**CI-Bot 智能聊天系统**

**软件需求规格说明书**

文件编号： QMS—PROC－RD02

版本：0.1

|  |  |  |  |
| --- | --- | --- | --- |
| 编写人 | 卢嵩 | 日期 | 2017-10-31 |
| 评审 |  | 评审号/日期 |  |
| 批准 |  | 状态 |  |
| 发布范围 | 北航软国重CI-Bot开发项目组 | | |

**修改历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **作者** | **修改内容** |
| **2017-10-31** | **0.1** | **卢嵩** | **新建** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

目录

[一、 引言 5](#_Toc497262089)

[1.1. 目的 5](#_Toc497262090)

[1.2. 背景 5](#_Toc497262091)

[1.3. 术语 5](#_Toc497262092)

[1.4. 预期读者与阅读建议 5](#_Toc497262093)

[1.5. 参考资料 6](#_Toc497262094)

[1.6. 需求描述约定 6](#_Toc497262095)

[1) 需求标识方法 6](#_Toc497262096)

[2) 本系统的需求跟踪粒度 6](#_Toc497262097)

[3) 本文档的需求级别定义 6](#_Toc497262098)

[4) 功能描述方法 7](#_Toc497262099)

[5) 界面描述规则 7](#_Toc497262100)

[二、 项目概述 7](#_Toc497262101)

[2.1. 系统功能 7](#_Toc497262103)

[2.2. 业务流程描述 8](#_Toc497262104)

[2.3. 数据流程描述 (可选) 9](#_Toc497262105)

[2.4. 用户的特点 9](#_Toc497262106)

[2.5. 运行环境要求 10](#_Toc497262107)

[2.6. 设计和实现上的限制 10](#_Toc497262108)

[三、 功能需求的描述 10](#_Toc497262109)

[3.1. 参加 TREC-LiveQA Track 10](#_Toc497262111)

[1) 流程要求 10](#_Toc497262112)

[2) 性能要求 11](#_Toc497262113)

[3.2. 接入微信聊天平台 11](#_Toc497262114)

[1) 平台注册 11](#_Toc497262115)

[2) 平台接入 12](#_Toc497262116)

[四、 非功能需求 12](#_Toc497262117)

[5.1. 系统性能要求 12](#_Toc497262120)

[1) 时间特性 12](#_Toc497262121)

[3) 精度要求 13](#_Toc497262122)

[4) 系统鲁棒性 13](#_Toc497262123)

[5) 容错性 13](#_Toc497262124)

[6) 可扩充性 13](#_Toc497262125)

[5.2. 系统安全及保密要求 13](#_Toc497262126)

[5.3. 系统备份与恢复要求 14](#_Toc497262127)

[5.4. 系统日志 14](#_Toc497262128)

[五、 外部接口说明 14](#_Toc497262129)

[六、 其他需求 14](#_Toc497262130)

[七、 需求变更识别 14](#_Toc497262131)

[八、 功能列表 14](#_Toc497262132)

[九、 附件 15](#_Toc497262133)

# 引言

## 目的

编写该文档的主要目的在于明确CI-Bot项目开发中的系统需求。通过本文档定义CI-Bot系统需要实现的功能范围，以求在项目组组员中达成一致的功能描述，并指导系统设计和编码，让用户对软件有一个简单的了解。

该文档首先给出了整个系统的整体网络结构和功能结构的概貌，试图从总体架构上给出整个系统的轮廓，然后又对功能需求、性能需求和其它非功能性需求进行了详细的描述。其中对功能需求的描述采用了UML的用例模型方式，主要描述了每一用例的基本事件流，若有备选事件流则描述，否则则省略。这些文字和图形都为了本文档能详细准确地描述用户的需求，同时也为用户更容易地理解这些需求的描述创造了条件。

该文档详尽说明了这一软件产品的需求和规格，这些规格说明是进行设计的基础，也是编写测试用例和进行系统测试的主要依据。同时，该文档也是用户确定软件功能需求的主要依据。

