XX

基本信息

- ✧ 民族：汉族　　年龄：24
- ✧ 手机：xxxxxxxx　　邮箱：beantsxu@gmail.com
- ✧ 博客：https://beants.github.io　　GitHub：https://github.com/beants

教育背景

- ✧ 2016 年-2020 年　　北京 xxxx 大学　　本科　　软件工程

自我评价

- ✧ 1 年半**简历解析**行业工作经验，熟悉面向对象程序设计，有足够的能力参与并组织项目组从项目需求分析、设计、开发到产品的部署，同时提升了沟通能力与英文文档阅读能力；
- ✧ 2 年半科研经验，3 个大学生科技创新计划项目（其中**国家级**一项），累计**发明专利**（公开号：CN109117129A）一项、**著作权**多项、论文 5 篇，积累了一定的**英文阅读能力**和丰富的项目开发经验。同时，熟悉开源项目开发流程，有**开源项目参与经验**，具有较强的编码能力和代码功底；
- ✧ 擅长数据处理相关工作，有**金融数据**和**文本数据**处理经验，有一定的**机器学习**经验，对数据比较敏感；
- ✧ 团队意识强，具备成熟的沟通技巧及能力，熟悉项目开发流程，具备项目管理经验，有多人项目协作经验和丰富的**独立开发**经验，曾多次独立完成 BS、CS 结构项目，在学校期间与深圳大学、香港城市学院、中科院信工所等机构有过项目交流经验。

工作/实习经验

✧ **2020 年 07 月-　　海搜科技（深圳）科技有限公司-（英特利普集团）　　算法工程师**

- 负责简历解析项目的项目结构设计、程序编写、主要算法编写等工作；
- 严格按照面向对象程序设计标准对项目架构进行设计，并对部分结构进行模块化设计，为后续的项目优化奠定了扎实的基础；
- 将原本只能解析文本文件的程序架构升级为支持多种文件格式，整体提升了简历解析项目的解析能力；
- 通过优化代码结构，优化算法时间复杂度，简化解析流程将简历的解析速度提高了 1 倍以上，已达到业界平均水平；
- 参与某搜索项目的设计与编写工作，主要是 ES 语句的构造与优化；
- 负责本项目的 Dockerfile、GitHub Action yml 等自动化脚本的编写，优化了部署流程；

✧ **2019 年 11 月-2020 年 06 月　　北京智谱华章科技有限公司　　平台部实习研发工程师**

- 负责某 Java 后端的重构工作，优化代码结构，增加可维护性；
- 负责 Go 后端部分接口工作，按照项目需求完成代码的编写工作；
- 负责部分 Scala 后端向 Go 后端的迁移重构工作；

✧ **2019 年 06 月-2019 年 09 月　　中科院信息工程研究所　　暑期实习生**

- 参与某涉密网络武器项目的实验性搭建工作；

- 阅读英文论文、项目源代码，参与讨论并优化平台架构与结构，复现项目架构；

✧ **2018 年 01 月-2018 年 09 月　　北京金信网银科技有限公司　　数据运营部数据处理实习生**

- 按照需求从互联网搜集公开社交网络的金融数据和以及其他可爬取的金融数据；
- 按照项目需求对获取到的数据进行清洗；
- 根据项目需求使用正则或机器学习等方法对数据进行二次加工并分析；

项目经验

✧ **2020 年 08 月-今　　简历解析项目　　负责人**

项目介绍：基于开源在 Github 上两个开源项目与自己编写的文档解析器，将各个格式的文档解析为项目自定义数据结构的文档；通过各种文本、字体、坐标、背景颜色等对文档文本进行重新组合排列，进而通过正则、机器学习算法等进行数据挖掘。

相关技术：Python3、Github、MongoDB、ElasticSearch、Redis、wsgiref、Jieba 以及各种开源 Python 库。

✧ **2019 年 12 月-2020 年 06 月　　基于知识图谱的虚假新闻检测工具设计与实现　　独立完成**

数据获取：知识图谱的数来自于**开源数据集 OwnThink**，包含约 2500 万个实体和 1.4 亿条实体属性关系，使用 Neo4J 数据库进行**知识图谱搭建**；虚假新闻数据集来源于 2019 智源&计算所-互联网虚假新闻检测挑战赛，其中包括约 4 万条数据；

数据预处理和存储：使用 jieba 工具包对虚假新闻数据集进行**分词**，计算 TF-IDF 值，并获取每条新闻的 TopK **关键词**，组成二维矩阵，然后在搭建好的知识图谱中获取两两关键词之间的**最短距离**，并填充到上述矩阵中，最后将矩阵列表存储到文件中，其中虚假新闻标记为 1，真是新闻标记为 0；

