校信息。并对所填写的高校或专业选择倾向进行推荐。

该软件目的在于帮助高考生或准高考生填写高考志愿，可以了解目标高校目标专业的招生要求，以此明确自己努力方向，也可以在志愿填写或者指定目标迷茫之际，根据系统推荐作为参考，明确志愿填报。

为了方便实用，该系统初步计划编写安卓端、IOS端、PC端与网页端四个版本，使得使用者可以随时随地得到高考志愿的推荐。首先开发的是网页端，在网页端推出后，进入试用期，了解用户的反馈，进行修改与完善，开发出最终版本，投入市场，再进行市场调查，了解用户对其评价，若反响良好，可以进一步打开该领域的市场，得到良好的收益回报，那便继续安卓端、IOS端与PC端的开发。

2.1.2系统性能

2.1.2.1精度   
定位精度小于等于10m。

2.1.2.2时间特性要求   
1. 响应时间：1秒

2. 更新处理时间：5分钟

3. 数据的转换和传送时间：即时

4. 计算时间：即时

2.1.2.3可靠性   
 系统应具有较高稳定性，综合可靠性包括从登录、显示到查询过程中所有环节正常运行的概率。系统处理目标容量大于等于1000批。

2.1.2.4灵活性

1. 操作方式上的变化：可以手动输入账号密码，也可保存到本地，也可通过其他软件如QQ，微信，手机号进行登录。

2. 同其他软件的接口的变化：只需更改变化接口

3. 精度的变化：字符形式存储，可以满足精度改变，只需更改前端输入形式即可。

4. 数据信息的变化：数据库全部采用字符格式存储，且数据库除了账号属性非空唯一外不设约束，在网页部分要求约束，能较好适应数据格式和内容的变化

2.1.2.5容纳能力

系统用户数：50,000,000

根据中国教育网数据，我国每年高考考生数量在900万到1100万之间，预计每年有50%的考生和学生家长会用到本系统，十年内的用户数量预估为5000万。

并发用户数：5,000,000

考虑到六七月份为用户并发高峰期，每年的活跃用户约为500万，我们设置500万的并发用户数量确保完全容纳。

2.1.3输入输出要求

输入信息主要包括几个部分：

用户输入院校信息：根据查找的院校自动从数据库中调出院校其他信息。

院校信息均为字符串格式。

排名信息均为字符串型存储。

用户个人信息：均为字符串格式。

2.1.4故障处理要求

服务器中数据可能丢失，为避免丢失应定时备份，当丢失时同样应即时恢复数据，以保证系统最快速的恢复的正常运行状态。  
 系统发生故障时应当回退到之前的页面并将错误信息发送至服务器。同时对用户显示错误处理。

|  |  |
| --- | --- |
| 故障 | 处理方式 |
| 网络连接断开 | 点击链接后转进页面显示“失去网络连接” |
| 必须输入内容为空 | 弹出子窗口显示“缺少必填项目” |
| 查询结果为空（输入信息有效） | 弹出子窗口显示“查不到合适的结果” |
| 输入信息无效 | 弹出子窗口显示“无效信息” |
| 客户端登录时间与服务器时间不符 | 弹出子窗口显示“时间显示错误” |
| 登录地点异常 | 弹出子窗口显示“您的登录地异常，账号存在风险” |

