

浙江大学

2020大学生科研训练计划申报书

项目编号:

项目名称: 基于NLP的英语句法分析工具

项目负责人: 嵇泓玮 学 号: 3180101874

院（系）: 生物医学工程与仪器科学学院

联系电话: 18888918861 电子邮件: jihongwei@zju.edu.cn

指导教师: 丁甯 职 称:

申报级别: 创新训练、 科技创新、 校级SRTP、 院级SRTP

浙江大学本科生院教务处

填 写 说 明

一、申报书要按照要求，逐项认真填写，填写内容必须实事求是，表达明确、严谨，首页只填负责人，“项目编号”一栏不填。

二、格式要求：申报书中各项内容以Word文档格式填写，表格中的字体为小四号楷体，行距为最小值20磅；表格空间不足的，可以扩展或另附纸张。

一、项目简介

项目概括	项目名称	基于NLP的英语句法分析工具						
	所属一级学科	计算机类						
	项目性质	<input type="checkbox"/> 基础研究 <input checked="" type="checkbox"/> 应用基础研究						
	项目来源	<input type="checkbox"/> 自主立项 <input checked="" type="checkbox"/> 教师指导选题 <input type="checkbox"/> 社会企业事业						
	申请经费	1000						
	起止时间	2020-05-01至2021-05-31						
项目状况		<input checked="" type="checkbox"/> 研发阶段 <input type="checkbox"/> 中试阶段 <input type="checkbox"/> 批量（规模）生产 （选项打√）						
项目申报人	姓名	嵇泓玮	性别		出生年月		入学年月	2018
	学号	3180101874			联系电话	18888918861	电子邮箱	jihongwei@zju.edu.cn
	院系专业	生物医学工程与仪器科学学院、测控技术与仪器						
项目组主要成员	姓名	联系电话		院系专业		年级	具体分工	
	房子琦	18888912375		计算机科学与技术学院、产品设计		2018		
	牛思旭	18888915774		电气工程学院、电气工程及其自动化		2018		
项目指导老师	姓名	联系电话		所在单位		职务/职称	主要研究方向	
	丁鼎	18758010537						
	近三年成果：国家级_0_等奖 _0_项，省部级_0_等奖_0_项							
	近三年科研经费_360_万元，年均 _120_万元							

项目主要内容简介	基于句法分析相关模型，完成一个英语句法学习的辅助工具：给定一个英语句子，用算法分析它的句法结构，然后把句法结构可视化。目的是通过可视化句法结构来更好地学习句法知识。
项目负责人参与科研情况	无
项目组成员参与科研情况	无

二、项目背景、目的及意义

（简要说明项目背景、意义和实施必要性，研究现状和发展动态，不超过1100字）

语法分为词法和句法，弄清句法对于语义的判断有着很重要的作用。但在现实生活中，以英语为例，不少人对句法并没有一个清晰的概念，甚至直接忽视语法，导致在碰到较为复杂的句子时，无法正确理解句子语义。因此一个进行句法分析的工具对英语学习会有所帮助。

句法分析是自然语言处理的基本模块，对句法进行分析虽然不是一个自然语言处理任务的最终目标，但它往往是实现最终目标的关键环节。

本项目希望通过了解自然语言处理中的句法分析相关模型，能够设计一个针对英语学习的句法分析工具，实现相关基础功能，从而实现更加清晰，更加准确的句法分析，对英语学习者进行帮助。

三、项目研究方案

（包括项目的主要内容、计划目标和拟解决的问题，思路方法、组织实施及进度安排，不超过1200字）

一、主要内容和思路方法

了解深度学习以及自然语言处理的基本知识，基于句法分析的不同实现模型，阅读相关文献，择优选择其中几种进行较为深入的研究，最终确定最佳的方法，读懂其实现的代码，基于此完成句法结构分析的功能，并将句法结构可视化。

二、进度安排

- 1) 借助网络资料，对神经网络和自然语言处理相关知识的学习。（2020年4月~2020年5月）
- 2) 阅读资料和查看文献完成知识储备的同时，关注一些相关的研究成果，进行模型的初步搭建。（2020年5月~2020年7月）
- 3) 搭建并训练模型，完成句法分析部分的测试，进行句法可视化。（2020年7月~2020年11月中期答辩前）
- 4) 产品的进一步功能实现及优化、测试、部署等。（2020年11月中期答辩后~2021年5月）

四、项目研究条件及创新之处

（已有研究基础，包含与项目有关的研究积累、已取得的成绩和已具备的条件，尚缺少的条件及解决办法，项目优势和风险，以及项目创新点等，不超500字）

研究条件：本项目主要研究句法分析以及可视化的相关模型。现已经有了许多比较热门的模型。仅在成分句法分析领域，就包括基于转移系统（transition-based）的模型、基于动态规划（chart-based）解码的模型、基于自顶向下贪心（greedy top-down）解码的模型和一些将预测树结构转化为预测序列（sequence to sequence）的模型。利用现有的关于神经网络、自然语言处理的资料课程，结合句法分析的论文和开源代码，可以为项目的进行打下基础。可视化方面可以参考Berkeley Parser的句法分析可视化效果。

创新之处：将句法分析以及模型可视化，应用于英语语法辅助学习。另外虽然现有模型也存在可视化，但是可视化效果对于英语学习者较难理解，本项目希望可以对可视化效果进行改进，让使用者没有NLP相关知识也能够理解句法。

五、项目预期成果

（包括知识产权成果，如论文成果、获奖成果、评议鉴定成果、推广成果、论著成果、专利成果、研制产品、开发软件，与毕设、学科竞赛等其他学习环节结合情况，或其他成果等，以及经济效益、社会效益等，不超130字）

一个的英语句法分析工具，实现句法结构分析与可视化等相关功能，且在速度、准确率等指标上符合要求，有一定的实用价值。

六、项目财务预算

（包括经费预算及经费支出明细等）

专用材料费 0.0 元；
印刷费与资料费600.0 元；
交通与差旅费 400.0 元；
出版费 0.0 元；
邮寄费 0.0 元；
评审费待确定。国创为300，省创 200 元。校级200元，院级自筹

七、项目组承诺

承 诺 书

以上所填内容真实可靠，本项目组承诺：该项目立项后，将严格遵守有关规定、遵守本申报书和预算表中规定的条款和内容，保证按计划进度完成项目任务。

项目组全体成员（签字）： 嵇泓玮 房子琦 牛思旭

年 月 日

八、指导老师意见

选题有意义，请尽快开展工作

指导老师（签字）：丁鼎

年 月 日

九、院（系）专家组意见

专家组组长（签字）：

年 月 日

十、学校审核意见

（盖章）：

年 月 日