# 小豆在线问答机器人项目logo

**项目成员（IdolGroup）：**

叶春毅

周世奉

康子博

王一汐

黄璐洁

孟凡月

## 1. 项目背景及意义

**EPG版本历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **说明** | **作者** |
| 2018.9.30 | 0.1 | 搜集在线问答机器人背景 | 叶春毅 |
| 2018.10.2 | 0.2 | 重新整理项目被背景 | 叶春毅 |
| 2018.10.20 | 0.8 | 撰写项目背景及意义初稿 | 康子博 |
| 2018.11.12 | 1.0 | 改进文本说明 | 叶春毅 |
|  |  |  |  |

**EPG审批记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **审批人** | **审批意见** | **审批日期** |
| 0.2 | 康子博 | 通过 | 2018.10.2 |
| 0.8 | 叶春毅 | 部分叙述可进行修改 | 2018.10.21 |
| 1.0 | 周世奉 | 通过 | 2018.11.12 |
|  |  |  |  |

#### (1) 项目背景

目前基于人工智能训练的机器人正在越来越广泛的应用于各种互联网服务，如客户服务或资讯获取。部分聊天机器人搭载了自然语言处理系统，应用了人工智能和云计算技术，如苹果公司IOS平台上的SIRI助手集成了搜索，知识库，问答推荐，语言模型等多种技术。这一类集成在系统上的智能机器人已经逐渐成为人们生活的助手，正被大众所接受和使用。反观许多公司官方网站的在线客服机器人，大多只能从消息中提取相应的关键字关键字，从数据库寻找合适的回答语句。遇到不能处理的情况，往往转接人工服务来处理。这给问题的处理效率造成了很大的限制用户的体验。

目前的聊天机器人根据商业应用目的不同，可以分为两种：纯粹以陪客户聊天打发时间为目的的娱乐型机器人、帮助用户获取所需信息和知识的信息获取型机器人。对于前者来说，它的唯一目的就是陪聊，在聊天的过程中，例如：Simsimi机器人会讲出各种卖萌、搞笑的话语，以“语出惊人”来娱乐用户。而对于后者，一般来说，用户对它的功能要求更为严格，主要表现在：机器人提供信息的相关性、可靠性和准确性，有时，用户可能还要求机器人去直接完成某件任务：如：医院挂号、订购商品、接通某位朋友的电话等等。

1960年外国的科学家就提出“计算机要能用自然语言与人沟通”的设想。随着科学技术的发展和移动互联网的日新月异，各大互联网公司推出了自己的智能聊天机器人，替代人工面对庞大的客服需求，使得聊天机器人成为人们关注的方向之一。国外聊天机器人的发展比较早，目前已经有很多成熟的产品，比如苹果的Siri, 用户可以给Siri发布定时、打开应用等指令。

微软的小冰可以与用户进行有趣的聊天。还有Google的Alice，IBM的Watson等。聊天机器人在中国也有广泛的应用，比如许多微信群，淘宝的客服，百度的度秘，科大讯飞的语音助手等都能提供智能聊天服务，极大地减少了人工成本。但是相较于国外，发展还是相对落后，主要原因是中文比较难处理，语料库的积累也比较少。在机器人的主要设计技术方面，目前比较主流的是基于人工模板和人工智能的两种聊天机器人。基于人工模板的聊天机器人制作最简单，主要是通过人工录入一些对话的模板，当有用户提问时，会从这些模板中找出匹配的答案返回给用户。这种技术的优点是开发简单，返回答案速度快，缺点是答案单一，人工成本高。基于人工智能的聊天机器人是利用一些算法来生成答案，答案的准确度取决于算法的好坏。这种聊天机器人有自我学习的能力，随着聊天次数的增多，答案也会越来越准确。这种聊天机器人的优点是前期不需要人工输入大量模板，答案较灵活，缺点是开发难度大，模型训练复杂。这种聊天机器人制作的主要难点在于语料库的积累，规则的匹配。

