编号：

****

**毕业设计(论文)开题报告**

|  |  |  |
| --- | --- | --- |
| 题 目： | 英语文本语篇关系提取 | |
|  | 算法设计 | |
| 院 （系）： | 计算机与信息安全学院 | |
| 专 业： | 软件工程 | |
| 学生姓名： | 薛天杭 | |
| 学 号： | 1400330129 | |
| 指导教师单位： | 计算机与信息安全学院 | |
| 姓 名： | | 黄桂敏 |
| 职 称： | | 教授 |

题目类型：软件开发

2019年12月3日

开题报告填写要求

1、开题报告作为毕业设计（论文）答辩委员会对学生答辩资格审查的依据材料之一。此报告应在指导教师指导下，由学生在毕业设计（论文）工作前期内完成，经指导教师签署意见审查后生效。

 2、开题报告内容必须用黑墨水笔工整书写，或按教务处统一设计的电子文档标准格式打印，禁止打印在其它纸上后剪贴，完成后应及时交给指导教师签署意见。

 3、学生查阅资料的参考文献应在10篇及以上（不包括辞典、手册）。

 4、有关年月日等日期的填写，应当按照国标GB/T 7408—94《数据元和交换格式、信息交换、日期和时间表示法》规定的要求，一律用阿拉伯数字书写。如“2010年9月20日”或“2010-09-20”。

