

对外API文档 - 简历信息抽取

功能介绍

简历信息抽取旨在从大量真实简历中提取出最有价值的信息，从而提高政府，企业，学校等筛选人才的效率。本服务提供的是基于 **UIE** 进行 **zero-shot** 的简历信息抽取模型，目前支持抽取的实体 schema 为 “姓名”，“国籍”，“民族”，“居住地”，“学历”，“专业”，“组织”，“职位”，由于是 zero-shot，所以后续的迭代支持对 schema 进行任意拓展和修改。

在线服务接口定义

服务路由

resume_info_extraction

入参说明

字段	类型	说明
input_data	List of String	每份简历的全文本

入参示例

```
1 input_data = [  
2     '''常建良，男，1963年出生，工科学士，高级工程师，北京物资学院客座副教授。1985年8月–1993年在国家物资局、物资部、国内贸易部金属材料流通司从事国家统配钢材中特种钢材品种的全国调拨分配工作，先后任科员、副主任科员、主任科员。1993年5月–1999年5月受国内贸易部委派到国内贸易部、冶金部、天津市政府共同领导组建的北洋(天津)钢材批发交易市场任理事长助理、副总裁。1999年5月–2010年4月任天津一德投资集团有限公司董事。2010年5月任天津一德投资集团公司董事、副总裁。''',  
3     '''陈学军先生：1967年5月出生，大学毕业，高级经济师。1986年7月进入无锡威孚高科技集团股份有限公司。历任公司采供处处长兼党支部书记，党委工作处处长。无锡威孚高科技集团股份有限公司第四届、第五届监事会主席，第六届董事会副董事长兼总经理，第七届董事会董事长。现任大股东无锡产业发展集团有限公司董事局董事、无锡威孚高科技集团股份有限公司党委书记。'''  
4 ]
```

返回说明

字段	类型	说明
----	----	----

姓名	string	"" 为未抽取到相关信息
国籍	string	"" 为未抽取到相关信息
民族	string	"" 为未抽取到相关信息
居住地	sngstr	"" 为未抽取到相关信息
学历	list of string	"" 为未抽取到相关信息
专业	list of string	"" 为未抽取到相关信息
组织	list of string	"" 为未抽取到相关信息
职位	list of string	"" 为未抽取到相关信息

返回示例

```
1  [
2      {
3          "姓名": "常建良",
4          "国籍": "",
5          "民族": "",
6          "居住地": "",
7          "学历": [
8              "学士"
9          ],
10         "专业": [
11             "工科"
12         ],
13         "组织": [
14             "北京物资学院",
15             "国家物资局",
16             "物资部",
17             "国内贸易部金属材料流通司",
18             "国内贸易部",
19             "国内贸易部",
20             "冶金部",
21             "北洋(天津)钢材批发交易市场",
22             "天津一德投资集团有限公司",
23             "天津一德投资集团公司"
24         ],
25         "职位": [
26             "高级工程师",
27             "客座副教授",
28             "科员",
```

```
29         "副主任科员",
30         "主任科员",
31         "理事长助理",
32         "副总裁",
33         "董事",
34         "董事",
35         "副总裁"
36     ]
37 },
38 {
39     "姓名": "陈学军",
40     "国籍": "",
41     "民族": "",
42     "居住地": "",
43     "学历": [
44         "大学"
45     ],
46     "专业": "",
47     "组织": [
48         "无锡威孚高科技集团股份有限公司",
49         "公司",
50         "无锡威孚高科技集团股份有限公司",
51         "无锡产业发展集团有限公司",
52         "无锡威孚高科技集团股份有限公司"
53     ],
54     "职位": [
55         "高级经济师",
56         "采供处处长",
57         "党支部书记",
58         "党委工作处处长",
59         "第五届监事会主席",
60         "第六届董事会副董事长",
61         "总经理",
62         "第七届董事会董事长",
63         "董事局董事",
64         "党委书记"
65     ]
66 }
67 ]
```



以下为内部内容

效果说明

本服务支持中文简历的实体抽取，目前实体的schema为“姓名”，“国籍”，“民族”，“居住地”，“学历”，“专业”，“组织”和“职位”共八类；在“新浪网金融企业高管简历数据集” - ResumeNER上直接进行zero-shot的效果如下：

	precision	recall	f1-score	support
CONT	0.87	0.96	0.92	28
EDU	0.85	0.76	0.80	112
LOC	0.10	0.17	0.12	6
NAME	1.00	0.99	1.00	112
ORG	0.92	0.90	0.91	553
PRO	0.70	0.79	0.74	33
RACE	1.00	1.00	1.00	14
TITLE	0.95	0.93	0.94	772
micro avg	0.92	0.91	0.92	1630
macro avg	0.80	0.81	0.80	1630
weighted avg	0.93	0.91	0.92	1630

临时API地址

http://192.168.50.121:8989/resume_info_extraction

请求代码示例

```
1 import requests
2 import json
3
4 if __name__ == "__main__":
5
6     url = 'http://192.168.50.121:8989/resume_task_extraction' # 服务地址
7     input_data = [
8         '''常建良，男，1963年出生，工科学士，高级工程师，北京物资学院客座副教授。1985年8月
9         '''陈学军先生：1967年5月出生，大学毕业，高级经济师。1986年7月进入无锡威孚高科技集
10     ]
11     resp = requests.request("POST", url, data=json.dumps(input_data))
12     print(json.dumps(resp.json(), ensure_ascii=False, indent=4))
```