数据分析：通过 **sklearn** 工具包中的多种算法模型进行训练，获取在该实验方法中，相同训练集数量的情况下，虚假新闻检测准确率随机器学习模型的变化，和在机器学习模型相同的情况下，训练集数量对模型虚假新闻检测准确率的影响，最后通过 F1 值进行评估。

成果：经过上述的处理之后，预测的正确率能达到 82%，基本符合预期，准确率与相同水平的论文相似，同时申请软件**著作权 1 项**，**发表论文 1 篇**。

✧ **2018 年 09 月-2018 年 12 月　　隐式编程规则过滤方法及装置　　第二作者**

使用 **sklearn** 和 **word2vec** 等工具对函数名及调用路径进行分析，其中包括**定义分词规则**，使用 text8 语料库进行训练，分析函数名间的相互**关联度**，把关联度超过阈值的函数名组合定义为**新的函数调用规则**，通过此规则分析其他程序的漏洞。

在导师的带领下，编写相关测试工具，并完成相关实验。同时作为**第二作者**，申请**专利《隐式编程规则过滤方法及装置》专利公开号：CN109117129A**，申请软件著作权 2 个。参与发表**论文一篇**“Mining Function Call Sequence Patterns Across Different Versions of the Project for Defect Detection.”（SATE 2018）

✧ **2019 年 03 月-2019 年 07 月　　恶意安卓应用检测　　第二作者**

数据获取：使用 Python 语言对小米应用商店进行爬取作为非恶意应用，其中下载器为自己编写的多线程下载器，恶意应用来源于 ArgusLab 的共享数据；

数据预处理和存储：使用 apktools 工具对 apk 文件进行反编译，通过分析反编译后的 smali 文件的语法分析出**官方函数库**的调用情况，用过分析 AndroidManifest.xml 文件获取该 应用调用的权限，上述两个指标都包括名称和数量；通过上述步骤对数据的处理，将名称和调用次数映射在从 Google 官网获取的官网的安卓系统的 API 包和权限名称，形成二维矩阵。同时进行数据的标注处理，将恶意应用标记为 1，非恶意应用标记为 0；

数据分析：通过 sklearn 工具包中的贝叶斯分类器和 TensorFlow 工具中的 DNN 对数据进行处理，并通过十折交叉进行评估。

成果：经过上述的处理之后，预测的正确率能达到 98%~99%之间。

✧ **2017 年 12 月-2018 年 12 月　　基于大数据的互联网信息综合分析平台　　独立完成**

项目目的：通过大数据技术处理互联网新闻信息，方便用户浏览针对某一特定事件的新闻。

数据获取：使用 Python 语言、 BeautifulSoup 库和正则表达式从各大互联网新闻平台爬取新闻数据；

数据预处理和存储：使用 Python 语言按照要求将爬取得数据清洗好并存入 MongoDB 数据库中；

数据处理：使用 Python 语言和 Sklearn 工具对数据进行处理，按照“分词->向量化->提取特征->训练->预测”的思路对数据进行处理，目的是将同一事件的新闻聚类，方便用户的阅读；

数据展示：使用 Flask 框架设计了网站，并严格遵循 **RESTful** 的设计原则；

成果：申报**国家级**大学生科技创新计划，并发表**论文两篇**：“基于大数据技术的新闻采集和事件分析系统的设计与实现”、“基于 Flask 框架的展示型网站的设计与实现”。申请**软件著作权：“新闻采集及时间提取系统 V1.0”、“新闻采集及时间提取系统 V2.0” 登记号：2018SR960946；**

职业技能

- ✧ 熟悉面向对象开发，有 Go 语言、Java 语言使用经验、熟练使用 **Python3**；
- ✧ 熟悉 Docker 的使用，以及 DockerFile、GitHub Action 等自动化脚本的编写；
- ✧ 熟悉 **MongoDB**、Mysql、**Redis** 等数据库；
- ✧ 有后端编写经验，会使用 **Flask**、**Django**、Spring Boot 等框架进行项目编写，了解 **RESTful** 设计原则，了解**幂等**设计原则；
- ✧ 有前端编写经验，熟悉 **Web**、**Android**、**Flutter**、**小程序**开发过程、熟悉 **Vue**、**Bootstrap** 框架、有 **HTML**、**CSS**、**JavaScript** 编写经验；
- ✧ 熟悉网页抓取原理及技术，熟悉正则表达式，熟练使用 **Python 爬虫** ，**Selenium** 等自动化测试工具，并对**反爬虫策略**有一定研究 ；
- ✧ 了解**机器学习**算法、**知识图谱**等，有新闻数据和金融数据和人才数据相关的**自然语言处理经验；**
- ✧ 熟悉**安卓反编译**过程，了解**漏洞挖掘**过程以及会使用简单工具；
- ✧ 熟练使用 **Git** 工具，有**开源项目**开发参与经验，有一定的**英语阅读**能力，可以阅读英文文档以及进行简单的英文交流；