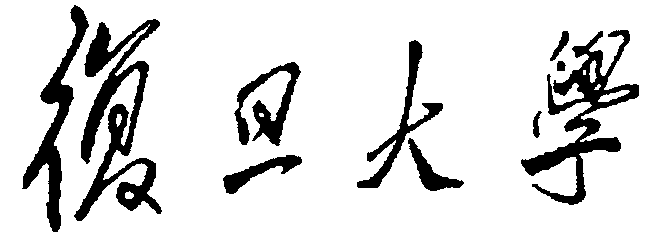
|  |  |
| --- | --- |
|  | 学校代码： 10246 |
|  | 学 号： 14210240086 |
|  |  |



|  |
| --- |
| 硕 士 学 位 论 文 |

（专业学位）

|  |
| --- |
| **面向残疾人社区的众包标注平台设计与实现** |

Design and Implementation of a CrowdLabel Platform For Disabled Community

|  |  |
| --- | --- |
| 院 系： | 计算机科学技术学院 |
| 专业学位类别( 领域 ) ： | 计算机技术 |
| 姓 名： | 罗东亮 |
| 指 导 教 师： | 卢 暾 副教授 |
| 完 成 日 期： | 2017年 3月 1日 |

指导小组成员名单

顾 宁 教 授

张 亮 教 授

卢 暾 副教授

丁向华 副教授

# 目录

[目录 I](#_Toc478729662)

[摘要 III](#_Toc478729663)

[ABSTRACT V](#_Toc478729664)

[第一章 绪论 1](#_Toc478729665)

[1.1 研究背景和问题描述 1](#_Toc478729666)

[1.2 国内外研究现状 3](#_Toc478729667)

[1.2.1 协同众包 3](#_Toc478729668)

[1.2.2 众包任务分配方式 4](#_Toc478729669)

[1.3 本文工作与贡献 5](#_Toc478729670)

[1.4 本文的组织结构 6](#_Toc478729671)

[第二章 相关工作 7](#_Toc478729672)

[2.1 众包 7](#_Toc478729673)

[2.2 众包平台 8](#_Toc478729674)

[2.2.1 InnoCentive 9](#_Toc478729675)

[2.2.2 Amazon’s Mechanical Turk 9](#_Toc478729676)

[2.2.3 Threadless 10](#_Toc478729677)

[2.2.4 威客 11](#_Toc478729678)

[2.3 小结 12](#_Toc478729679)

[第三章 面向残疾人社区的任务分配机制 13](#_Toc478729680)

[3.1 众包协同任务分配策略概述 13](#_Toc478729681)

[3.2 面向残疾人社区的众包任务池 14](#_Toc478729682)

[3.3 空闲优先分配策略生成算法 17](#_Toc478729683)

[3.4 效率优先分配策略生成算法 18](#_Toc478729684)

[3.5 小结 19](#_Toc478729685)

[第四章 面向残疾人工作者的众包管理设计 21](#_Toc478729686)

[4.1 管理权限设计 21](#_Toc478729687)

[4.1.1 平台管理员角色设计与管理权限 22](#_Toc478729688)

[4.1.2 项目创建者角色设计与管理权限 22](#_Toc478729689)

[4.1.3 项目管理员角色设计与管理权限 23](#_Toc478729690)

[4.2 基于角色的管理机制的实现 24](#_Toc478729691)

[4.3 任务结果比对机制 24](#_Toc478729692)

[4.4 小结 27](#_Toc478729693)

[第五章 众研平台的界面设计与实验评估 29](#_Toc478729694)

[5.1 需求分析 29](#_Toc478729695)

[5.2 众研平台环境配置 31](#_Toc478729696)

[5.3 模块设计与界面实现 31](#_Toc478729697)

[5.3.1 用户管理界面实现 32](#_Toc478729698)

[5.3.2 项目管理模块界面实现 33](#_Toc478729699)

[5.3.3 任务管理模块界面实现 34](#_Toc478729700)

[5.3.4 任务标注模块界面 36](#_Toc478729701)

[5.3.5 标注质量控制模块界面 37](#_Toc478729702)

[5.4 面向残疾人众包数据库选择 38](#_Toc478729703)

[5.5 众研平台的部署发布 39](#_Toc478729704)

[5.6 实验分析和评估 42](#_Toc478729705)

[5.6.1 数据集准备与评估方法介绍 42](#_Toc478729706)

[5.6.2 用户评价试验 43](#_Toc478729707)

[5.6.3 面向残疾人社区的任务分配机制效果评估 44](#_Toc478729708)

[5.7 小结 46](#_Toc478729709)

[第六章 总结与展望 47](#_Toc478729710)

[6.1 总结 47](#_Toc478729711)

[6.2 展望 47](#_Toc478729712)

[参考文献 49](#_Toc478729713)

[致谢 53](#_Toc478729714)

# 摘要

近年来，中国越来越重视残疾人的生活品质，我国的“十三五”计划十分重视残疾人的事业，而且还专门设立了章节规划发展残疾人的事业。就业是残疾人最大的保障，是残疾人建立安全感和生命尊严的主要方式，对其心智健康和融入社会都有重要意义。互联网尤其是网络众包的发展，为残疾人就业提供了一个重要的渠道，但是当前的众包标注模式缺乏有效的管理和运转流程，屏蔽了众包工作者之间的社会关系，无法有效利用残疾人社区强耦合的社会关系。残疾人众包工作者无法在众包标注工作中获得存在感，打击其通过众包就业的积极性。

为解决这一问题，本文从残疾人众包标注过程的管理以及运转流程控制的角度出发，设计并实现了面向残疾人社区的众包标注平台。论文的主要研究工作包括以下几个方面：

首先，提出了众包协同任务分配机制。通过对众包标注过程的详细分析，结合协同众包关于流程重构、能力值量化的概念，设计并实现了众包协同任务分配机制，并且根据唇语标注项目的实际需求实现了相应的分配策略生成方案。

其次，设计并实现了面向残疾人工作者的管理模型。本文通过对众包工作者参与标注工作的方式以及不同的职能进行详细分析，借鉴基于角色的访问控制设计思想，并结合项目实际进行过程中众包工作人员的反馈意见，设计并实现了面向残疾人众包标注过程的管理模型。

最后，设计并实现了面向残疾人社区的众包标注平台。通过分析残疾人众包标注需求，采用PostgreSQL数据库解决了标注任务需求和残疾人标注能力属性的异构问题，实现了相应的功能。

**关键字：**众包标注、协同众包、任务分配、管理控制

**中图分类号：**TP3

# ABSTRACT

In recent years, China puts more emphasis on the living quality of the disabled and the national 13th Five-Year Plan attaches great importance to undertakings for the disabled, which specially sets up a dedicated section about the development prospects of it for the next five years. As the best protection and the main way to establish security and dignity of life for the disabled, employment is of great significance to their mental health and social integration. Thankfully, Internet especially crowdsourcing developments provide an important channel for the employment of the disabled. To the best of our knowledge, the current crowdsourcing label models, however, lack effective managements and operation processes and shield the social relationship between crowdsourcing workers that leads to unable to utilize the strong coupling relation in the disabled community. What’s worse, the disabled crowdsourcing workers are difficult to get into existence during the label working which dampens the employment enthusiasm.

In order to solve this problem, the paper designs and implements a crowdsourcing label platform for the disabled community from the perspective of management and operation process control in the disabled crowdsourcing label process. The main work of this paper includes the following aspects:

Firstly, the paper proposes a socialized collaborative task allocation strategy. Through a detailed analysis of crowdsourcing label process, combining concepts about process reengineering and ability quantification in socialized collaboration, the paper designs and implements a socialized collaborative task allocation strategy. Furthermore, the paper also achieves the corresponding allocation strategy generation scheme to meet the actual needs of the lip-tagging project.

Secondly, the paper designs and implements a crowdsourcing management model for the disables workers. Drawing lessons from the role-based access control model, the paper conducts a detailed analysis on different working modes and functions of all the workers involved in the crowdsourcing label work. And then, combining the results with the workers’ feedbacks from the actual process of project, the paper designs and implements a management model for the disabled crowdsourcing label process.

Finally, the paper designs and implements a crowdsourcing label platform for the disabled community. On the basis of analyzing the needs of the disabled crowdsourcing label process, the paper solves the problem of heterogeneity of the label task needs and the disabled’s label ability property by using PostgreSQL database and finally implements the corresponding functions.

**Keywords:** crowdsourcing label, collaboration crowdsourcing, task allocation, management control

**Chinese Library Classification:**TP3

# 第一章 绪论

本章首先介绍了国内关于残疾人就业情况的现状和趋势；之后，总结和分析了众包平台在应对残疾人群体时存在的问题和面临的挑战；最后，介绍了本文的章节安排。

## 1.1 研究背景和问题描述

自2008年起，我国越来越重视改善残疾人的生活质量，为残疾人事业发展投入了很多人力物力资源。我国第十二个五年计划时期，尤其是我党的第十八次全国代表大会召开至今，国家在不断的完善残疾人的各项保障制度，建立保障残疾人日常生活的基本公共服务体系，改善了残疾人的生活质量。残疾人的就业人数也逐年增加，基本收入快速增长，越来越多的残疾人能够参与到社会的各行各业，实现了他们的事业梦想。

虽然残疾人就业情况是越来越好，但不得不承认的是，就目前而言，我国还是有很多的贫困残疾人，只能依靠国家政府的补助解决温饱的问题，尤其是农村残疾人，由于其受教育程度的影响，工作的收入十分低，甚至无法完成就业。目前国家的建设发展不平衡，城乡差距较大，对残疾人的教育、就业培训、康复服务等建设不够充分，对提高残疾人生活质量的建设还需要进一步的完善。

就业对于残疾人群体而言是最有力的保障，也是残疾人建立自信和尊严的手段，对保证残疾人心理健康、融入社会具有重要的意义。但是由于残疾人的身体原因，在生活、出行、工作等方面都受到一定限制，国内众多行业也存在对残疾人不同程度的歧视，导致其就业难度较大[1]。据相关统计，我国当前有8600万的残疾人，处于就业年龄段有3200多万，其中只有1/4实现了就业[2]。

互联网的普及，尤其是威客[3]、众包[4]（crowdsourcing）模式的出现和发展，为解决残疾人就业问题提供了一个重要的渠道[5]。很多残疾人只是肢体残疾，智力方面并无缺陷，这样的残疾人可以通过互联网进行社会交流和学习，最终通过互联网创业或者就业[5][6]。第四产业也就是信息产业的发展，创造了很多适合智力无缺陷的残疾人的工作，比如：信息采集、整理、图文标注等等，同时威客、众包模式的发展为发布这些网络任务提供了渠道，使残疾人通过互联网就业成为一种低成本又有一定的灵活性的方式。

通过对中国残疾人的调研[7]，发现目前已经有许多残疾人通过互联网尤其是QQ的方式从事打字、发帖等工作。但是，为了获取同样的工作，残疾人一般只能采取降低价格的方式，这就直接导致残疾人相对非残疾人同行收入更低，影响相关劳务市场的良性发展。在残疾人通过互联网工作的同时，也通过互联网作为社交媒介，在线上、线下形成了紧密联系的社会关系和网络社区，他们的群体意识更加强烈，具有比其他人群体更多的社会资本。然而传统的威客、众包模式以分散的独立个体为单位，不能很好的发挥残疾人具有强耦合关系和具有较高社会资本的优势，而且个人在成千上万人员参与的大规模众包平台中容易缺乏存在感和归属感，影响了众包的工作积极性。通过上述分析以及对相关工作的总结[8-12]，我们发现将传统众包模式应用于残疾人社区时会存在如下主要问题：

1. 缺乏面向残疾人社区的任务分配机制。对于传统众包平台，众包工作者通过直接查看所有任务列表来选择自己感兴趣的任务来完成，通常在任务未完成之前众包工作者都可以选择来完成该任务，这就造成了会有多个人同时在做同一个任务，最终只有部分众包工作者成功提交结果，浪费残疾人工作者的劳动力。面向残疾人社区的众包标注任务按照项目分类，同一项目的标注任务类型区分度不高，需要根据任务分配给残疾人标注者难易适中的任务，任务过难或者太简单都会打击残疾人参与标注工作的积极性。
2. 缺乏面向残疾人社区众包标注的管理功能。传统的众包模式缺乏管理监控功能，浪费大量的人力资源。众包需求方往往不能直接或者通过指派一些工作者对众包任务进行管理，没有充分利用残疾人之间的社会关系，导致生产过程中任务发布方与残疾人工作者之间没有及时有效的沟通，最终产品与任务发布方的需求可能产生偏差，打击残疾人工作者的自信。
3. 无法提供高效的界面。不同众包项目发布方的需求、任务描述等信息不尽相同，造成残疾人工作者无法直接在平台上完成相关标注工作，只能通过其他途径（例如QQ、网盘等）传输需要标注的数据，同时还需要提供相应的标注工具软件，影响残疾人的工作体验和参与众包的积极性。

