

Resume Project

目录

- [作者](#)
- [项目简介](#)
- [项目分工](#)
- [引用API](#)
- [关于各个文件的修改](#)
- [关于数据及最后结果的几点说明](#)

作者

刘宇宸， 刘昕彤， 苏畅

项目简介

Resume项目主要分为两个部分

1. 从json格式的resume数据中提取字段
2. 将字段导入MySQL数据库中

项目分工

Main.java由 刘宇宸 编写， 苏畅 整合
数据库的建立由 刘昕彤 完成

引用API

引用API以及下载地址

[_json-lib-2.2.3及附加](http://pan.baidu.com/s/1dDo4PzV) (<http://pan.baidu.com/s/1dDo4PzV>).

[_mysql-connector-java-5.1.30](http://dlwt.csdn.net/fd.php?i=816472178052927&s=b7e2e666ccf1ed1653598fb5c91635e9) (<http://dlwt.csdn.net/fd.php?i=816472178052927&s=b7e2e666ccf1ed1653598fb5c91635e9>).

关于各个文件的修改

- Main.java
主要代码由 刘宇宸 编写，后来 苏畅 添加了需要的sql语句部分
- JdbcTest.java
由 苏畅 编写，用于测试JDBC连接数据库，使用的jar是mysql-connector-java，后将代码在main.java中整合
- 数据库
 - 1.由于数据缺少的原因，数据库建表的sql语句在5-11行中的 NOT NULL 被删除
 2. MySQL 并不支持 VARCHAR 字节超过2000，所以将大文本用 TEXT 存储
 - 3.经过测试，有些学校的名字长度超过50，因此改为 VARCHAR(150)
 - 4.经测试发现，resume数据中含有一些不常见的学位信息，这些 ENUM 类型中没有设置，所以就将 ENUM 类型改成了 VARCHAR 类型

关于数据及最后结果的几点说明

- resume数据存在的问题
 - 1.没有姓名，联系方式，这些在数据库中都设成了NULL值
 - 2.部分简历中没有工作信息或者学校信息，这些也都设置成了NULL值
 - 3.部分工作，学习后面含有引号，双引号等不合法字符，采用了正则表达式过滤
 - 4.部分学校姓名后含有HTML的图片以及图片连接的代码，经测试这些对象是非法字符串，无法通过正则过滤；考虑到这些数据基本上20000条中会出现3条，频率很小，所以直接continue不再处理这些数据
 - 5.有些人的学校信息存在多条重复