#### (2) 项目意义

而本项目主要以与用户互动为主要目的，在此基础上再加以一些信息获取功能，将充分应用人工智能技术对不同需求进行可定制化的问答机器人服务，将对目前在线客服类的机器人带来智能化的体验。

问答系统一直是自然语言处理领域中非常热门的一个研究方向。人们可以向本系统提交用自然语言表达的问题，系统会返回紧凑相关又准确的答案，而不是像搜索引擎那样仅仅返回一个网页集合。本系统的意义是为问题找到精确的答案，而不是信息检索那样返回全文档或者最匹配的文章。更进一步，本系统不仅可以回答用户的某一个问题，还可以很人性化地与用户交流，了解用户的需求，就像用户的一个好朋友。

随着信息技术的发展，计算资源和存储资源在21世纪都得到了极大的进步，但人机交互方式并没有发生多大变化，尤其是用户获取信息的方法还不够有效。通过研究聊天机器人相关技术，尤其是聊天机器人技术中的对话理解问题，对于促进人机交互方式的发展有着重要的意义。

## 2. 项目可行性分析

**EPG版本历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **说明** | **作者** |
| 2018.10.1 | 0.1 | 寻找实现技术可行方案 | 周世奉 |
| 2018.10.3 | 0.2 | 进行经济可行性分析 | 周世奉 |
| 2018.10.10 | 0.3 | 社会可行性分析 | 王一汐 |
| 2018.10.15 | 0.8 | 撰写项目可行性分析文本 | 周侍奉 |
| 2018.10.20 | 1.0 | 梳理文档 | 叶春毅 |

**EPG审批记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **审批人** | **审批意见** | **审批日期** |
| 0.1 | 叶春毅 | 需要进一步讨论技术可行性 | 2018.10.2 |
| 0.2 | 叶春毅 | 通过 | 2018.10.4 |
| 0.3 | 叶春毅 | 通过 | 2018.10.10 |
| 0.8 | 叶春毅 | 部分技术描述过多，需结合项目说明 | 2018.10.15 |
| 1.0 | 王一汐 | 通过 | 2018.11.1 |

#### (1) 经济可行性分析

由于本项目开发人员都是本小组的成员，项目投入的人力不多。目前项目主要投入在本项目的网络建站上（服务器主机，域名），阿里云的优惠活动也一定程度上降低了投入成本。开源框架和协议给我们的开发带来了诸多便利。在进行大规模的语料训练时，需要使用较好的硬件资源做为支撑。因此，硬件资源成本和人力成本为本项目的主要成本。

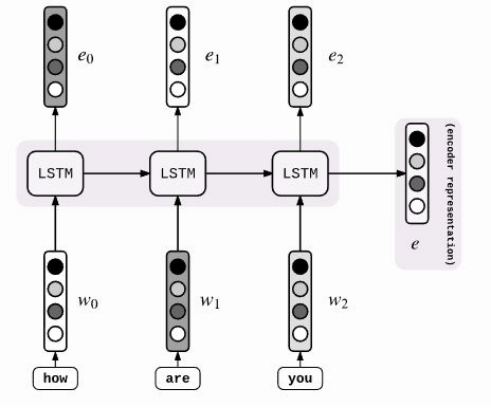
将项目进行推广和应用时，只需按照客户的需求进行定制化的改进，即可运用到不同的智能客服问答场景。项目可进行长期的维护工作，具有可观的经济收益。同时，在实现智能问答功能的同时，可附加部分互联网推广服务，进一步补充收益。

#### (2) 技术可行性分析

##### Seq2Seq模型

Seq2Seq技术，全称Sequence to Sequence，该技术突破了传统的固定大小输入问题框架，开通了将经典深度神经网络模型（DNNs）运用于翻译与智能问答这一类序列型（Sequence Based，项目间有固定的先后关系）任务的先河，并被证实在英语－法语翻译、英语－德语翻译以及人机短问快答的应用中有着不俗的表现。Seq2Seq解决问题的主要思路是通过深度神经网络模型（常用的是LSTM，长短记忆网络，一种循环神经网络）将一个作为输入的序列映射为一个作为输出的序列，这一过程由编码输入与解码输出两个环节组成。

