文智自然语言处理

API 文档

产品文档

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2018 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传 播全部或部分本文档内容。

【商标声明】

腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方 主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整 。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定 ，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

文档目录

API 文档
- 简介
- API概览
  - 调用方式
    - 请求结构
    - 接口鉴权
    - 公共参数
      - 返回值
        - 返回值结构
        - 错误码
        - 异步任务接口返回格式
- API调用快速上手指南
  - 通用API相关接口
    - 分词&命名实体识别API
    - 情感分析API
    - 文本分类API
    - 关键词提取API
    - 句法分析API
    - 同义词API
    - 纠错API
    - 转码API
    - 敏感信息识别API
  - 下载抽取API相关接口
    - 下载抽取API

版权所有：腾讯云计算（北京）有限责任公司

API 文档

简介

最近更新时间：2017-03-09 10:38:00

腾讯文智自然语言处理（Natural Language Processing）基于并行计算、分布式爬虫系统，结合独特的语义分析技术，一站满足NLP、转码、抽取、数据抓取等需求。基于文智API可实现搜索、推荐、舆情、挖掘等功能。文智同时支持定制化语义分析方案。

腾讯云文智中文语义平台以SDK模块的方式提供服务，多种编程语言都可以轻松使用。

版权所有：腾讯云计算（北京）有限责任公司

API概览

最近更新时间：2017-03-09 11:17:51

1. 通用API

接口	接口功能	Action ID
分词&命名实体识别API	提供智能分词、词性标注、命名实体识别功能	LexicalAnalysis
情感分析API	分析信息中情感的正向，负向及中性	TextSentiment
文本分类API	提供自动文本分类服务，目前能识别90多个类别	TextClassify
关键词提取API	为用户实现诸如新闻内容，评论等的关键词提取	TextKeywords
敏感信息识别API	识别信息的色情、政治敏感程度。	TextSensitivity
句法分析API	分析出句子中词与词之间的关系	TextDependency
同义词API	提供同义词查询服务	LexicalSynonym
纠错API	提供针对短文本的纠错功能	LexicalCheck
转码API	将PC端页面转换为适合在手机等移动端展示的页面	ContentTranscode

2. 下载抽取API

接口	接口功能	Action ID
下载抽取API	用户只需提供一个URL即可完成数据抓取	ContentGrab

版权所有：腾讯云计算（北京）有限责任公司

调用方式
请求结构

最近更新时间：2017-10-31 17:13:21

1. 服务地址

腾讯云文智服务使用的域名访问地址为：wenzhi.api.qcloud.com。

2. 通信协议

腾讯云API的所有接口均通过HTTPS进行通信，提供高安全性的通信通道。

3. 请求方法

同时支持 POST 和 GET 请求，需要注意不能混合使用。即如果使用 GET 方式，则参数均从 Querystring 取得；如果使用 POST 方式，则参数均从 Request Body 中取得，Querystring 中的参数将忽略。两种方式参数格式规则相同，一般使用 GET，当参数字符串过长时使用POST，请见各接口详细描述。

4. 字符编码

均使用UTF-8编码。

5. API请求结构

名称	描述	备注
API入口	API调用的 WebService 入口	https://wenzhi.api.qcloud.com/v2/index.php
公共参数	每个接口都包含的通用参数	详见 公共参数 页面
指令名称	API要执行的指令的名称，这里使用Action指定，例如Action=LexicalAnalysis	完整的指令请参见 API概览
指令参数	每个特定的指令需要的参数	详见每个指令的接口文档

接口鉴权

最近更新时间：2018-07-06 17:24:34

腾讯云 API 会对每个访问的请求进行身份验证，即每个请求都需要在公共请求参数中包含签名信息（Signature），以验证用户身份。签名信息由用户所执有的安全凭证生成，安全凭证包括 SecretId 和 SecretKey，若用户还没有安全凭证，则需要在腾讯云官网上自主申请，否则无法调用云 API 接口。

申请安全凭证

在第一次使用腾讯云 API 之前，用户需要在【腾讯云控制台】> 【API 密钥管理 】上申请安全凭证。安全凭证包括 SecretId 和 SecretKey，其中：

- **SecretId：**用于标识 API 调用者身份；
- **SecretKey：**用于加密签名字符串和服务器端验证签名字符串的密钥。

> **注意：**
> API 密钥是构建腾讯云 API 请求的重要凭证，使用腾讯云 API 可以操作您名下的所有腾讯云资源，为了您的财产和服务安全，请妥善保存并定期更换密钥，当您更换密钥后，请及时删除旧密钥。

申请安全凭证步骤：

1. 登录 腾讯云控制台。
2. 单击【云产品】，选择【管理工具】栏下的【云 API 密钥】，进入云 API 密钥管理页面。

3. 在 API 密钥管理 页面，单击【新建密钥】即可以创建一对 SecretId/SecretKey。

> **注意：**
> - 开发商帐号最多可以拥有两对 SecretId / SecretKey。
> - 被开发商添加为子用户的 QQ 帐号，在不同开发商控制台，可以申请不同的安全凭证。
> - 子用户的安全凭证，目前仅可调用部分接口的云 API。

生成签名串

有了安全凭证 SecretId 和 SecretKey 后，就可以生成签名串了。生成签名串的详细过程如下：

假设用户的 SecretId 和 SecretKey 分别是：
SecretId： AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
SecretKey： Gu5t9xGARNpq86cd98joQYCN3Cozk1qA

版权所有：腾讯云计算（北京）有限责任公司

> **注意：**
> 这里只是示例，请用户根据自己实际的 SecretId 和 SecretKey 和请求参数进行后续操作。

以腾讯云 CVM 为例，当用户调用腾讯云 CVM 的 查看实例列表 (DescribeInstances)接口时，其请求参数为：

参数名称	描述	参数值
Action	方法名	DescribeInstances
SecretId	密钥 ID	AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
Timestamp	当前时间戳	1465185768
Nonce	随机正整数	11886
Region	实例所在区域	ap-guangzhou
SignatureMethod	签名方式	HmacSHA256
InstanceIds.0	待查询的实例 ID	ins-09dx96dg

1. 对参数排序

首先对所有请求参数按参数名做字典序升序排列。（所谓字典序升序排列，直观上就如同在字典中排列单词一样排序，按照字母表或数字表里递增顺序的排列次序，即先考虑第一个“字母”，在相同的情况下考虑第二个“字母”，依此类推。）您可以借助编程语言中的相关排序函数来实现这一功能，如 PHP 中的 ksort 函数。上述示例参数的排序结果如下：

```
{
    "Action" : "DescribeInstances",
    "InstanceIds.0" : "ins-09dx96dg"
    "Nonce" : 11886,
    "Region" : "ap-guangzhou",
    "SecretId" : "AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA",
    "SignatureMethod" : "HmacSHA256",
    "Timestamp" : 1465185768,

}
```

