

北京神州泰岳软件股份有限公司

人工智能研究院

NLP 服务信息抽取

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。该说明旨在提供时间抽取、数值抽取两大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 时间抽取 API 调用说明

第二章 数值抽取 API 调用说明

第一章 时间抽取 API 调用说明

1. 接口描述

语言中存在大量表示时间的内容，如“第二天”、“2001 年”等，这些时间词语对于语言的理解有重要的意义。时间抽取服务提供自动识别文本中表示时间信息的功能。服务面向不同领域、不同场景的数据类型，结合时间信息上下文语境，与信息抽取技术，预先构建有效分析时间的语义模型。通过语义模型，提供检测出文本中精确表示时间的字符串内容，实现时间信息的有效识别与抽取功能。

2. 请求说明

请求参数

- ◇ HTTP 方法：POST
- ◇ 请求 URL：http://extracting.market.alicloudapi.com/shijianchouqu
- ◇ 请求参数：

参数	值
in	分析文本数据

3. 返回说明

返回参数

参数	是否必须	类型	值
out	是	String	时间抽取结果 (entityLable: 实体 标签 ,entityName: 实体名字)
success	是	boolean	请求成功与否
status	是	integer	请求返回码

返回示例

```
{
  "data": {
    "out": [{
      "entityLable": "time#时间词",
      "entityName": "25 日"
    },
    {
      "entityLable": "time#时间词",
      "entityName": "1 月 25 日"
    },
    {
      "entityLable": "time#时间词",
      "entityName": "1 月 24 日"
    },
    {
      "entityLable": "time#时间词",
      "entityName": "初五"
    },
    {
      "entityLable": "time#时间词",
      "entityName": "节后"
    },
    {
      "entityLable": "time#时间词",
      "entityName": "农历正月初三"
    }
  ]
},
  "success": true,
  "status": 200
}
```

第二章 数值抽取 API 调用说明

1. 接口描述

语言中表示数量的表达方式，如“20 吨”、“5000 万元”，是对于事物属性的量化描述。数量表达式抽取的功能就是将中文中表示数量的表达方式抽取出来。数量抽取采用规则的方式建立数量模型。同时建立匹配的数量表达式和消歧的数量表达式。输入文本后，先进行匹配，得到匹配内容，再根据匹配到正则的类别判断是否需要消歧，对于需要消歧的时间表达式通过前\后 n 个字符是否包含词表中的字符串，判断是否是数量表达式。

2. 请求说明

请求参数

- ◇ HTTP 方法: POST
- ◇ 请求 URL: <http://extracting.market.alicloudapi.com/shuzhichouqu>
- ◇ 请求参数:

参数	值
in	待分析文本数据

3. 返回说明

返回参数

参数	是否必须	类型	值
result	是	String	数值抽取结果
success	是	boolean	请求成功与否
status	是	integer	请求返回码

返回示例

```
{
  "data": {
    "out": [{
      "entityLable": "classifier#名量词",
      "entityName": "3000 多名"
    },
    {
      "entityLable": "classifier#名量词",
      "entityName": "3000 多名"
    }
  ]
},
  "success": true,
  "status": 200
}
```