

学校编码: 10384

分类号_____密级_____

学号: X2009221014

UDC_____

厦 门 大 学

工 程 硕 士 学 位 论 文

基于安卓系统的汉语智能聊天机器人

Android-based Chinese intelligent chat robot

姓名: 黄家兴

指导教师姓名: 陈毅东副教授

专 业 名 称: 计算机技术

论文提交日期:

论文答辩时间:

学位授予日期:

答辩委员会主席: _____

评阅人: _____

2013年__月__日

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

目 录

第一章 概述	1
第二章 Android 及其开发	6
2.1 Android 的简介	6
2.1.1 android 发展史	6
2.1.2 Android 应用程序组成【2】	7
2.1.3 Intent 的使用	9
2.1.4 系统 UI 事件的处理	11
2.2 Alice 聊天机器人的总体流程	11
第三章 Alice 与其接口	16
3.1 ALICE 相关基础	16
3.1.1 Alice 发展史【3】	16
3.1.2 AIML 人工智能标记语言	16
3.1.3 Alice 内部机理	18
3.2 ALICE 的接口分析与应用	21
3.2.1 Alice 接口分析	21
3.2.2 Alice 接口如何应用	22
第四章 基于自然语言的跨语言与语音聊天	25
4.1 自然语言处理概述	25
4.1.1 自然语言发展史	25
4.1.2 自然语言与语言理解概念	26
4.1.3 自然语言理解的过程【11】	27
4.1.4 语言理解方法	27
4.2 机器翻译	30
4.2.1 英语句法分析的研究【10】	30
4.2.2 中文分词的应用研究	31

4.2.3 机器翻译过程	32
4.2.4 有道翻译接口与应用	33
4.3 语音聊天	36
第五章 测试结果分析	39
第六章 总结	42
参考文献	43

Directory

The first chapter Outlines and its system architecture in figure	1
The second chapter Android and its development	6
2.1 Introduction of Android's	6
2.1.1 Android development	6
2.1.2 Android application components 【2】	7
2.1.3 The Intent to use	9
2.1.4 Processing system UI events	11
2.2 Alice chatbot the overall flow	11
The third chapter Alice and its interface	16
3.1 ALICE related basis	16
3.1.1 The Alice development 【3】	16
3.1.2 AIML markup language artificial intelligence	16
3.1.3 Alice internal mechanism	18
3.2 Alice's analysis and application interfaces	21
3.2.1 Alice interface analysis	21
3.2.2 How to apply Alice interface	22
The fourth chapter is based on natural language across language and voice chat	25
4.1 Summary of natural language processing (NLP)	25
4.1.1 Natural language development	25
4.1.2 Natural languages and language comprehension concept	26
4.1.3 Natural language understanding of the process 【11】	27
4.1.4 Language understanding method	27
4.2 Machine translation	30

4.2.1 English syntactic analysis research 【10】	30
4.2.2 The applied research of Chinese participle.....	31
4.2.3 Holdings machine translation process	32
4.2.4 Youdao translation interface and application	33
4.3 Voice chat	36
The fifth chapter Test results analysis.....	39
The sixth chapter Summarizes	42
Refs.....	43

摘 要

随着科学技术不断发展,人类社会生活不断的信息化,智能化生活对社会产生巨大影响。人们希望能通过自然语言与计算机进行交流,智能聊天机器人就是在这样的历史背景下诞生的新时代产物。它是一种通过自然语言同人进行交流的人机对话系统。美国苹果率先研发了一款名为Siri的智能聊天机器人。Siri 可以令iPhone4S变身为一台智能化机器人,利用Siri用户可以通过手机读短信、介绍餐厅、询问天气、语音设置闹钟等。Siri可以支持自然语言输入,并且可以调用系统自带的天气预报、日程安排、搜索资料等应用。还能够不断学习新的声音和语调,提供对话式的应答。因此我想在Android系统也开发一个与Siri类似的可以与人们智能对话的智能聊天机器人。

