

宋小远 中国科学院大学 硕士

移动电话: 15600616507 邮箱: songxiaoyuannkcs@163.com

教育工作经历

2016.07- 至今	中国科学院软件研究所	机器学习与数据分析	助理工程师
2013.09-2016.07	中国科学院大学	计算机软件与理论专业	硕士
2009.09-2013.06	南开大学	软件工程专业	学士

项目经验

2016.10 – 2017.01 上市公司知识图谱构建 核心成员 中国科学院软件研究所

项目简介 针对 A 股 2800 多家上市公司的年报进行半自动化解析, 根据年报中相关数据的表格, 提取出股东、高管情况、主营业务等主要数据。将数据进行清洗然后插入到 MySQL 数据库中。最后利用 ECharts 插件将这些数据在网页中进行可视化展示, 在展示的过程中还支持二次获取的功能。

承担工作 根据公司年报中表格所在的页数, 自动提取出该页中需要的数据并保存在 CSV 文件中。
对提取出的数据进行清洗, 插入到 MySQL 数据库中, 并且利用 ECharts 将数据可视化

涉及技术 Python JavaScript ECharts MySQL Neo4j

2016.11 – 2017.02 上市公司事件驱动解析 核心成员 中国科学院软件研究所

项目简介 在生成模型的基础上, 利用命名实体识别、语义角色分析等 nlp 技术将中文句子分词, 然后利用概率统计和朴素贝叶斯算法从上市公司的新闻中提取出发生的事件, 包括并购重组、发布财报、重大会议、天灾人祸等事件, 将事件信息组成机构化的数据保存在 MongoDB 数据库中进行事件引导。

承担工作 从 163 万条上市公司的新闻中提取出关于重大会议和公司财报的新闻, 并生成结构化的数据。
利用命名实体识别和语义角色分析技术对新闻中的每一句话进行分词, 根据分词结果提取事件。

涉及技术 Python Mongoddb nlp 朴素贝叶斯算法

2014.09 – 2016.05 Android 平台 NW.js 研发 核心成员 中国科学院软件研究所

项目简介 NW.js 是一个只用 Web 技术编写本地桌面应用的运行环境, 它利用 chromium 的 Webkit 对界面进行渲染, 利用 Node.js 的模块访问本地操作系统的数据和 API, 这对开源社区是一个很大的贡献。

承担工作 chromium 中的消息循环与 node.js 的 libuv 事件循环合并, 使两者在同一个线程中。
Node.js 的对象挂载到 chromium 中, webkit 中 context 的对象可以调用 Node.js 模块和函数。
Node.js 模块在 Android 平台支持性测试, 第三方模块在 Android 平台上编译及调用

涉及技术 Chromium Nodejs C++ V8 消息循环 设计模式

2014.09 -2015.01 核高基基础版新型桌面项目 子模块负责人 中国科学院软件研究所

项目简介 该项目是实现一个跨设备、跨平台的管理用户数据的新型桌面, 支持不同设备之间进程的通信, 支持用浏览器访问桌面, 对本地操作系统数据进行操作。

承担工作 前端数据管理器、通知服务器的设计与实现、Android 平台运行环境的研发。

涉及技术 node-webkit JavaScript nodejs git IPC

个人技能

熟悉机器学习常用算法, 了解它们的原理和应用场景, 对机器学习, 推荐系统具有浓厚的兴趣。

熟悉 C++、Python 语言, 了解 java, C, C#, JavaScript 语言, 掌握 SQL, Mongoddb 数据库。

英语六级, 熟悉数据结构, 算法, 熟悉 linux 操作系统以及常用指令。