

# 北京神州泰岳软件股份有限公司

## 人工智能研究院

### NLP 服务文档结构化

### API 调用说明

2019 年 4 月 2 日

说明版本: V1

## 前言

神州泰岳人工智能研究院，利用智慧语义认知技术、数据挖掘技术等优势，从事人工智能领域的基础研究。该说明旨在提供 PDF 表格提取服务在中文自然语言处理方面的基础语义服务。

本文由以下章节构成：

第一章 PDF 表格提取 API 调用说明

# 第一章 PDF 表格提取 API 调用说明

## 1. 接口描述

将给定的 PDF 文档中的表格识别并结构化抽取。

## 2. 请求说明

### 请求参数

- ◇ HTTP 方法: POST
- ◇ 请求 URL: <http://structured.market.alicloudapi.com/pdfbiaogetiqu>
- ◇ 请求参数: 请求 body 为非表单数据, 内容如下

```
{  
  "in": "待分析 PDF 文件 Base64 编码字符串"  
}
```

## 3. 返回说明

### 返回参数

参数	是否必须	类型	值
out	是	String	表格分析结果
success	是	boolean	请求成功与否
status	是	integer	请求返回码

## 返回示例

```
{
  "status": 200,
  "err": null,
  "data": {
    "out": [{
      "x0": 90,
      "y0": 87,
      "spreadOfNum": 0,
      "sprX": 0,
      "sprY": 0,
      "cellLists": [{
        "cells": [{
          "text": "保险金 \n",
          "row": 0,
          "column": 0,
          "rowspanNum": 1,
          "colspanNum": 1,
          "rowspan": false,
          "colspan": false,
          "rowAndCol": false
        },
        {
          "text": "领取人 \n",
          "row": 0,
          "column": 1,
          "rowspanNum": 1,
          "colspanNum": 1,
          "rowspan": false,
          "colspan": false,
          "rowAndCol": false
        }
      ]
    }
  ]
}
```

```
    },  
    "spread": false  
  }  
},  
"context": "com.dinfo.dispatcher.core.ModuleInvocation.invoke",  
"token": null,  
"success": true,  
"notSuccess": false  
}
```

```
    "审判人员": {  
      "title": "审判人员",  
      "content": "\n 审判员 黄 蓁",  
      "start": 1130,  
      "end": 1138  
    },  
    "裁判日期": {  
      "title": "裁判日期",  
      "content": "\n 二〇〇一年二月十九日",  
      "start": 1138,  
      "end": 1149  
    },  
    "书记员": {  
      "title": "书记员",  
      "content": "\n 书记员 陆剑佩",  
      "start": 1149,  
      "end": 1157  
    }  
  }  
}  
},  
"success": true,  
"status": 200  
}
```