## 背景

描述系统产生的背景，包括：

1. 需开发的软件系统的名称，和英文缩写（可选），项目编号（可选）；
2. 列出此项目的任务提出者、开发者
3. 软件系统应用范围、用户。
4. 产生该系统需求的原因或起源，如社会背景、市场发展、政策趋势、原有系统局限性

Crowd-Intelligence Bot（以下简称为CI-Bot）是为了将传统聊天系统和新型群智计算方法结合起来的研发项目。本项目由软件开发环境国家重点实验室（以下简称为软国重）的丁嵘老师提出，并交由软国重群体智能组进行开发。

该软件是一个独立的智能聊天系统，能够根据用户输入的信息实现自动化回复。原有的问答系统在回答的准确性以及问题的泛化性能上有所欠缺，而问答社区的回答往往缺乏用户间实时性的交互，因此，本系统主要着眼于实时而准确地回答互联网用户提出的问题。系统的终端用户主要为互联网用户，将来可能与微信等聊天平台进行连接。

## 术语

列出本文件中用到的专门术语、术语定义、外文首字母组词的原词组。

Text REtrieval Conference（简称TREC），是当今世界上影响力最大的文本检索会议之一，该会议每年都将举办名为LiveQA 的问答系统评测竞赛，该竞赛也是本系统计划参与的项目之一。

问答内容指用户通过系统前端输入的内容以及系统经过数据库和网络检索后返回的答案。

## 预期读者与阅读建议

|  |  |
| --- | --- |
| 预期读者 | 阅读建议 |
| CI-Bot项目组开发人员 | 仔细阅读概述，编写目的，文档约定，系统功能介绍和维度指标说明。 |
| 参加需求评审的人员 | 仔细阅读全部内容。 |
| 系统设计人员 | 仔细阅读全部内容。 |
| 系统测试人员 | 仔细阅读文档约定，系统功能介绍和维度指标说明。 |
| 用户 | 了解系统的预期功能和性能，并与分析人员一起对整个需求进行讨论和协商 |
|  |  |

## 参考资料

列出有关的参考资料，如：

1. 本项目经核准的计划任务书或合同、上级机关的批文；
2. 属于本项目的其他已发表的文件；
3. 本文件中各处引用的文件、资料、包括所要用到的软件开发标准。
4. 行业标准和规范。

列出这些文件资料的标题、文件编号、发表日期和出版单位，说明能够得到这些文件资料的来源。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 文档名 | 版本号 | 发表日期 | 来源 | 文档简称 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## 需求描述约定

### 需求标识方法

本系统采用层级化编号方法，具体形式为模块缩写＋序列号，如DB01、SEARCH01.01、WEB 01.01.02等。主要模块缩写参照表：

|  |  |  |  |
| --- | --- | --- | --- |
| 模块名 | 模块缩写 | 模块名 | 模块缩写 |
| 数据库模块 | DB | 网页模块 | WEB |
| 众包模块 | CI | 微信模块 | WX |

需求层次：分三个层次，用三位字符表示。第一层需求指主功能模块，第二层需求指功能模块的主功能点，第三层次指主功能点下的具体需求。

### 本系统的需求跟踪粒度

跟踪到第二层功能需求，即明确到各个功能模块的主要功能点。

### 本文档的需求级别定义

* 本文档统一规定对需求层次为二级以上（功能模板、主功能点）的定义优先级，三层需求依据二层需求的优先级执行。
* 本文档的优先级别分为：紧急、正常、缓
* 同时对于主功能点还描述实现的周期：一期、二期、三期