综上所述，残疾人就业是当今社会面临的一个突出问题，而互联网行业和网络众包的兴起为大规模实现残疾人就业提供了新的机会和渠道。然而现有的众包缺乏专业的运转流程分析管理，基本上只是任务发布的平台，平台也没有提供任务完成过程的环节，缺乏对任务的全流程的管控，一个组织即便最初是完全以有机形式存在的（通常会带有大量反金字塔体系宣言），但是随着时间推移该组织也会建立起自己的金字塔体系结构。因此，为满足残疾人能够积极的参与到网络众包工作当中，为残疾人就业问题提供渠道，急需一种具有专业的运转流程管理、集任务发布、分配、标注、结果检查一体的众包平台。

## 1.2 国内外研究现状

为解决面向残疾人社区的众包标注平台的需求，本节将对协同众包以及众包任务分配方法的已有研究工作进行分析和总结。

### 1.2.1 协同众包

传统的众包模式主要是通过自发参与和组织的方式开利用分散的社会资源的，但是这种方式缺乏专业的流程分析以及管理，会浪费大量的人力、时间，同时也缺乏对工作者的保障等问题。如果能够从协同的角度出发，通过构建完善的众包工作者角色以及对工作流程进行优化，就能够有效改善传统众包模式对人力资源的利用率[13- 17]。

在众包过程中，众包工作者之间的协作方式[13,25-27]主要有以下三种，第一种是众包工作者互相协作交流解决信任问题。用户需要在众包平台注册账号并填写银行账户信息等内容，用来完成众包任务并获取相应的报酬，平台开发人员一般从技术的角度去考虑用户的账号安全问题，然而如果该众包平台是由其亲友推荐的，那么用户更容易信任该平台的安全性。第二种是众包工作者互相分享任务相关的信息。众包工作者之间并不是完全独立地完成众包任务，他们通常会通过一些社交媒体，例如：SNS，电话，论坛等进行交流协作[28]，由于传统众包平台的API仅提供众包工作者和需求方之间的交流，而不支持众包工作者之间进行有效的分享信息，许多学者也在研究如何在众包过程中加入众包工作者已有的社会关系网[29]，同时这些学者的研究也表明比起令众包平台支持社交，直接利用众包工作者现有的社会关系网络的效果更好。第三种是让众包工作者协作完成任务。众包工作者通常会互相交流，并通过网络媒介重建其社会关系网，并通过所构建的社会关系网协作，提高工作效率和完善关系网络。

有相关研究表明通过众包模式解决复杂问题或者提高众包结果接受率，除了通过提高众包工作者的个人技能之外，还能够通过令众包工作者作为一个团队更加有效的协作。例如，早在2009年的时候，Little等人[30]就设计了一种用来完成手写信息识别任务的众包平台，它通过不断地向下一个众包工作者提供已完成的识别结果以及任务，直到手写信息完全识别为止的这种迭代的方式让多个众包工作者协作降低识别错误率。Ambati等人[13]提出了一种与此类似的协作工作流模型来完成众包翻译工作。关于众包协同问题研究主要是为了解决众包工作者之间协同完成任务，而面向残疾人社区的众包标注平台需要利用残疾人社区紧密的联系以及组织关系，加强对众包标注过程的管理。

### 1.2.2 众包任务分配方式

众包任务分配方式有很多种不同的分类方法[21][31]，通过总结和分析，本文划分为以下四类：

1. 自主领取方式。当任务被发布到众包平台上时，众包工作者可以查看任务的需求，根据个人的兴趣、技能选择想要完成的任务。比如AMT就是采取这种方式分配任务的。这种分配方式实质上是众包工作者即解决了任务，又完成了分配任务的工作。然而这种方式常常会出现多人重复完成同一任务，浪费了大量的人力资源，同时也无法保证每个残疾人工作者的基本工作量。
2. 人工分配方式。我们可以将分配任务的工作看做是一种众包任务，这项任务也需要由众包工作者来完成。这种分配方式有一定的局限性，即只有当完成任务分配工作的人数能够保证做任务的众包工作者及时的领取到任务，才能防止众包工作者长时间处于等待的状态，此外，还需要分配任务的工作人员了解待分配资源的各个方面特征。
3. 自动分配方式。如果能够对参与众包标注过程中的工作者以及任务的各项属性进行标准量化处理，那么就可以将任务分配工作交给计算机来完成。根据分配算法的原理，自动任务分配的算法大致分成了以下三类：基于图论的分配算法、整数规划分配算法、试探性分配算法[21]。根据任务分配的模式，则可以分为随机分配法以及根据优先级分配两类。但是如果仅仅采用自动分配方法，对任务分配处理的不够灵活，无法适应多变的实际需求。
4. 混合分配方式。通过组合或者结合两种或两种以上的任务分配方式，就能够做到合理、有效的分配任务，满足实际多变的分配需求。比如，可以先用自动分配的方式得到任务分配的不同方案，然后由众包工作者或者众包任务管理人员来确定任务分配方案。这种任务分配的方式能够根据实际众包任务的需求来选择不同的任务分配方式进行组合，得到合理、有效的分配方案。

关于众包的任务分配算法的研究，相关研究主要集中在自动分配方式[32-35]。Lee等人[36]提出了一种可以根据任务难度和众包工作者的历史工作经验及其评价进行匹配的方案。但是这种任务分配算法并没有给出可行的量化工作能力方法以及衡量任务难度的具体标准，在系统的初始状态以及对于匿名用户是没有历史工作经验的，因此该方法对于残疾人社区的众包标注平台并不适用。在文献[37]中，Liu等人提出了在生成众包任务分配方案的同时做到成本最低，文中指出对于工作人员是不能够完全信赖的，而对每一位众包工作人员的可信度进行打分，然后根据给定的预算成本求最优解。Roy等人[38]通过一个模糊因子对任务分配进行主观的衡量,但是并没有针对不同的实际应用检测模糊因子的有效性，而且该理论并没有考虑用户已认领任务的数量，这可能会导致任务分配过于集中，不利于应用的有效推进。Zhang等人[39]提出了一种基于多准则的动态任务分配算法（Multi-criter Dynamic Task Assignment），该算法根据众包标注工作者的个人技能、负载等情况得到待完成的工作，然后通过计时器对任务进行计时分配，在计时器归零前自主认领任务。上述任务分配算法生成的方案虽然能够解决任务难度匹配或者成本最优化的问题，但是无法解决面向残疾人社区的众包任务分配的需求：首先需要为每个残疾人分配一定量的基本标注任务量，其次在分配任务时还需要考虑当前用户未完成的任务量，最后要能够满足部分有能力的残疾人多劳多得的需求。

## 1.3 本文工作与贡献

本文工作主要是研究面向残疾人社区的众包标注平台，本文研发的面向残疾人社区的众包标注平台已经投入到实际应用当中。

面向残疾人社区的众包标注平台主要有以下需求，首先，需要对众包过程进行分析研究，设计并实现具有完整流程的众包标注平台。其次，针对残疾人需要设计并实现面向残疾人社区的众包任务管理模块，包括任务分配、领取、查看等功能。再次，由于析众包工作者在众包过程不同阶段的职责，因此需要分析和设计并实现全面的众包标注过程管理监控模块。最后，该众包平台应具备简洁美观的界面，便于相关用户使用。

基于上述需求，本文针对残疾人众包标注的研究工作如下：

1. 设计了面向残疾人社区的众包协同任务分配机制。建立面向残疾人标注工作者的用户模型，设置残疾人标注者的能力值，提出两种不同模式的任务分配方案生成算法。
2. 详细分析和研究面向残疾人社区的众包标注过程。面向残疾人社区的众包标注过程分为创建标注项目、发布标注项目、完成标注任务、验收标注结果、为残疾人支付报酬。
3. 设计了面向残疾人工作者的众包管理功能。构建众包项目创建者、项目管理员、任务管理员、标注工作者四种角色及其功能，借鉴基于角色的访问控制设计思想实现不同角色访问平台的不同功能模块。

## 1.4 本文的组织结构

本文主要分为六个部分，组织结构如下：

第一部分，介绍本文相关的研究背景以及研究意义，协同众包、众包任务分配方法的研究现状以及本文的主要研究内容。

第二部分，首先介绍众包的概念及其构成，其次从国内外相对比较成功的商业众包平台案例出发，介绍现有的众包平台是如何完成众包工作的，分析其成功之处以及应用到残疾人社区的众包标注工作的不足之处。

第三部分，详细描述了面向残疾人社区的众包协同任务分配机制，包括任务、众包工作者及其能力值的模型构建、众包协同任务池以及两种不同优先模式的任务分配方案生成算法。

第四部分，主要根据对协同众包的研究，分析众包过程工作者的不同职责，借鉴面向角色的访问控制思想，设计了面向残疾人社区的众包管理机制。

第五部分，介绍了众研平台的功能概要及其界面设计，并对本文所设计实现的众包任务分配机制以及众包管理机制进行了实验评估。

第六部分，总结本文的研究工作，并对研究中的不足进行反思，为未来的工作提供思路和方向。

# 第二章 相关工作

为了进一步的理解在线众包平台，以及挖掘当前在线众包平台在应对残疾人协作需求存在的问题和挑战，本章首节介绍了众包的概念及其构成，其次介绍并分析了当前主流的商业众包平台。

## 2.1 众包

众包（crowdsourcing）的概念是2006年由Howe[40]首次提出的，众包的定义随着众包的发展也在不断的扩充，指的是由个人、机构或者组织将需要解决的问题或者任务发布给社会大众来解决的一种在线活动。在众包的过程中，众包工作者也就是完成任务的社会大众能够获得金钱或者非金钱的报酬，而任务发布方则通过采集到的任务解决方案来获取相应的收益 [41-45]。众包的基本流程参与者一般包括：发包方，是指有任务需要发布解决的一方；中介，作为连接发包方与接包方的桥梁，通常是众包平台运营方；接包方，是指解决公开任务的一方。众包任务通常是由个人完成的[46]，但当工作需要协作时，众包或以大众生产的方式出现[15]。众包的概念虽然是2006年才提出的，与其类似的模式：悬赏、征集等早已存在，只是接包方受地理因素局限。众包是通过互联网发布任务，全世界的网民都能获取到任务信息。

所有研究中的众包基本上具有以下六个部分[47-53]，虽然具体实现有所差异，但其本质是相同的。

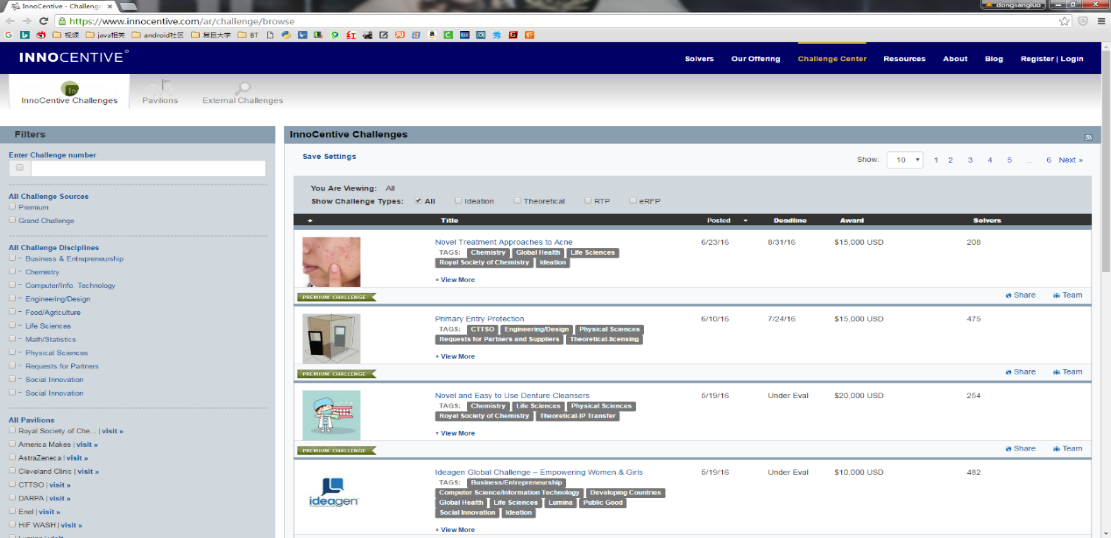
1. 用户。在众包过程的各个阶段均有不同职能的用户参与。众包的用户主要分为三类：众包发布方、众包参与者以及平台运营团队。这三类用户均可能是个人、企业法人或者组织、政府。众包发布方负责创建众包任务，将任务发布给众包参与者。而众包任务参与者进一步可以分为主动参与者（众包工作者）和被动参与者（寻求者），众包工作者直接或者间接的帮助发布方完成众包任务，寻求者以直接查看任务结果的方式获利。
2. 流程。流程是众包活动的重要组成部分。包括从众包开始到完成任务的所有活动。传统的众包流程包括：任务分析、任务发布、解决任务、提交解决方案、评估解决方案以及为获选解决方案支付报酬。伴随着众包形式的不断演变，众包的具体过程可能有所增加，比如在Freelancer.com[48]、互帮网[54]，众包工作者需要先申请任务，只有通过申请的用户才能够提交解决方案。相对的众包过程也可能会减少，比如信息可以自动通过用户安装在移动设备上的应用程序上传，而不需要众包工作者来提交结果。当任务需要多人同时协作完成时，众包的流程也会随之变化，在这种情况下，众包工作者在不同的阶段并且以不同的方式提供解决方案。
3. 任务。任务就是为了解决发包方问题而发布在众包平台上的工作。众包任务种类繁多，难度差异也较大。最简单的任务可以直接是提交现有的数据，比如用户当前位置或者当前位置的无线信号强度等，这种任务甚至可以不需要人工处理，直接由移动设备的传感器自动上传。需要人参与的任务的难易程度也不尽相同，简单到自然语言处理，如找出事件或者自然灾害的相关信息，复杂的情况，任务可能需要众包工作者具有一些专业的知识、技能，如开发应用程序。
4. 内容。内容是众包工作者在完成任务的过程中传输的数据信息。内容和任务的概念是不一样的，任务是众包工作者需要完成的工作，而内容则是众包工作者在参与众包任务过程中接受和产生的数据信息。
5. 平台。平台是连接众包发布者以及众包工作者的在线媒介。传统众包模式是由众包发布方负责开发和维护平台的，随着众包的发展，一些创业家察觉到了众包所带来的商业机会，众包平台就像交易市场一样，为发布方提供发布任务的地方，而众包工作者则可以在这里完成工作，这也就可以为众包平台运营团队在众包活动中提供收益。随着移动设备的发展，众包也出现了移动众包的模式。移动众包通过用户的移动设备发布任务、解决任务，提高了众包的可访问性、可达性和普遍性。同时也有一些众包平台与Twitter、Facebook等社交平台关联起来。
6. 报酬。对众包平台的一种简单分类方式就是分为商业性和公益性，公益性众包的报酬一般是精神的，而商业性众包的报酬一般是物质的。物质报酬又分为金钱和非金钱的。金钱报酬可以是现金或者支票的形式，而非现金报酬则可能是购物券、优惠券之类，或者通过给予积分等方式提高工作者的排名或者声誉等。

