

## 目录

1 引言.....	1
1.1 编写目的 .....	1
1.2 定义 .....	1
1.3 项目背景 .....	2
1.4 参考资料 .....	2
1.5 版本更新信息 .....	2
2 任务概述.....	3
2.1 开发目标 .....	3
2.2 需求概述 .....	3
2.3 开发环境 .....	3
3 总体设计.....	4
3.1 总体模块设计 .....	4
3.2 C/S架构 .....	6
3.3 设计规定 .....	7
3.4 设计约束 .....	7
3.5 设计策略 .....	8
4 模块功能.....	8
4.1 登录注册模块 .....	8
4.1.1登录模块.....	8
4.1.2注册模块.....	10
4.2“用户-专家”医疗问答模块 .....	11

4.3 问诊预约模块 .....	12
4.4 机器人问答模块 .....	13
4.5 信息管理模块 .....	15
5 数据库设计 .....	18
5.1 数据库表 .....	18
5.2 数据库关系图 .....	21
6 医疗问答机器人核心算法设计 .....	22
6.1 问答机器人简介 .....	22
6.2 Word2Vec简介 .....	23
6.3 语料库描述 .....	23
6.4 句子相似度计算方法 .....	24
7 界面设计 .....	25
7.1 界面设计总体规定 .....	26
7.2 登录与注册界面设计 .....	26
7.2.1 登录界面设计 .....	26
7.2.2 注册界面设计 .....	26
7.3 用户客户端界面设计 .....	27
7.3.1 用户提问界面介绍 .....	27
7.3.2 用户问诊预约界面设计 .....	28
7.4 专家客户端界面设计 .....	28
7.5 管理员客户端界面设计 .....	29
7.6 机器人交互界面设计 .....	30
7.7 用户/专家邮箱界面设计 .....	31

8 系统出错信息处理与安全设计 .....	32
8.1 主要出错信息处理设计 .....	32
8.2 安全保密性设计 .....	33
8.2.1 MD5加密技术 .....	33
8.2.2可信管理员.....	33
8.2.3 数据库安全性设计.....	33

# 1 引言

## 1.1 编写目的

本文档在《智能医疗问答系统需求分析规格说明书》的基础上，对智能医疗问答系统“疑问医答”的系统设计进行进一步的说明，解释系统中各个层次的设计考虑，为系统的开发工作和测试工作提供参考。本文档的叙事内容包括如何将系统划分为若干个模块，系统相关数据结构设计以及算法设计，预期读者为项目经理、软件开发人员、测试人员、系统分析员和系统管理员。

## 1.2 定义

**智能医疗问答系统：**用户与医疗领域专家进行交流，或与医疗智能机器人进行交互的平台，以Windows桌面客户端形式为用户提供服务。

**疑问医答：**本项目所开发智能医疗问答系统的名称。

**Python：**本项目主要使用的一门面向对象程序设计语言。

**PyCharm：**一款强大的Python IDE(集成开发环境)，具备编码协助，项目代码导航等提供用户开发效率的功能，本项目主要开发环境。

**Jupyter Notebook：**基于网页的交互计算应用程序，本项目中用于语言模型训练。

**My SQL：**本项目所使用的一个关系型数据库管理系统。

**MySQL Workbench：**一款专门为MySQL设计的可视化数据库管理软件。

**E-R图：**表示概念关系模型的一种方式，提供系统实体类型、属性和联系方法。

**JSON：**轻量级的数据交换格式，本项目将其用于客户端和服务端之间的通信。

**C/S架构：**即客户端/服务器模式，分为客户端和服务端两层，第一层是在客户端系统上结合了表示与业务逻辑，第二层是通过网络结合了数据库服务器。

**TCP连接：**一种面向连接的、可靠的、基于字节流的传输层通信协议，本项目中客户端与服务器的通信方式。

**Word2Vec：**又称词向量模型，自然语言处理领域常用的一种文本处理方法。用于将文本数据转化为结构化、可计算的向量，本项目对话机器人功能中进行语料库句子匹配的主要技术。

**医疗对话语料库：**用于Word2Vec模型的训练以及医疗机器人的基础对话库，来自阿里云天池数据集提供的Chinese Medical Dialogue Dataset 中文医疗对话数据集(<https://tianchi.aliyun.com/dataset/dataDetail?dataId=90163>)，属于公开数据集。

### 1.3 项目背景

医疗卫生服务直接关系到人民身体健康，互联网时代下，利用信息化技术进行医疗业务的治理与运营，是进一步提升医疗卫生现代化管理水平，优化资源配置，提高就医效率，降低服务成本的重要手段。基于此，智能医疗问答系统旨在建立用户与医疗领域专家的沟通平台，并基于前沿机器学习技术提供智能机器人问答服务，以满足用户的健康需求。

### 1.4 参考资料

- [1]. 张海藩,牟永敏.软件工程导论(第6版)[M].北京:清华大学出版社,2013
- [2]. 刘金岭,冯万利,张有东.数据库原理及应用SQL Server 2012[M].北京:清华大学出版社,2017
- [3]. 陈诗玉. Word2vec模型在聊天机器人中的应用研究[D]. 广东:暨南大学,2019.
- [4]. 马思丹. 基于加权Word2vec的微博文本相似度计算方法研究[D]. 陕西:西安电子科技大学,2019.
- [5]. 翟社平, 李兆兆, 段宏宇,等. 基于词法、句法和语义的句子相似度计算方法[J]. 东南大学学报（自然科学版）,2019,49(6):1094-1100.
- [6]. 陈丹华,王艳娜,周子力,等. 基于Word2Vec的WordNet词语相似度计算研究[J]. 计算机工程与应用,2022,58(3):222-229. DOI:10.3778/j.issn.1002-8331.2009-0090.
- [7]. 黄晓亮. 优化对话模型的在线医疗咨询系统设计与实现[D]. 江苏:南京邮电大学,2021.

### 1.5 版本更新信息

具体版本更新记录如下表：

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
1	2022-04-28	v1.0		

--	--	--	--	--

## 2 任务概述

### 2.1 开发目标

系统开发目标为设计一个功能实用，界面友好的智能医疗问答系统，系统主要为用户提供线上咨询功能；为专家提供答疑功能；为管理员提供便捷的管理模式。同时，系统还为用户提供了医生的线下问诊预约渠道以及基于医疗问答语料库的智能机器人问答服务。

### 2.2 需求概述

智能医疗问答系统所面向的客户主要分为三类：用户、专家以及管理员。系统需要为用户和专家提供注册账户，系统登录服务以及个人信息修改功能。对于用户，系统应当为其搭建与专家进行交互的良好平台，需要提供医疗问题咨询、线下问诊预约功能，系统同时提供机器人问诊功能，为用户提供更加快捷的解疑渠道；对于专家，需要相应地提供问题回复和预约处理功能，除此之外，专家也可以与机器人进行交互。

### 2.3 开发环境

表2-1展示了本小组开发智能医疗问答系统所用的配置信息，包括硬件信息和操作系统、编程语言、框架等软件信息。

表2-1：开发环境配置

类别	配置信息
硬件	处理器:AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx (8 CPUs) 内存: 8192MB RAM

软件	操作系统: Windows 10 Python 开 发 环 境 : PyCharm Community Edition 2021.2.2 + Jupyter Notebook 主要开发框架: PyQt5, Gensim 数据库: MySQL8.0 数据库软件: MySQL Workbench8.0 CE
----	--

### 3 总体设计

#### 3.1 总体模块设计

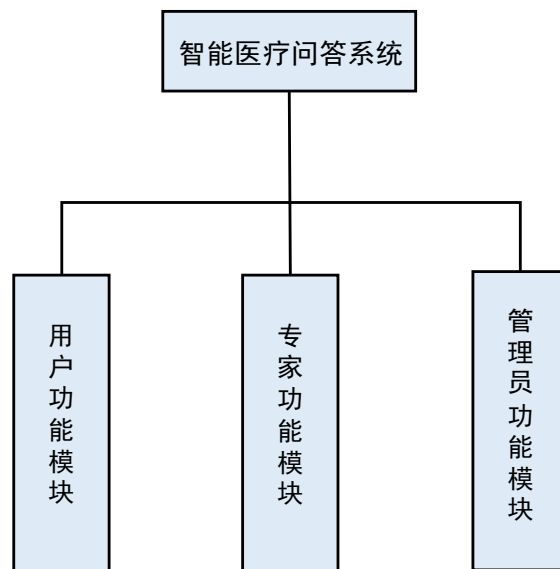


图3-1: 智能医疗问答系统功能划分

从总体上看, 本系统为用户、专家、管理员三种不同的对象提供了完全不同的功能。客户需要在注册和登录时需要选择自己的身份(用户、专家、管理员), 登录系统后将切换至对应的功能界面。