### 功能描述方法

本文档从以下几个方面对功能需求进行描述：

1. 业务定义/描述。
2. 适用的用户类型
3. 业务规则/业务要素。
4. 输入：提供所有与本功能有关的输入描述，包括：输入数据类型、媒体、格式、数值范围、精度、单位等。
5. 输出－提供与本功能有关所有输出的描述，包括：输出数据类型、方式、格式、精度、单位等，以及图形或显示报告的描述。
6. 业务操作流程
7. 描述正常业务流程，列举异常情况和处理流程。建议使用图示，并配合必要的文字说明
8. 约束条件/特殊考虑

列出在各个工作领域不需计算机化的功能并提供其原因以及特殊条件。

### 界面描述规则

　　界面描述使用VISIO的界面模型进行描述。

# 项目概述



## 系统功能

概述了产品所具有的主要功能。其详细内容将在系统功能需求和特性中描述，所以在此只需要概略地总结。很好地组织产品的功能，使每个读者都易于理解。

1. 建议以图表形式列出功能结构图，并加入必要文字说明。
2. 建议以列表形式列出功能分类，以及优先级，并加入必要文字说明。

本项目的开发任务旨在开发一个简单易用、交互性强的自然语言问答系统。融合了传统问答系统和群智众包平台的优势，一般用来回答互联网用户提出的不限领域的问题。

本软件初步计划为一个完整独立的系统，将来可能会接入微信平台。



图1 系统功能描述

## 业务流程描述

用文字或图形方式描述系统的主要业务流程，若引用了参考资料，应指明参考资料的简称与章节号或页码，以便复核与评审。



## 用户的特点

列出本软件的最终用户的特点，以及本软件的预期使用频度，确定可能使用该产品的不同用户类并描述它们相关的特征。有一些需求可能只与特定的用户类相关。一般来说至少有以下几类：

一般操作者：

系统管理者：

最终用户

本软件的用户可分为互联网用户以及系统管理员。

本软件的最终用户为互联网用户；他们没有经过软件的使用训练，需要系统提供步骤来指导用户完成网页的打开以及问题的发布等等。

系统管理员负责定期对系统进行简单的维护，解决常见的配置问题，并通过分析问答结果进行系统的优化。本系统将能够维持一天24小时运行，同一时间在线用户预计不会超过1000人。

## 运行环境要求

描述了软件的运行环境，包括硬件平台、操作系统和版本，还有其它的软件组件或与其共存的应用程序。

该系统采用B/S结构，他的运行环境分为后端服务器和网页客户端两部分。

系统的后端服务器拟部署在腾讯云服务器上，操作系统为Ubuntu 14.04 64-bit。数据库系统为sqlite，版本为。。。

系统的前端为多种形式呈现，计划初期搭建网页形式，在将来推出结合微信平台的综合性系统。网页形式需要浏览器版本为IE8以上，兼容其他常见浏览器Chrome以及Firefox。微信平台需要配合微信手机应用使用。

## 设计和实现上的限制

列举出在对软件需求规格说明中影响需求陈述的假设因素（与已知因素相对立）。这可能包括你打算要用的商业组件或有关开发或运行环境的问题对需求实现的影响，也可能是需求或业务规则对设计与实现方法的影响。可能还来自于经费、投资方面的限制，法律或政策方面的限制，或者可利用的资源和信息的限制。

软件需要在一年内（具体地说是TREC QA Track开始前）完成设计、开发、测试并成功在线运行。

服务运行在腾讯云的主机上，参与LiveQA Track时可能需要进行服务器迁移或数据迁移。

# 功能需求的描述

〔为每个确定的商业功能（需实现的功能）描述其定义、业务规则，详细叙述如何从输入转变到输出并且如何获得、处理和产生这些信息。这些内容在下列标题中有条理的阐述。

1. 业务定义/描述。
2. 适用的用户类型，指操作本功能所需的授权
3. 业务规则/业务要素。
4. 功能项的主要页面或是样式
5. 输入：提供所有与本功能有关的输入描述，包括：输入数据类型、媒体、格式、数值范围、精度、单位等。
6. 输出－提供与本功能有关所有输出的描述，包括：输出数据类型、方式、格式、精度、单位等，以及图形或显示报告的描述。
7. 业务操作流程

