**需求分析**

**1 引言**

* 1. **目的**

本文档描述一个智能语音推送系统的软件功能需求和非功能需求，其阅读对象是本项目的客户，开发和维护系统的开发团队成员。

* 1. **背景**

近些年来，人机交互的形式趋于多样化，人们已经不仅仅满足于与图形界面的交互，而是期待着语音交互。目前，具有语音交互的产品也有很多，比如最近热门的“Siri”、“小爱同学”、“天猫精灵”。通过调研了解到，这些产品虽然具有语音交互能力，但在现在“信息爆炸”的时代，不能满足人们对信息的需求而且缺乏个性化推送信息的能力。例如：用户想要得到瓜子二手车的数据，并做一个统计分析，只能在网上逐条浏览，效率低下。再例如：用户关注某一领域的新闻，当这一领域有新的动态时，需要点开界面去浏览，而不能直接收听，交互效果不好。为解决这一问题，开发一款语音爬虫软件来提升用户获取信息的体验就显得非常有必要。

* 1. **参考资料**

《树莓派开始 玩转linux》电子工业出版社周昕梓编

《PYQT4快速实战》电子工业出版社王硕 孙洋洋编

《一看就会的Python爬虫教程》 b站“青灯教育-马克老师”

**2 任务概述**

**2.1 目标**

本系统基于树莓派打造一款“智能语音推送系统”，该款系统不仅具有聊天对话的功能，而且具有搜索信息，推送信息的功能，相当于一个能进行语音播报的“搜索引擎”，满足人们对信息的需求。比如，用户下达指令说：“搜索新冠疫情动态”，那么系统将会语音播报最新的疫情新闻。当然，如果用户需要，也可以通过python Scripts用邮件发送疫情新闻。比如，用户需要获取瓜子二手车的数据，语音输入：“获取瓜子二手车信息”，系统将会自动爬取瓜子二手车的数据并列表显示。另外，本系统搭载有记录用户搜索信息的数据库，可以根据用户的喜好来自动推送信息。

**2.2 系统及用户的特点**

系统特点：

它是一个为方便用户获取信息的智能语音推送系统，它必须能够根据使用者的需要，及时提供所需要的信息，

以及完成所需操作。

用户特点：

用户类说明：

该系统的使用面向大众，范围十分广泛，任何人都可以从此系统中受益。用户可以把自己的要求等通过语音输入给该系统，系统随即与用户进行对话，并完成搜索信息、推送信息的功能。系统设计团队是本系统的维护人员。

**3 可行性分析**

用面向对象的分析方法根据需求提出一些模型，将这些模型实现为基本的python对象。语音识别部分调用百度语音识别的api，文字交互部分调用图灵机器人的api，语音合成部分调用百度语音合成的api，软件上通过三部分的调用以及pyaudio库的使用，配以硬件上语音捕获传感器的使用可基本实现语音交互。  
 爬虫上，利用python自带的request，bs4等请求库，对互联网数据进行模拟浏览器抓取。虽然当前大部分网站都具有反爬机制，但是我们可以在遵守robot.txt等协议的基础上，利用当前网页的User-Agent、Cookie“骗过”反爬机制进行数据爬取。爬虫利用语音识别传来的用户需求对互联网数据进行爬取，并可以通过显示屏与语音播报两种形式对用户进行反馈。  
 数据库上，用python内置的sqlite3模块，存储爬虫爬取的数据，方便用户的后续操作；同时存储用户输入的语句中提取出来的关键词。后续可通过识别提取的关键词来推荐用户近期感兴趣的信息，做到为用户量身打造的个性化推荐，以实现和完善智能人机交互的功能。  
 在表示层上，使用pyqt5制作图形界面与用户交互，通过各个控件显示用户输入以及爬取结果，并可供用户选择模式。