2.1.5其他专门要求

2.1.6.1可维护性

需要较好的可维护性，由于高考报名规则有改动的可能，在规则改动后需要尽快更新版本，因此院校推荐模块与其他模块应当尽量低耦合。

2.1.6.2安全保密

为保证数据库的安全，均使用纯字符串形式传输和存储。

账号密码泄露后需要能通过手机号码、邮箱找回。

2.2运行环境

Android：7.0及以上

IOS：9.0及以上

PC Client：Windows XP及以上，内存2G及以上

Web Browser：Internet Explorer、Microsoft Edge、Firefox、Google Chrome、360浏览器、猎豹浏览器

2.2.1设备

处理器：型号无要求，内存2G以上

外存：无要求

媒体及其存储格式：XLSX

输入及输出设备：鼠标、键盘、显示屏

数据通信设备的型号和数量：无要求

功能键及其他专用硬件：无要求

2.2.2支持软件

操作系统：Windows XP及以上、Android 7.0及以上、iOS 9.0及以上

编译程序：Windows系统使用Dev C++ 5.0以上，Android使用c4droid 4.0及以上

2.2.3接口

·用户接口

点击链接或按键：跳转指对应页面，错误则提示错误信息。

Esc：弹出菜单。

·外部接口

硬件输入：键盘向输入框输入字符；鼠标左键点击信息输入。

传输协议：https。

·内部接口

登录模块向服务器传输账号密码，查询模块向服务器传输输入的查询信息，服务器向各个界面模块传输对应的信息，如用户信息、院校信息、查询结果信息等。

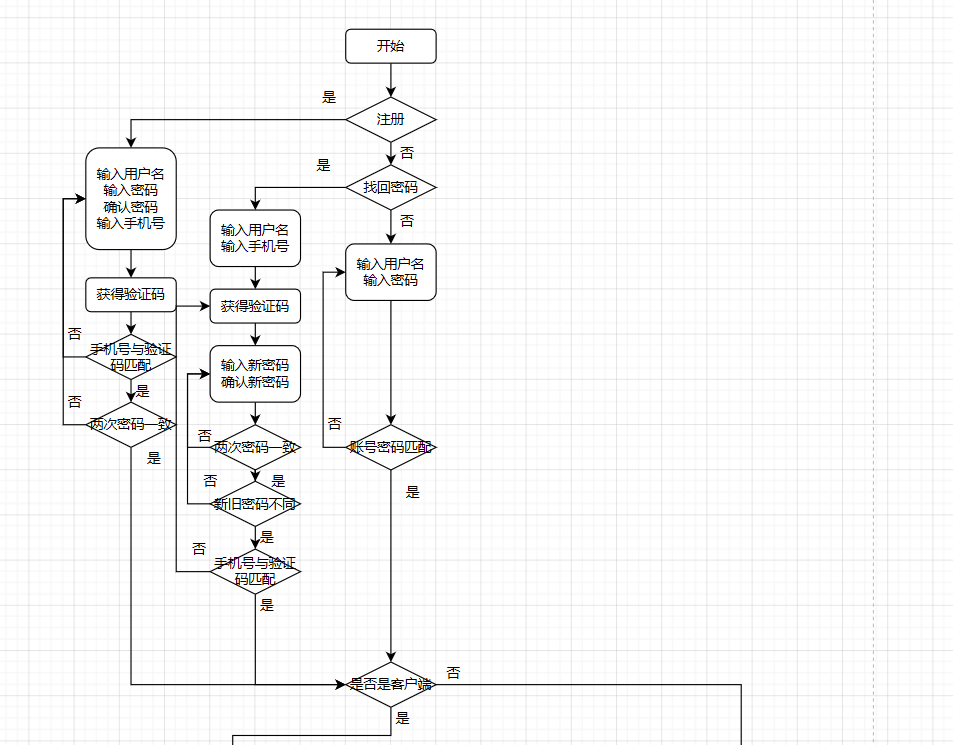
2.3基本设计概念和处理流程