描述正常业务流程，列举异常情况和处理流程。建议使用图示，并配合必要的文字说明

1. 约束条件/特殊考虑

列出在各个工作领域不需计算机化的功能并提供其原因以及特殊条件。



## 参加 TREC-LiveQA Track

### 流程要求

参加比赛的详细流程要求见官网：<http://trec.nist.gov/pubs/trec25/papers/Overview-QA.pdf>

以下做粗略的总结：

最终的成果需要以服务器的形式呈现，接受作为输入的包含问题标题，正文和类别的Yahoo Answer问题，回应以答案或可选的问题摘要。参赛者回答的问题将会在比赛期间被纪录下来。

训练的数据可以自行下载，其中包含了1000余个问题以及大约20个

<https://webscope.sandbox.yahoo.com/catalog.php?datatype=l%2520(set%2520L6)>

评级总共分为5档，对应分数为1~4，最终评判指标为平均分，i分以上问题个数，i分以上的准确率等指标。

### 性能要求

大赛的服务器将会每分钟为系统推送一个问题，系统需要在60s内进行作答，回答的长度不能超过1000个单词。

## 接入微信聊天平台

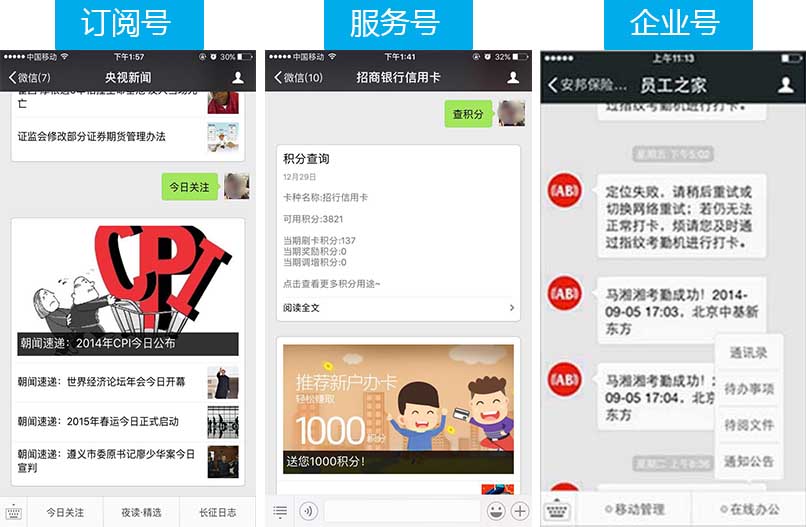
### 平台注册

[Buaacibot@163.com](mailto:Buaacibot@163.com)

Cibotpw

Buaa@cibotpw





1、订阅号：主要偏于为用户传达资讯（类似报纸杂志），认证前后都是每天只可以群发一条消息；

2、服务号：主要偏于服务交互（类似银行，114，提供服务查询），认证前后都是每个月可群发4条消息；

3、企业号：主要用于公司内部通讯使用，需要先验证身份才可以关注成功企业号。

温馨提示：

1）如果想简单的发送消息，达到宣传效果，建议可选择订阅号；

2）如果想用公众号获得更多的功能，例如开通微信支付，建议可以选择服务号；

3）如果想用来管理内部企业员工、团队，对内使用，可申请企业号；

4）订阅号可通过微信认证资质审核通过后有一次升级为服务号的入口，升级成功后类型不可再变；

5）服务号不可变更成订阅号。

### 平台接入

拟采用服务号形式进行平台接入

## 问答功能

### 提问

### 回答

# 非功能需求



## 系统性能要求

### 时间特性

由于LiveQA Track 对于回答问题的时间性有较高的要求，因此该系统在回答问题方面至少需要保证每分钟回答一个问题。

### 精度要求

任何一个系统的开发都应充分考虑各种系统错误，避免造成数据库系统不一致或损坏，因此在本系统中采用出错提示框形式为用户提供出错信息；另外对于一些关键性操作如管理员对数据库进行删除或修改记录操作时，也提供了确认机制。