本项目利用这一模型搭建中文对话机器人。



##### TensroFlow

TensorFlow™是一个基于数据流编程（dataflow programming）的符号数学系统，被广泛应用于各类机器学习（machine learning）算法的编程实现，其前身是谷歌的神经网络算法库DistBelief。Tensorflow拥有多层级结构，可部署于各类服务器、PC终端和网页并支持GPU和TPU高性能数值计算，被广泛应用于谷歌内部的产品开发和各领域的科学研究 。

TensorFlow由谷歌人工智能团队谷歌大脑（Google Brain）开发和维护，拥有包括TensorFlow Hub、TensorFlow Lite、TensorFlow Research Cloud在内的多个项目以及各类应用程序接口（Application Programming Interface, API）。自2015年11月9日起，TensorFlow依据阿帕奇授权协议（Apache 2.0 open source license）开放源代码。

本项目将这一技术运用于Seq2Seq模型的训练和测试。

##### AJAX

Ajax 即“Asynchronous Javascript And XML”（异步 JavaScript 和 XML），是指一种创建交互式网页应用的网页开发技术。Ajax = 异JavaScript 和 XML（标准通用标记语言的子集）。Ajax 是一种用于创建快速动态网页的技术。Ajax 是一种在无需重新加载整个网页的情况下，能够更新部分网页的技术。通过在后台与服务器进行少量数据交换，Ajax 可以使网页实现异步更新。这意味着可以在不重新加载整个网页的情况下，对网页的某部分进行更新。传统的网页不使用 Ajax，如果需要更新内容，必须重载整个网页页面。

本项目主要将这一技术运用于消息和数据库前后端交互。

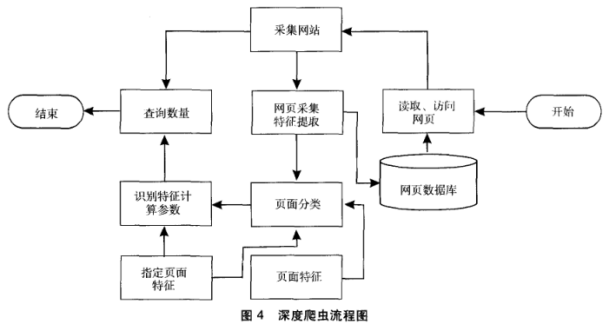


Ajax框架结构图

##### 网络爬虫

网络爬虫（又被称为网页蜘蛛，网络机器人，在FOAF社区中间，更经常的称为网页追逐者），是一种按照一定的规则，自动地抓取万维网信息的程序或者脚本。另外一些不常使用的名字还有蚂蚁、自动索引、模拟程序或者蠕虫。

本项目中其被用于对话中指定关键字的处理，实现信息的准确比对搜索。



网络爬虫原理图

#### (3) 社会可行性分析

问答机器人从简单的脚本化答案机器人走过了漫长的路。现在的问答机器人智能得足以吸引潜在客户，并确保人类客户服务和销售代表商不会浪费时间、资源等等。

智能问答机器人，是一种通过人类的语言同用户进行交流和对话的智能系统，目前广泛应用于诸多领域，如在银行的在线客服领域，客服机器人可以为客户解答一些简单的业务问题；在电商领域，机器人可以为客户推荐商品，为食客寻找美食，为游客查询最佳游玩路线等等。一款优秀的问答机器人不只是仅仅回答用户提出的某个问题，还可以进一步和用户进行人性化的交流，它能够像朋友一样理解用户的需求，按FaceBook的设想，在未来，问答机器人完全可能取代企业客服和各类APP软件。

目前的问答机器人根据商业应用目的不同，可以分为两种：纯粹以陪用户聊天打发时间为目的的娱乐型机器人、帮助 用户获得所需信息和知识的信息获取型的问答机器人。对于前者来说，它的唯一目的就是陪聊，在聊天的过程中，例如：Simsimi机器人会讲出各种卖萌、搞笑的话语，以“语出惊人”来娱乐用户。而对于后者，一般来说，用户对它的功能要求更为严格，主要表现在：机器人提供信息的相关性、可靠性和准确性。

