谨以此论文献给我的家人，我的导师，以及所有帮助和支持我的朋友们。

------------------侯海东

论文作者身份识别及认领服务的研究

学位论文答辩日期：

指导教师签字：

答辩委员会成员签字：

**独 创 声 明**

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含未获得（注：如没有其他需要特别声明的，本栏可空）或其他教育机构的学位或证书使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名： 签字日期： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解国家有关保留、使用学位论文的法律、法规和学校有关规定，并同意以下事项：

1、学校有权保留并向国家有关部门或机构送交本学位论文的复印件和磁盘，允许论文被查阅和借阅；

2、学校可以将本学位论文的全部或部分内容编入学校学位论文数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文；

3、学校可以基于教学及科研需要合理使用本学位论文。

需保密的学位论文在解密后适用本授权书。

学位论文作者签名： 导师签字：

签字日期： 年 月 日 签字日期： 年 月 日

论文作者身份识别及认领服务的研究

摘 要

关键词：

Research of author's identification and paper claim service

Abstract

Keywords:

目 录

[1 绪论 3](#_Toc501301552)

[1.1 研究背景和意义 3](#_Toc501301553)

[1.2 研究现状 3](#_Toc501301554)

[1.2.1 作者身份识别研究现状 3](#_Toc501301555)

[1.2.2 作者识别问题难点 3](#_Toc501301556)

[1.3 研究内容和主要工作 3](#_Toc501301557)

[1.4 论文内容安排 3](#_Toc501301558)

[1.5 本章小结 3](#_Toc501301559)

[2 测试数据集构建与特征选择 4](#_Toc501301560)

[2.1 测试数据集的构建 4](#_Toc501301561)

[2.1.1 文献记录的获取与整理 4](#_Toc501301562)

[2.1.2 测试数据集的选取 4](#_Toc501301563)

[2.2 评价指标 4](#_Toc501301564)

[2.3 作者识别的特征选择 4](#_Toc501301565)

[2.4 本章小结 4](#_Toc501301566)

[3 论文作者身份识别初步设计 4](#_Toc501301567)

[3.1 作者识别基本流程 4](#_Toc501301568)

[3.2 作者姓名提取与规范化 4](#_Toc501301569)

[3.2.1 作者姓名拆分 4](#_Toc501301570)

[3.2.2 作者姓名清洗 4](#_Toc501301571)

[3.2.3 作者姓名前后顺序分析 4](#_Toc501301572)

[3.2.4 作者姓名规范化 4](#_Toc501301573)

[3.3 作者所属单位提取 4](#_Toc501301574)

[3.3.1 编辑距离 4](#_Toc501301575)

[3.3.2 相似度 5](#_Toc501301576)

[3.3.3 单位名称对应字典 5](#_Toc501301577)

[3.3.4 作者单位提取方法描述 5](#_Toc501301578)

[3.4 实验结果与分析 5](#_Toc501301579)

[3.5 本章小结 5](#_Toc501301580)

[4 基于合作网络的重名作者消歧 5](#_Toc501301581)

[4.1 合作网络简介 5](#_Toc501301582)

[4.2 基于合作网络的作者消歧方法 5](#_Toc501301583)

[4.2.1 基于合作网络的作者表示模型 5](#_Toc501301584)

[4.2.2 论文作者合作网络构建 5](#_Toc501301585)

[4.2.3 合作者集与论文作者集相似性计算 5](#_Toc501301586)

[4.2.4 重名作者消歧 5](#_Toc501301587)

[4.3 实验结果与分析 5](#_Toc501301588)

[4.4 本章小结 5](#_Toc501301589)

[5 论文推送认领服务设计 5](#_Toc501301590)

[5.1 后台技术框架简介 5](#_Toc501301591)

[5.1.1 ASP.NET Web API 6](#_Toc501301592)

[5.1.2 Entity Framework 6](#_Toc501301593)

[5.1.3 Autofac 6](#_Toc501301594)

[5.1.4 AutoMapper 6](#_Toc501301595)

[5.2 推送认领服务分析 6](#_Toc501301596)

[5.3 系统详细设计 6](#_Toc501301597)

[5.3.1 认领平台 6](#_Toc501301598)

[5.3.2 推送认领 6](#_Toc501301599)

[5.3.3 补充认领 6](#_Toc501301600)

[5.4 本章小结 6](#_Toc501301601)

[6 总结与展望 6](#_Toc501301602)

[6.1 论文总结 6](#_Toc501301603)

[6.2 研究展望 6](#_Toc501301604)

[参考文献 6](#_Toc501301605)

# 绪论

## 研究背景和意义

## 研究现状

### 作者身份识别研究现状

### 作者识别问题难点

## 研究内容和主要工作

## 论文内容安排

## 本章小结

# 测试数据集构建与特征选择

## 测试数据集的构建

### 文献记录的获取与整理

### 测试数据集的选取

## 评价指标

## 作者识别的特征选择

## 本章小结

# 论文作者身份识别初步设计

## 作者识别基本流程

## 作者姓名提取与规范化

### 作者姓名拆分

### 作者姓名清洗

### 作者姓名前后顺序分析

### 作者姓名规范化

## 作者所属单位提取

### 编辑距离

### 相似度

### 单位名称对应字典

### 作者单位提取方法描述

## 实验结果与分析

## 本章小结

# 基于合作网络的重名作者消歧

## 合作网络简介

## 基于合作网络的作者消歧方法

### 基于合作网络的作者表示模型

### 论文作者合作网络构建

### 合作者集与论文作者集相似性计算

### 重名作者消歧

## 实验结果与分析

## 本章小结

# 论文推送认领服务设计

## 后台技术框架简介

### ASP.NET Web API

### Entity Framework

### Autofac

### AutoMapper

## 推送认领服务分析

## 系统详细设计

### 认领平台

### 推送认领

### 补充认领

## 本章小结

# 总结与展望

## 论文总结

## 研究展望

# 参考文献