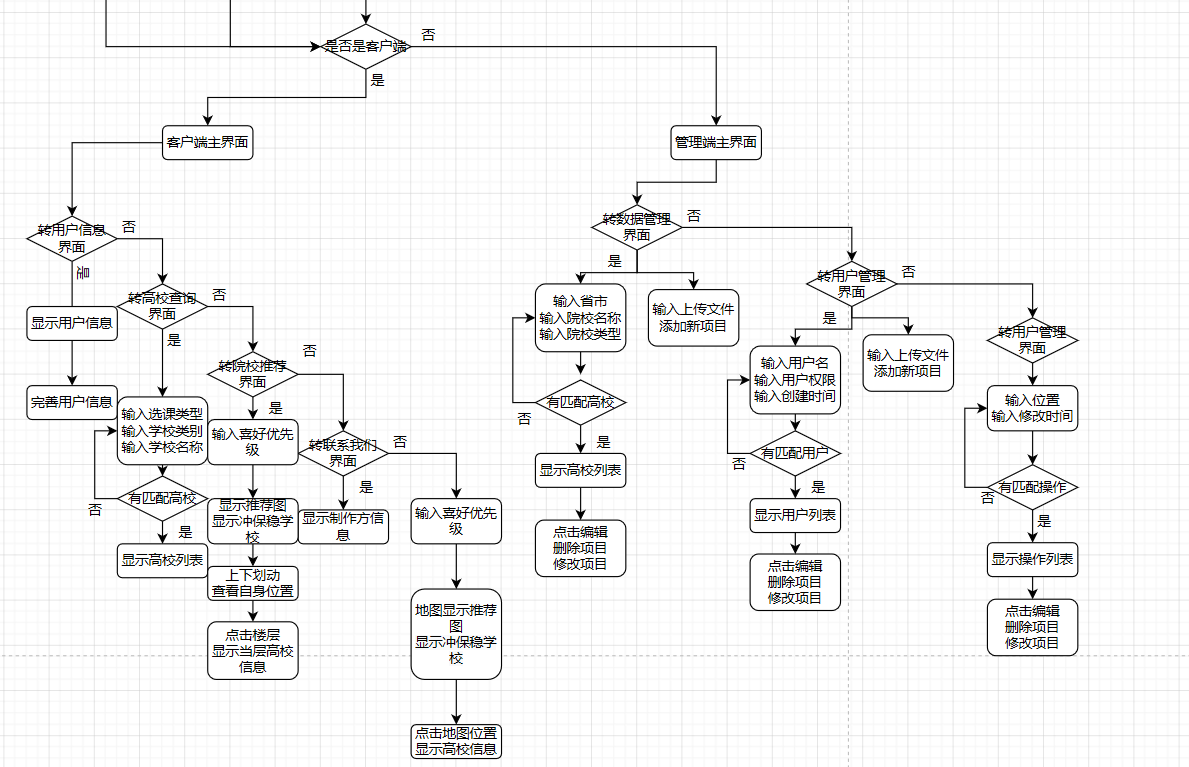
2.3.1 运行流程

管理员通过管理员账号密码，登入管理员账户并进行相关系统操作，最终退出管理员系统。

用户通过用户账号密码，登入用户账户并进行用户相关功能操作，最终退出用户系统。

流程图如下：





2.3.2数据处理设计

使用爬虫的方法获取高校数据。

·数据爬取流程

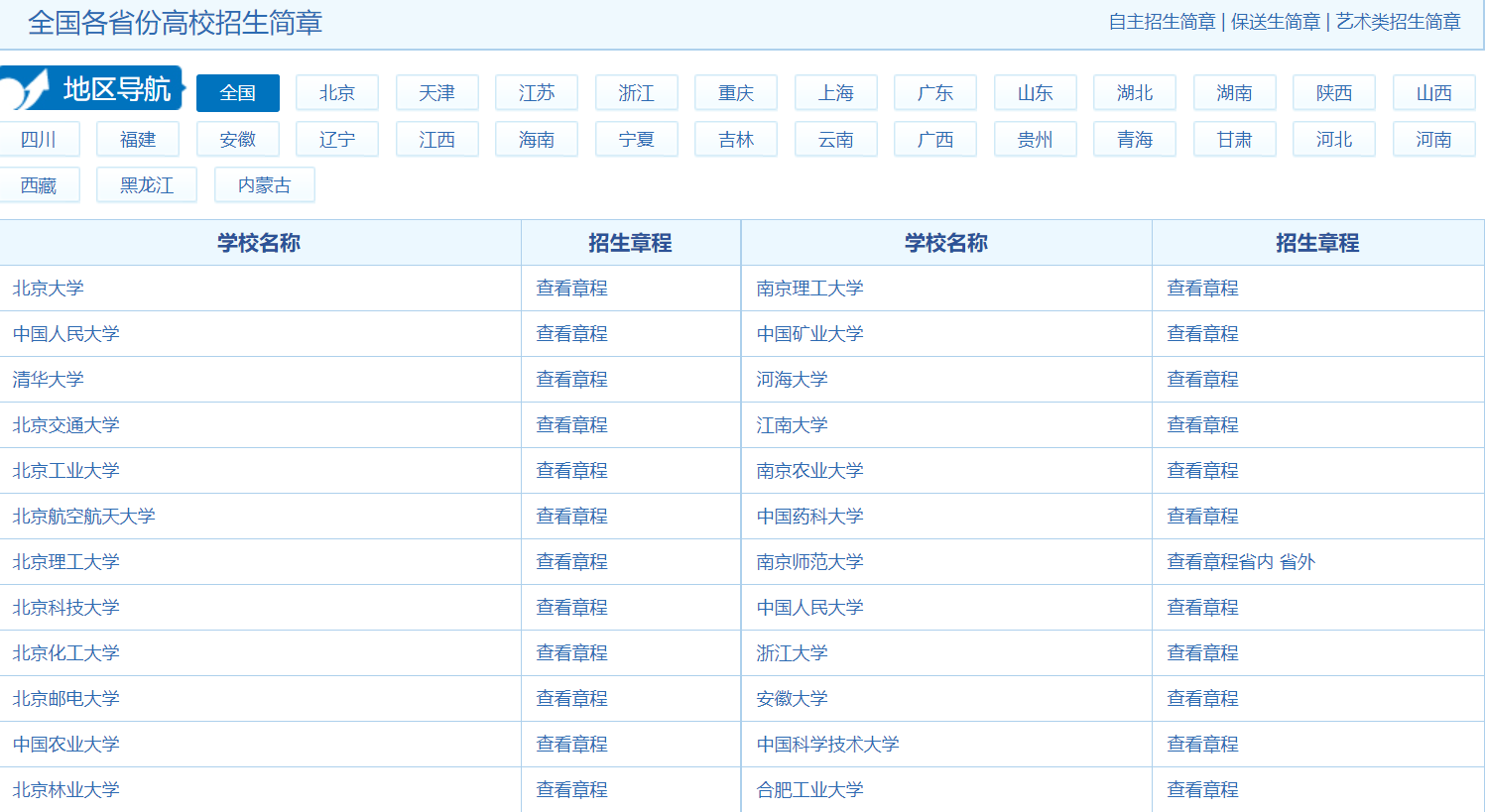
1、各省历年分数线；一本大学历年分数线；二本大学历年分数线；三本大学历年分数线

http://www.gaokao.com/baokao/lqfsx/gsfsx/

对爬到的数据进行预处理：提取出高校名称，省份，高校类型，以及该院校历年分数线文件路径等信息，以高校+省份为主键，保存到院校信息数据表。

2、各学校招生简章

http://www.gaokao.com/baokao/yxdq/zsjz/



爬取数据时需要逐个进入学校官网进行爬取，得到每个学校的高校隶属，高校所在地，院士数量，博士点数量，硕士点数量，通讯地址，联系电话，电子邮箱，学校网址，等信息，根据高校名称+省份的主键，存入数据表的对应表项中。

根据主键将院校简介文件路径，招生简章文件路径，专业目录文件路径也抓取下来存入院校信息数据表。

同时每年爬取各省份各大学历史录取名单以及所有学生志愿列表。



进入到各个专业的排名，需提取出专业的名称作为排名信息数据库中一个表的表头，获取排名表中高校和其对应的省内名次，全国名次，学科评级等，按降序写入专业排名表。

4、各种大学排名

http://www.gaokao.com/dxpm/zgdxpm/