## 3. 概要设计

**EPG版本历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **说明** | **作者** |
| 2018.10.15 | 0.3 | 撰写系统简述 | 叶春毅 |
| 2018.10.21 | 0.4 | 完成系统项目服务器配置 | 叶春毅 |
| 2018.10.22 | 0.6 | 列举4个项目项目目标 | 周世奉 |
| 2018.11.10 | 1.0 | 补充客服聊天室和管理端目标 | 叶春毅 |

**EPG审批记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **审批人** | **审批意见** | **审批日期** |
| 0.3 | 王一汐 | 通过 | 2018.10.15 |
| 0.4 | 周世奉 | 通过 | 2018.10.21 |
| 0.6 | 叶春毅 | 通过 | 2018.10.22 |
| 1.0 | 周世奉 | 通过 | 2018.11.12 |

#### (1) 系统简述

本项目拟建立一个以Seq2Seq语料训练模型为基础，辅以爬虫和API调用的机器人对话服务系统。针对不同需求进行可定制化的语料训练和关键词捕捉，拓展在线问答形式的功能性和服务性。同时进一步建立基于网页的网站后台管理系统，对访客进行管理，向客户提供客服服务。客服服务的形式以在线聊天室的形式呈现。

#### (2) 项目目标

主要实现如下功能：

①开发WEB和微信小程序前端

②实现天气查询的爬虫功能

③架设后台服务器，完善数据库和网站后台

④将Seq2Seq的模型和爬虫、第三方API调用的功能进行封装，形成独立的API

⑤构建客服在线聊天室

⑥在WEB管理端实现对访客和聊天室消息的基本管理

## 4. 详细设计

**EPG版本历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **说明** | **作者** |
| 2018.10.7 | 0.2 | 绘制系统数据流图 | 叶春毅 |
| 2018.10.7 | 0.3 | 介绍数据流图 | 康子博 |
| 2018.10.29 | 0.5 | 绘制系统数据库ER图 | 叶春毅 |
| 2018.10.30 | 0.6 | 介绍系统数据库结构 | 叶春毅 |
| 2018.10.11.5 | 1.0 | 补充系统需求 | 周世奉 |

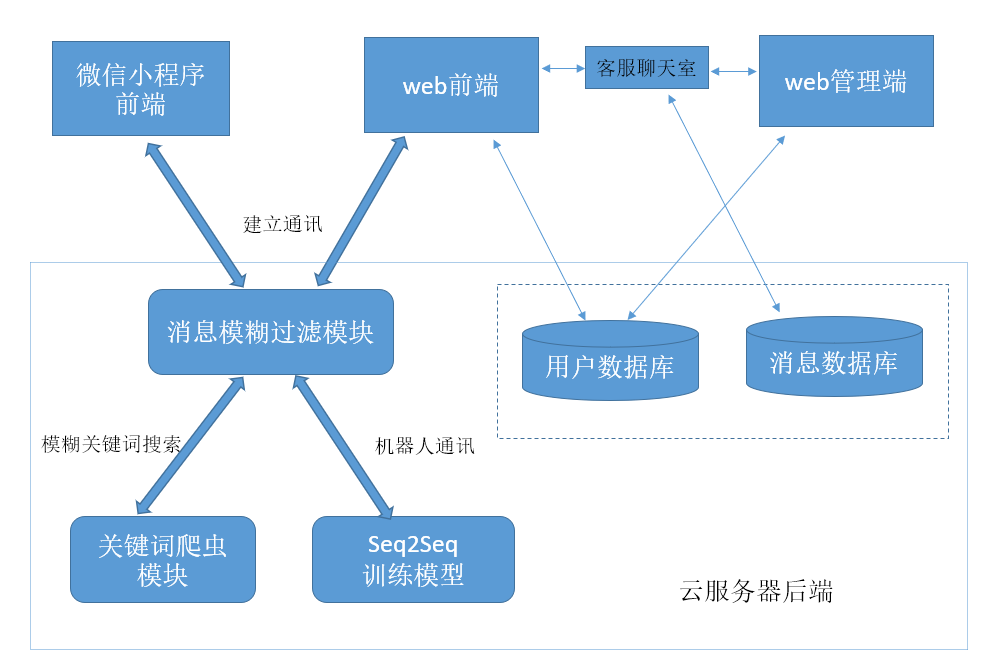
**EPG审批记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **审批人** | **审批意见** | **审批日期** |
| 0.2 | 康子博 | 可根据项目进程进一步改进数据流图 | 2018.10.8 |
| 0.6 | 康子博 | 通过 | 2018.10.30 |
| 1.0 | 康子博 | 通过 | 2018.11.5 |
|  |  |  |  |