## 2.2 众包平台

随着Web2.0技术的发展，网络众包获得了飞速发展，也获得工业界和学术界的广泛关注。尤其在美国，新的众包平台不断涌现，包括非盈利的志愿性质的解决问题的众包平台：例如摄影照片分享平台iStockphoto[55]、词条百科平台Wikipedia[56]等；以及盈利和收费性质的众包平台：例如科学问题平台InnoCentive[57]、AMT[58],upWork[59], CrowdFlower[60]等等。接下来，将对其中影响力较大的几个商业众包平台进行简单介绍，更好地了解商业众包模式。

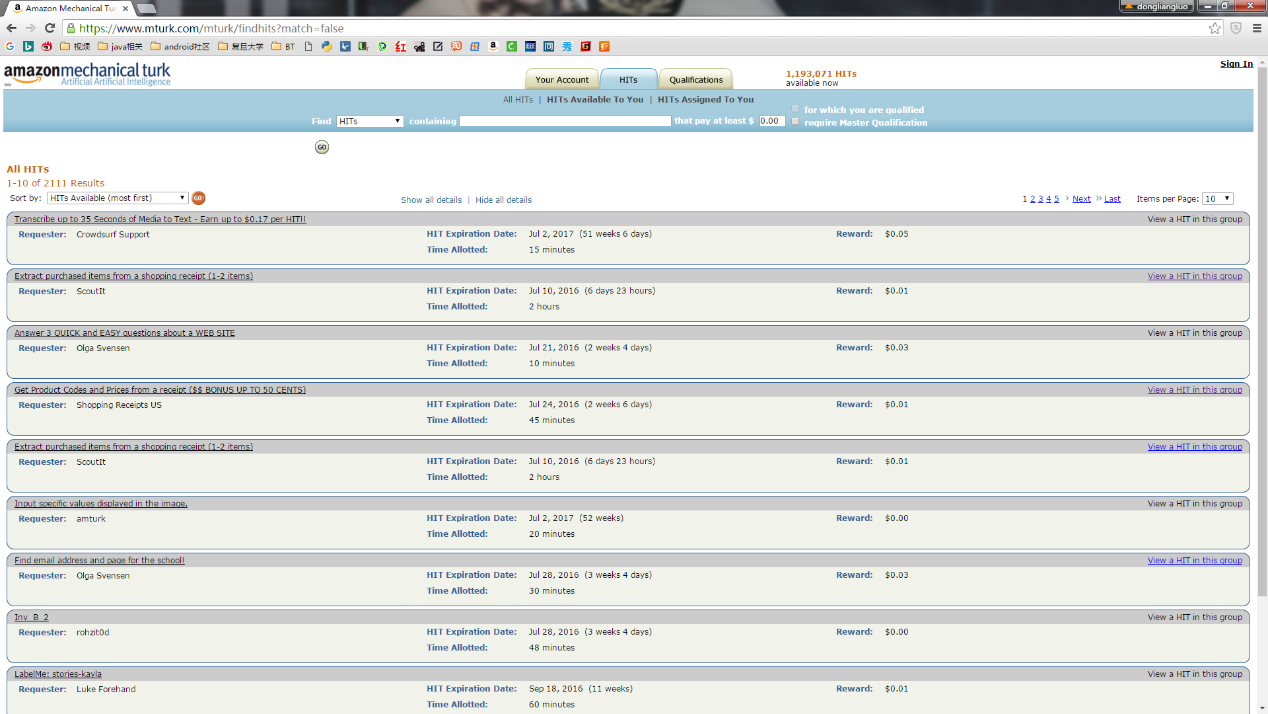
### 2.2.1 InnoCentive

InnoCentive是一家制药公司，把遇到的科研难题通过众包发布到互联网上来寻找解决方案，给予任务参与者一定的金钱激励。一些感兴趣的科研精英们在网站上注册成为“解决者”（Solver），提出了有效的解决方案，并拿走了奖金。这种模式发展成了一个良性循环，现在该网站已经成为了化学、生物领域十分重要的研发供求平台，有来自约200多个国家超过三十五万的科研精英注册成为InnoCentive的解决者，除了科研人员之外还有技师、学生、工程师等，其中65.8%的问题解决者拥有博士学位。与解决者相对应的就是“悬赏者”（Seeker），包括有一些世界著名的跨国公司，接近10%的世界五百强企业都有在该平台发布任务，其中就包括波音、杜邦、保洁等公司。悬赏者通过向缴纳InnoCentive会费以及悬赏任务来解决一些科研难题[61]。该平台主要面向企业科研任务，问题难度较高，需要相当的学识才能够提出有效的解决方案。

图2.1 InnoCentive问题浏览页面

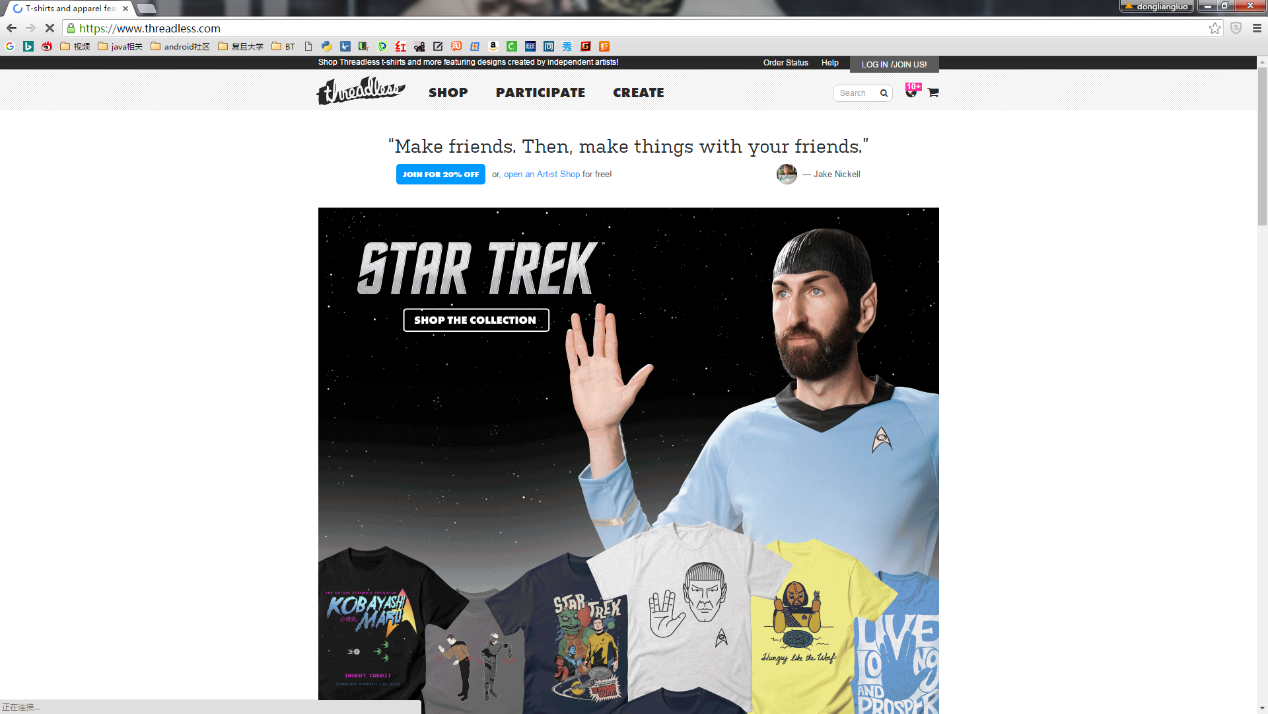
### 2.2.2 Amazon’s Mechanical Turk

AMT（Amazon’s Mechanical Turk）是Amazon在2005年创立的， AMT工作流程如下，首先需求方使用网站的API来发布任务的具体需求以及描述，生成相应的人工智能任务，众包工作者可以通过关键字快速搜索任务。需求方可以限制众包工作者的人数，还可以修改任务的报酬。然后众包工作者查看相关任务描述，完成具体任务并提交结果。最后由需求方审查任务结果，选择接受或者拒绝该方案，如果接受了所提交的结果，就需要支付一定的报酬。在整个过程中，众包工作者和需求方都是通过网站的API和所发布的任务间接沟通交流的，这也就导致需求方无法了解众包工作者的个人特征（比如工作者的性别、年龄、宗教信仰等），除此之外，众包工作者的社会关系也被屏蔽了（比如众包工作者的亲友是否也参与众包任务、是否有人帮助他完成任务、是否与其他工作者分享关于任务的信息等），无法使用残疾人标注工作者紧密的社会关系。

图2.2 Amazon’s Mechanical Turk任务浏览页面

### 2.2.3 Threadless

Threadless[62]是由Jake Nickell和Jacob DeHart于2000年成立的众包平台，总部位于美国芝加哥。Threadless每个星期会将艺术家（包括专业和业余）的约1000件设计发布平台，让公众进行打分投票。七天之后，工作人员得到最高评分的设计。根据众包的反馈，每个星期选取十件左右的设计印在服装或者其他产品上，并通过芝加哥的零售商以及网上商店销往世界各地。获选的设计师则获得该产品每月销售净利润的20%，以及250美元的礼品卡或200美元现金。如果设计重复被采用，则该设计师收到500美元的奖励。该平台主要面向艺术家、服装设计师等，对工作者专业技能要求较高。

图2.3 Threadless首页

### 2.2.4 威客

以上是一些国外比较成功的众包平台案例，2005年，刘峰在中国科学研究生院提出了威客（Witkey）[3]的概念：通过互联网这种媒介，将人的智慧、技能、知识或者技能变为物质或者非物质的收益，最终满足各方需求的一种新模式。威客模式主要是为了解决例如网页设计、LOGO设计、动画和影视的制作、软件开发服务、翻译、取名等工作、生活的各个方面的问题。在我国，已经有超过500万人参与威客模式，成交量已经超过1000万人民币[63]。

国内比较优秀的威客平台有：猪八戒网[64]、孙悟空威客网[65]、互帮网[54]等等。国内的威客平台虽然成熟度也很高，但是，平台发布的信息大多都是直接描述任务需求，众包工作者通过竞争来领取相关任务，在领取任务之后众包工作者无法直接通过网站去完成该任务，而是需要任务发布方通过其他手段将任务资源发送给众包工作者。