进入到中国762所大学的自然科学与社会科学排名，获取排名表中高校和其对应的省内名次，全国名次，学科评级等，按降序写入自然科学排名表、社会科学排名表。

5、各大学专业目录(还可以在本界面进行分数线查询)

<http://college.gaokao.com/schlist/>

点进学校界面，专业介绍中是专业目录





根据高校名称+省份的主键，将专业目录文件路径抓取下来存入院校信息数据表。

·数据转入

得到爬取到的各类信息表，可由管理员在管理端上传文件或直接上传到数据库，服务器接收到文件时按照主键和对应关系存入数据库。

·数据更新

每隔两个月对数据源进行一次重新爬取，并根据爬取结果更新数据库。在每年5月、6月、7月每半个月进行一次数据更新。

### **2.3.3数据库设计**

（1） 院校信息表

高校名称，省份，高校类型，高校隶属，高校所在地，院士数量，博士点数量，硕士点数量，通讯地址，联系电话，电子邮箱，学校网址，院校简介文件路径，招生简章文件路径，专业目录文件路径，历年分数线文件路径

类型：VARCHAR(100)

主键：高校名称+省份

1. 用户信息表

用户名，密码，手机号，邮箱

类型：VARCHAR(100)

主键：高校名称+省份

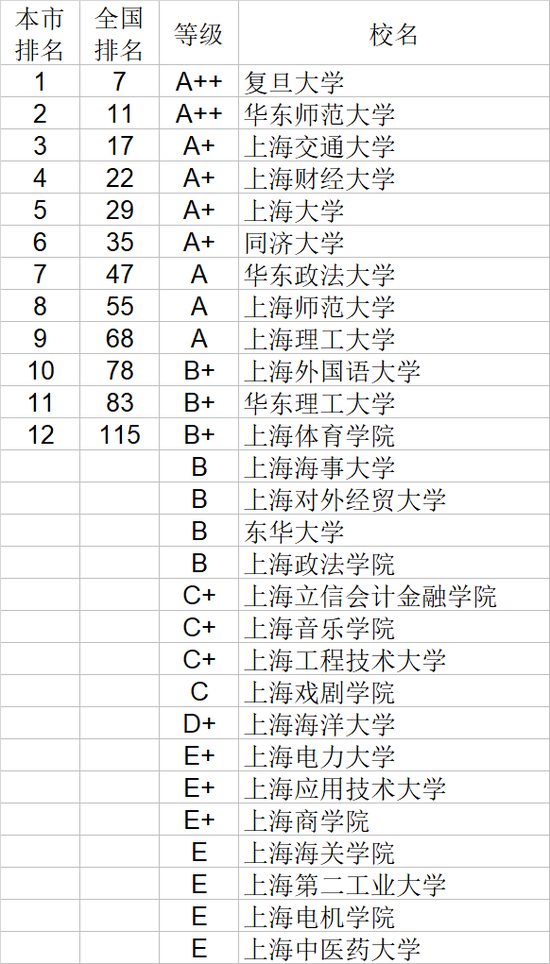
1. 社会科学综合排名信息表

高校名称，省份，本省排名，全国排名，等级

类型：VARCHAR(100)

主键：高校名称+省份

参考样例：



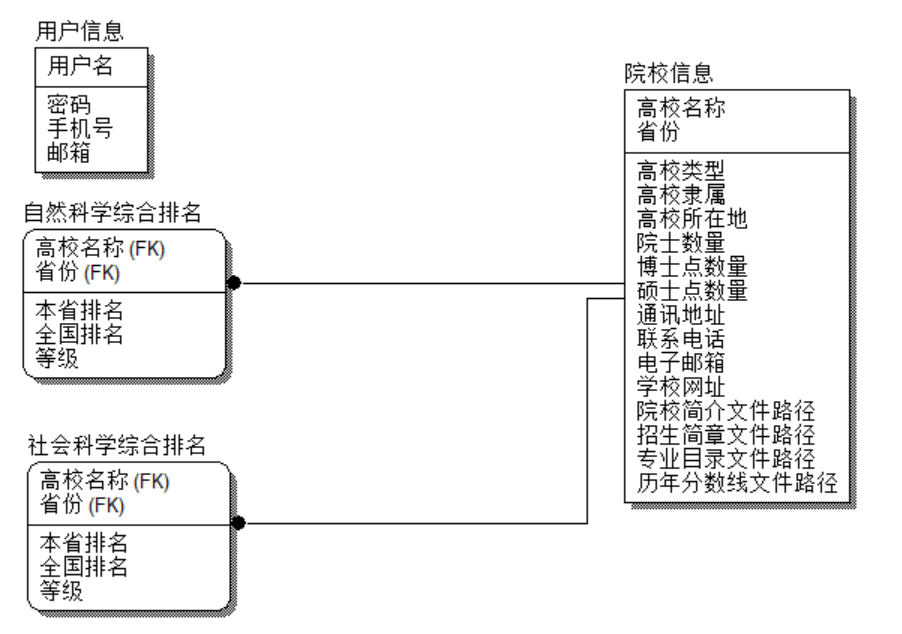
1. 自然科学综合排名信息表

高校名称，省份，本省排名，全国排名，等级

类型：VARCHAR(100)