使用其它程序设计语言开发时，可对上面示例中的参数进行排序，得到的结果一致即可。

2. 拼接请求字符串

此步骤将生成请求字符串。
将把上一步排序好的请求参数格式化成“参数名称”=“参数值”的形式，如对 Action 参数，其参数名称为 "Action"，参数值为 "DescribeInstances"，因此格式化后就为 Action=DescribeInstances。

> **注意：**
> 1. “参数值”为原始值而非 url 编码后的值。

> 2. 若输入参数的 Key 中包含下划线，则需要将其转换为“.”，但是 Value 中的下划线则不用转换。如 `Placement_Zone=CN_GUANGZHOU` ，则需要将其转换成 `Placement.Zone=CN_GUANGZHOU` 。

然后将格式化后的各个参数用 "&" 拼接在一起，最终生成的请求字符串为（请忽略文中的换行）：

```
Action=DescribeInstances
&InstanceIds.0=ins-09dx96dg
&Nonce=11886
&Region=ap-guangzhou
&SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
&SignatureMethod=HmacSHA256
&Timestamp=1465185768
```

3. 拼接签名原文字符串

此步骤将生成签名原文字符串。
签名原文字符串的拼接规则为：

> **请求方法 + 请求主机 +请求路径 + ? + 请求字符串**

参数构成说明：

- **请求方法：** 支持 POST 和 GET 方式，这里使用 GET 请求， 注意方法为全大写。
- **请求主机：**即主机域名，请求域名由接口所属的产品或所属产品的模块决定，不同产品或不同产品的模块的请求域名会有不同。如腾讯云 CVM 的查询实例列表（DescribeInstances）的请求域名为： `cvm.api.qcloud.com` ，具体产品请求域名详见各接口说明。
- **请求路径：** 腾讯云 API 对应产品的请求路径，一般是一个产品对应一个固定路径，如腾讯云 CVM 请求路径固定为 `/v2/index.php` 。
- **请求字符串：** 即上一步生成的请求字符串。

因此，上述示例的拼接签名原文字符串结果为（请忽略文中的换行）：

```
GETcvm.api.qcloud.com/v2/index.php?Action=DescribeInstances
&InstanceIds.0=ins-09dx96dg
&Nonce=11886
&Region=ap-guangzhou
&SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
&SignatureMethod=HmacSHA256
&Timestamp=1465185768
```

4. 生成签名串

此步骤生成签名串。

> **注意：**
> 计算签名的方法有两种：HmacSHA256 和 HmacSHA1 这里要根据您指定的签名算法（即 SignatureMethod 参数）

> 生成签名串。当指定 SignatureMethod 为 HmacSHA256 时，需要使用 HmacSHA256 计算签名，其他情况请使用 HmacSHA1 计算签名。

首先使用签名算法（HmacSHA256 或 HmacSHA1）对上一步中获得的 **签名原文字符串** 进行签名，然后将生成的签名串使用 Base64 进行编码，即可获得最终的签名串。

具体代码如下，以 PHP 语言为例，由于本例中所用的签名算法为 **HmacSHA256**，因此生成签名串的代码如下（使用其它程序设计语言开发时，可用上述示例中的原文字符串进行签名验证，得到的签名串与例子中的一致即可）：

```
$secretKey = 'Gu5t9xGARNpq86cd98joQYCN3Cozk1qA';
$srcStr = 'GETcvm.api.qcloud.com/v2/index.php?Action=DescribeInstances&InstanceIds.0=ins-09dx96dg&Nonce
$signStr = base64_encode(hash_hmac('sha256', $srcStr, $secretKey, true));
echo $signStr;
```

最终得到的签名串为:

```
0EEm/HtGRr/VJXTAD9tYMth1Bzm3lLHz5RCDv1GdM8s=
```

同理，当您指定签名算法为 **HmacSHA1** 时，生成签名串的代码如下：

```
$secretKey = 'Gu5t9xGARNpq86cd98joQYCN3Cozk1qA';
$srcStr = 'GETcvm.api.qcloud.com/v2/index.php?Action=DescribeInstances&InstanceIds.0=ins-09dx96dg&Nonce
$signStr = base64_encode(hash_hmac('sha1', $srcStr, $secretKey, true));
echo $signStr;
```

最终得到的签名串为：

```
nPVnY6njQmwQ8ciqbPl5Qe+Oru4=
```

签名串编码

生成的签名串并不能直接作为请求参数，需要对其进行 URL 编码。
如上一步生成的签名串为 `0EEm/HtGRr/VJXTAD9tYMth1Bzm3lLHz5RCDv1GdM8s=` ，则其编码后为 `0EEm/HtGRr/VJXTAD9tYMth1Bzm3lLHz5RCDv1GdM8s=` 。因此，最终得到的签名串请求参数 (Signature) 为： `0EEm/HtGRr/VJXTAD9tYMth1Bzm3lLHz5RCDv1GdM8s=` ，它将用于生成最终的请求URL。

> **注意：**
> 如果用户的请求方法是 GET，则对所有请求参数的参数值均需要做 URL 编码；此外，部分语言库会自动对 URL 进行编码，重复编码会导致签名校验失败。

鉴权失败

当鉴权不通过时，可能出现如下表的错误：

错误代码	错误类型	错误描述
4100	身份认证失败	身份验证失败，请确保您请求参数中的 Signature 按照上述步骤计算正确，特别注意 Signature 要做 url 编码后再发起请求。
4101	未被开发商授权访问本接口	该子用户未被授权调用此接口。请联系开发商授权，详情请查阅 授权策略。
4102	未被开发商授权访问本接口中所操作的资源	请问的资源参数中，存在未被开发商授权授权访问的资源，请在 message 字段中查看无权查看的资源 ID。 请联系开发商授权，详情请查阅 授权策略。
4103	非开发商的 SecretId 暂不支持调用本接口	子用户的 SecretId 不支持调用此接口，只有开发商有权调用。
4104	SecretId 不存在	签名所用的 SecretId 不存在，也可能是密钥状态有误，请确保 API 密钥有效且未被禁用。
4110	鉴权失败	权限校验失败，请确保您有使用所访问资源的权限。
4500	重放攻击错误	请注意 Nonce 参数两次请求不能重复，Timestamp 与腾讯服务器相差不能超过 2 小时。