ALICE是一个基于经验的人工智能聊天机器人,它通过了世界著名的图灵测验并作为“最像人类的计算机”两次获得AI科学界最高荣誉洛伯纳(Loebner)奖,但是它唯一的缺陷就是不支持中文聊天。本文通过分析人工智能聊天机器人ALICE 的知识组织结构和推理机制,针对其不支持中文聊天的缺陷,本文首先在android开发平台上利用android开发技术,设计并实现了基于ALICE系统的英文聊天接口,然后在这个基础上应用自然语言处理领域的技术加入翻译系统实现了中文互聊功能,最后加入语音识别和语音输出,对于语音识别(我们可以使用google的语音识别API进行语音的识别,将语音转成文字。语音输出,其实就是使用TTS,讲文字进行语音合成播放出来),来实现语音聊天功能。

综上所述,本文具有较好的选题意义,经过努力完成基于安卓系统的汉语智能聊天系统,基本实现支持汉语的智能对话。

关键字: ALICE;自然语言;Android

Abstract

With the development of science and technology, human social life of informatization unceasingly, intelligent life have a big impact to the society. People hope to be able to through natural language communicate with computer, intelligent chat robot was born in such a historical background of the new era. It is a natural language through submission for the man-machine dialogue in communication system. The United States apple first developed an intelligent chat robot called Siri. Siri can make the iphone 4 s as an intelligent machine, using Siri to users to read text messages through voice, introduced restaurant, inquired about the weather, set the alarm clock, etc. Siri can support natural language input, and the weather forecast system at can call search for information, schedules, etc. To keep on learning new sounds and intonation, also provide conversational responses. So I think in the Android system is also developing a similar to Siri can intelligent chat robot intelligent conversation with people.

ALICE is a chatting robot of artificial intelligence on the basis of experience, it through the Turing test and the world famous as \"the most like human computer\" twice received the highest honor in AI science Bernard (Loebner prize, but it's the only drawback is chat does not support Chinese. Based on the analysis of the ALICE chat robot of artificial intelligence knowledge organization structure and reasoning mechanism, aiming at its defect does not support chat in Chinese, this paper first on the android development platform using the android development technology, designed and implemented based on ALICE system English chat interface, and then on the basis of the application in the field of natural language processing technology to join translation system realize the function of each other to chat in Chinese, add voice recognition and voice output, for voice recognition (we can use the Google apis for voice recognition to speech recognition, speech into text. Speech output, is actually using the TTS, text to speech synthesis broadcast), to implement the voice

chat function.

To sum up, this paper has better selected topic significance, through the effort to finish chat Chinese intelligence system based on android system, basic implementation supports Chinese intelligent dialogue.

Key words: ALICE; Natural language; Android

厦门大学博士论文摘要库

第一章 概述

随着科技不断发展进步，人类进入了智能化社会，智能手机的出现给人类生活带来巨大改变。智能手机（Smartphone），是指“像个人电脑一样，具有独立的操作系统，可以由用户自行安装软件、游戏等第三方服务商提供的程序，并可以通过移动通讯网络来实现无线网络接入的这样一类手机的总称”。在所有智能手机操作系统中，他们之间的应用软件互不兼容。因为可以安装第三方软件，所以智能手机有丰富的功能。

智能手机的诞生，是掌上电脑（PocketPC）演变而来的。最早的掌上电脑是不具备手机的通话功能，但是随着用户对于掌上电脑的个人信息处理方面功能的依赖的提升，又不习惯于随时都携带手机和 PPC 两个设备，所以厂商将掌上电脑的系统移植到了手机中，于是才出现了智能手机这个概念。

智能手机同传统手机外观和操作方式类似，不仅包含触摸屏也包含非触摸屏数字键盘手机和全尺寸键盘操作的手机。但是传统手机都使用的是生产厂商自行开发的封闭式操作系统，所能实现的功能非常有限，不具备智能手机的扩展性。