**4 需求规定**

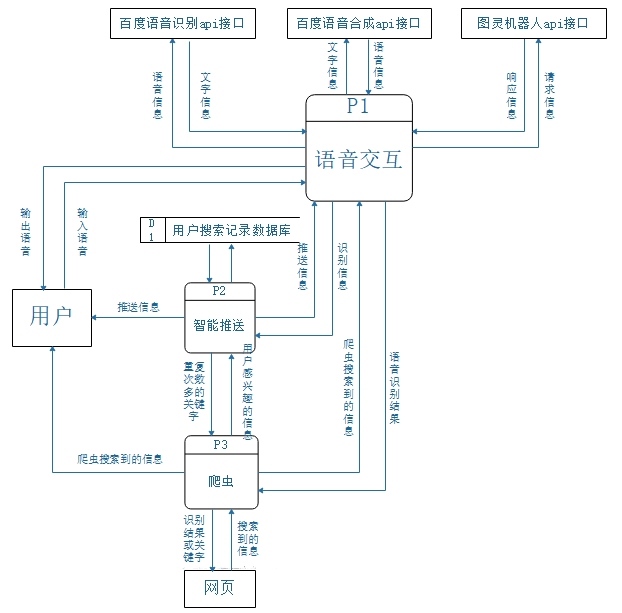
**4.1 软件功能说明**

功能划分

1. 语音交互模块
2. 爬虫模块
3. 智能推送模块（基于数据库）

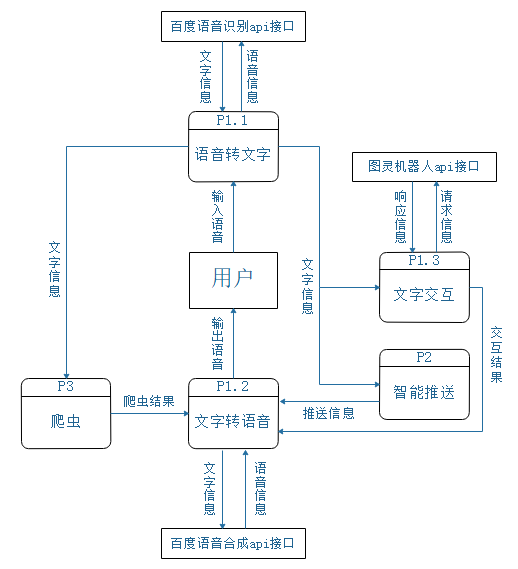
**5 数据流图**

根据系统功能以及各功能之间的联系，绘制出系统的顶层数据流图：

****

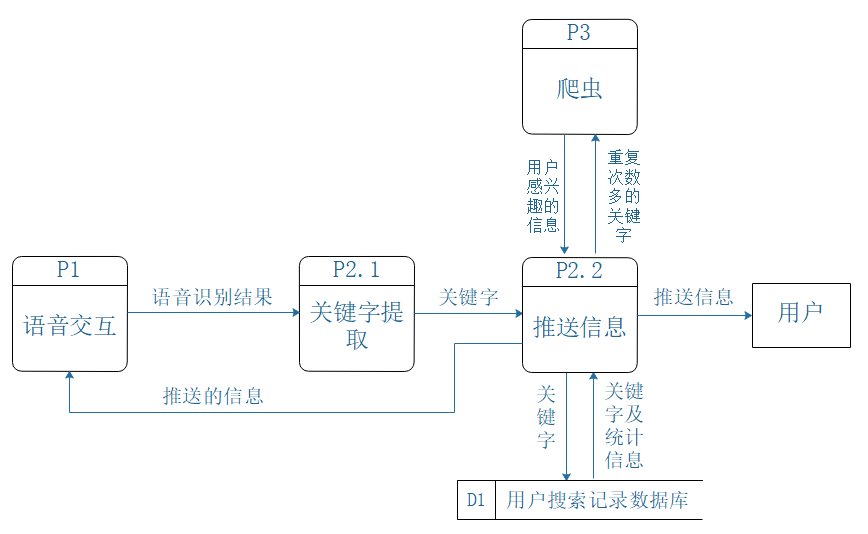
顶层数据流图仅从总体上反映了系统的信息联系，按照结构化分析方法，自顶向下，逐层分解，对顶层图进行细化。细化进行到数据流图中的每一个数据处理成为一个很容易理解的单一功能，且这个单一功能可以通过简单的逻辑表达式在数据字典予以说明。细化后的二层数据流图如下：

1. P1（语音交互）的细化图



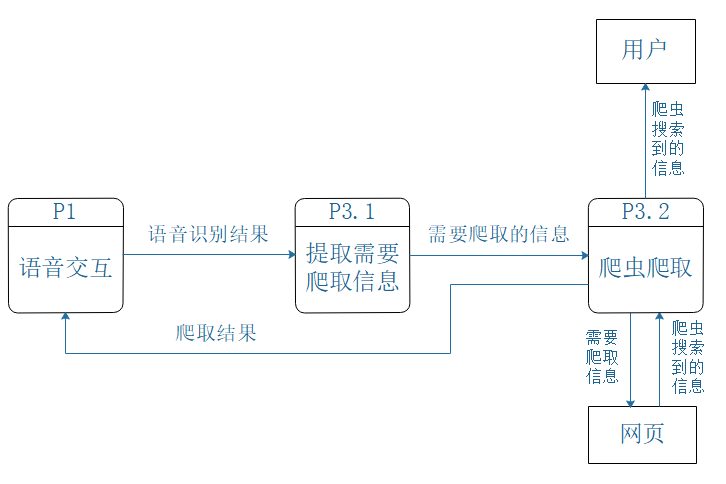
P1 语音推送细化图

1. P2（智能推送）的细化图



P2 智能推送细化图

1. P3（爬虫）的细化图



P3 爬虫细化图

**6 数据字典**

1. **数据流**

|  |
| --- |
| 数据流名称：输入语音  数据流别名：无  说明：用户输入的语音  数据流来源：用户  数据流流向：加工1.1（语音转文字）  数据流组成：语音内容+语音时间+日期 |

|  |
| --- |
| 数据流名称：输出语音  数据流别名：无  说明：系统输出给用户的语音  数据流来源：加工P1.2（文字转语音）  数据流流向：用户  数据流组成：语音内容+语音时间+日期 |

|  |
| --- |
| 数据流名称：推送信息  数据流别名：无  说明：系统根据用户喜好推送给用户的信息  数据流来源：加工P2.2（推送信息）  数据流流向：用户  数据流组成：推送信息编号+信息标题+信息内容+日期+来源 |

|  |
| --- |
| 数据流名称：爬虫搜索到的信息  数据流别名：无  说明：爬虫在网页上搜索到的信息  数据流来源：网页  数据流流向：用户  数据流组成：爬虫搜索到的信息编号+信息标题+信息内容+日期+来源 |