主键：高校名称+省份

1. 数据库ER图



2.3.4推荐算法设计

在本系统中，我们采用机器学习和权重相结合的推荐算法 。通过机器学习进行专业和地域推荐，通过权重进行学校推荐。在推荐算法中，我们要根据三种不同的优先级方式，进行不同的权重分配和预测，在每种方法内，我们要分别进行冲，稳，保三种方式的预测。首先要训练模型，再使用模型对每个人的请求作出预测。模型训练大致流程如下：

数据收集

数据集整理

BERT模型

地区预测模型

专业预测模型

我们根据专家评判等方式，对当年考题难度进行判断，并根据不同学校近五年的录取分数线及其走势，判断今年不同学校的预估录取分数线。我们将学校的分数线分为预估标准线（STD）和预估分数线两个概念，预估标准线为排除历年难度系数差别后，学校的预估录取分数线，我们认为该指标可以有效的反应学校的录取水平。预估分数线为结合本年的试题难度所生成的真是分数线，其计算方式为预估分数线与难度系数的乘积。预估分数线的具体算法如下：

STDn= F(STDn, STDn-1,……STD0)

其中STDn为预测年份，STDn-1……STD0为结合前n年的预估标准线情况。根据线性回归算法，推导出STDn，并根据难度系数判断当年该学校的预估分数线。

在预测模型方面，首先我们将学生的特征进行向量化，我们可以通过自然语言处理的方式，将学生的学校信息，地域信息，兴趣爱好等视为自然语言，对其进行特征化处理。然后通过自然语言处理模型如： BERT等进行训练并获得可以进行专业推荐的模型。我们可以将往年的学生数据作为数据集，这样可以保证数据的充足，得到更好的模型和预测结果。

具体实现过程中，我们将学生的学校所在城市，学校教学水平分别转化成长度为20的向量，将学生兴趣爱好，学生家庭背景转化为30的特征，将四个特征进行拼接得到代表每个学生长度为100的特征向量。在往年学生数据所构成的数据集中，我们讲每个学生最终所去到的省份和专业分别作为标签来训练神经网络。为了使得神经网络可以得到较为准确的结果，我们做对专业的细分，但不进行具体城市的细化。在训练过程中我们采用softmax进行预测，将softmax最终的返回值作为每个结果的权重，爬取各高校往年学生录取分数用于训练模型。模型包括学生的专业偏好预测和地区偏好预测。

模型训练算法如下：

|  |
| --- |
| Step1. 打开爬好的往年录取详细名录，提取province, major, score, university, rank, perference1, perference2, perference3, total\_score\_history  Step2. 构造(名录人数的90%)\*(20+20+30+30)特征向量，用做训练集；剩下的名录人数用作测试集  Step3. 特征化处理，投入BERT模型训练，每次训练完后使用测试集测试预测当年各高校各专业报名人数，并预测分数线，输出各测试集成员预测报名专业和此种情况下能被录取的概率，调节参数直到准确率达到87%以上，这将用于预测下一年全国各高校各专业在各省份的分数线和排名  Step4. 保存模型在数据库  Step5. 每年9月招生结束后运维人员再次更新模型 |

在权重方面，由于我们已经得到了学校的预估分数线，我们可以讲学校的预估分数线和学生的估分（或者实际得分，依据不同地区而定）进行对比，为学生推荐合适的冲（分数差30~10），稳（-10~10），保（-30~-10）的学校。在不同的冲，稳，保类别中，我们采用权重来判断专业优先，地域优先，院校优先。

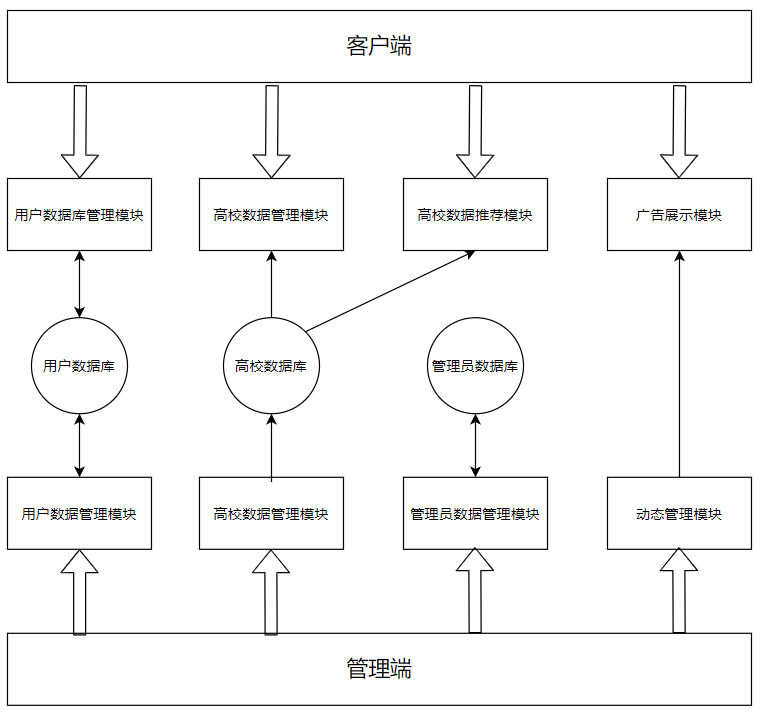
根据优先级，我们可以给三个不同的优先级分别以50，30，20的权重。根据机器学习获得的专业，地域信息，我们对不同学校的推荐专业，所在地域进行打分。我们对学校的院校排名，专业，地域打分进行加权，最终得到一个权重值，代表冲，稳，保三类中不同学校的优先级，对学生进行回复。