“智能手机（Smart Phone）”这个说法主要是针对“功能手机（Feature phone）”而来的，本身并不意味着这个手机有多“智能（Smart）”；从另一个角度来讲，所谓的“智能手机（Smart Phone）”就是一台可以随意安装和卸载应用软件的手机（就像电脑那样）。“功能手机（Feature phone）”是不能随意安装卸载软件的，JAVA 的出现使后来的“功能手机（Feature phone）”具备了安装 JAVA 应用程序的功能，但是 JAVA 程序的操作友好性，运行效率及对系统资源的操作都比“智能手机（Smart Phone）”差很多。

世界上第一款智能手机是 IBM 公司 1993 年推出的 Simon，它也是世界上第一款使用触摸屏的智能手机，使用 ROM-DOS 操作系统，只有一款名为《DispatchIt》第三方应用软件。它为以后的智能手机处理器奠定了基础，有着里程碑的意义。现在前沿智能手机系统有：Symbain S60、BlackBerry OS、Android、Mac OS、iphone ios、windows moblie 等。

随着智能手机的出现，人们又希望通过自然语言与智能手机进行交流，在这种需求下，美国苹果率先研发了一款名为 Siri 的智能聊天机器人。Siri 可以令 iPhone4S 变身为一台智能化机器人，利用 Siri 用户可以通过手机读短信、介绍餐厅、询问天气、语音设置闹钟等。Siri 可以支持自然语言输入，并且可以调用系统自带的天气预报、日程安排、搜索资料等应用。还能够不断学习新的声音和语调，提供对话式的应答。Siri 现在是 Beta 版本，因为 Apple 还在给她加入更多的功能和语言支持。Siri 现在支持三种语言：英语、法语和德语，其中英语支持三种口音：美式、英式和澳大利亚口音。

Siri 的出现给各大智能手机系统开了先例，进来 Android 系统也研制出聊天机器人如：小黄鸡、腾讯聊天机器人小 Q 等。这些聊天机器人都是支持汉语的聊天机器人，但是智能程度较低还有各种各样缺陷。网站上也出现了在线聊天机器人如，Cleverbot、Alice。

小黄鸡：“小黄鸡”公共主页用的是 simsimi 的开放 API，做出一个应用的。Simsimi，是韩国开发的智能聊天机器人。官方提供了开发 API，地址

<http://developer.simsimi.com/>。也就是说，这个应用的“灵魂”，人工智能部分，是官方做好了，设计小黄鸡完全不需要涉及什么复杂的算法，只要简单调用就行了。小黄鸡智能程度低经常冒出脏话粗话，但是它的用户界面做的比较好。



图一：小黄鸡聊天内容和界面

腾讯聊天机器人小 Q：这是一款以 QQ 聊天相似的聊天机器人，官方提供一个聊天机器人 QQ 号，我们只要下载 QQ 软件加这个 QQ 号就能把它当网友聊天。虽然他很受 Q 友的欢迎，特别是年纪比较小的中小学生，但是它也是存在不少问题，比如知识不够丰富很多，还会脏话连篇。

下面是引用一篇报道‘你好！’‘去你的，走开！’放暑假的儿子天天上网 QQ 聊天，而对方却不断说出一些“难听”的话。昨天上午，市民王女士称，腾讯最新推出的聊天机器人小 Q 妹妹竟然教导自己的儿子“说脏话”，为了避免孩子学坏，她只能删除了电脑上的 QQ 程序，彻底“屏蔽”小 Q 的出现。

小 Q “易学坏爱骂人”

“动不动就说些‘去死吧’，你说这样的机器人，我怎么敢让‘她’继续与孩子交流呢？”根据王女士提供的 QQ 号码，记者也在登录 QQ 后加了小 Q 为好友。在“寒暄”之前，与小 Q 的对话框跳出了一段“说明”：“小 Q 的部分回答内容来自网友的教育，不代表腾讯公司赞成其观点和立场。”