公共参数

最近更新时间：2017-10-31 17:14:51

公共参数是用于标识用户和接口鉴权目的的参数, 如非必要, 在每个接口单独的接口文档中不再对这些参数进行说明, 但每次请求均需要携带这些参数, 才能正常发起请求。

名称	类型	描述	必选
Action	String	接口指令的名称，例如: LexicalAnalysis	是
Region	String	区域参数，用来标识希望操作哪个区域的实例。可选: bj:北京 gz:广州 sh:上海 hk:香港 ca:北美	是
Timestamp	UInt	当前UNIX时间戳	是
Nonce	UInt	随机正整数，与 Timestamp 联合起来, 用于防止重放攻击	是
SecretId	String	由腾讯云平台上申请的标识身份的 SecretId 和 SecretKey, 其中 SecretKey 会用来生成 Signature 具体参考 接口鉴权 页面	是
Signature	String	请求签名，用来验证此次请求的合法性, 具体参考 接口鉴权 页面	是

一个典型的接口请求如下, Action=LexicalAnalysis表示分词&命名实体识别。

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=LexicalAnalysis
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &text=我爱洗澡
    &code=2097152
```

其中text、code为指令参数，其余为通用参数。

返回值
返回值结构

最近更新时间：2017-10-31 17:17:58

如无特别说明, 每次请求的返回值中, 都会包含下面的字段：

名称	类型	描述
code	Int	返回结果的错误码，0表示成功，其它值表示失败。具体错误码的含义可以参考 错误码 页面
message	String	请求结果

例如：

使用公共参数部分的示例请求：

```
https://wenzhi.api.qcloud.com/v2/index.php?
  Action=LexicalAnalysis
  &Nonce=345122
  &Region=sz
  &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
  &Timestamp=1408704141
  &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
  &text=我爱洗澡
  &code=2097152
```

可能的返回结果如下：

```
{
    "code": 0,
    "message": "",
    "combtokens": [
        {
            "cls": "短语",
            "pos": 0,
            "wlen": "8",
            "word": "我爱洗澡"
        }
    ],
    "tokens": [
        {
            "pos": 0,
            "wlen": "2",
            "word": "我",
            "wtype": "代词",
            "wtype_pos": 27
```

```
        },
        {
            "pos": 2,
            "wlen": "2",
            "word": "爱",
            "wtype": "动词",
            "wtype_pos": 31
        },
        {
            "pos": 4,
            "wlen": "4",
            "word": "洗澡",
            "wtype": "动词",
            "wtype_pos": 31
        }
    ]
}
```

错误码

最近更新时间：2017-03-09 11:05:40

错误码	含义说明
400	HTTP Method不正确
401	HTTP请求参数不符合要求
503	调用额度已超出限制
504	服务故障
4000	缺少必要参数，或者参数值格式不正确，具体错误信息请查看错误描述 message 字段。
4100	签名鉴权失败，请参考文档中鉴权部分。
4200	请求已经过期，请参考文档中请求有效期部分。
4300	帐号被封禁，或者不在接口针对的用户范围内等。
4400	请求的次数超过了配额限制，请参考文档请求配额部分。
4500	请求的 Nonce 和 Timestamp 参数用于确保每次请求只会在服务器端被执行一次,所以本次的 Nonce 和上次的不能重复, Timestamp 与腾讯服务器相差不能超过 2 小时。
4600	协议不支持，请参考文档说明。
5000	资源标识对应的实例不存在，或者实例已经被退还，或者访问了其他用户的资源。
5100	对资源的操作失败，具体错误信息请查看错误描述 message 字段，稍后重试或者联系客服人员帮忙解决。
5200	购买资源失败，可能是不支持实例配置，资源不足等等。
5300	购买资源失败，余额不足。
5400	批量操作部分执行成功, 详情见方法返回值。
5500	购买资源失败，用户资质审核未通过。
6000	服务器内部出现错误，请稍后重试或者联系客服人员帮忙解决。
6100	本版本内不支持此接口或该接口处于维护状态等。注意: 出现这个错误时, 请先确定接口的域名是否正确, 不同的模块, 域名可能不一样。
6200	当前接口处于停服维护状态，请稍后重试。

版权所有：腾讯云计算（北京）有限责任公司

异步任务接口返回格式

最近更新时间：2017-03-09 05:19:04

1. 普通异步任务接口返回格式

一次请求只能操作一个资源的异步任务接口，例如创建负载均衡，重置主机操作系统等。

名称	类型	描述	必选
code	Int	返回结果的错误码，0表示成功，其它值表示失败。	Yes
message	String	返回结果的错误信息	No
requestId	String	任务编号	Yes

2. 批量异步任务接口返回格式

一次请求能操作多个资源的异步任务接口，例如修改密码，启动机器，停止机器等。

名称	类型	描述	必选
code	Int	返回结果的错误码，0表示成功，其它值表示失败。	Yes
message	String	返回结果的错误信息	No
detail	Array	以资源ID为key, 返回对该资源操作的code,message,requestId	Yes

例如：

```
{
      "code":0,
      "message": "success",
      "detail":
      {
          "qcvm6a456b0d8f01d4b2b1f5073d3fb8ccc0":
         {
          "code":0,
          "message":"",
          "requestId":"1231231231231":,
         }
           "qcvm6a456b0d8f01d4b2b1f5073d3fb8ccc0":
         {
           "code":0,
           "message":"",
```

