

刘潇远

电话: +86-13668891761 | 邮箱: liuxy@smail.nju.edu.cn

教育经历

南京大学

09/2016-现在

- 理学学士，南京大学计算机科学与技术系，计算机科学与技术专业

技术栈

- C++: 熟练运用类、继承、多态、容器、模板等高级特性
- 其他语言: 能够使用 Java、Python 和 MATLAB 进行编程和 MySQL 进行数据库操作，了解数据库索引原理
- 数据结构与算法: 熟悉常用数据结构以及算法实现
- Linux: 熟悉大部分 Linux 操作
- PyTorch: 熟悉基础机器学习模型，能熟练搭建神经网络模型
- Hadoop: 熟悉 Hadoop 编程，能按照 MapReduce 框架实现分布式算法
- Git: 熟悉 Git Workflow
- Conda: 熟练使用 Conda 部署开发环境
- IDE: Visual Studio、PyCharm、Jupyter Notebook、IDEA、Eclipse、Navicat

项目经历

研究助理, Bioinformatics Laboratory, the State University of New York at Buffalo

07/2019-10/2019

- 使用 PyTorch 搭建神经网络模型并进行训练、调参和测试
- 将基因序列比对视作 NLP 中的文本相似度计算问题，使用 CNN 为基本结构的 Siamese Network，将基因序列映射为向量，之后以向量间的相似度作为基因间的相似度，在精确度上超越了传统算法
- 将 RNA 二级结构预测视为 NLP 中的序列标注问题，将 RNA 二级结构的点-括号表示法的表示结果作为 RNA 序列上每个核酸的标签，使用 LSTM+CRF 模型进行序列标注，训练出的模型与 the-state-of-the-art 相比，提升了 10%

参与实验室项目: 知识关联与时间推理类问题求解关键技术与系统——面向地理高考的问答系统（国家重点研发计划课题，863 项目）

11/2018-现在

- 地理知识库中存在地点冗余问题，使用离去点检测去除地点知识库中的冗余地点
- 将多个地点知识库进行合并，并对 MySQL 查询和数据储存进行了封装
- 地理知识库中缺少地点与地形区的对应，使用 Urllib 和对百度百科地理部分进行了爬取，并用 BeautifulSoup 对网页进行解析，之后从中提取有效信息补充进地理知识库

大学生创新项目: 基于语义知识图谱的实体搜索引擎研究

11/2017- 11/2018

- 使用 CoreNLP 对自然语言描述的搜索问题进行解析，利用解析结果进行答案类型检测
- 使用 Lucene 搭建了一套实体搜索引擎，改进重排序算法，将知识图谱搜索结果与传统搜索结果加权作为搜索结果最终打分，在 DBpedia v2 数据集上取得了比 BM25 算法更好的结果

邮件自动分类系统

06/2019- 07/2019

- 使用 Hadoop 实现分布式数据挖掘和机器学习算法，例如将邮件文本映射为向量的 TF-IDF，进行邮件分类的 K-Means, KNN, Naïve Bayes and Decision Trees 等数据挖掘和机器学习算法
- 使用 Hbase 将算法部署到集群运行并测试，并在 KNN 和 Naïve Bayes 算法上去的较好的分类效果

绘图软件

11/2018- 12/2018

- 使用 C# 编程语言实现了一个能对 2D 图像进行绘制和编辑的软件，主要实现的算法包括 DDA 直线算法、中点直线算法、Bresenham 圆绘制算法、Bezier 曲线、B 样条曲线、直线的中点分割算法、梁友栋算法

学生信息管理系统

12/2018-12/2018

- 使用 Qt 开发框架实现了表格式的学生信息管理系统，包含表格型应用的基础功能，如读写文件，信息导入，信息排序等
- 使用 QComboBox 实现了部分表格的下拉选项菜单，如学生的专业，须从表格的弹出菜单中进行选择

i386 模拟器

09/2017- 01/2018

- 在 Linux 平台上设计并使用 C 语言实现一个 i386 模拟器，包括一个虚拟 CPU，x86 指令集子集 n86，缓存调度算法，内存管理和设备管理
- 实现了包含 8 个通用寄存器、EFLAGS 寄存器、整数运算单元、逻辑运算单元和浮点数运算单元的虚拟 CPU
- 严格按照英特尔 i386 手册实现了 16 进制 n86 指令集的解码与执行，包括解析 OPERAND、ModR/M 字节和 SIB 字节进行操作数寻址，下一条指令地址的计算和 PC 的更新；还利用 C 语言宏精简指令实现，通过将大量相似操作替换为宏指令，极大地减少了工作量和避免了低级编码错误
- 实现了程序装载 Loader，包括 ELF 文件头解析和装载
- 模拟器的存储管理部分，模拟使用随机替换算法的 CPU Cache，实现了 80386 保护模式下的分段和分页机制
- 利用自陷实现系统调用，通过响应时钟中断实现对外部中断的响应，并在此基础上完成了串口的模拟、硬盘程序加载、键盘模拟和 VGA 的 MMIO

获奖经历

- 人民奖学金三等奖(2019)、人民奖学金社会工作奖(2018)、郑刚菁英奖学金(2018)、人民奖学金三等奖(2017)
- 南京大学环境协会优秀部长(2018)、南京大学计算机系优秀学生干部(2018)、