5、此页与开题报告封面进行双面打印，其他剩余内容可单面打印。

6、请确保最后一页（即“指导教师意见”所在页）单独成一页。

|  |
| --- |
| **1、毕业设计的主要内容、重点和难点等** |
| 修辞结构理论(RST)是英国学者W.C. Mann、S. A. Thompson等在从事语篇生成程序研究时提出,并在实践中不断完善的一种理论框架。它由两大部分组成,一个是“定义关系”,另一个是“认知图式”,两个部分相互关联。 一、定义关系 一个给定话语(或语篇,下同)中的两个命题之间存在着一种隐含的关系,这种关系在性质上也是一个命题,定名为“关系命题”。 定义关系依靠一种分析框架进行操作,这种分析框架由以下3个基本概念建立:核心性(nuclearity),指一对命题集合中某一命题的相对重要性。制约因素(constaints),指一对命题集合中命题存在的必要条件。效果(effect),指作者采用一种关系对特定话语所要产生的影响。Mann和Thompson认为,一个语篇可以分割为不同层次的语篇单元对,语篇单元对中的关系大多是非对称性的,即相互对应的两个语篇单元,对语篇整体而言具有不同的语义份量,其中语义份量相对重的那个单元称之为“核心成分” (nucleus),语义份量轻的那个单元则称之为“外围成分” (satellite)。 二、认知图式 认知图式是与定义关系密切相关的一个语义概念,指的是话语语义关系的一种抽象模式,可用图形表示。根据Mann和Thompson (1988)的观点,认知图式可决定一个语篇中所有成分的结构性排列,它由3个要素组成:一定数量的成分性语篇单元;语篇单元之间的关系说明;某个语篇单元(核心单元)与语篇整体的关系说明。 本次毕业设计我需要做的主要内容大致为以下几项： 1、收集相关处理文本数据，利用RST进行语篇关系分析； 2、设计基于RST的语篇关系提取算法； 3、完成软件编程并实现文本的语篇关系提取，提高文本语篇关系提取的精度。 4、完成相关设计并测试 5、撰写设计报告 本次毕业设计我所认为的重点主要有三个方面； 其一就是如何使用修辞结构理论即RST进行语篇关系分析；这一方面我相信在我深入了解修辞结构理论后，进行一定的学习可以做到。 其次就是如何对文本进行语篇关系的提取；在这一部分我准备参考自然语言处理（NLP）所采用的方法进行语篇关系的提取，通过学习借鉴NLP所使用的提取方式来完成自己想要达到的目标。 最后就是如何提高精度；在这一方面目前我暂时没有具体思路，具体实施时可以参考目前比较主流的方法应用到自己的程序中。 本次毕业设计我所认为的难点主要在对语篇关系的提取上，如何设计算法能使自己的程序高效、正确地提取出自己想要的文中的信息是我认为的在本次设计中比较难以实现的部分，需要通过一定的学习来扩充自己的知识才能完成这个目标。 |
| **2、准备情况（查阅过的文献资料及调研情况、现有设备、实验条件等）** |
| 根据题目查阅了有关修辞结构理论，自然语言处理，语料库等相关资料，具体查阅的参考资料如下： [1]冯德正,张艳,王艳.修辞结构理论在多模态语篇分析中的应用[J].当代修辞学,2016(05):49-58. [2]纪超杰. 基于深度学习和修辞关系的文本情感分析[D].南昌大学,2018. [3]杨丽纳.衔接连贯语篇建构的有效途径——基于修辞结构理论的视角[J].济源职业技术学院学报,2019,18(01):77-81. [4]练睿婷. 语篇标注语料库的建设研究[A]. 中国中文信息学会.第四届全国学生计算语言学研讨会会议论文集[C].中国中文信息学会:中国中文信息学会,2008:7. [5]于彦秋. 基于卷积神经网络的自然语言处理相关技术研究[D].国防科学技术大学,2016. [6]褚晓敏,奚雪峰,蒋峰,徐昇,朱巧明,周国栋.宏观篇章结构表示体系和语料标注研究[J/OL].软件学报:1-23[2019-12-07].https://doi.org/10.13328/j.cnki.jos.005868. [7]李成城. 夏斌.从修辞结构理论看翻译逻辑的处理——以Wolf Dream为例[J].英语广场,2018(08):43-44. [8]原颖. RST工具在环境评价导则翻译中的应用探究[D].山西大学,2015. 现有设备：计算机一台 |
| **3、实施方案、进度实施计划及预期提交的毕业设计资料** |
| 实施方案： 1、首先进行相关资料的查阅，通过查阅资料使自己对于修辞结构理论有一定的理解。 2、对自然语言处理原理进行一定的学习与探索，考虑其算法是如何实现的。 3、设计自己的基于RST的语篇关系提取算法，使之能实现一定的基础功能。 4、通过学习，强化自己所设计的算法，提高效率以及准确率。 5、完成相关设计并进行测试，确保自己的程序能完成预期 6、撰写设计报告 进度实施计划： 2019-11-25 ~ 2019-12- 3 ：查阅相关资料并完成开题报告； 2019-12- 4 ~ 2020- 1 -30 ：学习与修辞结构理论（RST）相关知识，初步确定整体制作方向； 2020- 1 -30 ~ 2020- 2 -15 ：学习NLP相关理论，尝试制作简单的demo验证自己学习的内容； 2020- 2 -16 ~ 2020- 3 -22 ：完成能够基本实现目标的算法设计以及程序的编写 2020- 3 -22 ~ 2020- 5 -11 ：测试程序记录数据并改进以及完成论文初稿； 2020- 5 -12 ~ 2020- 5 -25 ：参加软件验收，修改论文； 2020- 5 -26 ~ 2020- 6 - 6 ：论文第三方评阅，并根据其意见修改论文； 2020- 6 - 7 ~ 2019- 6 - 9 ：整理并提交设计材料 预期提交的毕业设计资料： 1.毕业设计说明书 1 份 2.英文原文及译文 1 份 3.毕业设计作品 1 份 |
| **指导教师意见** | |
| 该生课题内容适当，工作量较为饱满，以毕业设计课题“英语文本语篇关系提取算法设计”为中心查阅了相关的基本资料，掌握了相关的基本知识，对研究内容所涉及的研究领域进行了较为全面的研究，参考文献的格式、篇数及内容均符合要求。预期该学生能够按时按质顺利完成。  指导教师（签字）：  2019年12月4日 | |
| **开题小组意见** | |
| 同意  开题小组组长（签字）：  2019年12月5日 | |
| **院（系、部）意见** | |
| 同意    主管院长（系、部主任）签字：  2019年12月6日 | |