```
            "requestId":"1231231231232":,
          }
      }
}
```

> 注意：
> 资源全部操作成功，则最外层code为0
> 资源全部操作失败，则最外层code会返回5100
> 资源部分操作失败，则最外层code会返回5400
> 在第3种情况下，终端可以通过detail得到失败部分的操作信息。

API调用快速上手指南

最近更新时间：2018-06-11 15:59:03

1 介绍

腾讯文智自然语言处理（Natural Language Processing）基于并行计算、分布式爬虫系统，结合独特的语义分析技术，一站满足NLP、转码、抽取、数据抓取等需求。基于文智API可实现搜索、推荐、舆情、挖掘等功能。文智同时支持定制化语义分析方案。

腾讯云文智中文语义平台以SDK模块的方式提供服务，多种编程语言都可以轻松使用。在正式使用之前，您需要首先在腾讯云上注册文智账号。

2 API调用概述

2.1 通信协议和规则

详见《腾讯云通信协议和规则》

2.2 公共参数

公共参数是用于标识用户和接口鉴权的参数, 每次请求均需要携带这些参数, 才能正常发起请求。

名称	类型	描述	必选
Action	String	接口指令的名称，例如: LexicalAnalysis	是
Region	String	区域参数，用来标识希望操作哪个区域的实例。可选: gz:广州; sh:上海;hk:香港; 等部分云产品并非每个区域都有提供, 获取产品对应的地域列表可以使 用/v2/DescribeProductRegionList	是
Timestamp	UInt	当前UNIX时间戳	是
Nonce	UInt	随机正整数，与Timestamp联合起来, 用于防止重放攻击	是
SecretId	String	由腾讯云平台上申请的标识身份的SecretId和SecretKey, 其中SecretKey会用来 生成Signature 具体参考接口鉴权页面	是
Signature	String	请求签名，用来验证此次请求的合法性 具体参考接口鉴权页面	是

公共参数详见《腾讯云公共参数》

2.3 接口鉴权

接口鉴权方法详见《接口鉴权》

> 注意：在生成签名的过程中，需要将加密字符串中包含的"_"改写成"."，从而加密产生签名。

2.4 异步任务接口返回格式

详见《腾讯云异步任务接口返回格式》

2.5 错误码

详见《调用方式-返回值-错误码》

3 调用示例

这里将以一个简单的情感分析任务为例，介绍腾讯云SDK文智模块的使用。

3.1 首先请在腾讯云官方SDK下载地址

下载或更新最新版本的SDK（本次以php-sdk为例）

- 从 github 获取最新版本 SDK >>

3.2 修改demo.php文件，修改点如下

a) SecretId，SecretKey改为自己腾讯云上相应的值，这里查看：

http://manage.qcloud.com/capi/capiManage.php

b) $package=array('offset'=>0, 'limit'=>3); 改为：

$package = array("content"=>"李亚鹏挺王菲：加油！孩儿他娘。");

说明：这是文智情感分析接口的参数。

c) $a=$cvm->DescribeInstances($package); 改为：

$a = $wenzhi->TextSentiment($package);

说明：这是文智模块的相关接口，具体请查看接口列表：
http://cloud.tencent.com/wiki/v2/API

d) 其他所有地方的$cvm改为$wenzhi，即替换为文智模块。

修改后的demo.php如下：

```
<?php
error_reporting(E_ALL ^ E_NOTICE);
require_once './src/QcloudApi/QcloudApi.php';

$config = array('SecretId'       => '您在腾讯云上的SecretId',
        'SecretKey'      => '您在腾讯云上的SecretKey',
        'RequestMethod'  => 'POST',
        'DefaultRegion'    => 'gz');
```

```
$wenzhi = QcloudApi::load(QcloudApi::MODULE_WENZHI, $config);

$package = array("content"=>"李亚鹏挺王菲：加油！孩儿他娘。");

$a = $wenzhi->TextSentiment($package);

if ($a === false) {
    $error = $wenzhi->getError();
    echo "Error code:" . $error->getCode() . ".\n";
    echo "message:" . $error->getMessage() . ".\n";
    echo "ext:" . var_export($error->getExt(), true) . ".\n";
} else {
    var_dump($a);
}

echo "\nRequest :" . $wenzhi->getLastRequest();
echo "\nResponse :" . $wenzhi->getLastResponse();
echo "\n";
```

通用API相关接口
分词&命名实体识别API

最近更新时间：2018-01-23 18:13:57

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: LexicalAnalysis

提供智能分词（基本词和短语）、词性标注、命名实体识别功能。专业的团队对数据、模型、程序进行迭代更新以保证效果的不断提升。用户只需简单的调用相关API接口即可获取到所需结果，无需担心诸如新词发现、歧义消除、调用性能等词法分析难题。词法分析已经为应用宝搜索、微信公共账号搜索等业务提供支持，均取得了良好的效果。

2. 输入参数

参数名称	必选	类型	描述
text	是	String	待词法分析的文本
code	是	Int	text 的编码(0x00200000=utf-8) 目前文智统一输入为 utf-8
type	否	Int	取值 0 或 1，默认为 0。 0 为基础粒度版分词，倾向于将句子切分的更细，在搜索场景使用为佳。 1 为混合粒度版分词，倾向于保留更多基本短语不被切分开。

3. 输出参数

参数名称	类型	描述		
code	Int	错误码。0: 成功，其他值: 失败		
message	String	错误信息		
tokens	Array	分词结果 其中 Array 元素包含以下字段		
		word	String	切分出来的基础词
		pos	UInt32	该基础词在文本中的起始位置，单位为字节数
		wtype	String	基础词的词性，详细内容参照文末词性对照表
		wtype_pos	UInt32	基础词的词性的 ID，与“ wtype ”对应，同参照文末词性对照表

		wlen	UInt64	该基础词的长度，单位为字节数，如：“我”为 2 个字节，该值为 2	
combtokens	Array	命名实体识别结果 其中 Array 元素包含以下字段			
		word	String	实体词	
		pos	UInt32	该实体在文本中的起始位置，单位为字节数	
		wlen	UInt64	该实体词的长度，单位为字节数	
		cls	Int32	人名	100000010/100000011
				地名	100000012
				机构名	100000013

4. 示例

输入

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=LexicalAnalysis
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &text=我爱洗澡
    &code=2097152
```

输出

```
{
    "code": 0,
    "message": "",
    "combtokens": [
        {
            "cls": "短语",
            "pos": 0,
            "wlen": "8",
            "word": "我爱洗澡"
        }
    ],
    "tokens": [
        {
            "pos": 0,
            "wlen": "2",
```

```
            "word": "我",
            "wtype": "代词",
            "wtype_pos": 27
        },
        {
            "pos": 2,
            "wlen": "2",
            "word": "爱",
            "wtype": "动词",
            "wtype_pos": 31
        },
        {
            "pos": 4,
            "wlen": "4",
            "word": "洗澡",
            "wtype": "动词",
            "wtype_pos": 31
        }
    ]
}
```


5. 词性对照表

ID	英文简写	含义	示例
1	A	形容词	悲惨、宝贵啊、安恬
2	AD	副形词	保障、不便、诚挚
3	AN	名形词	便捷、惨烈、产业化
4	B	区别词	报复性、壁式、便民、半旧
5	C	连词	比如、而、还有
6	D	副词	不过、不管、必然
7	E	叹词	咦、呀、呀儿哟
8	F	方位词	顶端、侧后方、东南方
9	G	语素词	
10	H	前接成分	非、亚、前
11	I	成语	安于现状、哀鸿遍野、八仙过海
12	J	简称略语	爱委会、奥申委、巴控区