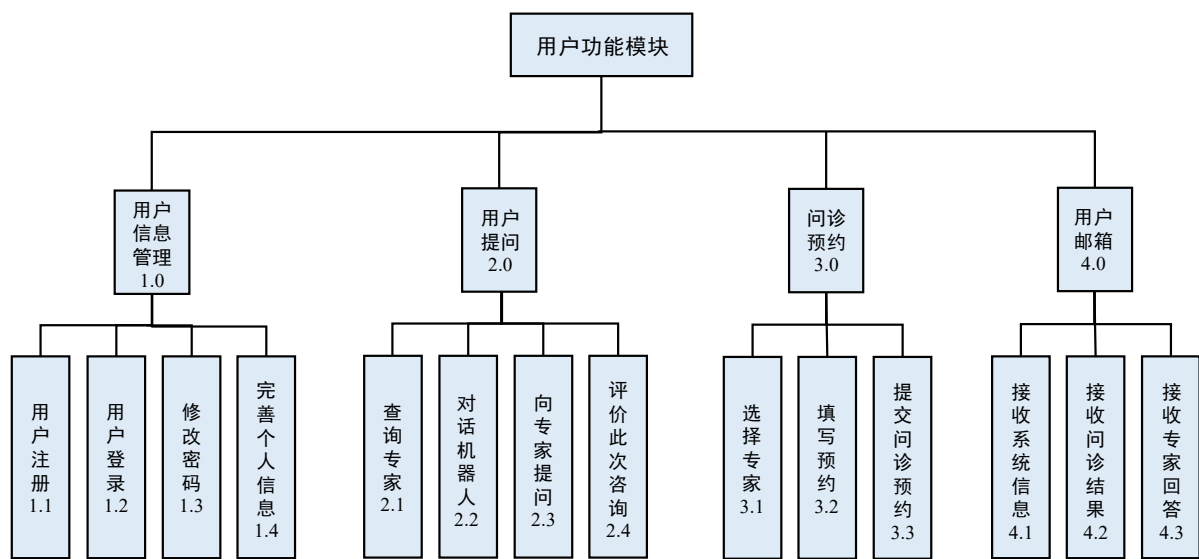


图3-2：用户功能模块概览

图3-2展示了用户功能模块的具体功能，主要分为用户信息管理、用户提问、问诊预约以及用户邮箱四大功能。其中用户信息管理为用户提供了基础的注册、登录以及密码和个人信息的修改与完善；用户提问功能包括查询专家、对话机器人、向专家提问以及评价此次咨询的业务，从而为用户提供方便快捷的医疗相关问题解惑，本功能为智能医疗问答系统的核心业务。问诊预约功能为用户提供线下问诊的渠道，有选择专家、填写预约和提交预约的功能。用户邮箱作为用户接受来自系统和专家相关信息和处理结果的平台。

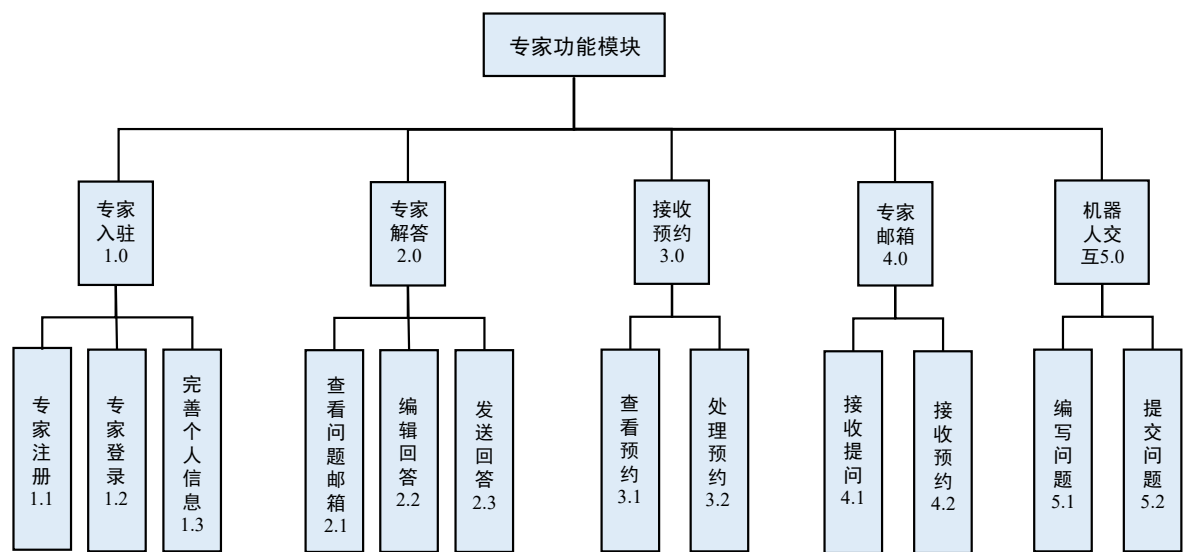


图3-3：专家功能模块概览

专家功能模块由专家入驻、专家解答、接受预约、专家邮箱、机器人交互



等相关功能组成。专家入驻的相关功能主要由注册、登录和完善个人信息功能，专家在注册时，需要填写用户名、密码(两次)、真实姓名、邮箱、单位、科室、性别、出生日期等信息；专家解答和接受预约则方便专家处理和回复来自用户的医疗问题与线下问诊预约申请。专家邮箱负责对用户的提问与预约进行收集。除此之外，专家可以与医疗机器人交互。

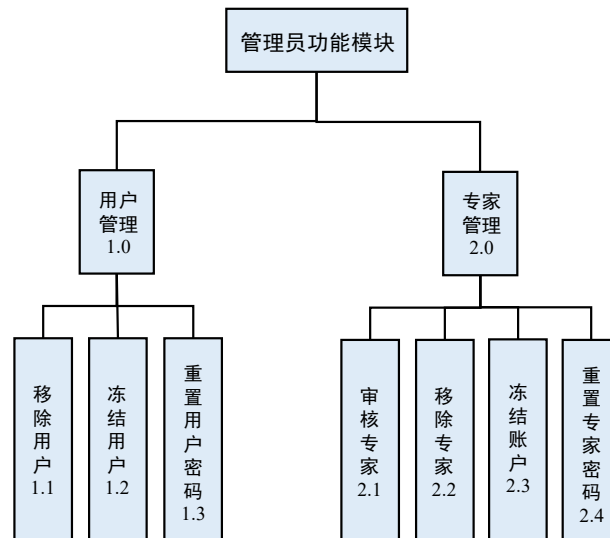


图3-4：管理员功能模块概览

管理员功能模块为系统管理员提供一个方便快捷的信息管理平台，可以实现对用户和专家的移除、冻结和重置密码等功能。同时，注意到专家的特殊性，专家注册后需要接受管理员的审核。除此之外，管理员还可以管理机器人的对话数据库，可以根据专家建议和用户评分进行数据库更新。

### 3.2 C/S架构

客户端-服务器（Client/Server）结构简称 C/S 结构，是一种网络程序开放架构，通常将该网络架构下的软件分为客户端和服务端。下图展示了本系统的C/S架构开放设计，具体而言，我们仅在客户端实现业务逻辑表达，包括交互界面，请求结果提示与可视化，也包括一小部分错误检测与提示(如检测用户输入十分合法等)。

具体而言，客户端的编程所需要实现的内容如下：

- 创建套接字，设置端口号和IP号；
- 连接服务器；
- 根据用户操作，发送对应相关业务的请求信息；

- 接收信息，进行信息提示或可视化；
- 处理相关事件，如部分错误检测与提示，风格切换等；
- 关闭套接字。

其中发送和接受信息一项则关系到本系统具体的业务实现。而相对应的，服务器端的编程框架如下：

- 创建套接字，设置服务器端口和ip地址；
- 开始监听来自客户端的请求信息；
- 接收请求信息并调用对应业务接口进行处理；
- 连接数据库，查询和处理数据库内容；
- 向客户端发送相关数据或错误信息；
- 关闭套接字。

