

孟 芑

github.com/pengmeng ◇ pengmeng.me

教育经历

西北大学 (Northwestern University)	2014年9月 - 2015年12月
计算机专业 硕士学位 GPA: 3.81/4.0	
核心课程: 机器学习、分布式系统、自然语言处理、数据挖掘 (Weka)	
北京科技大学	2010年9月 - 2014年6月
计算机科学与技术专业 学士学位 GPA: 3.64/4.0	

工作经历

百度北京实习	2015年7月 - 2015年9月
研发实习生	百度, 北京
<ul style="list-style-type: none">· 就职于百度指数, 负责数据处理、分析以及产品的升级和维护· 使用Python, Perl, Hadoop等技术完成了多个数据专题的数据处理工作并成功上线· 完成下一代产品中基于用户检索日志挖掘关联检索词模块的优化和升级, 将该模块运行效率提升一倍	

项目经历

分布式哈希表Kademlia	2015年3月 - 2015年6月
主要开发者	西北大学, 美国
<ul style="list-style-type: none">· 基于Go语言实现了分布式哈希表Kademlia类库, 并基于该库开发了Vanish数据自毁系统· 负责Kademlia系统中路由表的设计与实现以及系统核心功能的开发· 负责系统单元测试、模拟多节点环境测试的编写	
股票分析项目	2015年3月 - 2015年6月
主要开发者	西北大学, 美国
<ul style="list-style-type: none">· 设计实现了从教育数据库抓取华尔街时报新闻的爬虫系统, 支持根据关键词和时间范围抓取· 统计和分析了新闻数量、新闻标题情感和股票交易量、股票价格趋势之间的关系· 基于递归神经网络算法 (Recurrent Neural Network) 对新闻全文进行情感分析	
自然语言处理项目	2015年2月 - 2015年3月
主要开发者	西北大学, 美国
<ul style="list-style-type: none">· 分别设计实现了金球奖相关Twitter数据、Allrecipe网站菜单数据的解析程序· 成功从金球奖Twitter数据中提取出获奖者、颁奖者、最佳着装等关键信息· 完成了支持关键字检索和URL检索的菜单抓取程序、并根据给定规则对菜单进行语义转换· 使用Flask框架开发了简单的菜单分析Web应用	
网络爬虫框架	2015年3月 - 至今
开发者	西北大学, 美国
<ul style="list-style-type: none">· 设计实现了一个可扩展可定制的Python网络爬虫框架, 开源于Github· 实现了多个基本的爬虫组件, 例如支持Cookie的抓取器和不同类型的URL队列模块· 基于redis-py包开发了支持Python数据对象方法的Redis数据库接口, 用于URL队列的实现	

专业技能

编程语言	Python, JAVA; Go, Common Lisp, C
数据库	MongoDB, Redis, MySQL
其他工具	Hadoop, Vim, Git, Latex, Weka