#### (1)系统综合需求

对在线客服系统进行智能化处理，其核心为Tenorflow框架下建立的Seq2Seq模型，针对不同应用问答为访问者建环境下问答语料库进行模型训练。在此问答模型的基础上，利用模糊过滤技术，对前端的消息进行筛选，对部分关键词进行爬虫、第三方API调用处理，进一步扩展机器人的功能。系统数据库方面可讲数据库分为对用户信息的管理，以及对关键字查询的到的内容的存储。后者的主要目的是加快机器人消息响应，以提高系统对多用户并发性的性能。同时利用Ajax技术实现在线客服聊天系统

#### (2)数据流图介绍

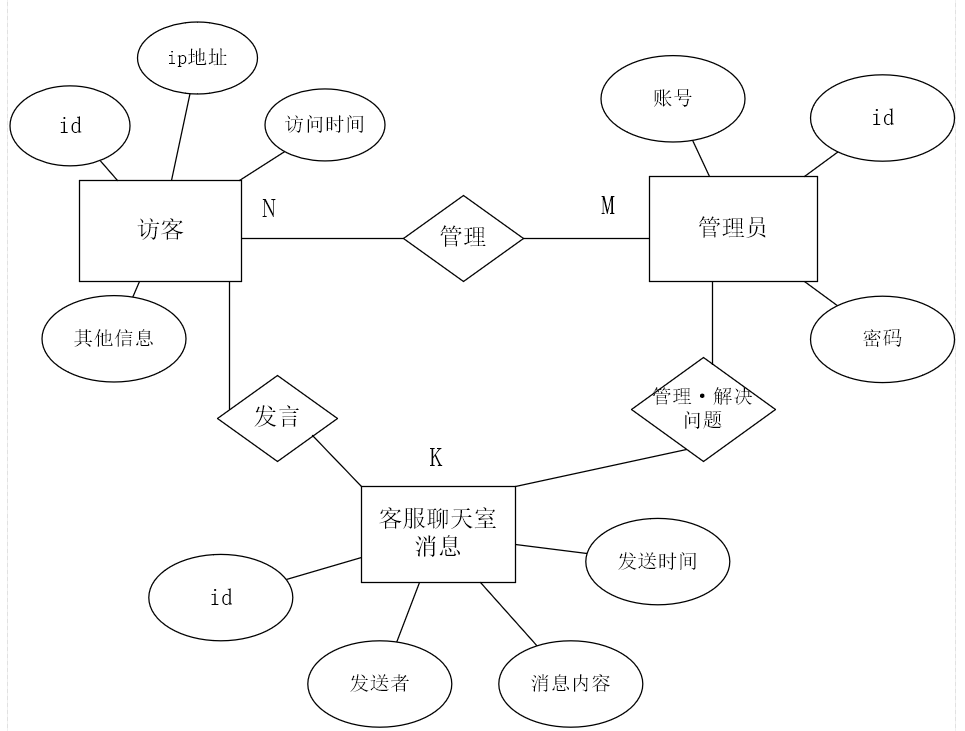


系统数据流图

项目的数据流图如上图所示。系统的前端，即用户对话界面采取的是web前端和微信小程序前端。微信小程序前端主要考虑到移动用户访问的便捷性。web管理端主要为管理员提供用户管理功能，向后台数据库进行基本的信息管理功能。同时针对对话情景中机器人不能解答的部分，访客可进入客服聊天室向其他在线访客或者管理员寻求帮助和回答。客服聊天室记录的对话消息由消息数据库记录。