本系统的每一个需要访问数据库信息的功能都有客户端和服务端共同实现，此时则需要建立客户端和服务器的通信机制。我们在客户端和服务端之间建立TCP连接，使用JSON数据格式封装数据，在发送数据的首位使用特定字符来标识该数据对应的系统业务信息，方便客户端和服务端调用相应的接口进行处理。我们将在第4章用时序图展示客户端与服务端之间具体的交互过程。

### 3.3 设计规定

编码规定：分为客户端(Client)和服务端(Server)两套程序,共同完成具体业务的实现。客户端基于PyQt5框架实现界面开发，服务端仅进行业务底层实现和数据库访问，不提供UI界面。编码时，为统一命名，具体功能的实现函数由功能对应的对象(用户、专家、管理员)和具体功能名称的组合命名，单词之间用下划线连接，部分单词使用缩写。

数据库设计规定：数据库的设计应当参照需求文档的相关说明和E-R图进行设计，数据库中对象和属性的名称以对应的英文单词或缩写进行命名。

界面设计规定：界面应当简洁大方，提供醒目的错误消息提示，方便用户快速上手。为方便使用者夜间使用，界面风格应支持切换夜间模式。

### 3.4 设计约束

#### (1) 需求约束

- 硬件和操作系统约束：在PC机上使用Windows7以上操作系统访问。

- 系统应当有具有较高的正确性、可靠性和运行效率。
- 为用户\专家的隐私安全提供一定保障，需要采用合适算法对密码进行加密。

## (2) 隐藏约束

- 用户、专家和管理员需要具备网络条件，网络延迟不超过1000ms。
- 用户和专家必须具备一定的智能设备使用能力设计策略。
- 管理员需要了解系统的业务逻辑、算法实现逻辑以及数据库设计。
- 管理员有权对存在违规行为的用户\专家实施冻结账号甚至删除信息。

## 3.5 设计策略

- (1) C/S架构：基于C/S架构进行设计，在客户端上实现业务逻辑，在服务器端进行相关模块的具体实现以及数据库访问，保证数据库安全。
- (2) 可扩展性策略：本系统采用模块化方法开发，将待开发的软件分解成若干个部分，新功能的添加影响到其他功能，具备良好的可扩展性。
- (3) 安全性：医疗信息严重涉及个人隐私，因此软件的开发需要考虑隐私安全。
- (4) 易用性：本系统需要有简洁友好的界面以及清晰的操作错误信息指引，以方便用户快速上手。

## 4 模块功能

本章将对本系统中几大核心功能模块进行详细设计，描述其基本信息并用时序图展示其运行过程中客户端、服务器和数据库之间的交互流程。为方便表述，我们这里对模块的分类是按功能分而不是按对象分类，这是由于某些功能比较相近，因此我们对其运作原理一起进行叙述。

我们主要对以下重点功能进行讲解：登录注册模块、“用户-专家”医疗问答模块、问诊预约模块、机器人问答模块以及信息管理模块。

### 4.1 登录注册模块

#### 4.1.1 登录模块

表4-1：登录模块基本信息

基本信息	描述
功能	实现用户、专家与管理员的登录
对象	用户\专家\管理员
输入	用户名、密码、身份(用户、专家或管理员)
输出	显示登录失败错误信息(用户名不存在、密码错误等)或跳转到对应界面；
客户端负责功能	<ul style="list-style-type: none"><li>• 发送请求信息；</li><li>• 接受响应信息；</li><li>• 显示登录失败信息；</li><li>• 跳转至对应界面；</li></ul>
服务器负责功能	<ul style="list-style-type: none"><li>• 接受和识别请求信息；</li><li>• 查询数据库进行错误检测；</li><li>• 发送响应信息；</li></ul>

用户、专家或管理员填写用户名、密码并选择对应身份时，客户端将信息打包发送至服务器。服务器根据数据头的标识，将其识别为登录请求后调用对应的函数处理，连接数据库。访问数据库以检查用户名是否存在以及密码是否正确，返回结果信息。客户端收到后如为错误信息则输出，显示检验通过则自动跳转到响应的界面。

下图展示了用户、专家或管理员登录过程中服务器、客户端以及数据库系统之间的交互。

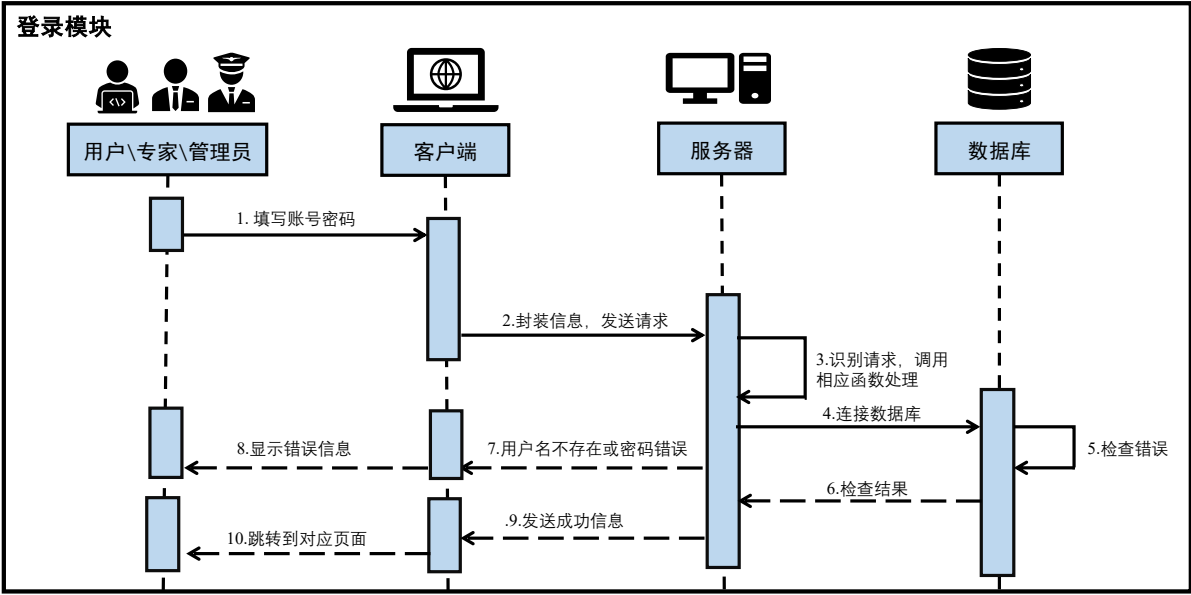


图4-1：登录功能

4.1.2注册模块

表4-2：注册模块基本信息

基本信息	描述
功能	实现用户与专家的注册
对象	用户\专家
输入	注册信息，包括用户名、密码、邮箱、身份(用户与专家)
输出	注册成功或错误信息(用户名或邮箱已存在)
客户端负责功能	<ul style="list-style-type: none"><li>• 部分错误检测；</li><li>• 发送请求信息；</li><li>• 接受响应信息；</li><li>• 显示注册结果</li></ul>
服务器负责功能	<ul style="list-style-type: none"><li>• 接受和识别请求信息；</li><li>• 部分错误检测；</li><li>• 连接数据库进行写入和查询；</li><li>• 发送响应信息；</li></ul>

用户\专家填写注册信息后，首先会在客户端检查部分错误(如密码长度，确认密码一行是否与之前一致，邮箱格式是否有效等),完全无误后发送至服务器。服务器根据数据头的标识，将其识别为注册请求后调用对应的函数处理，连接数据库。访问数据库以检查用户名和邮箱是否重复，若无误则写入数据库。写入成功后返回注册成功信息。

