

用户登入

介绍

用户登入

请求失败

name	required	type	description
message	False	String	
status	False	Integer	状态码

状态说明

message参数说明	status参数说明	http code
登入成功	200	200
登入失败	500	500
参数校验出错	422	422
登入失败次数太多，请稍后重试	400	400
用户名或密码错误	400	400
用户已禁用	400	400

请求

URL

- `http://host:port/auth/login`

方法

- `POST`

Headers

name	required	default	description
Content-Type	True	application/json	请设置成 <code>application/json</code>

Body

name	required	type	description
password	True	String	用户密码
username	True	String	用户名

name	required	type	description
verification	False	Object	验证码

响应

Headers

name	required	description
Content-Type	True	application/json

Body

name	required	type	description
access_token	False	String	access token
message	False	String	
need_update_password	False	Boolean	是否需要修改初始密码
nickname	False	String	用户昵称
refresh_token	False	String	refresh token
status	False	Integer	状态码
user_id	False	Integer	用户ID

示例

```
# python
import base64
from pprint import pprint

import requests

timeout = (5, 60)
response = requests.post(
    'http://host:port/auth/login',
    json=dict(
        username=base64.b64encode(b'username').decode(),
        password=base64.b64encode(b'password').decode()
    )
)
response.raise_for_status()
pprint(response.json())

# curl
curl -v -H 'Content-Type: application/json' -d '{"username":
"username", "password": "password"}' -X POST
```

```
'http://host:port/auth/login'
```

创建文档抽取或文字识别任务

介绍

异步方式请求成功

name	required	type	description
ids	False	Array[Integer]	抽取 ID 列表. Deprecated: 抽取ID详见items字段
message	False	String	
status	False	Integer	状态码
items	False	Array[Object]	抽取信息
total	False	Integer	抽取文件总数

请求失败字段说明：

请求失败

name	required	type	description
message	False	String	
status	False	Integer	状态码

状态说明

message参数说明	status参数说明	http code
创建任务成功	200	200
创建任务失败	500	500
同步任务只允许上传一份文件	400	400
上传文件包含不合法文件,上传失败!	400	400
参数校验出错	422	422
文件大小或页数超过上限	400	400

回调使用说明

1. 回调参数请使用 `form` 格式，以 `callback_args` 作为 key。
2. `callback_args` 值必须是 json 序列化后的 object 。

请求

URL

- `http://host:port/api/extracting/v2/task`

方法

- `POST`

Headers

name	required	description
Authorization	True	Bearer {jwt token}

Querystring

name	required	type	description
async_task	True	Boolean	async_task不允许为空, true: 异步任务, false: 同步任务
feature_type_id	True	Integer	卡片 ID
file_types	False	Array[Integer]	文件类型列表, 1: 电子文档, 2: 图像文档
language	False	String	语言代码
task_name	False	String	任务名称

Body

name	required	type	description
files	True	Array[Binary]	上传文件不允许为空, 同步任务只允许上传一份文件

响应

Headers

name	required	description
Content-Type	True	application/json

Body

name	required	type	description
content	False	String	
ids	False	Array[Integer]	抽取 ID 列表. Deprecated: 抽取ID详见items字段
items	False	Array[Object]	抽取信息

name	required	type	description
message	False	String	
name	False	String	
result	False	Object	
status	False	Integer	状态码
total	False	Integer	抽取文件总数

字段详情

items[*] 字段详情

name	required	type	description
create_time	False	String	创建时间
extract_task_id	True	Integer	抽取任务 ID
extract_version	False	Integer	抽取服务版本号
field_level	False		实体字段准确度
global_level	False	String	文档准确度
graph_define_level	False		组合字段准确度
id	False	Integer	
idps_version	False	Integer	当前idps系统版本
is_saved_by_business	False	Boolean	是否按业务存储文件
last_update_time	False	String	最近更新时间
origin_file	False	String	旋转前的原始文件备份
origin_name	True	String	文件原始名称
relation_extract_result	False		抽取结果
unique_name	True	String	文件在系统内的唯一名称

示例

curl

```
curl \
  -X POST \
  "http://host:port/extracting/v2/task?
  async_task=True&feature_type_id=1&file_types=1&language=example_string&tas
  k_name=example_string" \
  -H "Authorization: Bearer {jwt token}" \
```

```
-F "files=@测试.pdf" \  
-v
```

python

```
from pprint import pprint  
import requests  
  
timeout = (5, 60)  
  
params = {  
    "async_task": True,  
    "feature_type_id": 1,  
    "file_types": [1],  
    "language": "example_string",  
    "task_name": "example_string",  
}  
headers = {  
    "Authorization": "Bearer {jwt token}",  
}  
files = [  
    ("files", open('测试.pdf', 'rb')),  
]  
  
response = requests.post(  
    "http://host:port/extracting/v2/task",  
    timeout=timeout,  
    params=params,  
    headers=headers,  
  
    files=files,  
  
)  
response.raise_for_status()  
pprint(response.json())
```

获取抽取结果请求

请求

URL

- http://host:port/api/extracting/v2/{extract_id}

方法

- GET

Path

name	required	type
extract_id	True	Integer

Headers

name	required	description
Authorization	True	Bearer {jwt token}

响应

Headers

name	required	description
Content-Type	True	application/json

Body

name	required	type
code	False	Integer
item	False	Object
message	False	String
status	False	Integer

