ResumeSDK简历解析系统安装文档

[一、服务架构介绍 0](#_Toc461539969)

[二、安装文件介绍 1](#_Toc461539970)

[三、环境依赖 1](#_Toc461539971)

[1、硬件环境 1](#_Toc461539972)

[2、操作系统 1](#_Toc461539973)

[3、Java环境 1](#_Toc461539974)

[4、Python环境 1](#_Toc461539975)

[三、安装步骤 2](#_Toc461539976)

[1、创建www用户 2](#_Toc461539977)

[2、部署文件解析服务 2](#_Toc461539978)

[1）创建模块环境 2](#_Toc461539979)

[2）修改配置 2](#_Toc461539980)

[3）修改文件夹权限： 2](#_Toc461539981)

[4）启动&停止服务： 3](#_Toc461539982)

[3、部署简历解析服务 3](#_Toc461539983)

[1）拷贝安装包 3](#_Toc461539984)

[2）选择字符编码 3](#_Toc461539985)

[3）获取签名文件 3](#_Toc461539986)

[4）启动&停止服务 3](#_Toc461539987)

【注意事项】

1）服务需要有官方授权的签名文件方能正常启动，自主安装前请先阅读《安装文档》中获取签名文件部分；

2）授权签名文件和服务器硬件关联，因此请不要自行更换或升级服务器硬件，以免影响服务的正常运行；

# 一、服务架构介绍

ResumeSDK简历解析系统由简历解析和文件解析两部分组成，其中：

1. 简历解析模块：基于python开发，将输入的简历解析为结构化字段，默认端口为2015，通过uwsgi提供应用服务；
2. 文件解析模块：基于java开发，从附件提取出文本，默认端口为2014，通过tomcat提供应用服务，被简历解析模块所调用；

为提高解析效率，两个服务需同机部署。

基本结构图如下所示，



# 二、安装文件介绍

1. requirements.txt：python依赖包list；
2. check\_ucs.py：ucs编码方式检测脚本；
3. resume\_parser.zip：简历解析安装包；
4. resume\_file\_parser.war：文件解析安装包；

# 三、环境依赖

## 1、硬件环境

硬件配置越高越好，以下为推荐的最低配置需求：

1. CPU：X86-64、4核及以上；
2. 内存：16G及以上；
3. 硬盘：100G及以上；
4. 网络：能联网；（以便安装python相关的lib包及调用百度OCR解析）

## 2、操作系统

1. Linux系统（centos/redhat等，推荐CentOS 6.5及以上），带gcc（v4.6及以上）；

## 3、Java环境

1. Java8：参考官网进行安装；
2. Tomcat7：<https://tomcat.apache.org/download-70.cgi>，参考官网进行安装；

## 4、Python环境

1. 安装Python2.7：
   1. 确保bz2已安装，如未安装，则运行
      * CentOS6：yum install libbz2-devel
      * CentOS7：yum install bzip2-devel
      * Ubuntu/Debian：apt-get install libbz2-dev
   2. 参考相关文档安装python
2. 安装pip：（已安装则忽略）
   1. 从<https://pypi.org/project/pip/#files> 下载pip-x.y.z.tar.gz; (x.y.z为实际版本号，比如10.0.1)
   2. 运行tar –xvzf pip-x.y.z.tar.gz
   3. 进入pip-x.y.z/，运行python setup.py install
3. 安装第三方包
   1. 运行pip install –r requirements.txt；（requirements.txt随安装文档提供）
   2. 安装过程若有error则会退出，通过下述方式进行解决：
      1. 定位安装错误的包，比如chardet，通过运行pip install chardet进行安装；
      2. 重新运行pip install –r requirements.txt，直至所有包均安装成功；

