

个人信息

- | | |
|--|--|
|  姓 名: 陈伟 |  电 话: 151-2002-2401 |
|  工作经验: 3 年 |  邮 箱: cw525bl@163.com |
|  工作性质: 全职 |  现住城市: 北京市 |
|  求职意向: Python 爬虫工程师 |  期望月薪: 面议 |

教育背景

学校名称: 河北科技大学 就读时间: 2012.09-2016.06 学历学位: 本科

专业技能

熟练 MySQL 关系型数据库的基本操作

熟练 MongoDB, Redis 非关系型数据库的常用操作

熟练使用 Scrapy 框架

熟练爬虫 Re 正则, Xpath 和 BS4 的使用

熟练使用分布式爬虫 scrapy-redis 组件

熟练使用 selenium 组件

熟练使用 Requests, urllib 第三方组件

熟悉 HTML, CSS, JavaScript, jQuery, VUE 前端开发技术

熟悉 docker 容器的常用操作

熟练使用 Numpy, Pandas 等数据分析库的使用

精通 Django, Flask 等 Web 框架熟练

熟练使用 Django REST Framework 框架实现 API 接口

熟练使用 Git 常用指令

熟悉 linux 系统常用操作命令

工作经历

公司名称: 北京传臻科技有限公司 职位名称: Python 爬虫工程师 在职时间: 2016.9-至今

- ♥ 主要负责分析爬取相关网站, 网页数据 API 的特点和规律
- ♥ 负责网页信息搜集并抓取页面信息
- ♥ 利用 MySQL, MongoDB 实现数据的存储
- ♥ 负责业务需求分析, 系统统计, 业务数据的收集整理
- ♥ 负责 SDK 文档接口的开发

项目经验

项目名称： 快讯网站的股市数据抓取

项目职位： Python 爬虫工程师

软件环境： Python3.6 + MySQL

硬件环境： Windows10

开发工具： Pycharm

项目描述： 为了解决客户需求，通过爬取东方财富网，新浪等相关的需求目标网站获取股市实时数据，以及对后期每日更新的数据实现日增式爬虫。并对数据进一步的清洗处理，以便满足用户的需求。

技术描述：

- ♥ 分析网页请求参数，实现数据的抓取
- ♥ 解决网站反爬机制，
- ♥ 优化爬虫代码，加快数据采集效率及时更新数据
- ♥ 实现日增式爬虫，并且负责爬虫的后期维护工作

项目名称： ai 伙伴

项目职位： Python 爬虫工程师

软件环境： Python3.6 + Redis + MongoDB + Flask

硬件环境： Windows10

开发工具： Pycharm

项目描述： 基于智能音箱智能玩具的发展现状，以及项目发起人的个人情感和社会责任，目的是促进儿童的健康成长，防止过早沉迷手机，促进亲情沟通，促进孩子的多方面健康成长。基于语音识别、图像识别、自然语言处理和 Flask 框架，使玩具成为孩子健康成长的小伙伴，使玩具实现了智能对话、做游戏、讲故事、点播歌曲等功能

技术描述：

- ♥ 基于 Flask 搭建后端框架
- ♥ 基于 Python 分布式爬虫爬取数据
- ♥ 基于 Redis 和 MongoDB 实现数据存储
- ♥ 基于百度 AI 实现语音合成和语音识别
- ♥ 基于图灵机器人实现智能信息回复
- ♥ 基于 pypinyin, gensim, jieba 实现自然语言的处理(NLP)

项目名称： 旅行酒店数据的抓取

项目职位： Python 爬虫工程师

软件环境： Python3.5 + MongoDB

硬件环境： Windows10

开发工具： Pycharm

项目描述： 随着旅游行业的发展，出门在外住宿条件是每一个客户所关心的问题。通过爬取携程旅行等相关的需求目标网站获取酒店实时数据，以便满足用户的需求。

技术描述：

- ♥ 分析网页请求参数，解析实现数据的抓取
- ♥ 解决网站反爬机制，
- ♥ 优化爬虫代码，加快数据采集效率及时更新数据
- ♥ 将数据存储于 MongoDB 数据库中

项目名称： 国家裁判文书网数据抓取 **项目职位：** Python 爬虫工程师

软件环境： Python3.5 + MySQL

硬件环境： Windows10

开发工具： Pycharm

项目描述： 采集国家裁判文书网的案件信息，将相关信息按照规定整理后存入 MySQL 数据库中，并针对后期更新的文书做增量式爬虫。

技术描述：

- ♥ 分析目标网站页面源码结构，设计爬虫逻辑，实现数据爬取
- ♥ 解决网站反爬机制
- ♥ 利用正则表达式，BeautifulSoup, xpath 方法找出需求信息
- ♥ 优化爬虫代码并做增量式爬虫
- ♥ 将数据存储到 MySQL 数据库中

项目名称： 住房数据的抓取 **项目职位：** Python 爬虫工程师

软件环境： Python2.7 + MySQL

硬件环境： Windows10

开发工具： Pycharm

项目描述： 为了便于用户选择合适的房源，通过爬取我爱我家，链家等相关的需求目标网站获取租房数据，以便满足用户的需求。

技术描述：

- ♥ 用抓包工具抓取出目标网站的基本信息
- ♥ 基于 Python 分布式爬虫爬取数据
- ♥ 利用正则表达式，BeautifulSoup, xpath 方法找出需求信息
- ♥ 按照所需格式对需求信息进行简单处理
- ♥ 将数据存储到 MySQL 数据库中

兴趣爱好

- ♥ 技术广度探索者
- ♥ 书籍深度畅游者
- ♥ 运动不懈追求者
- ♥ 数据世界崇尚者
- ♥ 翻山越岭旅游者
- ♥ 台球俱乐娱乐者

自我评价

- ★ 有很好的团队合作精神，工作中积极求上进，有良好的沟通能力。
- ★ 对技术的无限热爱与追求，对技术学习有满腔热情，更喜欢钻研新技术。
- ★ 有良好的代码风格和书写规范，良好的学习能力和抗压能力，能够胜任新的工作。
- ★ 对数据知识体系有足够的动力。
- ★ 对机器学习伟大的实践充满向往。