随后，在与小 Q 的聊天中，小 Q 不时爆出一些粗口，有些内容甚至涉及色情、暴力。而遇到一些机器人不知如何回答的问题时，系统自动提示记者“如何教育小 Q”的方式。

在与小 Q 的简单“套招”之后，记者登录了腾讯网上的 QQ 论坛，只见从六月中旬机器人“出世”之后，关于小 Q 骂人的帖子便在论坛上层出不穷，而内容更是从小 Q 的惊人之语衍生到了如何让小 Q 骂出连篇脏话的“聊天秘笈”，有些“经典”招数回帖率甚至高达 300 多。

腾讯会管好“大嘴巴”

由于小 Q 是机器人，其语言系统不可能十分丰富，因此才邀请用户一起参与到教育小 Q 的过程中。昨天下午，腾讯公司公关部负责人陈小姐告诉记者，小 Q 目前仍是一个试运行的产品，腾讯今后会对小 Q 进行更为严格的审查。

为了保证小 Q 的思想健康，腾讯对聊天机器人采取的是系统加人工的严格审查机制，但由于腾讯用户庞大，因此有些不良内容的删除可能会存在一些遗漏或滞后，腾讯已经考虑到这个问题，并会针对这个情况商议一个具体措施。”

Cleverbot：是一个在线的聊天机器人，含有一点人工智能的高科技成份。难能可贵的是不仅支持英文聊天还支持中文，不是中文界面，而是中文交流；但是只

能用汉语拼音，而不认汉字导致聊天很别扭。



图二：Cleverbot 聊天界面

它的大部分回答都是毫不相干的，不过有时也会显示出一些智能的回答，比如问“Chun ge chun ye men”，它的回答是“Tie xue zhen han zi.”，这句足以说明它有深厚的知识积累。还有下面这段对话，貌似还能感觉出它的情绪：

问：ni shi ri ben ren?

答：Bu shi.

问：ni shi ri ben ren?

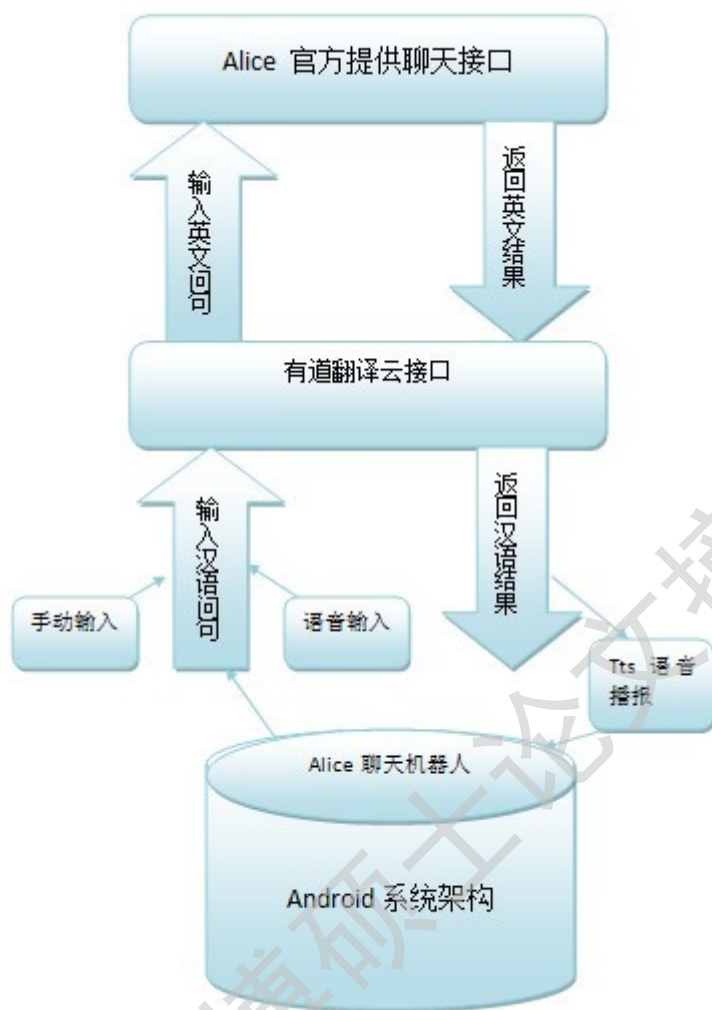
答：Ni nai nai shi ri ben ren.