版权所有：腾讯云计算（北京）有限责任公司

ID	英文简写	含义	示例
13	K	后接成分	级、期、症
14	L	习用语	安于现状、按需分配、八旗子弟
15	M	数词	八旬、百儿八十、半拉
16	N	名词	电视、小说、程序
17	NR	人名	阿布穆萨、陈宏子、王雪
18	NRF	姓	赵、张、李
19	NRG	名	
20	NS	地名	阿根廷、北京、李家屯
21	NT	机构团体	公牛队、澳联社、财政部
22	NZ	其他专名	阿贾克斯、阿里朗、阿波罗
23	NX	非汉字串	B2C、 .net、 3D
24	O	拟声词	叮咣、叭哒、丁丁当当
25	P	介词	除了、对于、经
26	Q	量词	盎司、便士、帕
27	R	代词	鄙人、别国、俺
28	S	处所词	场外、村口、杯中
29	T	时间词	春秋、半夜、北汉
30	U	助词	着、的、地
31	V	动词	安排、瞭望、哀悼
32	VD	副动词	标准化、濒临、鼎力
33	VN	名动词	罢免、拜年、败北
34	W	标点符号	，；。…
35	X	非语素字	鸳、枇、蚣
36	Y	语气词	嘛、呢、呀
37	Z	状态词	本本分分、不羁、半死
38	AG	形语素	地、鼎、瘆

ID	英文简写	含义	示例
39	BG	区别语素	赝、雄、当
40	DG	副语素	陡、务、劲
41	MG	数词性语素	丁、戊、辰
42	NG	名语素	餐、采、旙
43	QG	量语素	票、闾、仞
44	RG	代语素	伊、或、汝
45	TG	动语素	雍、今、金
46	VG	量语素	脧、崇、僈
47	YG	语气词语素	咧、耳、尔
48	ZG	状态词语素	

情感分析API

最近更新时间：2017-03-09 10:42:49

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: TextSentiment

在舆情监控、话题监督、口碑分析等商业分析领域有非常重要的应用价值。

2. 输入参数

参数名称	必选	类型	描述
content	是	String	待分析的文本（只能为utf8编码）
type	是	Int	（可选参数，默认为4） 1：电商；2：APP；3：美食；4：酒店和其他。

3. 输出参数

参数名称	类型	描述
code	Int32	错误码。0：成功，其他值：失败
message	String	错误信息
positive	Double	正面情感概率
negative	Double	负面情感概率

4. 示例

输入

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=TextSentiment
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
```

版权所有：腾讯云计算（北京）有限责任公司

```
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &content=双万兆服务器就是好，只是内存小点
```

输出


```
{
  "code": 0,
  "message": "",
  "negative": 0.138263002038002,
  "positive": 0.8617370128631592
}
```

文本分类API

最近更新时间：2017-03-09 02:18:27

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: TextClassify

为用户提供自动文本分类服务，平台已对文本分类的模型算法进行了封装，用户只需提供待分类的文本数据，而不必关注具体的实现，通过平台就能得到提供文本的所属类别。目前平台能识别类别囊括了求职招聘、影视、音乐、健康养生、财经、广告推广、犯罪、政治等90多个类别，且算法支持快速迭代更新已有类别及增加新类别。

2. 输入参数

参数名称	必选	类型	描述
title	否	String	文章标题，编码格式utf8
content	是	String	正文内容，编码格式utf8
secd_nav	否	String	二级导航，编码格式utf8
url	否	String	文章对应的url

3. 输出参数

参数名称	类型	描述		
code	Int	错误码，0：成功，其他值：失败		
message	String	错误信息		
classes	Array	文本分类结果，其中Array元素包含以下字段(一般取conf最大的为分类结果)		
		class	String	类别
		class_num	UInt32	类别编号
		conf	Double	分类置信度

文本分类API支持的类别列表如下：
（下表类别不全）

分类代码	值	说明
E_PTC_CATEGORY_UNKNOWN	0X00000000	未知分类
E_PTC_CATEGORY_TRAVEL	0X00000001	旅游 旅游攻略，景点介绍等
E_PTC_CATEGORY_GAMES	0X00000002	游戏 游戏下载，攻略，介绍，新闻等
E_PTC_CATEGORY_CHARACTERS	0X00000003	人物 主要是人物介绍、人物访谈类页面
E_PTC_CATEGORY_SPORTS	0X00000004	体育 体育赛事，运动等
E_PTC_CATEGORY_MUSIC	0X00000005	音乐 音乐相关，乐器等
E_PTC_CATEGORY_MOVIE	0X00000006	影视 电影，电视剧，电视节目
E_PTC_CATEGORY_SOFTWARE	0X00000007	软件 软件介绍，下载
E_PTC_CATEGORY_LITERATURE	0X00000008	文学 小说，文言文，散文等相关页面
E_PTC_CATEGORY_FOOD	0X00000009	美食 菜谱，食品制作，介绍等
E_PTC_CATEGORY_INTERACTION_HEALTH	0X0000000A	健康 医疗、保健、健身等
E_PTC_CATEGORY_INTERACTION_MEDICINE	0X0000000B	医药 药品介绍，药方等，其他是健康类别
E_PTC_CATEGORY_RETAILSHOP	0X0000000C	商铺 公司黄页、店铺首页等
E_PTC_CATEGORY_FINANCE	0X0000000D	财经 股票证券、理财、银行
E_PTC_CATEGORY_CAR	0X0000000E	汽车 汽车，租车等
E_PTC_CATEGORY_HOUSE	0X0000000F	房产 房地产，租房，装修等
E_PTC_CATEGORY_COMIC	0X00000010	动漫 动画，漫画
E_PTC_CATEGORY_INTERACTION_EDUCATION	0X00000011	教育 学校、学科、考试，培训等
E_PTC_CATEGORY_TECHNOLOGY	0X00000012	科技 手机、电子数码、互联网技术等
E_PTC_CATEGORY_MILITARY	0X00000013	军事 国防，军队，战争等
E_PTC_CATEGORY_WEATHER	0X00000014	天气 天气预报，气候等
E_PTC_CATEGORY_INTERACTION_ADVERTISE	0X00000016	广告
E_PTC_CATEGORY_SOCIAL_ASSEMBLE	0X00000017	群体聚集
E_PTC_CATEGORY_SOCIAL_DISASTER	0X00000018	自然灾害
E_PTC_CATEGORY_SOCIAL_ACCIDENT	0X00000019	交通事故
E_PTC_CATEGORY_SOCIAL_OFFENSE	0X0000001A	刑事犯罪

版权所有：腾讯云计算（北京）有限责任公司

E_PTC_CATEGORY_SOCIAL_FORCE	0X0000001B	暴力执法
E_PTC_CATEGORY_INTERACTION_JOB	0X0000001C	求职招聘
E_PTC_CATEGORY_SOCIAL_FOODSAFETY	0X0000001D	食品安全
E_PTC_CATEGORY_SOCIAL_POLLUTION	0X0000001E	环境污染
E_PTC_CATEGORY_SOCIAL_EPIDEMIC	0X0000001F	疾病疫情
E_PTC_CATEGORY_SOCIAL_FINANCE	0X00000020	金融安全
E_PTC_CATEGORY_POLITICAL_SENSITIVE	0X00000021	敏感 政治
E_PTC_CATEGORY_POLITICAL_CORRUPTION	0X00000022	贪腐
E_PTC_CATEGORY_POLITICAL_CULT	0X00000023	非法组织
E_PTC_CATEGORY_POLITICAL_REACTION	0X00000024	反动言论
E_PTC_CATEGORY_POLITICAL_MODEL	0X00000025	先进事迹
E_PTC_CATEGORY_INTERACTION_INSPIRATION	0X00000028	心灵鸡汤
E_PTC_CATEGORY_POLITICAL_OTHER	0X0000002A	其它政治类
E_PTC_CATEGORY_OTHER	0X0000FFFF	其他

4. 示例

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=TextClassify
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &content=腾讯入股京东
```