字段详情

item 字段详情

name	required	type	description
assessor_id	False	Integer	Deprecated: 旧字段，调用 v2 接口请勿使用该字段
assessor	False	String	Deprecated: 旧字段，调用 v2 接口请勿使用该字段
content	False	String	纯文本内容
create_time	False	String	创建时间
doc_id	False	Integer	Deprecated: 旧字段，调用 v2 接口请勿使用该字段
doc_name	False	String	文件原始名称

name	required	type	description
doc_status	False	Integer	Deprecated: 旧字段，调用 v2 接口请勿使用该字段
doc_type	False	Integer	文件类型, 1: 电子文档, 2: 图像文档
docType	False	Integer	Deprecated: 旧字段，调用 v2 接口请勿使用该字段
enable_img_render	False	Boolean	
extract_task_id	True	Integer	抽取任务 ID
extract_version	False	Integer	抽取服务版本号
feedback	False	Boolean	是否反馈给算法模型的训练任务
field_level	False	Object	实体字段准确度
global_level	False	String	文档准确度
graph_define_level	False	Object	组合字段准确度
id	False	Integer	
idps_version	False	Integer	
img_path	False	String	
is_saved_by_business	False	Boolean	是否按业务存储文件
language	False	String	多语种抽取-语言类型
last_update_time	False	String	最近更新时间
name	False	String	文件原始名称
ocr_extract_type	False	String	ysOCR抽取类型
operator_id	False	Integer	Deprecated: 旧字段，调用 v2 接口请勿使用该字段
operator	False	String	Deprecated: 旧字段，调用 v2 接口请勿使用该字段
origin_file	False	String	旋转前的原始文件备份
origin_name	True	String	文件原始名称
result	False	Object	抽取结果
rich_content_file_name	False	String	富文本文件路径
shortcut	False	Array[Object]	快捷键配置
status_name	False	String	抽取状态

name	required	type	description
status	False	Integer	抽取执行的状态, 0: 未处理, 2: 文本转换失败, 3: 文本转换成功, 4: 预标注失败, 5: 预标注成功, 6: 文档转换中, 7: 抽取中, 20001: 文件检查失败, 8: 已审核
support_img_render_selection	False	Boolean	
tag_type_id	False	Integer	文档类型 ID
task_id	False	Integer	Deprecated: 旧字段, 调用 v2 接口请勿使用该字段
task_name	False	String	Deprecated: 旧字段, 调用 v2 接口请勿使用该字段
unique_name	True	String	文件在系统内的唯一名称

item.shortcut[*] 字段详情

name	required	type	description
id	True	Integer	字段 ID
name	True	String	字段名称
shortcut	True	Array[]	快捷键组合

示例

```
{
  "status": "# 抽取结果状态: 成功"OK", 失败"ERROR",
  "doc_prob": "# 抽取置信度
  "pdf_path": "# 文件地址
  "fuzzy_tag_list": "# 模糊抽取字段 (暂时无用)
  "tag_list": "# 抽取字段
    [after_treatment: null # 字段后处理结果
    color: [242, 54, 21] # 字段颜色, RGB格式:
    confidence: 1 # 置信度
    index: 1355 # 在文件中的位置
    ppr: null # 字段后处理结果
    tag_id: 11 # 字段id
    tag_name: "电话" # 字段名称
    terms_id: 11 # 字段id
    value: "443115095705." # 字段抽取内容
    word: "443115095705." # 字段抽取内容
  ],
  "relation_extract": "# 关系抽取
    [{
      "children":
        [{
```

```
        "children":
        [{
            "id":5567, # 组合内的字段id
            "name":"BlackTeaLatte", # 字段名称
            "type":"tag", # 无用
            "value":[{}] # 同上面tag_list内容, 组合内字段的具体信息
        }],
        "confidence":1, # 字段置信度
        "type":"card" # 无用
    }],
    "id":5576, # 组合字段id
    "name":"Latte",# 组合字段名称id
    "node_index":1, # 组合字段排序
    "type":"group" # 无用
}]
}
```

示例

curl

```
curl \
-G \
"http://host:port/extracting/v2/{extract_id}?" \
-H "Authorization: Bearer {jwt token}" \
-v
```

python

```
from pprint import pprint
import requests

timeout = (5, 60)

headers = {
    "Authorization": "Bearer {jwt token}",
}

response = requests.get(
    "http://host:port/extracting/v2/{extract_id}",
    timeout=timeout,

    headers=headers,

)
```

```
response.raise_for_status()
pprint(response.json())
```

下载高亮 PDF 文件

介绍

请求

URL

- `http://host:port/api/extracting/v2/{id}/export/highlight`

方法

- `GET`

Path

name	required	type
<code>id</code>	True	Integer

Headers

name	required	description
Authorization	True	Bearer {jwt token}

响应

Headers

name	required	description
Content-Type	True	<code>application/json</code>

Body

name	required	type	description
<code>code</code>	False	Integer	
<code>item</code>	False	Object	文件名，下载地址
<code>message</code>	False	String	
<code>status</code>	False	Integer	

示例

curl

```
curl \
  -G \
  "http://host:port/extracting/v2/{id}/export/highlight?" \
  -H "Authorization: Bearer {jwt token}" \
  -v
```

python

```
from pprint import pprint
import requests

timeout = (5, 60)

headers = {
    "Authorization": "Bearer {jwt token}",
}

response = requests.get(
    "http://host:port/extracting/v2/{id}/export/highlight",
    timeout=timeout,

    headers=headers,

)
response.raise_for_status()
pprint(response.json())
```