图2.4 互帮网首页

## 2.3 小结

本章节对国内外主要的在线众包平台进行了简单的介绍，这些众包平台主要针对的群体是具有一定专业知识或技能的大众群体，也没有充分发挥群体组织关系的优势，比较缺乏针对残疾人群体的考虑和设计。后文将以此为基础结合关于协同众包、众包任务分配方法的相关研究，来设计和实现众包项目的任务分配机制。

# 第三章 面向残疾人社区的任务分配机制

前两章介绍了关于协同众包和众包任务分配方法的研究现状，然后对一些国内外比较成功的众包平台进行了调研，通过对现有众包平台、众包任务分配算法、协同众包的调研，提出了面向残疾人的协同任务分配机制。

## 3.1 众包协同任务分配策略概述

在设计面向残疾人社区的任务分配机制之前，首先需要一个能够全面支持面向残疾人标注需求的任务、用户模型。在面向残疾人标注任务分配过程中，需要综合考虑任务的多样性特点以及难度、优先级、冗余度等属性，残疾人工作者的综合技术能力、以及非技术能力（比如，性格、信誉等）对任务如何分配也有着至关重要的影响。

**众包任务、残疾人工作者定义**

将任务模型定义为一个五元组，该五元组表现为：

**Task=<TaskID,Information, Requirement, Recipients, Preference>**

接下来对五元组的每一个维度进行相应的说明。

**TaskID**为任务的唯一标识符，在进行任务分配及结果检查时均以该标识符作为唯一的检索条件。

**Information**表示的是任务的基本信息，包括任务的创建时间、所属的项目、需要标注的具体内容或者链接、标注说明信息、任务的当前完成状态等信息。

**Requirement**表示众包任务的需求信息，比如对众包工作者的分析决策能力、独立解决问题能力、环境知识获取与应用能力、创新力、计划组织能力等的要求。标注任务对众包工作者的技能需求不尽相同，主要式由任务发布方通过问卷的方式得到需求信息，当然也可以对众包工作者技能无要求。

**Recipient**表示任务分配的众包工作者们的ID。

**Preference**表示任务的设置，包括任务的冗余度、众包工作者自主领取任务时采取的分配方式、任务的优先级等信息。

将众包残疾人工作者定义为一个四元组，其表现为：

**User=<UserID, Information, Role, Capability>**

**UserID**为残疾人工作者的唯一标识符，在进行任务分配时以该标识符作为唯一检索条件。

**Information**表示残疾人工作者的基本信息，包括注册时间、注册邮箱地址、工作者名称、众包工作者QQ号码、残疾人认证等信息。保存众包工作者的基本信息一方面是为了能够标记一些用户；另一方面是为用户提供账户找回功能，防止用户忘记密码。

**Role**表示残疾人工作者的角色，在众包标注过程的不同阶段，用户有着不同的职责或者说具有不同的用户：在众包任务发布阶段，需要有众包需求方和平台的工作人员合作来完成任务的发布、做任务界面的展示等设计（或者根据模板发布任务）；发布任务之后，需要对任务进行分配（可选阶段），在任务分配阶段可能需要有项目管理员来对任务进行合理安排分配；做任务阶段，需要有工作者来认领并完成任务，提交结果；有了结果根据需要还需要安排管理员来检查结果的可信度或者准确性，防止众包工作者提交了无意义的结果。

**Capability**表示残疾人工作者的能力信息，包括工作者的独立解决问题能力、环境知识获取能力与应用能力、计划组织能力、标注相关经验、认真负责程度等信息[66]。其中标注相关经验通过完成任务自动增加，认真负责程度则是通过项目管理员来对其评价打分，其余能力通过评测的方式获取相关考核分数。

**面向残疾人社区的任务分配策略**

项目采取将自动分配与人工分配相结合，并混合主动认领的方式来完成任务分配。首先，对于任务中需要指派给特定的残疾人工作者，或者需要分配给指定的若干残疾人工作者的，平台提供随机分配方案供选择，有人工完成分配。其次，剩余的任务，则是由残疾人工作者自主认领，为了提高任务的完成质量以及改善残疾人工作状态，与自动分配算法相结合，为残疾人工作者分配最适合自身条件的任务。当用户主动领取任务时，采用如下的策略分配任务：

1. 对于注册用户，查看是否有尚未完成的已分配任务，如果有，则按照任务的优先级返回已分配的任务，否则，转至下一步。
2. 检查任务池状态，如果任务池为空，则从数据库中获取未完成的任务，初始化任务池，否则，转至下一步。
3. 再次检查任务池状态，如果任务池为空，则表示任务已经全部分配或完成，无剩余任务可领取，跳转至项目首页，否则，转至下一步。
4. 检查用户登录状态，如果是匿名用户，则按照优先级从任务池中分配最低任务需求的任务，否则，转至下一步。
5. 查找并返回最接近当前用户能力的任务。

## 3.2 面向残疾人社区的众包任务池

如果每次残疾人主动领取任务，众包平台都要查询一次数据库，会严重影响残疾人用户的用户体验，残疾人工作者又具有抱团的特性，会进一步导致过于频繁、大量的数据库请求，有可能会令数据库连接非常缓慢，导致平台响应时间过于漫长，严重影响残疾人工作者用户体验，打击残疾人通过众包标注就业的积极性。如果残疾人主动领取的任务难度普遍高于自己所能完成的难度，其参与众包标注的积极性也会严重受挫。因此，需要众包协同任务池来缓存众包标注任务，在残疾人标注工作者领取任务时，能够快速的返回难度适宜的标注任务。对于众包协同任务池中任务对象的临时存放方式采用的是平衡二叉树的方式。采取平衡二叉树主要是由于残疾人标注者主动认领任务需要在较短时间得到任务，减少残疾人标注者的等待时间，同时又可以根据标注任务的难度的综合评分以及残疾人能力的综合评分分配难度适宜的任务。

|  |
| --- |
| 代码3.1 任务池数据结构 |
| class TaskPull: //面向残疾人社区的众包任务池  def \_\_init\_\_(self, task):  self.hight = 0 //池深度  self.comprehensiveValue = task.getComprehensiveValue() //任务难度  self.leftChild = None //左子池  self.rightChild = None //右子池  self.tasks = [task] //众包协同任务列表 |

其中hight表示以此节点为根的众包协同任务池深度，在本文中规定，众包协同任务池空时的深度为-1，只有一个根节点的池的深度为0，以后每多一层深度加1，comprehensiveValue表示任务的综合难度评价，由任务的Requirement计算得到，leftChild表示任务池节点的左子池，也就是Requirement值比当前节点值小的节点子树，rightChild表示Requirement值比当前节点值大的节点子池，tasks是存放相同难度任务的列表。

在进行面向残疾人社区的众包任务池初始化操作时，需要向任务池中插入从数据库中获取的未分配且未完成任务。插入任务过程中，首先根据任务的需求信息计算得到任务综合难度评价，然后根据综合评价的得分查找节点位置，如果插入位置为空节点，则直接创建新的节点保存任务，然后对池进行调整，使其保持平衡，如果插入位置已经有节点，那么再根据任务的优先级向任务池节点的任务序列中插入该任务，同优先级则直接按照时间先后顺序插入。

|  |
| --- |
| 代码3.2向任务池插入任务 |
| 输入：众包任务task，面向残疾人社区的众包任务池self  输出：面向残疾人社区的众包任务池  def \_insert(self, task):  if self is None : //池空  self = new TaskPull(task)  return self  if self.comprehensiveValue > task.getComprehensiveValue():  self.leftChild.\_insert(task) //难度比当前池节点小，插入左子池  if 2==self.leftChild.\_height()-self.rightChild.\_height() ://维持池平衡  if task.getComprehensiveValue()<  self.leftChild.comprehensiveValue :  self.\_SingRotateLeft()  else  self.\_DoubleRotateLR()  elif self.comprehensiveValue < task.getComprehensiveValue() :  self.rightChild.\_insert(task) //难度比当前池节点大，插入右子池  if 2==self.rightChild.\_height()-self.leftChild.\_height()://维持池平衡  if task.getComprehensiveValue() >  self.rightChild.comprehensiveValue :  self.\_SingRotateRight()  else  self.\_DoubleRotateRL()  else self.\_addTask(task) //难度和当前任务池节点难度相同，加到任务列表  self.hight = Math.Max(self.leftChild.\_height(),  self.rightChild.\_height())//调整池深度 |

当用户主动领取任务，首先根据用户的能力信息Capability计算出综合能力评价值，然后根据用户的综合能力评价值在任务池中查找任务难度最接的任务池节点，再按照任务优先级在任务节点的任务序列中返回相应的任务，最后如果当前任务池节点的任务序列已经不存在可分配任务，那么删除该任务池节点并调整任务池，保持平衡。

|  |
| --- |
| 代码3.3 查找最接近用户能力的任务 |
| 输入：残疾人标注者user，面向残疾人社区的众包任务池self  输出：众包任务  def \_find(self, user):  if self is None :  init(self);//初始化任务池  if user.getComprehensiveValue()>self.comprehensiveValue ://残疾人能力综合评价比当前任务池节点难度高  if self.rightChild is not None :  return self.rightChild.\_find(user)//查找难度更适合的任务  if user.getComprehensiveValue()<self.comprehensiveValue : //残疾人能力综合评价比当前任务池节点难度低  if self.leftChild is not None :  return self.leftChild.\_find(user) //查找更容易的任务  return self.\_getTask() //返回难度最合适的任务 |

## 3.3 空闲优先分配策略生成算法

上一节通过面向残疾人社区的任务池相关算法完成了任务的主动认领方式，对于任务自动分配方案生成，在这里首先我们采用的是一种较为简单的自动生成方案：空闲优先任务分配方案。

空闲优先的任务分配算法，通过结合人工分配，由众包项目管理人员点选该分配方案生成按钮，通过对众包工作者的负载进行简单的判断，其中负载的量化指标选择上，主要是以已分配的任务数量来进行量化。生成任务分配方案后，由众包管理人员决定是否按照该方案分配任务。也就是结合人工分配与自动分配的一种任务分配方法。主要算法思想如下：

|  |
| --- |
| 算法3.1空闲优先分配策略生成算法 |
| 输入：任务集合S，众包工作者集合I,众包工作者已分配任务数集合E  输出：任务分配结果   1. sort list S ,order by priority; 2. for i = 1 to S.length 3. j=1 4. if j > I.length 5. j=1 //从第一个众包工作人员开始继续分配工作 6. if E[j]<=0 //当前众包工作者没有已分配任务 7. Alloc(I[j],S[i++]) //将S[i]分配给I[j],准备分配第i+1个任务 8. else E[j]--; //当前众包工作人员负载-1， 9. end |

该自动生成任务分配方案算法的优势在于：首先，通过计算众包工作者的负载，合理安排众包工作者的负载，分配的总任务数量基本一致，相对于简单随机生成的分配方案更为合理。其次，相对于复杂的考虑任务复杂度、用户经验值等多种信息生成分配方案的方式，空闲优先的任务分配方式简单，需要评估的指数很少，只需要考虑当前众包工作者已分配任务数量,也就是只需要对任务遍历一遍，生成分配方案的速度快。而且对于同一个众包标注项目而言，往往会出现任务难度区分并不大，任务对众包工作者要求较低，比如标注图片中人的眼耳口鼻、四肢等任务。

## 3.4 效率优先分配策略生成算法

上一节主要介绍了一种简单的自动任务分配方案，这种简单的任务分配方案生成速度快，对于众包项目较为简单的标注任务而言生成的分配方案比较合理，能够较为平均的分配任务给众包工作者，每个众包工作者被分配的任务数量相当。但是，并不适用于要求任务尽快完成的需求，同空闲优先的任务分配算法类似，这里也采用自动生成方案之后由人工决定是否按照生成的方案进行分配，具体算法如下：

|  |
| --- |
| 算法3.2效率优先分配策略生成算法 |
| 输入：任务集合S，众包工作者集合I,众包工作者历史工作效率集合E（提交任务结果与分配任务的时间差总和除以被分配任务数得到）  输出：任务分配结果   1. sort list S ,order by priority; 2. E\_sum= Sum(E) //所有候选人的工作效率总和 3. k = MaxIndex(E) //工作效率最高候选人 4. for i = 1 to S.length 5. Ni = Round(S.length/E\_total\*Ei) //每个候选人需要分配的任务数量 6. if sum(Ni) <> S 7. Nk = Nk+S.length- sum(Ni) //剩余任务数加到工作效率最高候选人的任务数中 8. j = 1 9. for i = 1 to S.length 10. if N[j] > 0 11. Alloc(I[j],S[i]) 12. Nj— 13. j++ 14. else 15. j++//当前众包工作人员需要分配任务量足够 16. end |

该自动生成任务分配方案算法的优势在于：通过计算众包工作者的效率，根据其完成任务的速度来分配任务数量，使得任务的总体完成速度能够最快。而且，在分配任务的阶段，依次将任务分配给不同的候选人，使得优先级较高的任务能够优先开始。