志愿推荐算法如下：

|  |
| --- |
| Step1. 取得界面传递的用户信息username, position, score, rank, preference1, preference2, preference3（3个preference可能是地区、专业、院校中的一个）  Step2. 将信息转换为20+20+30+30特征向量vec\_stu  Step3. 加载模型进行计算，得到今年各学校各专业预测分数线和排名  Step4. 权重按照preference的顺序从高到低排列，按照模型预测的分数线将所有university\_major按照成功概率排序，选取成功概率在90%-100%的进入保组third\_group，成功概率在70%-90%的进入稳组second\_group，成功概率在5%-70%的进入冲组first\_group  Step5. 按照preference1在冲稳保三组中各选择不多于5个志愿，再按照preference2在冲稳保三组中各选择不多于3个志愿，再按照preference3在冲稳保三组中各选择不多于1个志愿（志愿可重复）。结果按照成功概率排名，保存在result中  Step6. 向前端返回结果result供显示 |

2.4结构

2.4.1 模块图



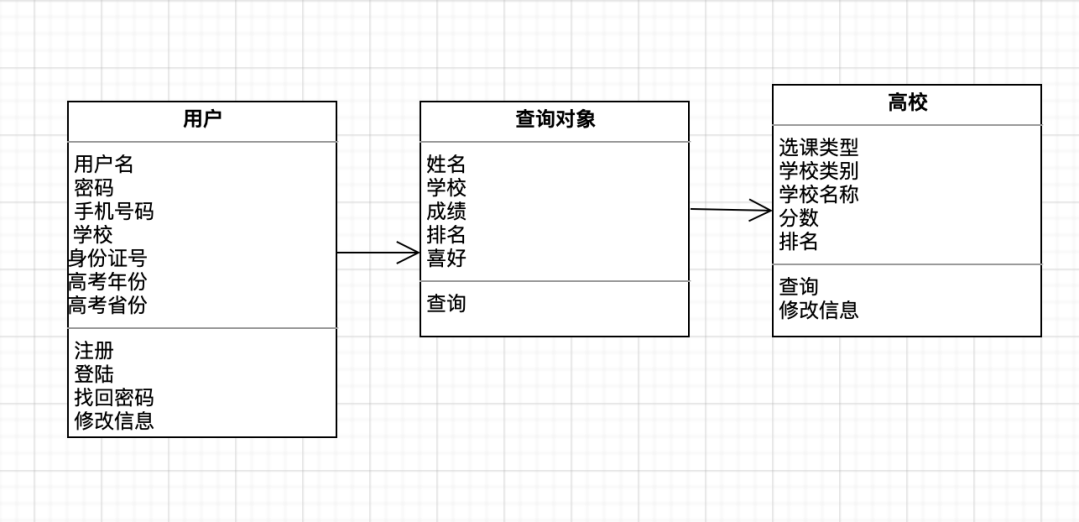
用户数据库和管理模块：用于验证用户的登录信息，区别不同的用户，控制和调整用户权限等。

高校数据库和管理模块：管理端可以定期对高校数据库进行更新，客户端对高校数据库进行查询，获得用户需要的查询内容并生成院校推荐。

管理员数据库和管理模块：用于验证管理员的登录信息，对操作管理端的管理员信息进行存储。

动态管理模块和广告模块：广告模块为客户端中自动展示的内容，管理端的动态管理模块可以对广告位进行管理和更新。

2.4.2 类图



2.5功能需求与系统模块的关系

客户端：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 用户数据库管理模块 | 高校数据管理模块 | 高校数据推荐模块 | 广告展示模块 |
| 登陆 | √ |  |  |  |
| 注册 | √ |  |  |  |
| 找回密码 | √ |  |  |  |
| 完善用户信息 | √ |  |  |  |
| 高校筛选 |  | √ | √ |  |
| 高校信息查询 |  | √ | √ |  |
| 给出推荐方案 | √ | √ | √ |  |
| 可视化方案 |  |  | √ |  |
| 广告展示 |  |  |  | √ |

管理端：

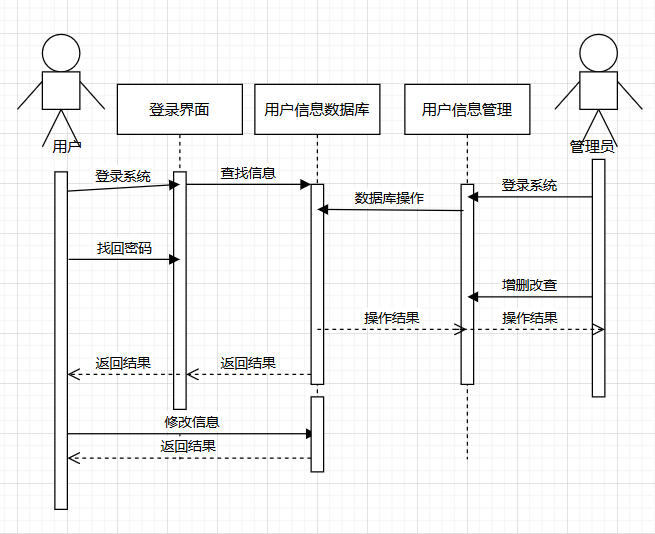
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 管理员数据库管理模块 | 高校数据管理模块 | 用户数据管理模块 | 动态管理模块 |
| 管理员登陆 | √ |  |  |  |
| 管理员注册 | √ |  |  |  |
| 管理员找回密码 | √ |  |  |  |
| 高校信息修改 |  | √ |  |  |
| 用户信息修改 |  |  | √ |  |
| 动态管理 |  |  |  | √ |

2.6人工处理过程

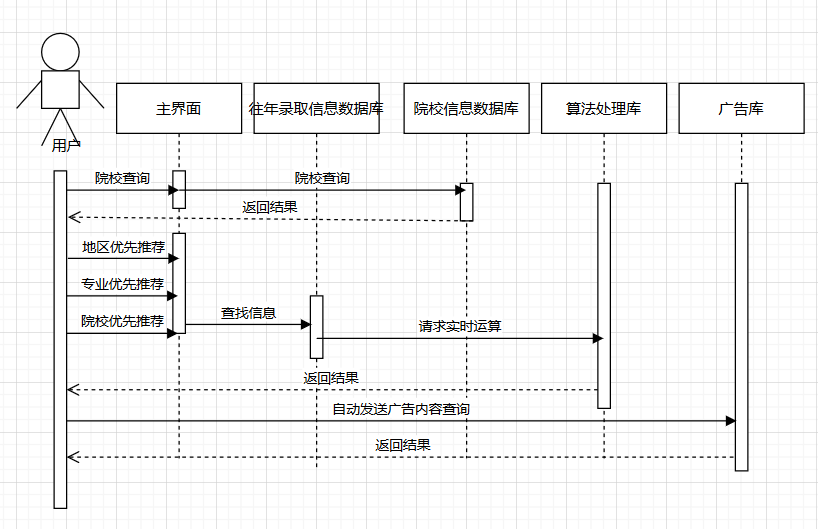
1. 若用户需要权限变更，需人工处理
2. 广告内容的更新需要工作人员手动更新
3. 每年9月1日，管理员会手动更新新高考数据和模型