输出

```
{
    "code": 0,
    "message": "",
    "classes": [
        {
            "class": "未分类",
            "class_num": 0,
            "conf": 0.291
```

```
            },
            {
                "class": "旅游",
                "class_num": 1,
                "conf": 0.291
            },
            {
                "class": "科技",
                "class_num": 18,
                "conf": 0.419
            }
        ]
    }
```

关键词提取API

最近更新时间：2017-03-09 10:20:22

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: TextKeywords

基于关键词抽取平台, 为用户实现诸如新闻内容关键词自动提取、评论关键词提取等提供基础服务。支持用户自定义词典，提高在垂直领域的抽取效果。目前已经接入的业务包括腾讯新闻客户端、手机腾讯网等。

2. 输入参数

参数名称	必选	类型	描述	
title	是	String	新闻标题	
channel	否	String	新闻频道(选填 不填默认是科技)	
			CHnews_news_sports	体育新闻
			CHnews_news_ent	娱乐新闻
			CHnews_news_astro	星座新闻
			CHnews_news_auto	汽车新闻
			CHnews_news_cul	文化新闻
			CHnews_news_digi	数码新闻
			CHnews_news_finance	财经新闻
			CHnews_news_game	游戏新闻
			CHnews_news_house	房产新闻
			CHnews_news_lad	时尚新闻
			CHnews_news_mil	军事新闻
			CHnews_news_ssh	社会新闻
			CHnews_news_tech	科技新闻
			CHnews_news_others	其它

版权所有：腾讯云计算（北京）有限责任公司

content	是	String	新闻正文

3. 输出参数

参数名称	类型	参数描述			
code	Int32	错误码, 0: 成功, 其他值: 失败			
message	String	错误信息			
keywords	Array	关键词提取结果，其中Array元素包含以下字段			
		score	Double	关键词权重	
		keyword	String	关键词提取服务对文本标记的标签	
		type	String	keyword	文本中出现的关键词
				topic	话题

关键词提取API错误码详细信息如下：

错误码	含义说明
400	HTTP Method不正确
401	HTTP请求参数不符合要求
503	调用额度已超出限制
504	服务故障

4. 详细示例

示例业务详细信息如下表：

	参数名称	参数描述	必选	参数值示例
腾讯云公共参数	Action	方法名	是	TextKeywords
	SecretId	SecretId	是	AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
	Timestamp	当前时间戳	是	1408704141
	Nonce	随机正整数	是	345122
业务参数	title	新闻标题	是	Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外

				套、码数：SM、P330
	channel	新闻频道(选填不填默认是科技)	否	
	content	新闻正文	是	Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330

下面以上述业务为例，详细说明"关键词提取API"接口的使用方法。

4.1 接口鉴权

示例要调用服务的数据为：{"title":"Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330","content":"Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330"}
则上述业务的参数列表如下：

```
{
        'Action' : 'TextKeywords',
        'Nonce' : 345122,
        'Region' : 'sz',
        'SecretId' : 'AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA',
        'Timestamp' : 1408704141,
        'title': 'Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330',
        'content': 'Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330'
    }
```

根据上述参数列表进行签名，得出的数字签名为：HgIYOPcx5lN6gz8JsCFBNAWp2oQ（示例），详细的数字签名的生成方法请参照：《腾讯云接口鉴权》。

注意：

1）在生成签名的过程中，需要将加密字符串中包含的"_"改写成"."，从而加密产生签名；

2）鉴权时，需要将参数列表按key进行排序：字典序，同时大写在前。

4.2 API调用

根据上一步（6.4.1）中得到的数字签名，以POST请求为例构造请求URL，将数字签名加入到参数Signature中。

```
https://wenzhi.api.qcloud.com/v2/index.php?
  Action=TextKeywords
  &Nonce=345122
  &Region=sz
  &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
  &Timestamp=1408704141
  &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
  &title=Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330
  &content=Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330
```

执行上述操作之后，会将数据{"title":"Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330","content":"Dior新款，秋冬新款娃娃款甜美圆领配毛领毛呢大衣外套、码数：SM、P330"}发送给API接口，进行相应

分析。

注意：在发送请求过程中，不能将参数字符串中包含的"_"改写成"."。

上述指令返回的数据结构如下：


```
{
        "code": 0,
        "message": "",
        "keywords": [
            {
                "keyword": "p330",
                "score": 0.2800000011920929,
                "type": "keyword"
            },
            {
                "keyword": "dior",
                "score": 0.2784992158412933,
                "type": "keyword"
            },
            {
                "keyword": "毛领",
                "score": 0.2746416926383972,
                "type": "keyword"
            }
        ]
    }
```

句法分析API

最近更新时间：2017-03-09 10:52:31

1.接口描述

域名：wenzhi.api.qcloud.com
接口名: TextDependency

句法分析出句子中词与词之间的关系，可用于提取句子主干，提取句子核心词等。句法分析可以更好的理解句子，从而在机器翻译、自动问答、知识抽取等领域都可以应用。

2. 输入参数

参数名称	必选	类型	描述
content	是	String	待分析的文本（只能为utf8编码）

3. 输出参数

参数名称	类型	描述	
id	Int	节点id	
father_id	Int	父节点id	
dep_rel	String	SBV	主谓关系
		VOB	动宾关系
		IOB	间宾关系
		FOB	前置宾语
		DBL	兼语
		ATT	定中关系
		ADV	状中结构
		CMP	动补结构
		COO	并列关系
		POB	介宾关系
		LAD	左附加关系