## 3.5 小结

本章主要为了解决残疾人众包任务分配问题，首先构建任务、众包工作者以及项目的模型，设计基于平衡二叉树的任务池解决混合任务分配任务方式下，自动、人工结合的任务分配方式与众包工作者主动认领时对同一任务实例重复分配的问题。并根据实际项目的需求构建了两种自动分配与人工分配结合的分配策略生成算法：空闲优先和效率优先的分配策略生成算法。

# 第四章 面向残疾人工作者的众包管理设计

面向残疾人社区的众包标注过程同时具有平台管理员、项目创建者（众包需求方）、残疾人项目管理员、残疾人任务解决者四种角色。各个角色的功能权限各不相同，无法采用传统的单一维度的权限管理。因此需要设计能够满足残疾人众包标注过程中众包工作者的不同管理访问权限问题。

## 4.1 管理权限设计

针对残疾人社区强耦合的特点，残疾人在日常生活交流时会分享众包的工作经历，希望能够快速的访问平台进行演示。同时，能够上网参与众包工作的残疾人总人数相对于实验室服务器的性能来说并不多，为了尽可能的避免系统升级等问题给残疾人工作者带来的不便，因此，本文所设计开发的面向残疾人的众包平台采取了B/S的构架。

平台采用了B/S构架，那么就需要有针对残疾人用户不同角色的权限管理方案。首先，对于面向残疾人社区的众包标注过程，众包工作者种类多样。面向残疾人社区的众包工作者除了残疾人工作者之外，还有非残疾人工作者。根据具体的职能又可以划分为众包平台管理员、标注项目需求方、标注项目管理员、标注任务管理员以及标注工作者四种类型。例如，众包平台管理员需要能够查看平台注册用户是否为残疾人用户以及对该用户进行评价等功能，而其他工作者则不能够访问相应的功能。标注项目创建者可以选择删除所创建的项目，而其他工作者是不能够删除该项目的。其次，同一个众包工作者在众包标注过程的不同阶段所起的职责不尽相同。例如，残疾人工作者可以管理标注项目的同时，完成部分标注任务，或者在标注任务完成之后审核任务结构，平台需要保证这样的用户能够能够同时正确、完整的访问众包平台相关的功能。最后，众包平台前端页面的菜单、功能按钮所包含的内容需要根据不同的众包工作者变化。众包用户不能够访问的部分隐藏起来，让界面显示的内容简洁。面向残疾人社区的众包标注过程具有如上的特点，通过借鉴基于角色的访问控制[67][68]，本文将平台提供的管理功能看作是一种系统资源，不同角色的管理权限问题就可以看做是一种访问控制问题。首先我们需要分析众包标注平台的管理功能资源，然后对众包过程中的众包工作者划分级别或者角色，最后分配对应的访问权限给各个不同角色的众包工作者。

### 4.1.1 平台管理员角色设计与管理权限

这里所描述的平台管理员角色是对标注众包过程的用户而言，而非系统管理员，系统管理员主要负责众包平台本身的运营、维护、安全防护等，平台管理员主要是在对众包过程中的用户、信息等进行访问和管理。

面向残疾人社区的众包标注过程需要对残疾人工作者的用户信息进行管理。为了能够了解众包残疾人工作者的一般情况、发放薪资，平台管理员需要能够查看、编辑残疾人工作者的部分相关信息，包括查看和导出所有残疾人工作者的列表（用户的昵称、全名、邮箱、残疾类型、标注工作者的各项能力值等）、统计残疾人工作者在平台上所做的操作信息、对残疾人工作者的能力值进行评价或者修改等。

面向残疾人社区的众包标注过程需要统计残疾人标注工作者的活动行为。为了以后能够对残疾人参与众包标注过程的行为等进行全面有效的分析，平台管理员需要查看最近用户注册量的统计信息，查看众包平台的活跃用户数目统计信息，查看用户平均访问时间间隔，活跃用户中非注册用户的数目统计信息，项目发布的任务量的统计信息，完成任务数量的统计信息等。

面向残疾人社区的众包标注过程需要更换平台管理员。由于残疾人社区群体的人数较多，而且残疾人工作者参与众包过程时具体职责也不是一成不变，平台管理员角色的分配和回收不能完全由系统管理员来完成，需要平台管理员角色能够通过搜索用户的全名或者昵称来赋予其他残疾人工作者平台管理员的角色，以及将转移到其他角色的残疾人用户从平台管理员角色中剔除。

面向残疾人社区的推广众包标注项目需要及时更换。不同的标注项目时效性不同，平台需要进行推广的标注项目也不是一成不变，为了让残疾人标注工作者更方便的选择适合自己的标注任务，平台管理员需要对标注项目进行分类、推广（优先显示）以及修改分类信息等功能。

### 4.1.2 项目创建者角色设计与管理权限

面向残疾人社区的众包标注平台接收到社会人士、企业或者政府的标注任务需求时，需要创建相应的标注项目，就需要有项目创建者的角色来完成相应的功能。

面向残疾人社区的众包标注项目需要对项目相关信息进行管理。当残疾人标注者参与标注工作时，不是简单机械的完成标注工作，他们希望能够通过标注工作体现自己的价值，了解所完成的相关任务是具有意义的。因此，需求方自身或者通过相关众包工作人员创建标注项目时，需要能够添加或者修改项目的名称、简介、任务描述等信息，除此之外，为了方便残疾人标注者选择自己感兴趣的或者擅长的标注工作，需要对项目进行分类。需求方本身也需要查看项目所发布、分配以及完成任务的情况。

面向残疾人社区的众包标注项目需要对用户进行管理。为了更好的管理标注项目，需要添加项目管理员、任务管理员来帮助完成项目的管理工作，或者是移除一些不够认真负责的管理员。有的标注项目本身具有一些信息安全需求，需要通过设置来限制匿名、非认证用户的访问，只允许通过平台认证的残疾人标注工作者来访问相关数据完成标注工作。

面向残疾人社区的众包标注项目需要对标注任务进行管理。众包标注项目的需求不同，所要完成的标注任务各自迥异，针对不同的标注任务能够设置不同的任务展示界面和结果提交的格式。具体的众包标注任务也需要人为的上传到平台，对于残疾人标注者提交的结果也要能够导出到Excel，供需求方使用。此外还需要对残疾人标注者领取任务的次序应该要能够和任务的优先级保持一致，当需求方对任务的优先需求改变时，标注项目中发布的具体任务优先级能够随之改变。

### 4.1.3 项目管理员角色设计与管理权限

众包标注项目，不可能完全依靠项目创建者一个人进行管理。具体标注任务的数量通常是非常巨大的，又众多残疾人标注者分别完成部分标注任务并提交结果，控制数量巨大的标注任务以及检查相应的标注结果的正确性工作量是十分巨大的，这就需要有众包项目管理员来协助项目创建者来完成相应的工作。

面向残疾人社区的众包标注项目管理员需要辅助项目创建者共同管理当前项目。在创建初期，项目创建者赋予残疾人工作者当前项目管理员的角色，项目管理员需要帮助项目创建者上传标注任务，添加或者修改残疾人标注者完成标注工作的界面，以及提交结果的格式，设置只能够允许通过平台验证的残疾人标注者查看完成标注任务。通过设置让不同的残疾人标注者完成同一个标注任务来对结果进行比较，提高任务结果的准确性，同时对残疾人标注工作者的相应技能进行评价，间接的控制残疾人标注工作者领取的标注任务难度。

设置标注任务需要重复完成的次数，浏览查看残疾人标注所提交的结果，设置具体标注任务的

对于项目管理员，其主要职责就是辅助项目发布者共同管理当前项目下发布的任务，设置任务的调度方式：广度优先（任务剩余需要完成次数、优先级依次）、深度优先（按任务优先级依次）；设置任务的优先级；设置任务冗余度；查看任务状态；导入任务；设置定时自动导入任务；导出任务答案；设置是否允许匿名用户参与；编辑任务展示界面；浏览任务相关信息；查看任务结果、查看所属完成任务数量的统计信息；评价、修改用户的能力信息。

## 4.2 基于角色的管理机制的实现

根据前一节对众包用户的角色功能（权限）的设计，首先，我们建立了平台管理功能表，用来存放管理机制的最小单位：功能（权限），比如查看所有用户的列表、修改项目名称、对项目进行分类、设置任务的优先级等。其次，建立角色功能表，用来控制具体角色所具有的功能权限。最后，在第三章中以及具体说明了用户表与角色的关联方式。通过这三张表以及具体的信息来完成对角色的管理权限功能的控制，在前台页面则直接根据是否有权限访问该功能模块来显示或者隐藏具体的控制模块视图。

图4.1 数据库结构图

## 4.3 任务结果比对机制

首先从任务表以及任务结果表中获取项目的任务提交者ID，任务提交者全名、任务的完成状态、任务冗余度、任务分配对象以及与任务提交者相对应的答案：

|  |
| --- |
| 代码4.1 获取任务以及任务结果列表 |
| 输入：项目编号project\_id  输出：任务及其结果列表  def browse\_tasks(project\_id):  sql = text('''SELECT browse\_task.id,"user".fullname as fullname,  browse\_task.state,browse\_task.n\_answers,  browse\_task.task\_answers,browse\_task.recipients  FROM( SELECT task.id,task\_run.user\_id ,task.state,  task.n\_answers,task\_run.info as task\_answers,  task.recipients FROM task LEFT OUTER JOIN task\_run  ON(task.id=task\_run.task\_id)  WHERE task.project\_id=:project\_id ) browse\_task  LEFT OUTER JOIN "user" ON (browse\_task.user\_id="user".id)  ORDER BY browse\_task.id''')//数据库中获取数据  results = session.execute(sql, dict(project\_id=project\_id))  tasks = []  for row in results:  if len(tasks) > 0 and tasks[-1].get('id') == row['id']:  pre\_task = tasks[-1]  pre\_task['userName'].append(row['fullname'])//用户名  pre\_task['details'].append(row['task\_answers'])//结果  else:  task = dict(id=row.id, userName=[row['fullname']],  recipients=[], state=gettext(row.state),  n\_answers=row.n\_answers ,details=[])  if row['recipients']://接收当前任务的用户  for id in row['recipients']:  task['recipients'].append(db.session.query(User)  .get(id).fullname)  task['details'].append(row['task\_answers'])  tasks.append(task)  return tasks |

结果一致性方面检查是按照如下方式进行的，首先判断由数据库返回的任务结果数组长度，如果任务结果唯一，那么在网页上显示未检查：

|  |
| --- |
| 代码4.1 对比任务结果序列是否一致 |
| 输入：任务及其结果value，数字类型结果误差允许范围precent  输出：是否一致  function resultFormatter(value) {  if (value.length < 2) { //任务结果未提交或者只有一个  return "未检查";  }else{  var result = "一致"  var precent = document.getElementById("precent");  for(var i= 1,l=value.length;i<l;i++) {  $.each(value[i], function (key, val) {  if(!isNaN(val)){ //数字类型  if(Math.abs(value[0][key]-val)> precent \*val){  result= "不一致";//百分比比较  return false;  }  }else if(val !== value[0][key]) {//字符类型直接比较  result= "不一致";  return false;  }  });  }  return result;  }  } |

任务结果序列是按照Json格式存储的，同一个任务的不同众包解决者提交的结果系列的内容以key-value方式存放，对比采用JsDiff的Javascript包来完成，显示则是以灰色代表相同部分，绿色表示的部分代表两者比较前者比后者多出的内容，以红色表示后者比前者多出的内容。

|  |
| --- |
| 代码4.2 多序列比对结果显示 |
| 输入：任务及其结果row，序号index  输出：格式化显示结果  function detailFormatter(index, row) {  var result = JsDiff.diffJson(value[0], value[1]);//比较两结果差异  var display = document.createElement('p');  result.forEach(function (part) { //显示颜色  green for additions //绿色前者比后者多的内容  red for deletions //红色后者比前者多的内容  grey for common parts //灰色相同部分  var color = part.added ? 'green' :  part.removed ? 'red' : 'grey';  var span = document.createElement('p');//前端显示  span.style.color = color; //颜色  span.appendChild(document.createTextNode(part.value));  display.appendChild(span);  });  return display.innerHTML;  } |

## 4.4 小结

本章主要介绍了基于角色的管理机制设计，通过对众包过程的四个不同角色的职能进行分析，明确各自对平台服务功能的访问权限，并且通过提供对序列比对任务结果检查算法，提出了一种完善的面向残疾人社区的标注众包管理机制。

# 第五章 众研平台的界面设计与实验评估

前两章主要介绍了面向残疾人众包标注过程的任务分配机制和管理模块设计，基于前两章的工作，本章主要分析了众包标注平台的需求，实现相应模块和界面，完成众研的开发，最后将其发布到生产环境供残疾人使用。