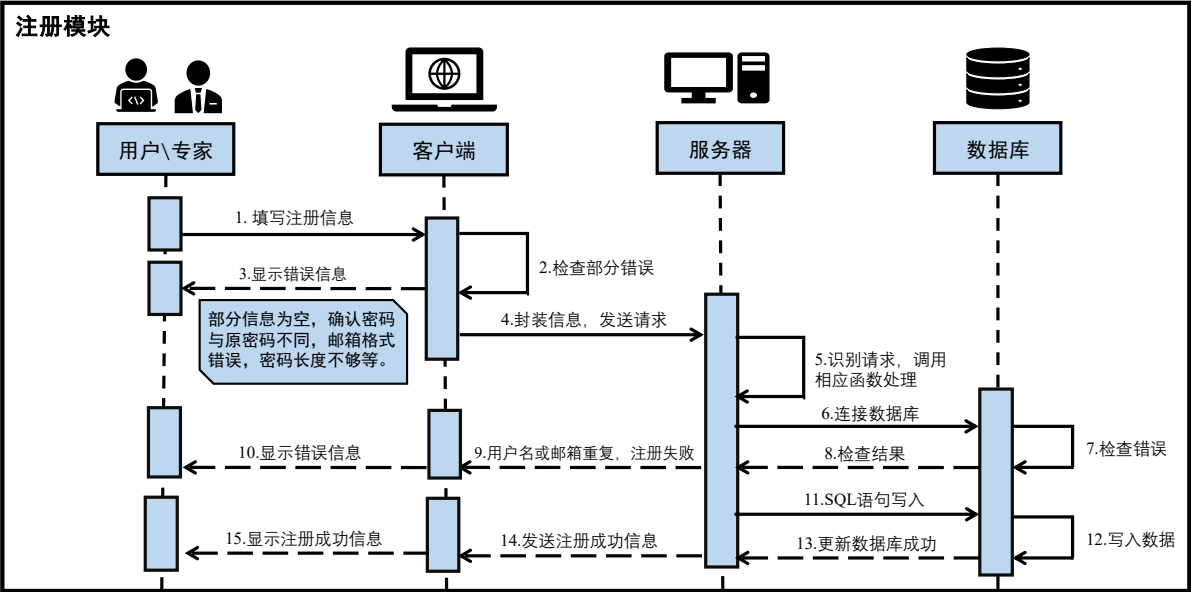


图4-2：用户\专家注册功能

4.2 “用户-专家” 医疗问答模块

医疗问答是本系统最核心的一块内容，可分为“用户-专家”问答以及“用户-机器人”问答和“专家-机器人”问答。我们这里展示用户与专家之间的问答过程。

表4-3：“用户-专家” 问答模块基本信息

基本信息	描述
功能	实现用户提问，专家解答
对象	用户与专家
输入(用户端)	用户问题
输入(专家端)	对用户问题的回复

输出(用户端)	专家答复
客户端负责功能	<ul style="list-style-type: none"><li>• 发送请求信息;</li><li>• 接受响应信息;</li><li>• 显示答复</li></ul>
服务器负责功能	<ul style="list-style-type: none"><li>• 接受和识别请求信息;</li><li>• 发送至专家邮箱系统;</li><li>• 转发回复;</li></ul>

如图4-3所示，用户选择专家填写问题后由客户端发送到服务器，再由服务器发送至专家邮箱(这里专家邮箱具体编程时由建立数据库表实现，方便起见这里没有画出)，专家填写回复后从专家客户端发送至服务器，最后由服务器转发至客户端，客户端为用户显示答复。

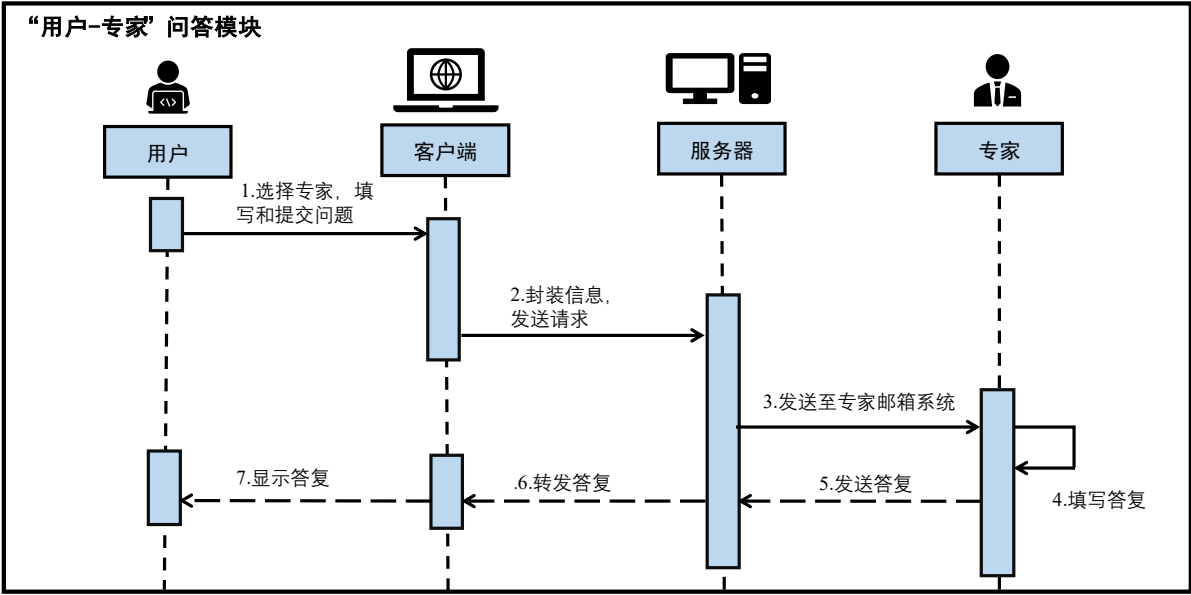


图4-3: “用户-专家”问答模块

4.3 问诊预约模块

表4-3: 问诊预约模块基本信息

基本信息	描述
功能	实现用户申请预约功能，专家回复

	功能
对象	用户与专家
输入(用户端)	用户预约申请，包括时间和描述
输入(专家端)	对用户申请的回复，可接受或拒绝
输出(用户端)	专家答复
客户端负责功能	<ul style="list-style-type: none"> <li>• 发送预约信息；</li> <li>• 接受响应信息；</li> <li>• 显示预约结果</li> </ul>
服务器负责功能	<ul style="list-style-type: none"> <li>• 接受和识别信息；</li> <li>• 发送至专家邮箱系统；</li> <li>• 转发回复；</li> </ul>

问诊预约的流程与之前的问答流程基本一致。需要指出，用户在填写预约信息时需要给出具体的时间约定以及对病状的简要描述。专家的处理预约时可以选择接受或者拒绝。图4-4为整个预约过程的时序图：

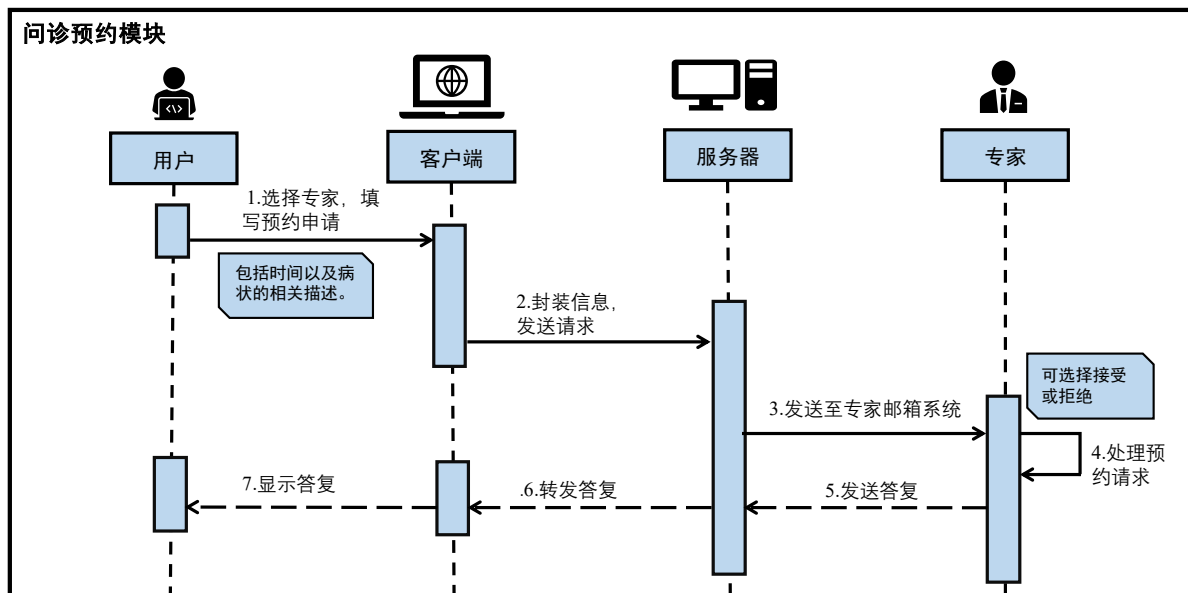


图4-4：问诊预约模块

#### 4.4 机器人问答模块



表4-5：机器人问答模块基本信息

基本信息	描述
功能	提供与医疗机器人进行交互的功能
对象	用户与专家
输入	医疗相关问题
输出	机器人的答复
客户端负责功能	<ul style="list-style-type: none"><li>• 发送相关问题；</li><li>• 接受响应信息；</li><li>• 显示机器人答复</li></ul>
服务器负责功能	<ul style="list-style-type: none"><li>• 接受和识别信息；</li><li>• 运行机器人模型得到结果；</li><li>• 发送答复结果；</li></ul>