问：ni shi ri ben ren?

答：Ni cai shi ri ben ren , ni quan jia dou shi ri ben ren.

ALICE 是一个基于经验的人工智能聊天机器人，在2000年和2001年先后两次通过了世界著名的图灵测验并作为“最像人类的计算机”两次获得AI科学界最高荣誉洛伯纳(Loebner)奖。由于智能程度如此高的机器人，虽然Wallace博士说原理可以支持中文【1】但是并没有人去这么做，所以我就想模仿小黄鸡的设计思路，利用现有较高智能的Alice机器人系统，通过官方提供的接口做一个支持中文聊天的智能机器人Alice.，也就是Alice机器人的核心部分不需要我们开发，直接链接官方接口就行了，再加上利用现有机器翻译技术，可以直接链接有道翻译云接口直接作出支持汉语聊天的Alice智能机器人。

根据这个思想做出系统架构图如下：



图三 Alice 聊天机器人架构图

本文先在第一章给出了论文整体概述及其论文系统架构思路与架构图，然后在后面三章分别介绍了该系统涉及各个领域的技术：Android开发技术、Alice系统与接口应用、自然语言处理技术(自然语言理解、机器翻译、语音聊天)。最后对该系统做出测试分析与总结该系统有缺陷及今后可以改进的地方。

第二章 Android 及其开发

2.1 Android 的简介

2.1.1 android 发展史

关于 Android 系统的发展史,首先我们要先说说 Android 系统这个名字的由来。Android 第一次是出现在法国作家利尔亚当在 1886 年发表的科幻小说《未来夏娃》中,作者将外表像人类的机器起名为 Android,这也就是 Android 名字的由来。

Android 系统一开始并不是由谷歌研发的,Android 系统原本公司的名字就叫做 Android,谷歌公司是在 2005 收购了这个仅成 22 个月的高新科技企业。从此 Android 系统开始由谷歌接手研发,Android 系统的负责人以及 Android 公司的 CEO 安迪·鲁宾做为谷歌公司工程部的副总裁,继续负责 Android 项目的研发工作。

2007 年 11 月 5 日这天,谷歌公司正式向外界展示了这款名为 Android 的操作系统,并且在当天谷歌宣布建立一个全球性的联盟组织,该组织由 34 家手机制造商、软件开发商、电信运营商以及芯片制造商组成。这一联盟将支持谷歌发布的手机操作系统以及应用软件,并共同开发 Android 系统的开放源代码。

在 2008 年的 GoogleI/O 大会上,谷歌提出了 Android HAL 架构图,在同年 8 月 18 号,Android 获得了美国联邦通信委员会(FCC)的批准,在 2008 年 9 月,谷歌正式发布了 Android 1.0 系统,这也是 Android 系统最早的版本。

2009 年 4 月,谷歌正式推出了 Android 1.5 系统的手机,从 Android 1.5 版本开始,谷歌开始将 Android 的版本以甜品的名字命名,Android 1.5 命名为 Cupcake(纸杯蛋糕)。

2009 年 9 月份,谷歌发布了 Android 1.6 的正式版,并且推出了搭载 Android 1.6 正式版的手机 HTC Hero G3,凭借这出色的外观设计以及全新的 Android 1.6 操作系统,HTC Hero G3 成为当时全球最受欢迎的手机。Android 1.6 也有一个有趣的甜品名称,它被成为 Donut(甜甜圈)。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库