		RAD	右附加关系
		IS	独立结构
		HED	核心关系
postag	String	词性	
word	String	词	
code	Int	0表示成功，非0表示失败	
message	String	失败时候的错误信息，成功则无该字段	

句法分析API错误码详细信息如下：

错误码	含义说明
400	HTTP Method不正确
401	HTTP请求参数不符合要求
503	调用额度已超出限制
504	服务故障

4. 详细示例

示例业务详细信息如下表：

	参数名称	参数描述	必选	参数值示例
腾讯云公共参数	Action	方法名	是	TextDependency
	SecretId	SecretId	是	AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
	Timestamp	当前时间戳	是	1408704141
	Nonce	随机正整数	是	345122
业务参数	content	待分析的文本（只能为utf8编码）	是	双万兆服务器就是好，只是内存小点

下面以上述业务为例，详细说明"句法分析API"接口的使用方法。

4.1 接口鉴权

示例要调用服务的数据为：{"content":"双万兆服务器就是好，只是内存小点"}
则上述业务的参数列表如下：

```
{
        'Action' : 'TextDependency',
        'Nonce' : 345122,
        'Region' : 'sz',
        'SecretId' : 'AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA',
        'Timestamp' : 1408704141,
        'content': '双万兆服务器就是好，只是内存小点'
    }
```

根据上述参数列表进行签名，得出的数字签名为：HgIYOPcx5lN6gz8JsCFBNAWp2oQ（示例），详细的数字签名的生成方法请参照：《腾讯云接口鉴权》。
注意：
1）在生成签名的过程中，需要将加密字符串中包含的"_"改写成"."，从而加密产生签名；
2）鉴权时，需要将参数列表按key进行排序：字典序，同时大写在前。

4.2 API调用

根据上一步（1.4.1）中得到的数字签名，以POST请求为例构造请求URL，将数字签名加入到参数Signature中。

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=TextDependency
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &content=双万兆服务器就是好，只是内存小点
```

执行上述操作之后，会将数据{"content":"双万兆服务器就是好，只是内存小点"}发送给API接口，进行相应分析。
注意：在发送请求过程中，不能将参数字符串中包含的"_"改写成"."。
上述指令返回的数据结构如下：

```
 {
    "code": 0,
    "message": "",
    "keywords": [
        [
            {
                "dep_rel": "ATT",
                "father_id": 2,
                "id": 1,
                "postag": "m",
                "word": "双"
            },
            {
                "dep_rel": "ATT",
```

```
            "father_id": 3,
            "id": 2,
            "postag": "n",
            "word": "万兆"
        },
        {
            "dep_rel": "SBV",
            "father_id": 4,
            "id": 3,
            "postag": "n",
            "word": "服务器"
        },
        {
            "dep_rel": "HED",
            "father_id": 0,
            "id": 4,
            "postag": "v",
            "word": "就是"
        },
        {
            "dep_rel": "CMP",
            "father_id": 4,
            "id": 5,
            "postag": "a",
            "word": "好"
        },
        {
            "dep_rel": "WP",
            "father_id": 4,
            "id": 6,
            "postag": "w",
            "word": ","
        },
        {
            "dep_rel": "COO",
            "father_id": 4,
            "id": 7,
            "postag": "c",
            "word": "只是"
        },
        {
            "dep_rel": "ATT",
            "father_id": 9,
            "id": 8,
            "postag": "n",
            "word": "内存"
        },
        {
            "dep_rel": "VOB",
            "father_id": 7,
            "id": 9,
            "postag": "n",
```

```
                "word": "小点"
            }
        ]
    ]
}
```

同义词API

最近更新时间：2017-03-09 11:08:50

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: LexicalSynonym

为用户提供同义词查询服务，搜索团队通过全网数据挖掘出海量同义词，并持续对数据、模型等进行迭代更新，保证同义词的效果始终与时俱进。用户也可以通过提供产品专有的数据，与我们合作打造专属的同义词库。同义词服务作为搜索引擎检索串理解的基本功能，目前已经应用在视频、音乐、应用宝、群搜、商圈等数百个产品中。

2. 输入参数

参数名称	必选	类型	描述
text	是	String	待分析的文本（目前文智统一输入为utf-8）

3. 输出参数

参数名称	类型	描述				
code	Int32	错误码, 0: 成功, 其他值: 失败				
message	String	错误信息				
query	String	输入的原文本				
syns	Array	同义词分析结果，其中Array元素包括以下字段				
		word_ori	Object	原词信息，包含以下字段		
				idx_beg	Int32	起始位置
				idx_end	Int32	终止位置
				text	String	提取的原词文本
		word_syns	Array	原词对应的同义词列表，其中Array元素包含		
				text	String	同义词文本
				conf	Double	同义词置信度

同义词API错误码详细信息如下：

错误码	含义说明
400	HTTP Method不正确
401	HTTP请求参数不符合要求
503	调用额度已超出限制
504	服务故障

4. 详细示例

示例业务详细信息如下表：

	参数名称	参数描述	必选	参数值示例
腾讯云公共参数	Action	方法名	是	LexicalSynonym
	SecretId	SecretId	是	AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
	Timestamp	当前时间戳	是	1408704141
	Nonce	随机正整数	是	345122
业务参数	text	待分析的文本（目前文智统一输入为utf-8）	是	周杰伦结婚

下面以上述业务为例，详细说明"同义词API"接口的使用方法。

4.1 接口鉴权

示例要调用服务的数据为：{"text":"周杰伦结婚"}
则上述业务的参数列表如下：

```
{
        'Action' : 'LexicalSynonym',
        'Nonce' : 345122,
        'Region' : 'sz',
        'SecretId' : 'AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA',
        'Timestamp' : 1408704141,
        'text': '周杰伦结婚'
    }
```

根据上述参数列表进行签名，得出的数字签名为：HgIYOPcx5lN6gz8JsCFBNAWp2oQ（示例），详细的数字签名的生成方法请参照：《腾讯云接口鉴权》。

注意：

1）在生成签名的过程中，需要将加密字符串中包含的"_"改写成"."，从而加密产生签名；

2）鉴权时，需要将参数列表按key进行排序：字典序，同时大写在前。

4.2 API调用

根据上一步（4.4.1）中得到的数字签名，以POST请求为例构造请求URL，将数字签名加入到参数Signature中。

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=LexicalSynonym
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &text=周杰伦结婚
```

执行上述操作之后，会将数据{"text":"周杰伦结婚"}发送给API接口，进行相应分析。

注意：在发送请求过程中，不能将参数字符串中包含的"_"改写成"."。

上述指令返回的数据结构如下：

```
{
        "code": 0,
        "message": "",
        "syns": [
            {
                "word_ori": {
                    "idx_beg": 0,
                    "idx_end": 2,
                    "text": "周杰伦"
                },
                "word_syns": [
                    {
                        "conf": 0.2000000029802322,
                        "text": "jay"
                    },
                    {
                        "conf": 0.2000000029802322,
                        "text": "周董"
                    }
                ]
            }
        ]
    }
```