医疗机器人智能问答业务为本系统的一大特色所在。我们假设当前已经有一个问答机器人可以直接实用(具体如何实现我们将在第6章叙述)，其在服务器端开始运行时就已经完成部署并成功载入相关语料库。当用户或专家在客户端发送问题后，服务器接收后运行机器人模型得到答复，并发送答复，客户端显示相应结果。下图为该流程的时序图：

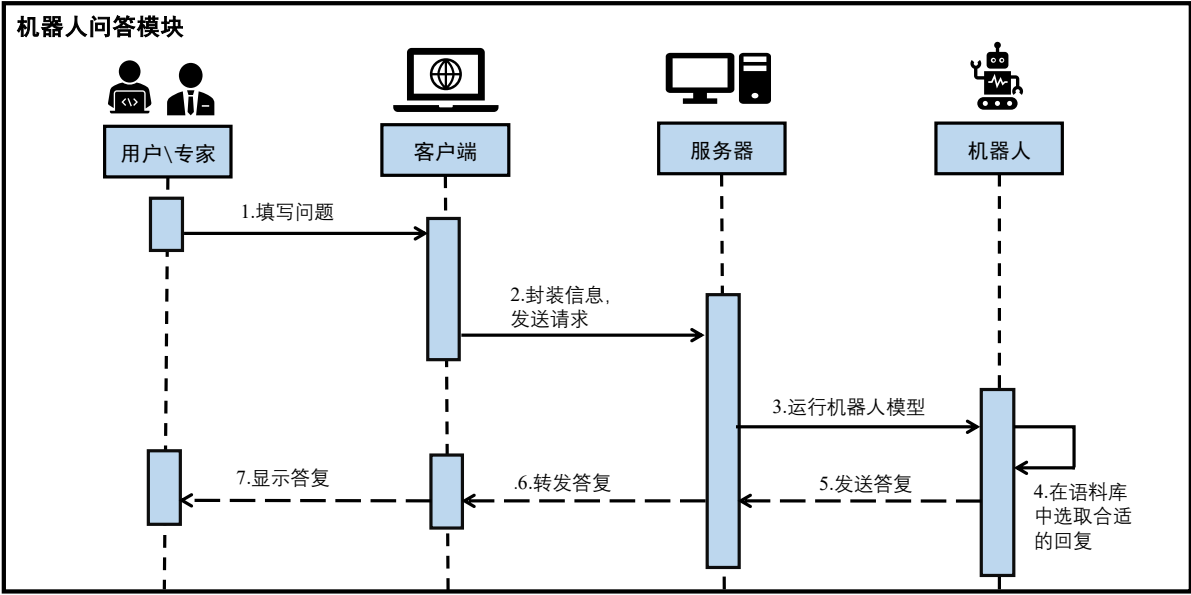


图4-5：机器人问答模块

4.5 信息管理模块

本系统的信息管理功能包括用户与专家对自身信息的完善和管理，也包括管理员对用户和专家的信息管理。具体功能如下表所示：

表4-6：信息管理模块的相关功能

对象	功能
用户	修改个人信息
	修改密码
专家	修改个人信息
	修改密码
客户端	审核专家
	添加用户或专家记录
	冻结用户或专家

	删除用户或专家
--	---------

需要指出，以上功能在具体实现中都表现为对数据库中记录的修改。我们以管理员删除直接信息为例。首先，管理员通过UI界面提供的操作按钮进行删除信息提交，服务器接收到请求后使用SQL语句向数据库中删除信息，之后发送相关数据提示删除成功或失败。

表4-7：管理员对专家\用户的信息删除模块基本信息

基本信息	描述
功能	实现管理员对专家\用户信息的删除操作
对象	管理员
输入	专家\用户信息的选中，删除按钮的点击
输出	删除结果
客户端负责功能	<ul style="list-style-type: none"> <li>• 发送删除请求；</li> <li>• 接受响应信息；</li> <li>• 显示响应结果；</li> </ul>
服务器负责功能	<ul style="list-style-type: none"> <li>• 接受和识别信息；</li> <li>• 发送SQL语句至数据库中；</li> <li>• 发送删除结果；</li> </ul>

下图展示了管理员对用户专家的删除功能具体实现过程。其他信息管理功能的实现也是类似的，这里不再赘述。

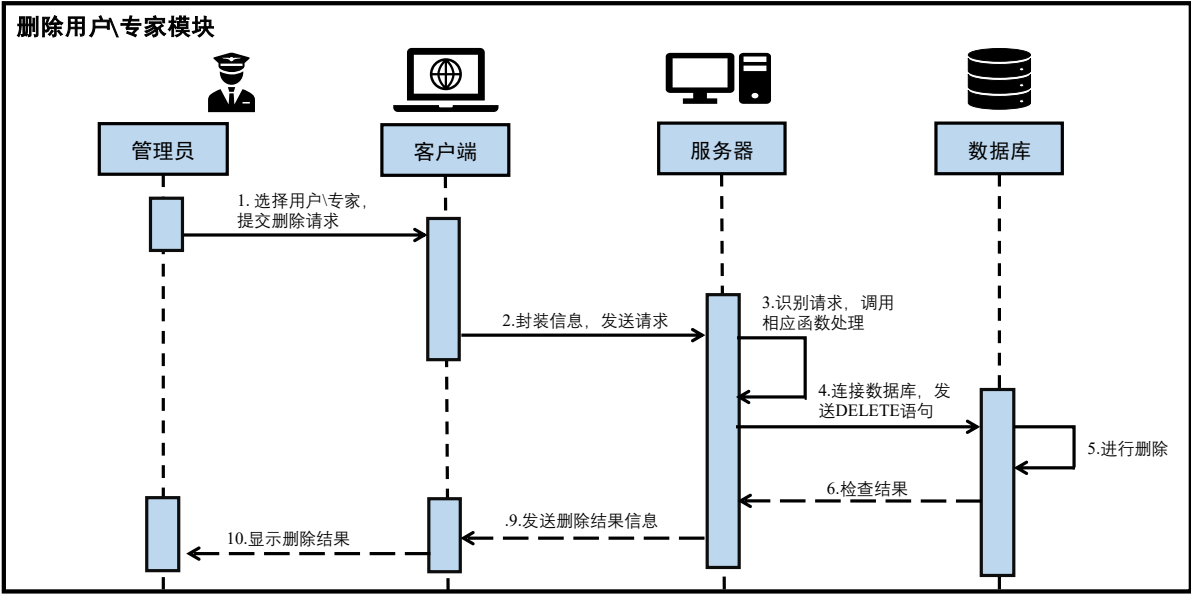


图4-6：管理员对用户\专家信息的删除

4.7 用户\专家邮箱模块

用户通过邮箱来参考自己已发出的问题和预约以及对应的结果，而专家则通过邮箱来查看发给自己的问题和预约。

表4-8：用户\专家邮箱模块基本信息

基本信息	描述
功能	对用户展示已发送的邮件和预约及其处理结果。对专家展示发送给自己的邮件和预约及其处理清空。
对象	用户\专家
输入	用户\专家邮箱界面刷新按钮的触发信号
输出	医疗问答信息或预约信息
客户端负责功能	<ul style="list-style-type: none"><li>• 提供刷新按钮</li><li>• 发送信息请求；</li><li>• 接受响应信息；</li><li>• 显示响应结果；</li></ul>

服务器负责功能	<ul style="list-style-type: none"><li>接受和识别信息；</li><li>发送SQL语句至数据库中；</li><li>发送医疗问答或预约信息表结果；</li></ul>
---------	--