3．时序设计

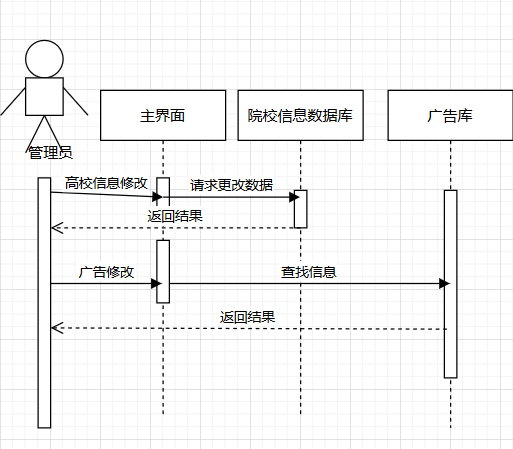
3.1登录时序图



3.2用户功能时序图



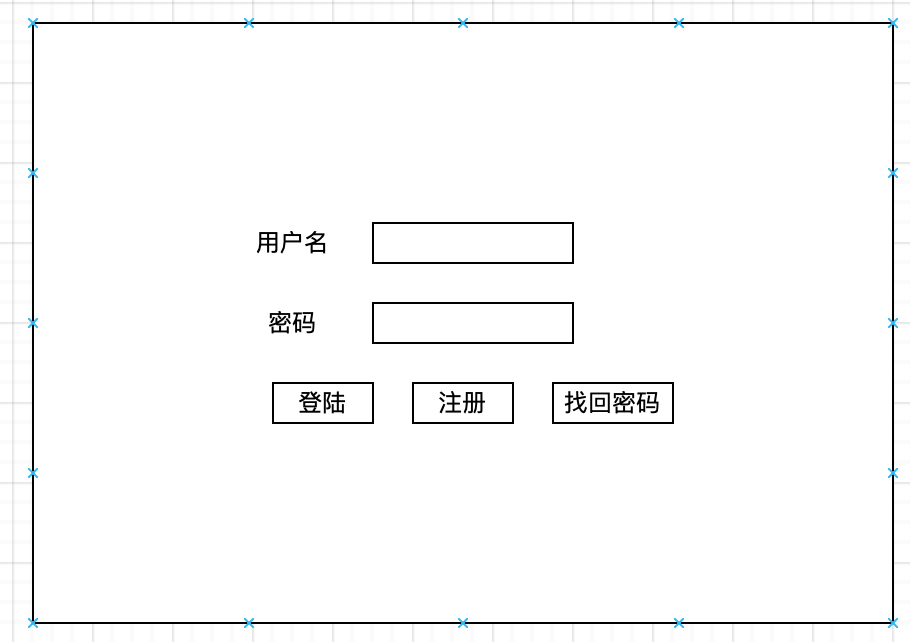
3.3管理端功能时序图



4．接口和操作

4.1客户端

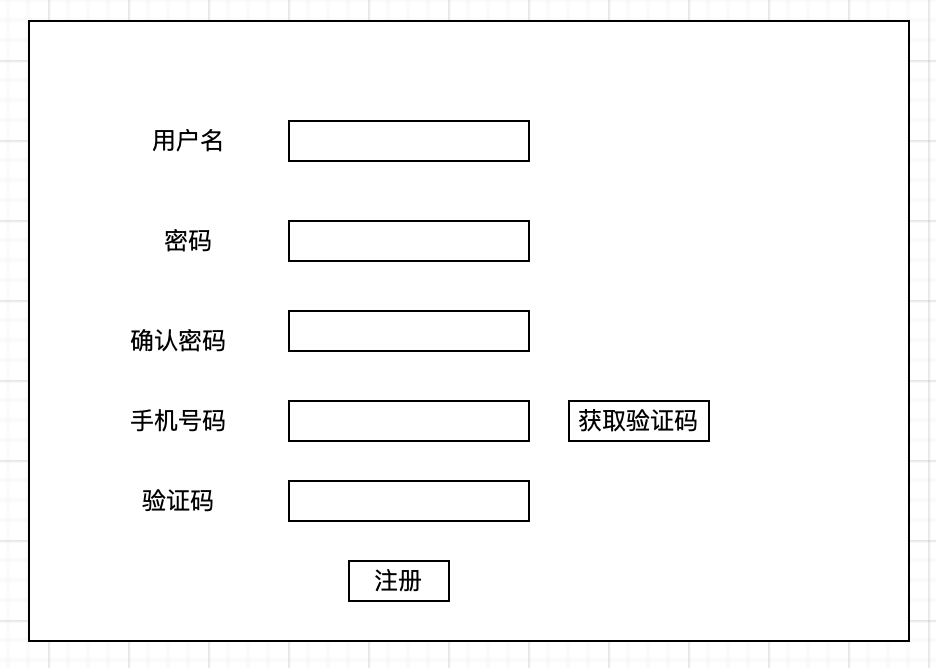
* 登录页面



用户输入账号密码，服务器查询用户数据库验证账号密码的正确性，并返回验证结果。

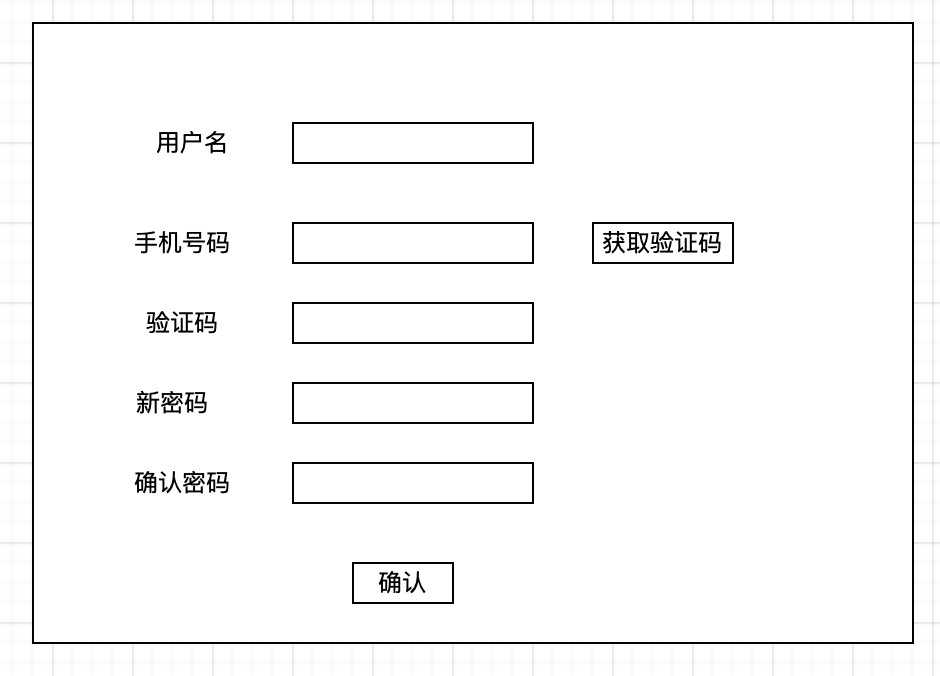
用户点击按钮进行注册界面和找回密码的跳转。

* 注册页面



用户输入账号密码，服务器在数据库中新建一项记录，点击获取验证码时服务器发送验证码，点击注册跳转回登录界面。

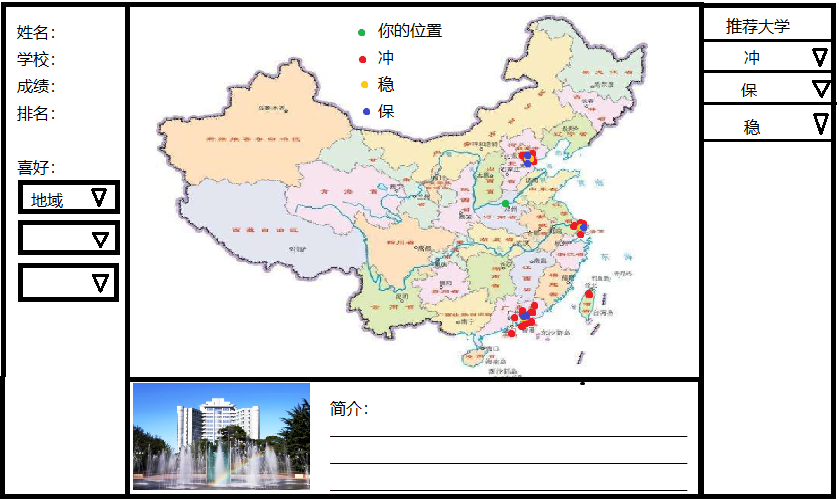
* 找回密码页面



用户输入信息，获取验证码，确认后发给服务器，若无误，更换密码并跳转到登录界面。

* 主页面/院校优先页面

院校优先界面要求在中国地图上显示三个志愿中的所有学校，可以不填满，冲稳保以颜色标出，并在右边栏显示列表详细信息和成功率，分别推荐五个。左边栏三个喜好为优先度最高的学校，必须在图中显示，替换掉学校所在的冲稳保档次所在院校。



用户点击不同子页进行跳转。

* 地域优先界面

地域优先界面要求中间左侧背景为当地景观，右侧从上到下列出喜好地区冲稳保学校，冲稳保以颜色标出，并在右边栏显示列表详细信息和成功率，分别推荐五个。左边栏三个喜好为优先度最高的城市，可以不填满，必须在图中显示至少2个当地院校，替换掉学校所在的冲稳保档次所在院校。中间以成功概率85%为上下分界，标注“你的85%成功概率线”。