## 5.1 需求分析

面向残疾人社区的众包标注平台需求，除了平台系统本身需求之外，还有非功能性需求，比如尽可能的改善残疾人工作者的用户体验，避免残疾人工作者长时间疲劳用眼等。本文主要研究的是面向残疾人社区众包标注平台系统本身的需求。

图5.1 系统整体功能概要图

图5.1主要展示了面向残疾人众包标注平台系统的功能概要，面向残疾人的众包标注平台用户分平台管理员和众包用户两类。平台管理员除了负责平台的更新维护，还有认证、管理残疾人工作者，对分类不正确的项目进行修改分类，推广具体的标注项目等。而众包用户既有残疾人工作者也有一些匿名用户来完成一些公益的标注任务。针对图5.1中展示的系统功能概要，本文将面向残疾人众包标注平台功能划分为用户管理模块、众包标注项目管理模块、众包标注任务管理模块、众包任务标注模块以及众包标注质量控制模块。

面向残疾人众包标注过程的众多用户以及不同工作者的不同功能需求，需要有用户管理模块来完成相应的工作。用户管理模块提供了用户管理、角色分配、角色功能管理的功能。对残疾人用户提供注册账号、修改个人资料（邮箱、密码等）、密码找回、删除个人信息以及提交残疾人认证申请的功能。为了保证残疾人工作者能够受到平台所提供的服务和保障，需要平台管理员通过用户管理提供的功能认证残疾人工作者，同时统计查看标注平台用户的个人信息、对残疾人工作者在平台上的操作行为进行跟踪记录，以便进一步分析残疾人工作者的行为模式，改善残疾人工作者的工作体验。标注任务需求方发布任务时由角色分配功能自动分配其项目创建者的角色，项目创建者通过角色分配模块分配项目管理员和任务管理员的角色给指定用户协助其完成众包标注过程。

面向残疾人众包标注平台通过项目管理模块为需求方提供管理功能。为使残疾人标注者能够根据自己的喜好选择感兴趣或者擅长的标注任务，项目创建者和项目管理员通过项目管理模块对项目进行分类，增加项目的详细信息。通过为项目写博客的方式可以向残疾人标注者说明标注任务的具体需求以及如何完成标注工作等。

众包标注任务管理模块包括任务的上传、查看、分配、导出任务的功能。面向残疾人的众包标注任务的数量比较多，项目管理员具有删除项目的权限，所以分配过多项目管理员影响项目的安全性，仅靠项目创建者和项目管理员不足以有效的管理标注任务，需要任务管理员来协助对标注任务进行管理。标注任务需求方需要更具残疾人标注者的工作量发放薪资，为保障残疾人都能够获取最低保障工资，任务管理员需要通过社会化任务分配机制向申请参与当前标注项目工作的残疾人工作分配最低工作量。除完成分配的任务外，残疾人工作者也可以选择通过任务管理模块额外认领标注任务以获取额外的收入。

众研平台通过任务标注模块来控制不同标注项目的具体任务展示界面，以及残疾人工作者完成标注任务的界面，提交标注结果。众研任务标注模块为众包标注任务提供了图片标注、语料标注的模板，众包需求方可以通过使用模板快速发布众包标注任务，当模板无法满足标注任务的需求时，还可以使用标注模块提供的任务展示编辑器设计编写标注界面，供残疾人标注者完成标注任务。

为提高众包标注结果的质量，众研平台增加了标注质量控制模块。针对残疾人众包标注任务对标注结果准确度的需求，当残疾人标注者标注区分度较小的任务时结果的误差较大，需要让不同的残疾人标注者完成同一个标注任务，通过对比多个标注数据得出较为理想的标注结果，根据最终接受的标注结果与每个标注者结果的差异，对残疾人工作者的标注能力进行评价。

## 5.2 众研平台环境配置

明确面向残疾人众包标注平台的各个模块功能需求之后，接下来需要完成的就是基础开发环境的搭建。众研的开发采用的是Flask框架，Flask是由Python语言开发的Web框架，同时应用Flask-Login、Flask- SQLAlchemy、Flask-Cache等Flask的扩展我们搭建并发布了自己的众包工作平台：众研。开发环境需求如下：

Ubuntu 14.04 LTS

Python >= 2.7.6, <3.0

PostgreSQL >= 9.3

Redis >= 2.6

pip >= 6.1

其他软件安装步骤如下：

sudo apt-get install git-core

sudo apt-get install postgresql postgresql-server-dev-all libpq-dev python-psycopg2

sudo apt-get install python-virtualenv

sudo apt-get install python-dev build-essential libjpeg-dev libssl-dev swig libffi-dev dbus libdbus-1-dev libdbus-glib-1-dev

到此众研平台的基本环境就搭建完毕。

## 5.3 模块设计与界面实现

搭建好众研平台的基本开发环境后，接下来需要完成的就是实现具体的各个功能模块以及界面的设计。功能模块的开发主要都是业务逻辑的开发，本章第一节已经对各个模块的业务逻辑进行了详细的说明，直接实现即可。界面的设计我们采用了Bootstrap帮助我们完成界面的布局，对于功能模块的界面，尽量将众包工作者不具有的功能全部隐藏起来，以减少残疾人工作者的误操作。

### 5.3.1 用户管理界面实现

用户管理主要用来管理众包用户的相关数据，通过基于角色的权限控制和项目以及任务的访问控制来管理用户对众研平台的模块的访问。该模块首先具有一些常规的用户管理职能，例如：众包用户的登陆、注册、登出、修改个人信息、账户找回、修改密码等功能，登陆注册界面如下图5.2所示。

图5.2 用户登录注册界面

平台管理员需要通过用户管理模块来导出导出用户的相关信息以及为用户分配或者移除管理员的角色，但是该功能使用的频率较低，在设计界面时将这两个功能按钮放在管理员功能汇总界面的最后。

图5.3 平台管理员角色职能汇总界面

在众研平台中，设定了平台管理员、项目发布者、项目管理员、任务解决者四类用户角色，分别进行管理。

图5.4 平台管理员角色管理界面

项目创建者添加删除项目管理员角色、以及项目管理员添加删除项目管理员与平台管理员角色管理界面类似，都是通过搜索用户的全名或者昵称来赋予用户相应的管理权限的。

### 5.3.2 项目管理模块界面实现

众研项目管理模块主要是用来创建、修改和删除项目的，创建项目的功能是对所有注册用户开放的，所有在众研平台注册的用户均可以创建众包项目，成为项目创建者，创建好项目之后就可以在我的项目中看到当前用户所发布的所有项目。项目管理员可以通过点击项目进入项目的管理界面，也可以通过点击位于平台界面右上角的用户下拉标签页中“我的项目”进入项目管理界面。

图5.5 用户下拉标签页

对于项目创建者，则可以看到关于项目的设置界面，如图5.6所示。项目设置界面具有编辑项目详情、任务设置、删除项目以及为项目写一篇博客等按钮，其中任务设置标签则是指向了任务管理模块界面。非项目创建者是无法看到项目设置界面的。

图5.6 项目设置界面

### 5.3.3 任务管理模块界面实现

从图5.7任务管理界面中可以看到具体项目的任务标签页下具有导入任务、导出任务、任务演示、设置、浏览、自动导入任务、任务派发、查看的功能按钮。导入任务具有从本地CSV文件、网络CSV文件链接、EpiCollectPlus项目、谷歌电子表格的方式，以及提供的声音模式识别、地理编码、图片模式识别、视频模式识别、抄写文件等模板方式。导出任务可以将任务或者任务结果以CSV或Json的格式导出到本地供需求方使用。任务演示是指向任务标注模块界面的链接。

图5.7 任务管理界面

下图5.8展示了任务设置的子界面，在任务设置界面任务管理员可以完成根据具体项目的需要修改残疾人标注者自动认领的任务顺序、改变任务的优先级、删除任务及结果、设置任务的冗余度的操作。

图5.8 任务设置

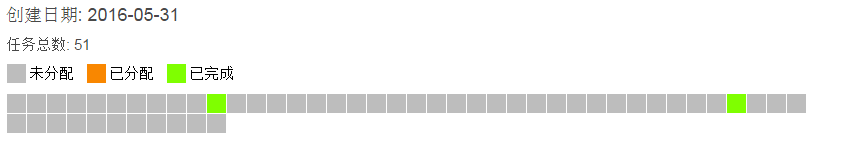
图5.9展示了查看任务界面，以图像化的方式展示任务的分配、完成状态的信息。

图5.9 查看任务

### 5.3.4 任务标注模块界面

众研平台需要为众包标注需求方提供能够个性化设置的任务展示、标注的界面，为方便部分常见的标注任务能够简单快速的发布，任务标注模块提供了图像识别、语音识别、视频识别、地理位置标识、PDF文字识别的众包标注任务模板，如图5.10任务标注模板所示。

图5.10 任务标注模板

众研平台提供的任务标注模板不能够完全满足多样的残疾人众包标注任务需求，为了能够更加灵活的设计标注任务展示以及残疾人工作者标注界面，众研平台还提供了通过编写WEB应用前端代码HTML+CSS+JavaScript的方式。在唇语标注项目案例中，众研就通过这种方式设计残疾人标注工作者完成标注工作的界面，前端代码编辑界面如图5.11所示。

图5.11 自定义任务演示

### 5.3.5 标注质量控制模块界面

图5.12展示了众研平台的标注质量控制模块的界面，残疾人众包标注任务难度相对较低，只有当标注任务的区分度较小时，需要多人重复完成统一标注任务，由标注任务管理员通过标注质量控制模块提供的多结果比较功能对所提交的标注结果进行检查。众研标注质量控制模块主要是为了帮助标注任务管理员查看提交的任务结果、多个结果的比对是否一致，以及查看具体任务的连接功能，界面设计通过表格的方式显示。

图5.12 质量控制模块界面

## 5.4 面向残疾人众包数据库选择

选择数据库是个长期的决策，因为在系统完成并上线之后再想要做数据库的迁移是非常困难的，所以最好就是在平台开发之前就选择最合适的数据库。MySQL和PostgreSQL成为了最后需要抉择的产品。

首先，当众研发布了众包标注项目通知时，残疾人通常都是抱团完成标注任务的，短时间内用数量众多的残疾人访问众研平台，完成标注任务需要从数据库中获取标注任务的相关信息，提交标注结果的时间也较为集中，数据库需要在高并发读写，负载较大的情况下，性能指标不能明显下降影响残疾人标注者的体验。MySQL数据库在这种情况下，性能指标有明显下滑的情况，而PostgreSQL能够维持较高的性能。

其次，为支持异构的面向残疾人社区的众包标注任务展示需求，众研平台需要提供自定义编写WEB应用前端代码HTML+CSS+JavaScript的方式来完成，具体项目的展示页面也不尽相同，在数据库中需要保存该展示页面的代码，前端代码的长度较长，MySQL数据库字段最大长度限制为8K左右，其无限长的TEXT字段类型只能作为外部数据来访问，功能限制很多，相对而言PostgreSQL的 TEXT字段类型没有长度限制能够直接访问，而且SQL语法支持正则表达式、支持全文检索，优势明显。

再次，面向残疾人社区的众包标注任务对残疾人标注者的能力需求不尽相同，残疾人的能力标签需要根据标注项目的不同而改变，同时标注任务的结果格式也不仅相同，为了灵活设置导出的任务结果格式，众研平台采用Json格式来存储残疾人标注者的能力值标签以及提交的标注任务结果，而相对于一些固定的属性例如用户名、电子邮箱、是否为认证残疾人等则需要传统关系型数据结构来存储。PostgreSQL既具有传统关系型数据库的数据结构，又能够直接存储数组与Json，与众研平台的需求非常吻合，而MySQL不支持Json格式的字段类型。

最后，众研平台不仅仅只是为了让残疾人参与众包标注工作，更重要的是改善残疾人参与众包标注过程，让残疾人能够体现自己的能力和价值，建立自尊。平台需要通过分析记录的残疾人参与众包标注过程的操作行为、完成任务的质量等信息，来达到次目的，而PostgreSQL支持复杂SQL编程，具有非常丰富的统计函数以及语法的支持，还可以用多种汇编语言来写存储过程，对R语言的支持也很好。而MySQL只能采用SQL来编写存储过程，也不支持许多分析函数。

通过以上对比，在众研平台中，我们选择了PostgreSQL数据库来存储数据。

## 5.5 众研平台的部署发布

残疾人工作者在接收到标注任务的通知时，通常都是抱团一起登陆众研平台完成相关的标注工作的，需要将众研平台发布在相应的生产环境供残疾人使用。Python应用的发布方式有很多种，性能基本都能够满足残疾人社区的访问量需求，本文将平台使用nginx+uwsgi+supervisord的方式部署在Linux系统上。nginx是一个http服务器，uwsgi是http服务器与Python应用程序之间进行数据交换的服务程序,supervisord是用Python实现的一款进程管理工具。supervisord会把管理的应用程序转成daemon程序，可以方便的通过命令开启、关闭、重启，而且它管理的进程一旦崩溃会自动重启，这样就可以保证程序执行中断后的情况下有自我修复的功能。nginx+uwsgi+supervisord的部署方式相对比较简单，能够快速部署众研平台，而且相关资料也十分丰富。