以下为用户或专家刷新邮箱页面时的时序图。需要指出，用户和专家在登录时系统会自动发送邮箱信息和预约信息请求。

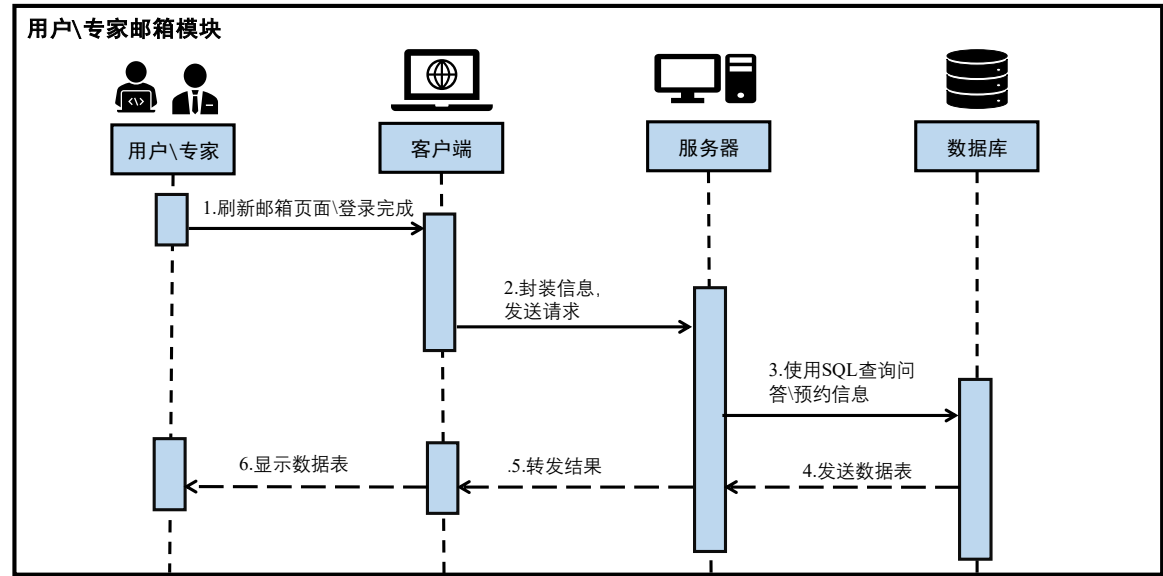


图4-7：用户\专家邮箱模块

5 数据库设计

5.1 数据库字段

根据本项目的具体业务。我们需要使用的表有用户表(User),专家表(Expert),管理员表(Administrator),医疗问答表(QA)以及问诊预约表(Appointment)。本节将展示本系统所使用的数据库表及其具体属性和相关设置。

1) 表User

表5-1：数据库表User属性

属性	类型	长度	非空	主键	自增
----	----	----	----	----	----

Id	INT	11	True	True	True
Username	VARCHAR	45	True	False	False
Password	VARCHAR	45	True	False	False
Email	VARCHAR	45	True	False	False
Sex	VARCHAR	45	True	False	False
Birthdate	DATE		True	False	False
Active	VARCHAR	45	True	False	False

## 2) 表Expert

表5-2: 数据库表Expert属性

属性	类型	长度	非空	主键	自增
Id	INT	11	True	True	True
Username	VARCHAR	45	True	False	False
Name	VARCHAR	45	True	False	False
Email	VARCHAR	45	True	False	False
Password	VARCHAR	45	True	False	False
Sex	VARCHAR	45	True	False	False

Birthdate	DATE		True	False	False
Department	VARCHAR	45	True	False	False
Score	DOUBLE		True	False	False
Active	VARCHAR	45	True	False	False

### 3) 表Administrator

表5-3：数据库表Administrator属性

属性	类型	长度	非空	主键	自增
Id	INT	11	True	True	True
Username	VARCHAR	45	True	False	False
Password	VARCHAR	45	True	False	False

### 4) 表QA

表5-5：数据库表QA属性

属性	类型	长度	非空	主键	自增
Id	INT	11	True	True	True
Userid	INT	11	True	False	False
Expertid	INT	11	True	False	False
Question	VARCHAR	200	True	False	False

Response	VARCHAR	200	True	False	False
The date	DATE		True	False	False
Score	DOUBLE		True	False	False

### 5) 表Appointment

表5-6: 数据库表Appointment属性

属性	类型	长度	非空	主键	自增
Id	INT	11	True	True	True
Userid	INT	11	True	False	False
Expertid	INT	11	True	False	False
Description	VARCHAR	200	True	False	False
The date	DATE		True	False	False
Result	VARCHAR	45	True	False	False

## 5.2 数据库关系图

基于前文的分析, 一个医疗问答(QA)中包含着提出问题的用户和回复问题的专家的用户名, 除此之外, 一个问诊预约(Appointment)也包含了一个发起预约的用户和处理预约的专家的用户名。因此这两个表中的用户名均为外键, 需要由User表和Expert表约束。数据表间关系如下:



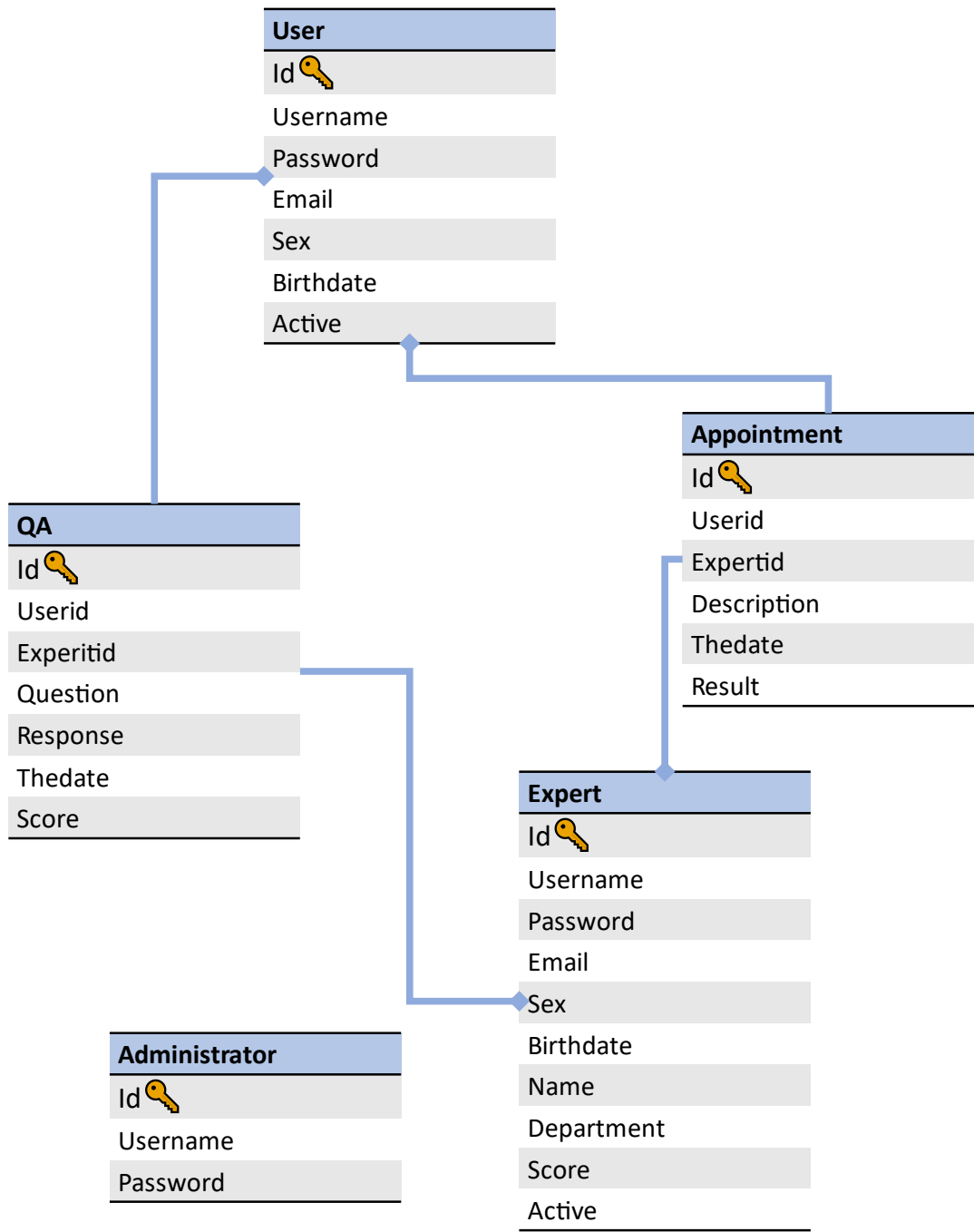


图5-1：数据库关系图

## 6 医疗问答机器人核心算法设计

### 6.1 问答机器人简介

从对话机器人的实现逻辑来看，可以分为检索式机器人和生成式机器人，检索式机器人需要首先建立好一个对话语料库，在接受到用户输入信息后，立

刻通过某些指定的算法进行句子相似度匹配，找出最合适的答句返回给用户，而生成式机器人则直接基于大型语料数据集训练深度生成模型，并基于训练好的模型的计算结果直接输出答句。对于问答机器人，需要针对特定领域的问题进行解答，因此我们希望严格控制回答的内容，基于语料库的检索式机器人比较符合这一点。而生存式机器人由于深度神经网络的难解释性，不适用于特定任务由于是医疗问答这种事关生命健康安全任务。因此本项目使用检索式的问答机器人实现方法。