# 三、安装步骤

设tomcat安装路径为：/usr/local/apache-tomcat-7.0.70；

设简历解析模块计划安装路径：/data/apps/；

若实际安装路径有别于上述路径，则下述替换下述安装步骤中对应的路径为实际路径。

## 1、创建www用户

1. 运行命令：useradd www

注：resume\_parser和resume\_file\_parser两个服务请以用户www来运行，否则会碰到访问权限异常问题。

## 2、部署文件解析服务

### 1）创建模块环境

1. 创建路径：mkdir –p /data/apps/resume\_parser/data/tmp\_dir;
2. 进入/usr/local/apache-tomcat-7.0.70/webapps/，创建文件夹resume\_file\_parser；
3. 解压war包：

拷贝resume\_file\_parser.war至resume\_file\_parser/文件夹下，运行jar –xvf resume\_file\_parser.war进行解压；

### 2）修改配置

1. 修改local\_path配置项（可选）：

改/usr/local/apache-tomcat-7.0.70/webapps/resume\_file\_parser/WEB-INF/classes/parser.properties的配置项为：local\_path=/data/apps/resume\_parser/data/tmp\_dir；

1. 修改tomcat的端口号为2014：

修改/usr/local/apache-tomcat-7.0.70/conf/server.xml，http对应的端口号为2014：



3）**关于图片解析**

ResumeSDK使用百度OCR的接口进行图片解析，因此如果需要对图片类型简历进行解析：

a）请保持网络顺畅；

b）请申请百度ocr通用文字解析版：https://cloud.baidu.com/product/ocr/general，得到APP\_ID/API\_KEY/SECRET\_KEY，

替换tomcat/webapps/resume\_file\_parser/WEB-INF/classes/parser.properties下的ai/ak/sk三个配置字段即可。（百度提供每天50000的免费调用量）

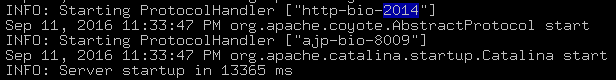
### 3）修改文件夹权限：

1. chown –R www:www /usr/local/apache-tomcat-7.0.70;

### 4）启动&停止服务：

1. 启动服务：su –c “sh /usr/local/apache-tomcat-7.0.70/bin/startup.sh” www

若logs/catalina.out下输出如下信息，则启动成功：



1. 停止服务：su –c “sh /usr/local/apache-tomcat-7.0.70/bin/shutdown.sh” www

## 3、部署简历解析服务

### 1）拷贝安装包

1. 拷贝resume\_parser.zip至部署路径：/data/apps/；
2. 运行unzip resume\_parser.zip进行解压；

### 2）确认字符编码

1. 拷贝check\_ucs.py脚本至部署路径：/data/apps/；
2. 运行python check\_ucs.py得到ucs编码方式，请确保输出为UCS4，否则python需要按UCS4编译方式进行重新安装；

### 3）获取签名文件

1. 进入/data/apps/resume\_parser/，运行：python ./bin/app/app\_resume.py，此时运行不成功，但会输出类似以下的digest信息：



1. 发送该digest信息给ResumeSDK方，从ResumeSDK方得到ra.cert.enc许可签名文件；
2. 将ra.cert.enc文件放置/data/apps/resume\_parser/data/pem\_dir目录下；

注：签名文件通过采集机器指纹进行加密生成，每台机器对应一个签名文件。因此签名文件只发送一次，请确保所选择的机器为最终要部署的机器！

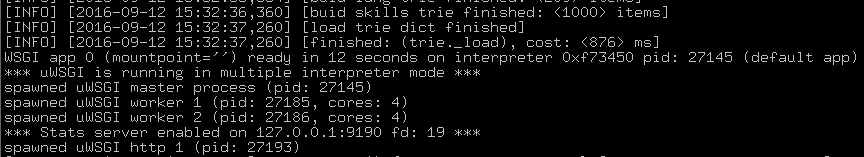
### 4）启动&停止服务

启动前，先进入/data/apps/resume\_parser路径：

方式一（以方式运行）：

1. 启动服务：运行sh run.sh start

当出现log/uwsgi\_resume\_parser.log如下输出信息时，则启动成功：



1. 停止服务：运行 sh run.sh stop

方式二（以flask方式运行）：

1. 启动服务：sh start.sh
2. 停止服务：sh stop.sh