1. **数据项**

|  |
| --- |
| 数据项名称：推送信息编号  数据项别名：无  说明：推送信息的唯一标识  类型：字符串  长度：8  取值范围及含义：1-4位（0...9999）为日期编号，5-8（xxxx0001...xxxx9999）位为当前阶段所有信息按时间排列的顺序编号 |

|  |
| --- |
| 数据项名称：爬虫搜索到的信息编号  数据项别名：无  说明：爬虫搜索到的信息的唯一标识  类型：字符串  长度：8  取值范围及含义：1-4位（0...9999）为日期编号，5-8（xxxx0001...xxxx9999）位为当前阶段所有搜索到的信息按时间排列的顺序编号 |

|  |
| --- |
| 数据项名称：语音时间  数据项别名：无  说明：用户输入或系统输出语音的时间  类型：时间日期  取值范围及含义：‘1000-01-01 00：00：00’到‘9999-12-31-23-59-59’ |

|  |
| --- |
| 数据项名称：日期  数据项别名：无  说明：输入语音、输出语音、推送信息、搜索到的信息对应的时间  类型：时间日期  取值范围及含义：‘1000-01-01’到‘9999-12-31’ |

1. **数据文件**

|  |
| --- |
| 数据文件名称：用户搜索记录数据库  说明：用户搜索信息的关键字的记录  数据文件组成：信息编号+关键字+类别+日期+备注  组织方式：按信息编号从小到大排列  存取方式：顺序 |

1. **数据加工**

|  |
| --- |
| 数据加工名称:语音转文字  加工编号：1.1  说明：将用户输入的语音信息转成文字形式  输入数据流：输入语音  输出数据流：文字信息  加工逻辑：通过百度语音识别api接口将用户的语音信息转换成文字形式 |

|  |
| --- |
| 数据加工名称:文字转语音  加工编号：1.2  说明：将爬虫结果以及智能推送的文字信息转换成语音  输入数据流：爬虫结果、推送信息  输出数据流：输出语音  加工逻辑：通过百度语音合成api接口将获得的文字信息转换成语音信息 |

|  |
| --- |
| 数据加工名称:文字交互  加工编号：1.3  说明：根据文字信息进行文字交互  输入数据流：文字信息  输出数据流：交互结果  加工逻辑：通过向图灵机器人api接口发送请求信息获得响应信息 |

|  |
| --- |
| 数据加工名称:关键字提取  加工编号：2.1  说明：根据语音识别的结果进行关键字提取  输入数据流：语音识别结果  输出数据流：关键字  加工逻辑：基于语音识别结果搜取关键字 |

|  |
| --- |
| 数据加工名称:推送信息  加工编号：2.2  说明：根据提取出来的关键字向用户推送信息  输入数据流：关键字  输出数据流：推送信息  加工逻辑：根据关键字，在用户搜索记录数据库中搜索关键字及统计信息，将重复次数多的关键字利用爬虫方法获得用户感兴趣的信息 |

|  |
| --- |
| 数据加工名称:提取需要爬取的信息  加工编号：3.1  说明：根据语音识别结果提取需要爬取的信息  输入数据流：语音识别结果  输出数据流：需要爬取的信息  加工逻辑：基于语音识别结果提取需要爬取的信息 |

|  |
| --- |
| 数据加工名称:爬虫爬取  加工编号：3.2  说明：根据需要爬取的信息进行爬虫爬取  输入数据流：需要爬取的信息  输出数据流：爬虫搜索到的信息  加工逻辑：根据需要爬取的信息在网页中爬取信息 |

**7 运行环境及规定**

**7.1 设备及分布**

客户端主机基本配置为：树莓派3b+

**7.2 支撑软件**

操作系统

树莓派raspbian操作系统（linux内核）

数据库管理系统

SQLite

**7.3 程序运行方式**

本系统属于独立运行程序

**8 开发时间和资金估算**

任务/阶段任务时间（周） 人员（人） 资金预算（元）

项目调研及规划

需求分析：

系统设计：

程序开发：

系统测试：

总计：

**9 实现方案**

需求分析阶段,产出：需求计划书。

概要设计阶段，产出：概要设计报告。

详细设计阶段，产出：详细设计报告。

编码阶段，产出：系统编码实现。

测试分析，产出：系统测试报告。