用户的ID、管理员的ID: 必须以连续的9位纯数字表示，例如123456789等；

问题与回答的ID: 以hash值表示，必须唯一

报文的形式：为post报文，且必须是以特定开头+内容的形式呈现，具体对应关系 /q指问题，/a指回答， /u指用户

### 系统鲁棒性

为保证系统能够完成LiveQA竞赛，系统应具有较强的鲁棒性。例如：系统能够每天连续运行24小时，连续7天正常使用不出现宕机的情况。并发能力为200（假设能），也不会出现死机或者崩溃的状况。

### 容错性

对于某些没有问题的答案，或者可能引发Bug的特殊情况，需要考虑周全，不能因此出现宕机的情况。

### 可扩充性

系统的初步搭建可能以服务器+网页的情况出现，将来还可能跟微信或者其他平台进行连接扩展。最终客户端需要实现可以在任何终端上使用，系统管理员可以为系统添加额外的功能。对于开发者来说，该系统可以进行模块替换。

## 进度需求

## 系统安全及保密要求

指定可以访问各自功能的用户群，如需要可指定用户访问权和安全包（如有），以便有效控制系统访问和数据访问。

确认审核记录和所有有关报告及接受人。阐述是否任何违反系统访问的内容都需要监控，以及以什么方式监控。

列明所有安全需求，例如数据加密，信息验证等。

如用户单位对安全保密的要求，对使用方便的要求，对可维护性、可补充性、易读性、可靠性、运行环境可转换性的特殊要求等。

可维护性：在使用过程中，系统管理员可以对系统进行简单的系统维护；

可补充性：在使用过程中，由于学校的需求，可以在系统显示添加功能以及添加后台的数据库；

易读性：程序可应对不同文化水平的使用者；

可靠性：系统必须保证使用者的信息的安全；

## 系统备份与恢复要求

1. 指定每种信息类型的保存期；
2. 阐述在保存期过后需要实施的行为，例如：转移到计算机外部的介质中，或删除它们。
3. 如转移到计算机外的介质中，叙述存储期及贮存介质的类型。例如：磁带、磁盘、报告等。
4. 环境异常时，系统恢复策略描述。

系统按照每周一次的频率对数据库中的数据进行备份，每个备份将在一个月后删除；

软件系统崩溃，要求系统可以有系统管理员扩大在线人数；

无论系统正常或有问题，系统管理员都可以使用任何终端登陆；

系统出问题，不会影响问题，回答以及用户数据库中的所有信息；

## 系统日志

1. 日志内容、记录策略
2. 日志的保存时长、保存策略
3. 日志内容的访问控制

# 外部接口说明

外部接口包括：硬件接口、软件接口、通信接口，每个接口需考虑以下内容：

1. 接口描述，包括接口类型、接口特点（如版本、名称、来源等）
2. 接口与本系系统的输入输出关系
3. 技术方面的约束
4. 转换的安全考虑

# 其他需求

[对其它需要描述但未在本模板中列出的需求，在此进行说明，如果某个这样的需求比较重要，可以单独用新的一节来描述。

这样的需求可能包括，数据库需求、法律需求、国际准则、重用目标等。]

# 需求变更识别

识别并定义在将来可能会变化的需求

# 功能列表

罗列本需求中的功能点、需求编号、需求内容、优先级与内容描述。必要时独立成立做为本需求的附件。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 功能点 | 子功能 | 需求编号 | 优先级 | 内容描述 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 附件

附件可能包括各个模块的具体的功能需求描述、需求跟踪表，或者系统的词汇表、待确定问题列表，以及其它所有能够成为需求基线内容的正式文档。