1. **nginx安装配置**

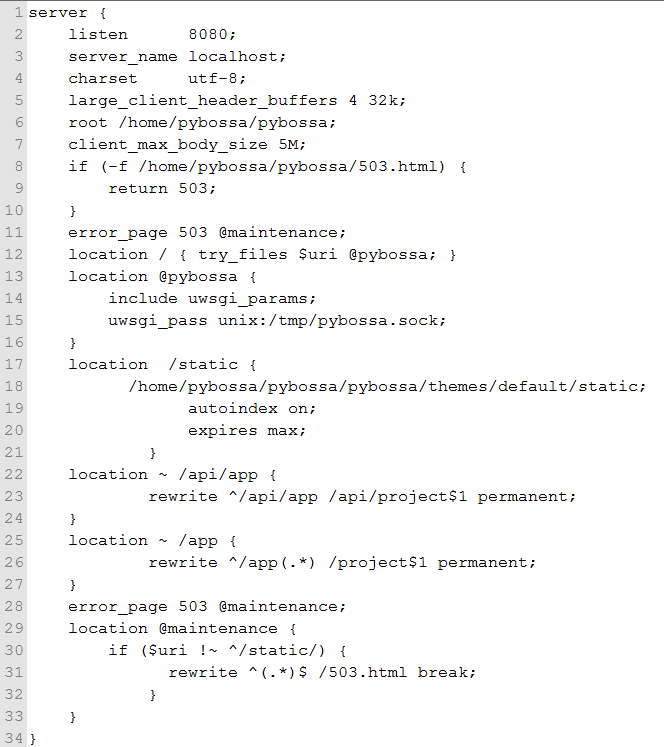
安装nginx

sudo apt-get install nginx

启动nginx服务器：

sudo service nginx start

配置nginx

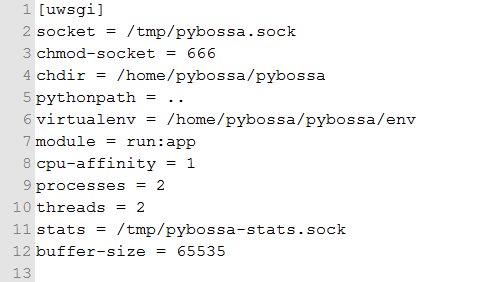
 进入/etc/nginx/sites-enabled/文件夹生成包含以下内容的文件：

1. **uwsgi安装配置**

安装uwsgi

sudo apt-get install uwsgi

编写uwsgi配置文件



nginx和uwsgi都配置好了之后启动nginx和uwsgi了。

启动nginx

sudo service nginx restart

启动uwsgi

uwsgi wsgi.ini

若一切正常的话就可以在终端上看到uwsgi的启动信息了。

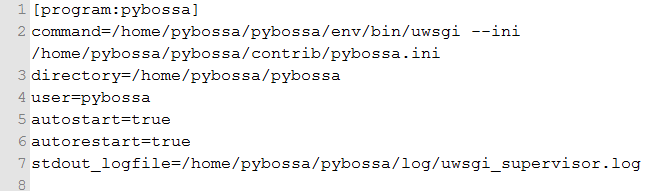
1. **进程监控**

安装supervisor

sudo apt-get install supervisor

生成默认配置文件

echo\_supervisord\_config > supervisord.conf

加入监控程序的配置

启动supervisord

supervisorctl reload

检查uwsgi进程是否正常运行

ps aux|grep uwsgi

如果一切正常，此时应该可以看到uwsgi进程。尝试kill掉uwsgi进程supervisor会重新启动一个新的uwsgi进程。

sudo killall uwsgi

再次通过ps aux|grep uwsgi查看发现有新的uwsgi进程在运行。至此就成功的将众研平台部署到了生产环境。

## 5.6 实验分析和评估

针对本文提出的面向残疾人社区的众包标注平台，本文设计若干相关实验用于验证和对比本文所提出的标注平台是否能够达到预期的效果并且能够比传统众包标注平台更好的解决第一章问题描述所提到的相关问题需要进一步的验证和评估，因此本文所实现的面向残疾人社区的众包标注平台除了完成基本需求之外，其使用性能也是衡量系统可用性的一个重要指标，除此之外，对系统的性能进行评估还具有如下几个方面的意义，首先在使用新的众包标注平台时，通过相关性能测试可以分析平台的性能特点，调整软硬件设置，最大限度的发挥软硬件性能，其次在平台评估过程可以及时发现性能变化点，针对性能瓶颈进行改进提高，最后在平台使用过程中也可以更好的获取用户需求。

### 5.6.1 数据集准备与评估方法介绍

针对本文要进行的实验分析，本文选择唇语标注项目的3146个图片标注任务以及身体部位标注项目的6709标注任务作为测试数据。每个唇语标注任务包含一个图片序列，其内容是一个人读八位数字验证码的口型，标注其中每个读数在该图片序列中的起止位置标号。身体部位标注任务包含一张图片，其内容是一个人的全身照，需要标注图片中人物的头、颈、左右肩、左右手腕、左右手肘以及胸部。这两个项目是面向残疾人社区的众包标注任务中具有代表性，身体部位标注任务相对较为简单，任务之间的难度区分并不大，而优先级别和准确度需求有一定的差别，唇语标注任务的难度相对较高，任务的图片序列数量差别较大，对残疾人标注者的能力需求也有较大差别。

标注任务数据准备完成之后，接下来需要做的就是对众包标注者的能力进行评估。对众包工作者能力的评估方法主要有两种，一种是在众包工作者参与众包任务之前，利用能力评测试题对众包工作者的能力进行分析评估，另一种是通过在众包任务中加入部分测试任务，通过对工作者提交测试任务的结果与正确结果进行比较，评估众包工作者的各项能力指标。本文采取综合上述两种评估方法对众包标注者进行评估，将评测指标划分为三大类，基本素质类、工作态度类、工作能力类。基本素质类，占综合评分的30%，评价选取学习能力、理解能力、解决问题能力三项指标，通过评测试题得到具体的评分。工作态度类别占综合评分的20%，选取积极性、责任心两项指标，由众包项目管理员对其进行评分。工作能力类别评价占综合评分的50%，选取工作数量、速度、质量三项指标，工作数量以及速度由系统计算并更新，质量指标根据项目管理员以及任务管理员对残疾人标注者提交结果的采纳率得到。

本文提供的面向残疾人社区的众包标注平台的评估主要有一下几个方面标准，首先面向残疾人社区的众包标注平台是否既能够满足所有残疾人标注者一定的任务数量又能在一定程度上令残疾人标注者多劳多得，其次面向残疾人社区的众包标注平台是否能够有效解决还有未完成任务时不同残疾人标注者重复做同一任务，最后面向残疾人社区的众包标注平台是否能够充分发挥残疾人群体的社会关系的优势，对众包过程进行有效管理，提高众包的效率。

针对面向残疾人社区的众包标注平台的分析评估主要包括两个方面的评估。一个方面是针对面向残疾人工作者的众包管理效果评估，通过对参与评的实验室同学以及残疾人工作者的采访调查评估管理功能的效果，并且更具采访调查结果对面向残疾人工作者的众包管理模块进一步的完善。另一个方面是针对面向残疾人社区的任务分配机制的评估，通过分析参与评测的标注者完成任务的情况，验证面向残疾人社区的任务分配机制的效果。

### 5.6.2 用户评价试验

标注用户在残疾人工作者中随机抽选出18人，并且邀请了8位学生参与本文的用户评价试验，标注任务选取了身体部位标注项目的六千多个任务。参与实验的残疾人工作者和学生都或多或少的拥有图片标注任务的经验。这些候选人被要求完成发布在AMT、本地标注软件以及众研平台上的标注任务并进行打分。评价的打分制为1-5分，其中1分为最差，5分最好。每个候选人完成一定数目的标注任务之后会被要求做出评价。评价主要关注连续做标注任务的用户体验、用户误操作状况、众包标注过程的总体体验评价。表5.1展示了众研平台的用户打分评价结果，而与AMT、本地标注软件的对比试验则可以通过图5.13获得。通过对参与评价的用户做问卷调查发现，相对于AMT和本地标注软件，用户认为在众研平台连续标注任务更加方便，只要提交结果就能进入下一个任务，而且分配的任务难度相对更加合理；众研平台提供了丰富的用户角色以及相对应的功能，能够对众包标注的全过程进行有效的管理，实时监控任务完成进度，并能够对提交的标注结果进行初步的检查，有效节省了任务结果验收时间。结合表5.1、图5.13以及调研的结果可以看出，相对于AMT和本地标注软件，本文设计实现的众研平台具有较强的吸引力和竞争力。

表5.1 面向残疾人社区的众包标注平台用户评价

|  |  |  |
| --- | --- | --- |
| **评价标准** | **评价均值** | **评价标注差** |
| **连续做标注任务体验** | 4.12 | 0.291 |
| **误操作状况** | 4.35 | 0.282 |
| **总体体验评价** | 4.51 | 0.276 |

图5.13 用户评价对比结果

### 5.6.3 面向残疾人社区的任务分配机制效果评估

对众包标注任务的合理分配，既能够合理安排残疾人标注者的工作量，保证残疾人工作者能够获得基本的收入保障，又能够让有能力的残疾人标注者多劳多得，获得额外的收入。同时，根据残疾人标注者的能力分配与之相匹配的标注任务，在一定程度上也能够提高任务结果的质量。本小节主要针对标注者的完成标注任务数量、平均完成时间、标注任务结果的接受比率进行评估对比，根据对比的结果来评估本文所提出的对面向残疾人社区的任务分配机制的效果。

图5.14 标注任务总数

从图5.14可以看出，当用户使用本地软件标注时，最多只能是完成预先分配给他的全部任务，无法根据自己的情况选择多做任务。而用户在AMT上做任务则不能保证都能够做到一定数量的任务，如果用户觉得累需要休息或者有紧急情况需要处理，则只能做非常少的任务，而在众研平台，在项目发布之后可以由任务管理员直接分配一定数量的任务给参与标注的工作者，剩余的任务给希望多劳多得的用户来完成。因而，众研平台能够解决标注着完成一定基础工作量的同时令部分标注着多劳多得。

图5.15 平均标注时间

从图5.15可以看出，用户在本地软件和众研的标注任务完成时间相近，而且本地软件会略微快一些，主要原因是标注任务发布方首先将所有标注任务随机分成26份，然后由众包标注者申请，发布方同意申请之后将标注的数据以及预先开发出的标注软件发送给标注者，标注者在本地完成已经分配好的任务，而这里只统计了用户完成标注任务的时间，因此本地软件的平均标注任务时间较众研略短。而对于AMT，标注者需要用过浏览所有标注任务的列表来选择完成标注任务，而且通过调研还发现有不同的标注者同时做同一个标注任务，最终只有先做完的标注者成功提交了结果的情况，因此AMT相对本地软件以及众研平台的标注任务耗时较多。

图5.16 标注任务结果的接受比率

从图5.16可以看出相对于本地软件和AMT，用户在众研平台标注的结果的接受率较高，直接说明了用户在众研平台标注的结果更加准确，表明用户在众研平台领取的任务难度更加合理。

结合图5.14、图5.15、图5.16的结果看出相对于AMT和本地标注软件，本文设计实现的面向残疾人社区的任务分配机制对任务的分配更加合理。同时，通过由管理人员选择采用自动生成的任务分配方案，能够保证残疾人的基本工作量，让残疾人标注者自主认领剩余的部分未分配任务，满足部分残疾人多劳多得的意愿。

## 5.7 小结

本章主要介绍了与课题小组成员共同设计开发的面向残疾人社区的众包平台：众研系统的搭建，给出了系统的框架结构、界面设计、平台部署的具体实现。通过在评估实验表明，该系统能够满足对残疾人标注项目的需求。

# 第六章 总结与展望

## 6.1 总结

本文中，我们设计并实现了面向残疾人社区的众包标注平台——众研。众研平台为残疾人提供了包含较为完善的管理、组织以及运转流程控制的通用众包标注平台，同时也可以用作各种众包相关的算法的研究。工作总结如下：

1. 调研了相关研究工作。众包过程一般包括用户、流程、任务、内容、平台和报酬六个部分。协同众包的研究现状，对众包活动的角色、流程、信息以及人的能力进行重构，提高众包效率。
2. 实现了众包协同任务分配机制。混合人工分配、自动分配以及自动认领的方式，有效的避免了同一任务重复分配、分配模式单一等问题。构建众包协同任务模型同时根据具体项目的需求设计并实现了两种任务分配方案的自动生成算法。
3. 实现了面向残疾人社区的众包管理功能。详细介绍了每个角色功能模块的设计。通过基于角色的功能权限控制，减少了系统框架与特定功能组件之间的依赖，提高了平台角色的可扩展性，使得系统对众包过程不同阶段不同职责的用户权限管理更加具有可扩展性。
4. 设计并实现了具有从创建项目、发布任务、完成任务到结果验收的完整标注流程的平台原型系统。通过选用PostgreSQL数据库，提高系统的可扩展性以及通用性，支持各种形式的标注任务以及各种格式的结果保存方式。采用了Bootstrap的风格，界面美观简洁，改善众包工作人员的用户体验。