## 6.2 Word2Vec简介

对于检索式机器人，其需要在语料库中选取“最为合适的答句”，一个基本的思路是将用户输入的问句与问答语料库中的问句一一进行比对，找到相似性最大的句子并输出对应的答句。因此核心问题是度量句子间的相似性，我们使用Word2Vec模型(词向量模型)来进行实现。

Word2Vec是自然语言处理领域的一种重要技术，用于将非结构化的，不可计算的文本信息转化为结构化的，可计算的向量。除此之外，关联度较高的两个词(Word2Vec的一个基本假设是在一段文本中位置相近的词关联度较高)，使用Word2Vec生成的向量在几何空间中会比较接近。

本项目语言模型的训练基于Gensim框架实现，由于Word2Vec仅需要提供一段文本即可无监督地进行训练，我们使用中文医疗对话语料库作为训练集，使用随机梯度下降算法进行训练。

## 6.3 语料库描述

本项目所用的中文医疗对话语料库为公开数据集，可直接在阿里云天池数据集(<https://tianchi.aliyun.com/dataset/dataDetail?dataId=90163>)下载。本项目主要将其用于Word2Vec模型的训练以及作为医疗问答机器人的基础对话库。

数据集提供了6个科室的相关医疗问答内容，总计大约有80万条对话信息。下表展示了其中的部分对话信息。

表6-1：中文医疗对话语料库部分内容展示

科室	问题	回答
----	----	----

儿科	小儿肥胖怎样治疗？	当孩子患上肥胖症的时候家长可以增加孩子的运动量和控制他的饮食来改变症状，给孩子在承受范围内安排孩子游泳，慢跑等运动，并且多吃一些蔬菜，像西红柿，黄瓜等，不能给孩子吃像夏威夷果，曲奇饼干等高热量的零食，要帮助孩子改正吃完不喜欢动，或者暴饮暴食的习惯，在治疗小儿肥胖期间如果情况严重，建议家长先带孩子去医院检查一下孩子肥胖症的原因在针对性的治疗。
儿科	出现扁桃体发炎的小孩会有哪些症状	当孩子扁桃体出现炎症的时候，他的咽喉部和耳朵均有明显的疼痛感，影响正常吞咽，孩子伴有一定发烧的情况，严重的话可烧到40摄氏度，要重视，而且孩子还会有乏力，头疼乃至全身痛，以及寒战等表现，此时扁桃体也是处于红肿、充血的状态，还可能出现一些严重的并发症，比如像肾炎啊，败血症这类的，直接威胁生命，可见儿童扁桃体炎是非常严重的一个疾病，甚至可能威胁到孩子的生命，家长务必要及时带孩子治疗。
内科	高血压该治疗什么？	高血压患者首先要注意控制食盐摄入量，每天不超过六克，注意不要吃太油腻的食物，多吃新鲜的绿色蔬菜水果，多吃有机食物，注意增强体育锻炼，增加身体素质，同时压力不要过大，精神不要紧张，效果不佳的话，可以积极配合以降压药物控制血压治疗，情绪平时保持平和，不易激动。

在当前版本开放的对话机器人科室为儿科。我们将以上数据集作为本系统智能对话机器人的基础语料库，并在软件经营用户与专家问答业务时不断根据评分和专家建议来扩大语料库规模，提高智能机器人的性能。

## 6.4 句子相似度计算方法

使用Word2Vec处理后，所有的单词都变成结构化、可计算的向量。因此句

子相似度的计算转化为对向量相似度的计算。如下伪代码展示了对话机器人基于Word2Vec模型计算句子相似度并搜索最优答句的过程。由于Word2Vec是对单词进行处理，我们引入Jieba分词工具将句子分为单词，之后再每个单词用Word2Vec转化为向量，并累加作为该句子的句向量。句子之间的相似度利用余弦相似度进行度量。

---

**算法** 对话机器人搜索最合适的答句

---

**Input:** 客户发送的问句 $q^*$ , 机器人对话语料库 $D$ , 训练好的Word2Vec模型 $f$ , 词向量维数 $n$

**Output:** 最合适的答句 $a^*$

```

1:  导入Jieba分词工具;
2:  初始化最大相似度 $c^* = 0$ , 最合适答句 $a^* = null$ , 问句句向量 $v$ 为 $n$ 维0向量;
3:  使用分词工具对 $q^*$ 进行分词, 得到字符串数组记为 $q'$ ;
4:  for  $w$  in  $q'$ 
5:       $v = v + f(w)$  ; // 将句子中每个词的向量进行累加作为句子的向量
6:  end for
8:  for  $q, a$  in  $D$  //从数据集逐个中取出问句 $q$ 和答句 $a$ 
9:      使用分词工具对 $q$ 进行分词, 分词后得到字符串数组记为 $\hat{q}$ ;
10:     设 $v'$ 为 $n$ 维0向量; ;
11:     for  $w$  in  $\hat{q}$ 
12:          $v' = v' + f(w)$ ; // 将句子中每个词的向量进行累加作为句子的向量
13:     end for
14:      $c = \frac{\langle v', v \rangle}{\|v\| \|v'\|}$ ; //基于余弦值度量句子相似度。
15:     if  $c > c^*$ 
16:          $c^* = c$ ;
17:          $a^* = a$ ;
16:     end if
17: end for
18: return  $a^*$ 

```

---

## 7 界面设计

本章我们将阐述我们的预期界面设计风格以及部分核心模块对应UI界面的预期框架设计。

## 7.1 界面设计总体规定

本系统的风格以简洁清新为主要设计风格。并且考虑默认模式与夜间模式之间的切换，方便用户夜间使用。默认模式以白色为主色，边框部分使用青色，夜间模型则主色为黑色。操作按钮的摆放应当清晰合理，方便用户快速上手。

## 7.2 登录与注册界面设计

### 7.2.1 登录界面设计

图7-1展示了登录界面的设计图，客户在进行登录时需要输入用户名，密码并选择身份(用户、专家或管理员)。除此之外，界面是左下角为风格切换按钮，当前初代版本仅提供默认模式和夜间模式的切换。右下角为返回登录界面按钮，该按钮做为进入其他页面后返回键。

登录界面设计

智能医疗问答系统

用户名

密码

用户 ▼

登录 注册

☐ 默认模式 ☐ 夜间模式

返回登录界面

图7-1：登录界面设计

### 7.2.2 注册界面设计

**注册界面设计(用户)**

☐ 用户注册    ☐ 专家入驻

用户名

密码

确认密码

邮箱

性别

出生日期

一键注册

☐ 默认模式    ☐ 夜间模式

返回登录界面

图7-2: 注册界面设计(用户)

图7-2和图7-1展示了注册页面的预期设计图。用户在进行注册时需要提供的信息有用户名、密码(重复确认一次)、邮箱、性别以及出生日期。而专家则提供用户名、真实姓名、密码(重复确认一次)、邮箱、性别和出生日期。页面上有“用户注册”和“专家入驻”之间的切换按钮。注册后,可使用返回登录界面按钮返回并进行登录。