前端的消息通过ajax框架和服务器后端进行通讯。来自web前端和微信小程序前端的语句文本由消息模糊过滤模块进行接收和处理，带有制定关键词的消息调用爬虫爬取相关信息，并返回处理过的消息。而基本的对话功能的消息则于训练的模型进行通信。

#### (3) 数据库设计



系统数据库 E-R图

管理员通过账号和密码在管理员端登入WEB管理端，对访客和客服聊天室消息进行管理，为客服聊天室中的访客解答问题。

客服聊天室有专有的数据表记录对话内容，记录访客和管理员的对话信息，消息中记录了发送者、消息内容和发送时间。

访客的信息被记录在访客数据库中，记录了访客的ip地址、访问时间以及其他信息。

本系统对访客与机器人的对话消息暂不予以记录，日后可进行功能的扩展。

## 5. 系统界面功能设计

**EPG版本历史**

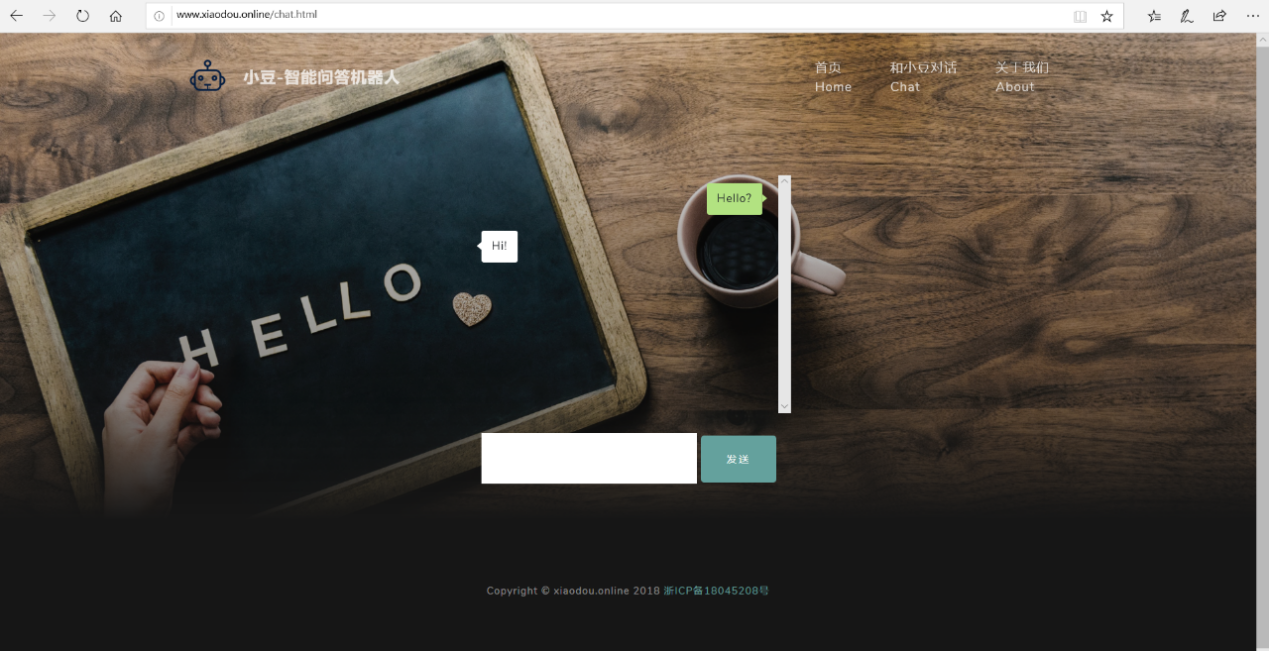
|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **说明** | **作者** |
| 2018.9.30 | 0.1 | WEB端基础框架 | 王一汐 |
| 2018.10.7 | 0.3 | WEB端页面改进 | 王一汐 |
| 2018.10.15 | 0.4 | 微信小程序开发 | 黄璐洁 |
| 2018.10.20 | 0.6 | WEB管理端开发 | 王一汐 |
| 2018.11.10 | 1.0 | WEB端页面信息调整，增加客服聊天室页 | 王一汐 |