## 6.2 展望

论文提出的面向残疾人社区的众包标注平台系统虽然已经按照计划完成，但是仍然还有许多方面需要改进，为了跟好的为残疾人服务，使残疾人能够充分利用众包标注平台获取相应的报酬保证持续收入，对平台进行推广，需要更加完善平台。后续相关工作列举如下：

1. 对残疾人辅助设配进行调研，在残疾人标注任务的过程中针对不同类型的残疾人，配合辅助器具协助各类残疾人参与众包标注工作。
2. 进一步探究各种标注任务的界面设计，做到更加高效的发布任务，避免通过编写代码的方式来完成标注任务界面的设计。
3. 实现薪资报酬的支付功能，结合任务分配以及结果验收，在项目创建者接受任务结果后，直接由平台为残疾人发放报酬。

# 参考文献

1. 廉串德, 罗秋月, 徐桂花. 残疾人网络创业就业现状、问题及建议[J]. 中国市场, 2014(12):71-73.
2. “互联网+为残疾人就业做加法.”
3. 刘锋．威客(Witkey)的商业模式分析[D].中国科学院研究生院硕士论文,2006
4. 杰夫.豪.众包:大众智慧缘何推动商业未来.北京:中信出版社,2009,6-9.
5. 林筱芳, 杨乐, 王鹏. 互联网+带来新机遇——残疾人网络就业创业专家企业对话会纪事[J]. 中国残疾人, 2015(8):32-33.
6. 张国霞. 互联网环境下残疾人就业模式与途径创新[J]. 中国电子商情:科技创新, 2014(3):165-166.
7. P. Liu, X. Ding, and N. Gu, “‘Helping Others Makes Me Happy’: Social Interaction and Integration of People with Disabilities.,” in Proceedings of the 19th ACM conference on Computer Supported Cooperative Work, 2016.
8. Arshad, N.H, Salleh, S.S., Aris, S.R.S., Janom, N., and Mastuki, N.,2013. “Formulating Cohesive Digital Ecosystem of Micro Sourcing Business Process in Malaysia”, Proc. Science and Information Conference (SAI), October, London, pp. 768-773.
9. Habibah, Noor, et al. "Micro Sourcing Strategic Framework for Low Income Group." International Journal of Advanced Computer Science & Applications 4.6(2013).
10. Arshad, Noor Habibah, et al. "Enablers In Implementing Micro Sourcing For The Low Income Group In Malaysia." International Journal of Computers & Technology (2014).
11. Hirth, Matthias, T. Hoßfeld, and P. Tran-Gia. "Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms." Mathematical & Computer Modelling 57.11–12(2013):2918-2932.
12. Zazaleena Zakariah, Norjansalika Janom, and Noor Habibah Arshad. "Business model of crowdsourcing: Review paper." IEEE, 2015.
13. Gray, Mary L., et al. "The Crowd is a Collaborative Network." ACM Conference on Computer-Supported Cooperative Work & Social Computing ACM, 2016.
14. Peng, Xin, M. Ali Babar, and C. Ebert. "Collaborative Software Development Platforms for Crowdsourcing." IEEE Software 31.2(2014):30-36.
15. Vashi Ambati, Stephan Vogel, Jamie Carbonell. 2012.Collaborative workflow for crowdsourcing translation. In Proceedings of the ACM conference on Computer supported cooperative work (CSCW ‘12), 1191-1194.
16. Mladenow, A., et al. "Collaboration and Locality in Crowdsourcing." International Conference on Intelligent NETWORKING and Collaborative Systems IEEE, 2015.
17. 杨子羲. 面向社会化协同的任务分配体系研究与应用. Diss. 重庆大学, 2015.
18. [Online]. Available:http://web.esna.com/blog/what-is-social-collaboration
19. Carr, and F. David. "Social Collaboration For Dummies." For Dummies (2013).
20. (美)麦卡菲(McAfee, and A. P. ). 企业2.0:企业社会化协作趋势与工具. 机械工业出版社, 2011.
21. [Online]. Available:https://en.wikipedia.org/wiki/Social\_collaboration
22. 李红蕾. 面向社会化协同工作的威客社区设计与实现. Diss. 重庆大学, 2012.
23. 魏建良, and 琚春华. "基于社会化标注的用户协同模型研究." 情报学报 31.3(2012):281-288.
24. Rahman, Habibur, et al. "Task Assignment Optimization in Collaborative Crowdsourcing." (2015):949-954.
25. Huang S W, Fu W T. Don't hide in the crowd!: increasing social transparency between peer workers improves crowdsourcing outcomes[C]// Sigchi Conference on Human Factors in Computing Systems. 2013:621-630.
26. Pan Y, Blevis E. A survey of crowdsourcing as a means of collaboration and the implications of crowdsourcing for interaction design[C]// Collaboration Technologies and Systems (CTS), 2011 International Conference on. IEEE, 2011:397-403.
27. Stewart O, Huerta J M, Sader M. Designing crowdsourcing community for the enterprise[C]// ACM SIGKDD Workshop on Human Computation, Paris, France, June. 2009:50-53.
28. Huberman B A, Romero D M, Wu F. Social Networks that Matter: Twitter Under the Microscope[J]. First Monday, 2008, 14(1):2009.
29. Wang W, Jiang J, An B, et al. Toward Efficient Team Formation for Crowdsourcing in Noncooperative Social Networks.[J]. IEEE transactions on cybernetics, 2016.
30. Little G. TurKit: Tools for iterative tasks on mechanical turk[J]. 2009:252-253.
31. Yang, Zhe, et al. "OnTac: Online task assignment for crowdsourcing." IEEE International Conference on Communications IEEE, 2016.
32. Cheng, Peng, et al. "Task Assignment on Multi-Skill Oriented Spatial Crowdsourcing." IEEE Transactions on Knowledge & Data Engineering 28.8(2015):2201-2215.
33. Boutsis I, Kalogeraki V. On Task Assignment for Real-Time Reliable Crowdsourcing[J]. 2014:1-10.
34. Xu, Wei, et al. "DATA: A double auction based task assignment mechanism in crowdsourcing systems." The, International ICST Conference Oncommunications and NETWORKING in China 2013:172-177.
35. Fan, X., P. Yang, and Q. Li. "Fairness Counts: Simple Task Allocation Scheme for Balanced Crowdsourcing Networks." International Conference on Mobile Ad-Hoc and Sensor Networks 2015:258-263.
36. Lee S, Park S, Park S. A quality enhancement of crowdsourcing based on quality evaluation and user-level task assignment framework[J]. Applied Mathematics & Information Sciences, 2014, 9(2):60-65.
37. Liu, Xiangyang, H. He, and J. S. Baras. "Trust-aware optimal crowdsourcing with budget constraint." IEEE International Conference on Communications IEEE, 2015.
38. Roy, Senjuti Basu, et al. Task assignment optimization in knowledge-intensive crowdsourcing. Differentiable optimization and equation solving :. Springer, 2003:467-491.
39. Zhang, Hao, and M. Sugiyama. "Task selection for bandit-based task assignment in heterogeneous crowdsourcing." Technologies and Applications of Artificial Intelligence IEEE, 2015.
40. J. Howe, "The Rise of Crowdsourcing.", Wired magazine, vol. 14, no. 6,pp. 1-4, 2006.
41. Vargo, S.L., Lucsh, R.F. Evolving to a new dominant logic for marketing. Journal of Marketing.2004(68):1～17
42. Howe, J .Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business. New York, Crown Publishing Group.2008
43. Kleeman F., Voss G.G., Rieder K, Un(der)paid innovators: the commercial utilization of consumer work through crowdsourcing. Science, Technology and Innovation Studies.2008.4（1）：5~26
44. Oliveira, F, Ramos, I. Santos, L. Definition of a crowdsourcing innovation service for the European SMEs. Current Trends in Web Engineering. 2010(6385):412~416
45. Estellés-Arolas, E. ,González-Ladró-de-Guevara. Towards an integrated crowdsourcing definition. Journal of Information Science. 2012: 1~14
46. D. C. Brabham, "Crowdsourcing as a Model for Problem Solving an Introduction and Cases.", Convergence: the International Journal of Research into New Media Technologies, vol. 14, no. 1, pp. 75-90, 2008.
47. Chittilappilly, Anand Inasu, L. Chen, and S. Ameryahia. "A Survey of General-Purpose Crowdsourcing Techniques." IEEE Transactions on Knowledge & Data Engineering 28.9(2016):2246-2266.
48. [Online]. Available: www.Freelancer.com.
49. Fan, X., P. Yang, and Q. Li. "Fairness Counts: Simple Task Allocation Scheme for Balanced Crowdsourcing Networks." International Conference on Mobile Ad-Hoc and Sensor Networks 2015:258-263.
50. Alfarrarjeh, Abdullah, T. Emrich, and C. Shahabi. "Scalable Spatial Crowdsourcing: A Study of Distributed Algorithms." IEEE International Conference on Mobile Data Management IEEE, 2015.
51. Amato, Ariel, et al. "Divide and conquer: atomizing and parallelizing a task in a mobile crowdsourcing platform." Annals of the Phytopathological Society of Japan 70.22(2013):21-22.
52. Guo, Xintong, et al. "Brief survey of crowdsourcing for data mining." Expert Systems with Applications 41.17(2014):7987-7994.
53. Silberman, M. Six, L. Irani, and J. Ross. "Ethics and tactics of professional crowdwork." XRDS 2010:39-43.
54. 互帮网.[Online]. Available: http://www.bangcn.com/
55. [Online]. Available: www.istockphoto.com.
56. [Online]. Available: www.wikipedia.org.
57. [Online]. Available: www.innocentive.com.
58. [Online]. Available: www.mturk.com/mturk/welcome.
59. [Online]. Available: www.upwork.com.
60. [Online]. Available: www.crowdflower.com.
61. [Online]. Available:http://www.xianjichina.com/news/details\_12385.html
62. [Online]. Available: www.threadless.com.
63. [Online]. Available:http://cq.qq.com/a/20110629/000629.htm
64. 猪八戒网.[Online]. Available: www.zbj.com.
65. 孙悟空威客网.[Online]. Available: http://www.72swk.com
66. 李勇军, 郭基凤, and 缑西梅. "软件“众包”任务分配方法." 计算机系统应用 24.2(2015):1-6.
67. 马铭, 刘玲, and 田龙. "B/S模式下基于角色的用户权限设计." 安阳工学院学报 11.2(2012):48-52.
68. Li, Ninghui, J. C. Mitchell, and W. H. Winsborough. "Design of a Role-Based Trust-Management Framework." IEEE Symposium on Security & Privacy IEEE, 2002:114-130.

# 致谢

光阴转瞬即逝，转眼之间就到了毕业之际，在复旦大学计算机学院学习的这三年时光里收获颇丰，在毕业论文完成之际，请容许我对协同信息与系统实验室的各位老师、同学表达最衷心的感谢。

首先，非常感谢我的导师卢暾老师，在实验室的这三年学习生活期间不仅在科研学习中给予非常耐心的指导，而且也身体力行地教导着我的工作态度，卢老师这种认真细致的治学态度、刻苦钻研的精神让我学到了很多。

感谢顾宁老师对我的指导与教诲，顾宁老师严谨的工作科研态度是我今后工作学习的榜样，而且顾宁老师也为我们实验室的同学提供了良好的科研工作环境和生活上的各种帮助。

感谢丁向华老师对我的指导和鼓励，丁老师您求实执着的科研精神是我今后工作学习的榜样。

感谢协同信息与系统实验室的全体团成员！他们乐观向上的生活态度、良好的生活习惯让我感受到这个集体浓浓的学术氛围、大家庭般的温馨。

最后，感谢我家人对我的关心、鼓励和支持。

**复旦大学**

**学位论文独创性声明**

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。论文中除特别标注的内容外，不包含任何其他个人或机构已经发表或撰写过的研究成果。对本研究做出重要贡献的个人和集体，均已在论文中作了明确的声明并表示了谢意。本声明的法律结果由本人承担。

作者签名： 日期：

**复旦大学**

**学位论文使用授权声明**

本人完全了解复旦大学有关收藏和利用博士、硕士学位论文的规定，即：学校有权收藏、使用并向国家有关部门或机构送交论文的印刷本和电子版本；允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其它复制手段保存论文。涉密学位论文在解密后遵守此规定。

作者签名： 导师签名： 日期：