用户点击不同子页进行跳转。

* 专业优先界面

专业优先界面部分专业喜好优先级，喜好均为平行。中间列出冲稳保院校，后跟专业和成功率，点击可以查看院校详情。



用户点击不同子页进行跳转。

* 高校查询界面

选择选课类型，学校类别，学校名称，如图查询得高校历年分数线，位次，专业目录和官网等详细信息。



* 联系我们界面

显示官方联系方式。

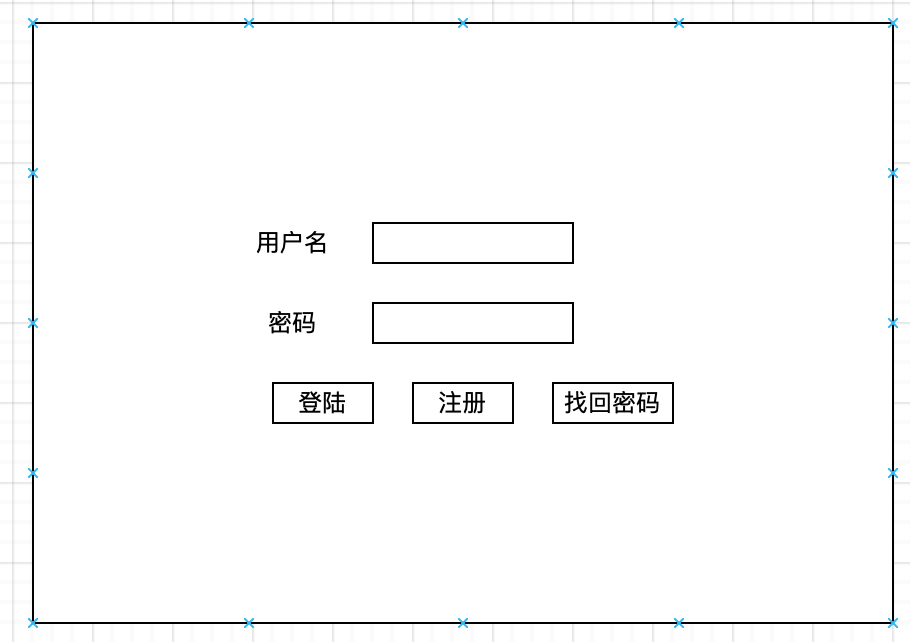


* 广告位

在用户界面最下面一行显示等宽度，高为显示界面1/5的广告位置，用户点击广告位图片可跳转到广告详细内容。

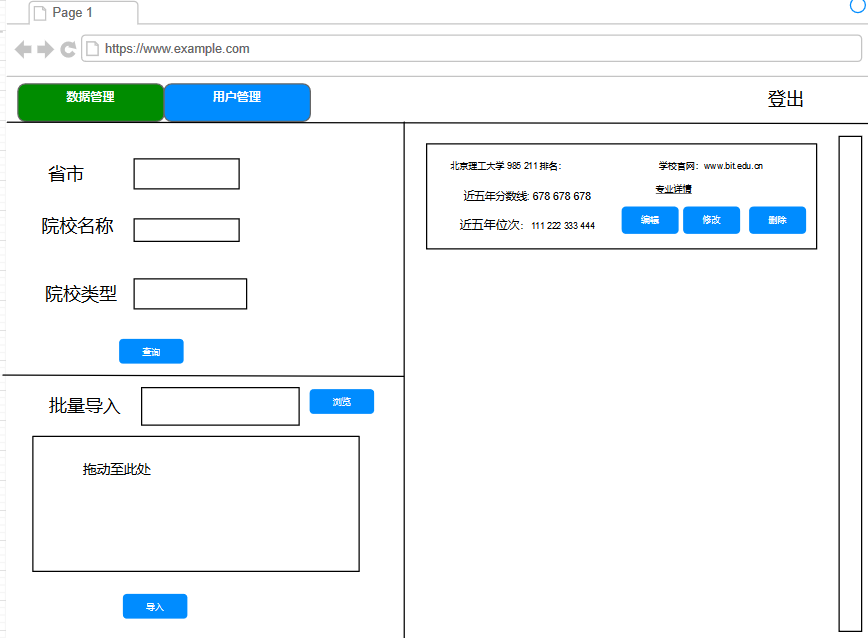
4.2管理端

* 登录页面



与客户端相似，管理员输入账号密码，服务器在管理员数据库中验证账号密码正确性和管理员权限。管理员输入信息进行注册，服务器在数据库中新增记录，点击获取验证码时服务器发送验证码。管理员账号需要root权限账号事先输入数据库，不允许他人注册。

* 主界面（院校管理界面）

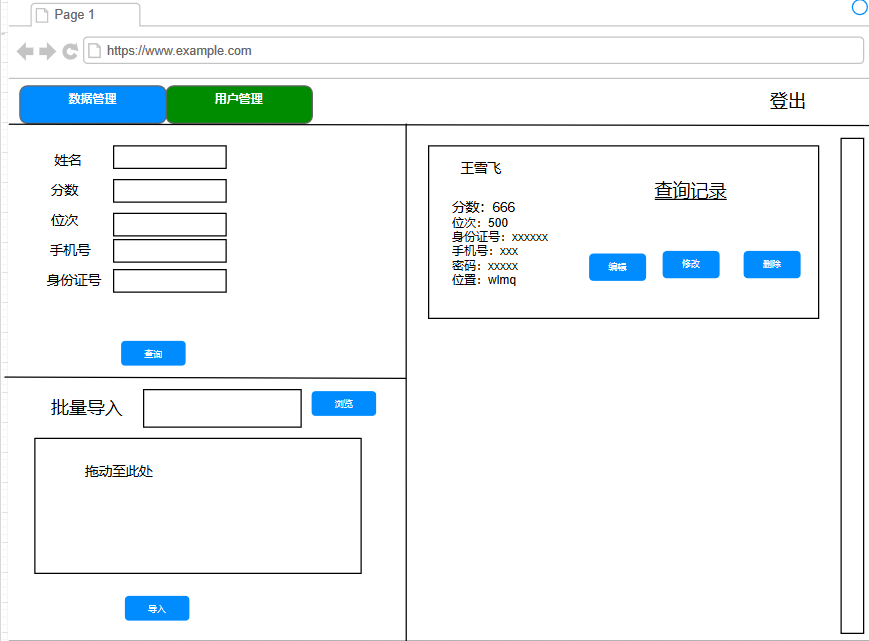


输入省市/院校名称/院校类型中的0-3种查询可得符合条件院校，若为0个输入查询全部，可以在右边编辑、修改、删除。左下可以批量导入院校信息，支持浏览和拖动上传。

上传文件功能：管理员上传对应格式的文件，服务器接收到文件对内容进行预处理，再根据文件中的内容更新数据库。

用户点击不同子页进行跳转。

* 用户管理界面



输入省市/院校名称/院校类型中的0-3种查询可得符合条件院校，若为0个输入查询全部，服务器根据目标信息在数据库中进行查找和筛选。可以在右边编辑、修改、删除。左下可以批量导入院校信息，支持浏览和拖动导入。

上传文件功能：管理员上传对应格式的文件，服务器接收到文件对内容进行预处理，再根据文件中的内容更新数据库。

用户点击不同子页进行跳转。

5．系统出错处理设计

5.1出错信息

|  |  |
| --- | --- |
| 错误类型 | 错误含义 |
| URL错误 | 用户在浏览器输入了错误的网址。 |
| 数据库查找无结果 | 根据用户设置的筛选条件，在数据库中进行查找时，没有符合条件的数据。 |
| 数据库查找无响应 | 用户发送查找请求，由于同时访问数据库的请求过多，发生了超时。 |
| 数据库写入出错 | 用户在注册账号时，由于用户名重复等情况无法将账号信息写入数据库。 |
| 数据库比较出错 | 用户在登录时账号密码错误。 |
| 数据库宕机 | 由于各种原因导致的数据库无法访问或数据丢失。 |
| 服务器宕机 | 由于各种原因导致的服务器无法访问。 |

5.2补救措施

|  |  |
| --- | --- |
| 错误类型 | 解决方法 |
| URL错误 | 返回404.html。 |
| 数据库查找无结果 | 提醒用户重新设置筛选条件。 |
| 数据库查找无响应 | 提醒用户当前系统繁忙，请稍后再试。 |
| 数据库写入出错 | 提醒用户修改用户名。 |
| 数据库比较出错 | 提醒用户账号密码错误，引导找回密码。 |
| 数据库宕机 | 每天晚上进行动态增量转储；每周进行一次动态海量转储；每月进行一次静态海量转储；设置日志文件，则事物故障可有系统自动完成处理，系统故障可由系统重启后自动完成处理；采用检查点恢复技术，设置重启动文件，则发生介质故障时可以根据日志文件、转储的副本和重启动文件进行数据库恢复。 |
| 服务器宕机 | 准备2个网站空间，他们存放的内容相同，而ip不同，并且机房的地理位置不同。这样2个主机，同时宕机的可能性就大大降低了。第一时间发现宕机问题后，可以迅速的通过修域名记录，指向目前正常的网站空间。 |

5.3系统维护设计

每天晚上进行动态增量转储；

每周进行一次动态海量转储；

每月进行一次静态海量转储；

设置日志文件，则事物故障可有系统自动完成处理，系统故障可由系统重启后自动完成处理；

采用检查点恢复技术，设置重启动文件，则发生介质故障时可以根据日志文件、转储的副本和重启动文件进行数据库恢复；

准备2个网站空间，他们存放的内容相同，而ip不同，并且机房的地理位置不同。一旦服务器宕机，可以迅速的通过修域名记录，指向目前正常的网站空间。