

小析智能定制化简历解析产品方案设计书

1. 项目介绍：针对CGL模板的简历进行信息化提取，通过本地部署方式接入上海弘玑提供的服务器
2. 解决方案：
 - a. 方案宗旨：解决CGL简历模板中工作简历的混淆及缺失问题
 - b. 方案概述：上海弘玑额外提供的50-100份简历样本作为训练集，小析智能将建立定制化简历解析模型组，从识别、提取、判断以及逻辑等方面进行优化，在小析简历解析产品对CGL简历模板时调用适用的模型组，从而优化CGL简历的工作经历版块解析效果，部分个例情况无法进行优化（个例无法优化情况请参考附件，但附件无法涵盖所有情况，需具体案例具体分析）。如解析其它类型简历时，仍可按小析标准化产品解析简历。
附赠功能：项目经历整段提取。由于项目经历无法优化的个例较多（部分已在附件列举），小析将以附赠形式，使用CGL简历解析模式时，在原有项目经历版块的基础上，额外添加“整段项目经历”的字段，整段提取CGL简历中所有项目经历内容。
 - c. 主要内容：本地部署小析的定制化简历解析产品后，通过指定参数对不同平台的简历进行解析。在CGL模板的简历解析，可支持一种CGL简历解析模式，调用参数为`parse_mode=ogl`，其余参数仍可参照官网进行，若解析其它类型的简历仍可通过参数`parse_mode`不同模式，具体参数类型请参考官网开发文档（<http://wiki.xiaoxizn.com/#parser-api-base64>）。
 - d. 涵盖服务：上海弘玑定制化简历解析产品支持多种文档格式的CGL简历，如遇到图片格式或图片形式的PDF则需要额外OCR支持。
 - e. 开发周期：4-6周
 - f. 所需材料：由上海弘玑提供50-100份简历作为训练集，额外50份CGL简历作为验证集
3. 预期效果（验收标准）：通过CGL简历解析模式解析验证集简历，除附件中提及的个例以外，准确率达到85%或以上。
 - a. 准确率计算方式：正确提取CGL简历字段数占简历可提取的总字段数（不含项目经历及整段项目经历）的比例。具体可提取的字段类型可参考小析简历解析的结果列表（<https://wiki.xiaoxizn.com/#parser-result>）。
4. 方案价格：
 - a. 本地部署定制化简历解析（永久版），请与商务对接。

附件

短期内无法优化的情况，但无法列举所有情况，具体案例具体分析。

1. OCR对图片格式和含图片格式的PDF格式的简历中文字提取不准确导致的解析错误。

2. 在公司/项目名称及工作/项目描述中，中英文夹杂且比较长

a. 例1：刘宏志-CGL推荐报告.docx

项目名称：华为 SAP HANA TDI 解决方案设计

b. 例2：刘磊-CGL推荐报告.docx

项目名称：Q2C ~JAVA Web 开发项目

职位：Developer

时间：2015.11-2016.04

公司：上海惠普有限公司

项目描述：维护型项目,采用 Spring MVC + AngularJS,主要负责维护代码,修改 bug。

项目名称：AVS - Web 自动化测试项目

时间：2011.04-2015.04

公司：上海惠普有限公司

项目描述：自动化框架开发与维护

项目名称：ATPCO - Web 自动化测试项目

时间：2010.07-2011.05

公司：印孚瑟斯技术(中国)有限公司

项目描述：

3. 无项目关键词且项目名称不明确难进行识别（即需要项目经历标志关键词或者项目名称：xxx等项目关键词）

a. 例：张彪-CGL推荐报告.docx

项目经验

核电大型变压器在线监测与智能诊断平台集中研发项目 2021.03-至今

前端工程师

所在公司：杭州安脉盛智能技术有限公司

项目描述：在线监测与智能诊断平台集中研发项目

项目职责：负责前端部分架构及开发工作

项目业绩：

田湾电机状态监测项目 2020.10-至今

前端工程师

所在公司：杭州安脉盛智能技术有限公司

项目描述：为江苏核电多台设备开发 phm 定制化系统

项目职责：负责前端部分架构及开发工作

项目业绩：

旋转机械状态无线监测技术 2020.08-至今

前端工程师

所在公司：杭州安脉盛智能技术有限公司

项目描述：旋转机械状态无线监测技术产品化研发项

4. 在工作经历后跟着“项目”二字，会被识别为工作经历一部分，对整份简历的版块划分有所影响

a. 例：刘洪明-CGL推荐候选人.docx



2006. 07-2011. 02

NTT DATA

技术 leader

工作地点：上海

工作职责：

➢ 技术组长, 写代码

2003. 07-2006. 06

中国软件

程序员

工作地点：

工作职责：

➢ 代码开发

项目：

销售运营数据分析 BI 平台

高级架构师 2021.07-至今

内容

业务概要：

1. 行业大盘数据分析,给运营参谋部提供全国各个区域各个城市,细化下钻到一个 1-2 公里范围的网格的店铺,商品,用

5. 项目经历写在工作经历里面嵌套着没法处理，项目名中大量空格会导致没法正确提取项目名

a. 例：许勇强CGL推荐报告.docx

时间：2017.07—2020.06

公司：比亚迪汽车工业有限公司

职位：机械结构工程师

（一）项目名称：比亚迪云巴 II 代道岔/曲线道岔研发设计

时间：2018.08-2020.06

工作职责：

- 1、项目规划:在云巴道岔(I 代)的基础上,与团队成员定义项目研发设计方案。
- 2、项目管理:依据项目定义,制订项目研发计划,明确团队各阶段的具体任务,跟进项目进程。
- 3、设计开发:运用 UG、CAD 等工具建立三维方案,设计工程图纸,运用 ANSYS、ADAMS 等仿真软件完成结构件仿真计算,把控整体方案一致性和准确定。
- 4、文件输出:运用 FEMA 理论分析项目安全性,撰写 BOM、工艺规程等技术规范,以及专利申请等文档资料。
- 5、安装调试:输出项目安装指导书,指导单体设备安装与调试,参与项目联调,并提供项目售后技术支持。

6. 开源项目格式混乱无法提取

a. 例：彭铮-CGL推荐报告.docx

开源项目	+	
Loggrove		
对本地、远程日志文件进行阅读、轮询、关键词匹配、统计、监控、告警、图表展示的日志系统。		
Demo: http://8.140.146.105:6218/ 用户名: guest 密码: guest123		
Dormer		
支持指标数据收集、函数计算、用户与组织空间、树形授权与定义聚合图形的监控系统。		
Demo: http://8.140.146.105:3221/ 用户名: guest 密码: dormer123		
测试打数: 执行 <code>echo "app.test.your-nickname.value 5 `date +%s`" nc 8.140.146.105 2003</code> , 打数成功后在“指标->app”页面, 指标树中逐级查找指标。		
Loggrove		