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **审批人** | **审批意见** | **审批日期** |
| 0.1 | 叶春毅 | 通过 | 2018.10.1 |
| 0.3 | 康子博 | 网页打开速度较慢 | 2018.10.7 |
| 0.4 | 叶春毅 | 小程序后台功能进一步改进 | 2018.10.15 |
| 1.0 | 叶春毅 | 通过 | 2018.11.12 |

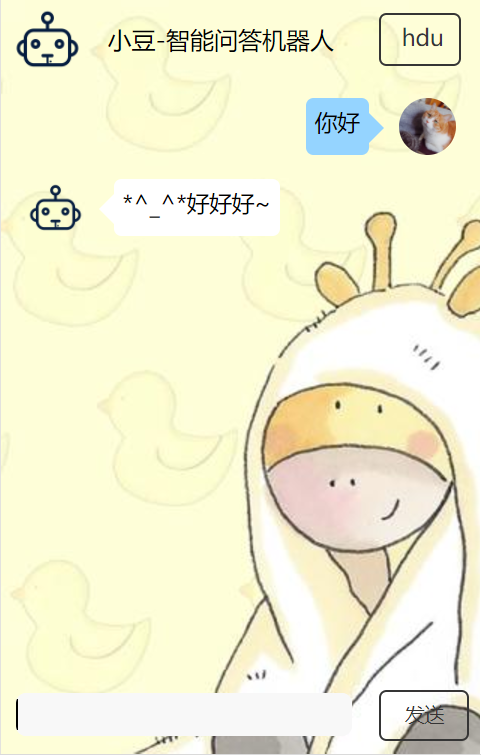
**EPG审批记录**

#### Web前端和微信小程序设计

小豆主页网址www.xiaodou.online



以上为小豆聊天机器人网页版的对话界面，整个主页分为首页，对话，信息三个不部分。



以上为小豆聊天机器人微信小程序版的对话界面，右上角附有HDU官网链接按钮，可在输入框中与后台进行通讯。

#### 管理端设计

管理员端网址www.xiaodou.online/admin



管理端分为三个子页面，分别为基本信息、用户对话和访客管理。用户对话分页是管理员进入聊天室的子页面，管理员在访客管理页实现对系统访客和客服聊天室消息的管理。

## 6. 项目功能测试

**EPG版本历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **说明** | **作者** |
| 2018.10.15 | 0.1 | 基本对话功能，调用后台 | 周世奉 |
| 2018.10.20 | 0.4 | 管理员端获取数据端查询信息 | 周侍奉、叶春毅 |
| 2018.10.30 | 0.6 | 完善管理员端功能 | 叶春毅 |
| 2018.11.15 | 0.8 | 上线客服聊天室功能 | 叶春毅 |
| 2018.11.20 | 0.9 | 接入模糊过滤和天气爬虫 | 康子博 |
| 2018.11.29 | 1.0 | 整体功能测试 | 孟凡月 |