**注册界面设计(专家)**

☐ 用户注册    ☐ 专家入驻

用户名

真实姓名

密码

确认密码

邮箱

性别

出生日期

科室

一键注册

☐ 默认模式    ☐ 夜间模式

返回登录界面

图7-3: 注册界面设计(专家)

## 7.3 用户客户端界面设计

### 7.3.1 用户提问界面介绍

图7-4展示了用户提问界面的设计,用户可以在此界面实现对专家的提问。提供上方的表格框可以查看专家信息,用户在进行提问时,可以先提供关键词

搜索，比如可以在关键词类型中选中姓名，之后输入专家姓名来快速查找。在表格中选中专家后，并在下方输入框中输入问题点击发送则完成提交。

用户提问界面设计

专家相关信息显示，点击可选中信息

医疗相关问题输入框

关键词类型 V

输入关键词

点击

发送

重置

☐默认模式 ☐夜间模式

返回登录界面

图7-4：用户提问界面设计

### 7.3.2 用户问诊预约界面设计

图7-5展示了用户问诊预约的设计界面，与问答界面不同的是，进行问诊是需要提交预约日期。其他相关操作包括查找和提交与问答界面类似，这里不再赘述。

用户问诊预约界面设计

专家相关信息显示，点击可选中信息

病状描述

关键词类型 V

输入关键词

点击搜索

日期 V

发送

重置

☐默认模式 ☐夜间模式

返回登录界面

图7-5：用户问诊预约界面设计

## 7.4 专家客户端界面设计

**专家处理用户提问界面设计**

与专家相关的用户提问信息显示，点击可选中

此处用于填写回复

回复

重置

☐默认模式 ☐夜间模式

返回登录界面

图7-6：专家处理用户提问界面设计

对应于用户的提问界面和问诊预约界面，专家客户端设置了答复界面。图7-6展示了专家处理用户提问界面的设计图，可以在用户提问信息中选中并在下框中填写答复。图7-7展示了专家处理问诊预约界面的设计，专家只需在表格中选中响应的设计，并根据右端按钮选中接收或拒绝此次预约。

**专家处理问诊预约界面设计**

与专家相关的问诊预约信息显示，点击可选中

接收预约

拒绝预约

☐默认模式 ☐夜间模式

返回登录界面

图7-7：专家处理问诊预约界面设计

## 7.5 管理员客户端界面设计



用户信息管理界面设计

用户名

密码

确认密码

邮箱

性别

▼

出生日期

▼

添加

删除用户

激活用户

冻结用户

重置密码

☐默认模式

☐夜间模式

返回登录界面

用户信息展示表格

图7-8：用户信息管理界面设计

管理员客户端的核心功能在于对用户信息和专家信息的管理，图7-8展示了用户信息管理的界面预想图。管理员可以在用户信息展示表格中查看和选中响应的用户，通过下方按钮进行删除、激活、冻结和重置密码等操作。除此之外，管理员可以直接在界面中添加用户信息。图7-9展示了专家的信息管理界面，其设置与用户信息管理界面相似。另外，管理员对用户和专家信息的修改需要实时地反映在表格中，因此在每次修改后都需要自动进行表格信息的刷新。

专家信息管理界面设计

删除专家

激活专家

冻结专家

重置密码

用户名

密码

邮箱

出生日期

▼

性别

▼

真实姓名

确认密码

所在科室

添加专家

☐默认模式

☐夜间模式

返回登录界面

专家信息展示表格

图7-9：专家信息管理界面设计

7.6 机器人交互界面设计

图7-10展示了机器人交互界面的设计，专家和用户可在此界面与机器人交了

医疗问题。机器人回对提问进行实时答复。页面右上角的空缺地方可以用于放置机器人形象。

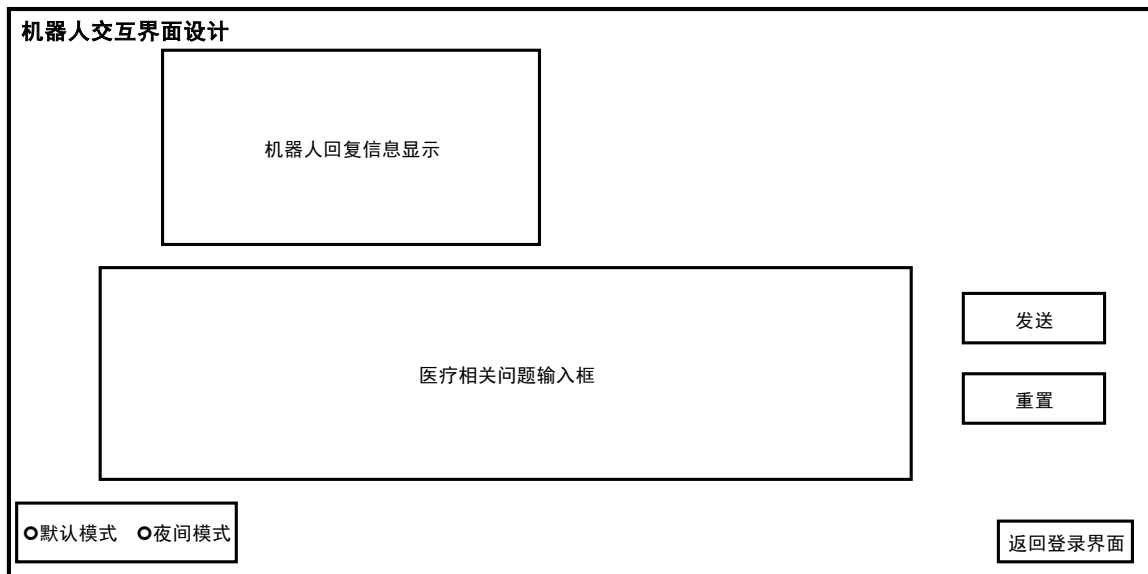


图7-10：机器人交互界面设计

## 7.7 用户/专家邮箱界面设计

每一个用户和专家都在系统中有一个用于接收信息的邮箱模块，用户可以在用户邮箱中查看自己发送的问题和预约申请以及对应的回复结果，专家可以在邮箱中查看发送给自己的问题和预约申请。图7-11展示了用户和专家的邮件界面设计。



图7-11：用户/专家邮件界面设计

## 8 系统出错信息处理与安全设计

### 8.1 主要出错信息处理设计

由于用户和专家并不一定能快速上手我们的系统，因此出现错误操作是合理的。开放小组在设计阶段预计了部分可能客户的出错信息，并有针对性地进行处理功能的相关设计，下表展示了部分错误以及处理方式：

表8-1：出错信息及其处理手段

错误描述	处理手段
注册时用户名为空	在客户端检测，并提示具体错误
注册时密码为空	在客户端检测，并提示具体错误
注册时两次密码输入不同	在客户端检测，并提示具体错误
注册时输入邮箱格式无效	在客户端检测，并提示具体错误
密码长度不超过8位	在客户端检测，并提示具体错误
登录时输入不存在用户名	在服务器端检测，客户端提示具体错误
注册时输入已注册用户名\邮箱	在服务器端检测，客户端提示具体错误
登录时用户名为空	在服务器端检测，客户端提示具体错误
登录时输入用户名不存在	在服务器端检测，客户端提示具体错误
登录时密码错误	在服务器端检测，客户端提示具体错误

向专家提交问题时输入为空	在客户端检测，并提示具体错误
向机器人提交问题时输入为空	在客户端检测，并提示具体错误
提交预约时未填写日期	在客户端检测，并提示具体错误
提交预约时未填写病状描述	在客户端检测，并提示具体错误
已冻结用户\专家执行登录	在服务器检测，拒绝登录

注：管理员添加用户/专家信息时的易错问题与注册时的问题相同，这里不再赘述。

## 8.2 安全保密性设计

### 8.2.1 MD5加密技术

本系统中对于用户、专家、管理员的密码均使用MD5加密技术处理。该技术在客户端实现，且存入数据库时使用机密后的编码进行存储，登录时比对密码也是使用加密后编码进行比对。

### 8.2.2 可信管理员

由于管理员对整个系统具有最高控制权和管理权，能够重置用户和专家密码，能够删除用户和专家信息。因此系统在实际应用时需要选拔最为可信的管理员。

### 8.2.3 数据库安全性设计

考虑到医疗业务事关生命健康安全，数据库中存储着大量的用户私密信息，考虑到可能存在黑客入侵，系统崩溃等突发事件，通过设置数据库备份的手段，预防数据丢失。