**EPG审批记录**

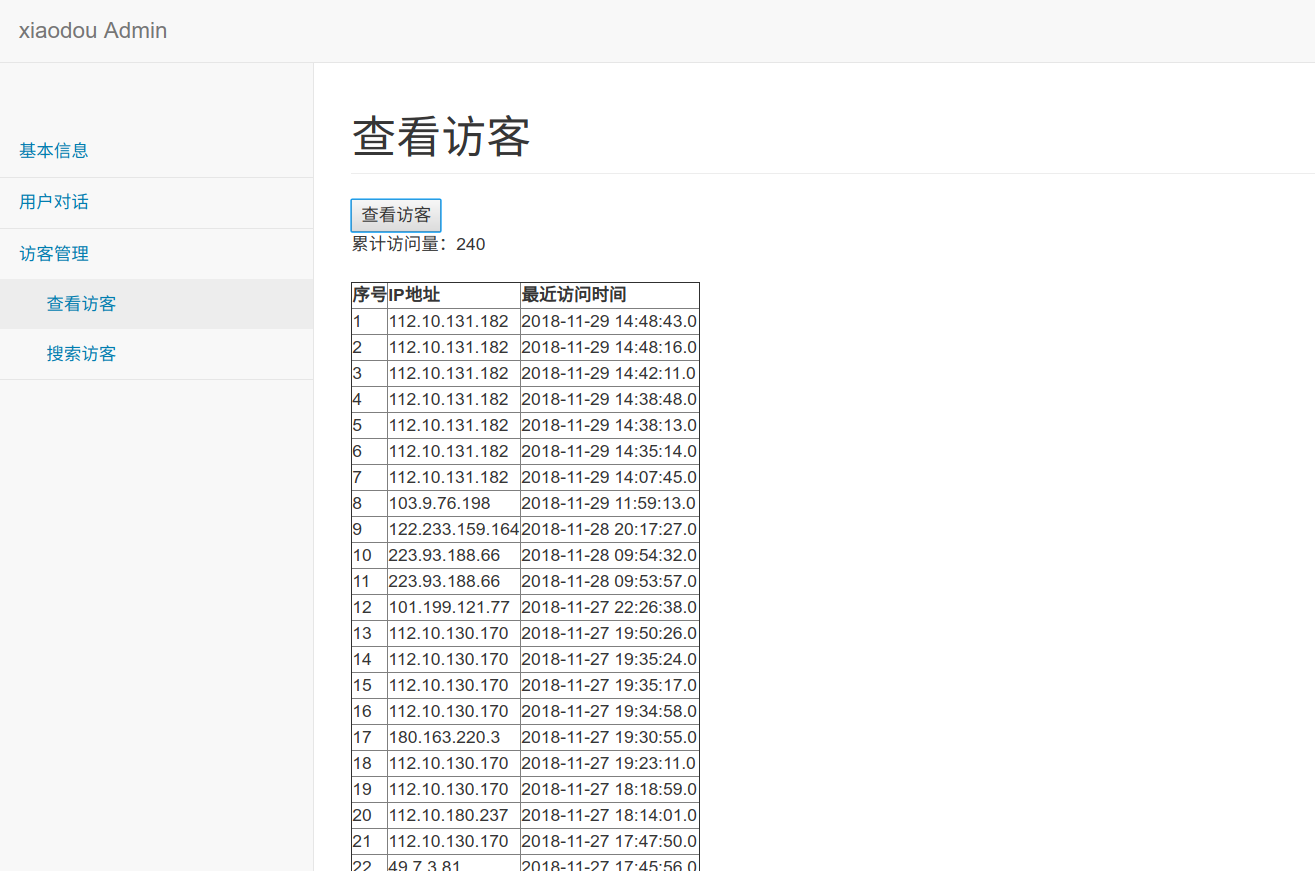
|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **审批人** | **审批意见** | **审批日期** |
| 0.1 | 叶春毅 | 通过 | 2018.10.15 |
| 0.4 | 康子博 | 前端可进一步美化 | 2018.10.20 |
| 0.8 | 康子博 | 通过 | 2018.11.16 |
| 0.9 | 叶春毅 | 通过 | 2018.11.25 |

(1) 用户端机器人对话测试



实现基本问答通讯。

(2) 管理端主要功能测试





可以查询到近期用户访问的时间和IP地址。

(3) 客服聊天室测试



用户可在聊天室向管理员或其他用户寻求帮助，提问。