纠错API

最近更新时间：2017-03-09 10:54:05

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: LexicalCheck
能够实现对短文本的自动纠错功能，长文本的自动纠错也即将推出。用户只需要提供业务数据和日志, 无需关注技术细节和更新流程, 就可以享受到业务自身定制的纠错服务, 甚至不提供业务数据，享受通用的纠错服务。目前已经接入的业务包括音乐、视频、应用宝、云搜等，评测效果均好于竞品。

2. 输入参数

参数名称	必选	类型	描述
text	是	String	待纠错的文本（utf-8）

3. 输出参数

参数名称	类型	描述
code	Int32	错误码, 0: 成功, 其他值: 失败
message	String	错误信息
text	String	纠错后词
text_annotate	String	标红的纠错后词
conf	Double	纠错置信度

纠错API错误码详细信息如下：

错误码	含义说明
400	HTTP Method不正确
401	HTTP请求参数不符合要求
503	调用额度已超出限制
504	服务故障

版权所有：腾讯云计算（北京）有限责任公司

4. 详细示例

示例业务详细信息如下表：

	参数名称	参数描述	必选	参数值示例
腾讯云公共参数	Action	方法名	是	LexicalCheck
	SecretId	SecretId	是	AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
	Timestamp	当前时间戳	是	1408704141
	Nonce	随机正整数	是	345122
业务参数	text	待纠错的文本（utf-8）	是	睡交吃饭

下面以上述业务为例，详细说明"纠错API"接口的使用方法。

4.1 接口鉴权

示例要调用服务的数据为：{"text":"睡交吃饭"}
则上述业务的参数列表如下：

```
{
        'Action' : 'LexicalCheck',
        'Nonce' : 345122,
        'Region' : 'sz',
        'SecretId' : 'AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA',
        'Timestamp' : 1408704141,
        'text': '睡交吃饭'
    }
```

根据上述参数列表进行签名，得出的数字签名为：HgIYOPcx5lN6gz8JsCFBNAWp2oQ（示例），详细的数字签名的生成方法请参照：《腾讯云接口鉴权》。

注意：

1）在生成签名的过程中，需要将加密字符串中包含的"_"改写成"."，从而加密产生签名；

2）鉴权时，需要将参数列表按key进行排序：字典序，同时大写在前。

4.2 API调用

根据上一步（3.4.1）中得到的数字签名，以POST请求为例构造请求URL，将数字签名加入到参数Signature中。

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=LexicalCheck
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
```

```
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &text=睡交吃饭
```

执行上述操作之后，会将数据{"text":"睡交吃饭"}发送给API接口，进行相应分析。

注意：在发送请求过程中，不能将参数字符串中包含的"_"改写成"."。

上述指令返回的数据结构如下：

```
{
        "code": 0,
        "message": "",
        "conf": 1.3,
        "text": "睡觉吃饭",
        "text_annotate": "睡觉吃饭"
    }
```

转码API

最近更新时间：2017-03-09 11:06:20

1.接口描述

域名：wenzhi.api.qcloud.com
接口名: ContentTranscode

转码分为两大类：网页转码和网页名片。网页转码将在PC机上展示的二维页面转换为适合在手机等移动端设备上展示的一维页面，方便用户在移动端阅读。网页名片将页面简化为主体图片、标题、摘要的组合，以“卡片”的形式展示给大众，适合做页面的分享、收藏、推广等。用户只需要提交网页的 url ，就能获取我们的转码服务，方便、快捷。当前，网页转码已为公司QQ、qzone、微云、微博、正文吧等平台提供服务。

2.输入参数

参数名称	必选	类型	描述
url	是	String	网页地址
to_html	是	Int32	1 : html 0 : xml

3.输出参数

参数名称	类型	描述
code	Int32	错误码, 0: 成功, 其他值: 失败
message	String	错误信息
content	String	转码结果

转码API错误码详细信息如下：

错误码	含义说明
400	HTTP Method不正确
401	HTTP请求参数不符合要求
503	调用额度已超出限制
504	服务故障

版权所有：腾讯云计算（北京）有限责任公司

4.详细示例

示例业务详细信息如下表：

	参数名称	参数描述	必选	参数值示例
腾讯云公共参数	Action	方法名	是	ContentTranscode
	SecretId	SecretId	是	AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
	Timestamp	当前时间戳	是	1408704141
	Nonce	随机正整数	是	345122
业务参数	url	网页地址	是	www.163.com
	to_html	1 : html 0 : xml	是	1

下面以上述业务为例，详细说明"转码API"接口的使用方法。

4.1 接口鉴权

示例要调用服务的数据为：{"url":"www.163.com","to_html":1}
则上述业务的参数列表如下：

```
{
        'Action' : 'ContentTranscode',
        'Nonce' : 345122,
        'Region' : 'sz',
        'SecretId' : 'AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA',
        'Timestamp' : 1408704141,
        'url': 'www.163.com',
        'to_html': 1
    }
```

根据上述参数列表进行签名，得出的数字签名为：HgIYOPcx5lN6gz8JsCFBNAWp2oQ（示例），详细的数字签名的生成方法请参照：《腾讯云接口鉴权》。

注意：

1）在生成签名的过程中，需要将加密字符串中包含的"_"改写成"."，从而加密产生签名；

2）鉴权时，需要将参数列表按key进行排序：字典序，同时大写在前。

4.2 API调用

根据上一步（9.4.1）中得到的数字签名，以POST请求为例构造请求URL，将数字签名加入到参数Signature中。

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=ContentTranscode
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &url=www.163.com
    &to_html=1
```

执行上述操作之后，会将数据{"url":"www.163.com","to_html":1}发送给API接口，进行相应分析。

注意：在发送请求过程中，不能将参数字符串中包含的“”改写成“.”。

敏感信息识别API

最近更新时间：2017-03-09 11:17:52

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: TextSensitivity

识别信息的色情、政治等敏感程度。可用于敏感信息过滤，舆情监控等。

2. 输入参数

参数名称	必选	类型	描述
content	是	String	待分析的文本（只能为utf8编码）
type	是	int	区分敏感词类型:1表示色情，2表示政治

3. 输出参数

参数名称	类型	描述
code	Int32	错误码。0：成功，其他值：失败
message	String	失败时候的错误信息，成功则无该字段
sensitive	Double	敏感的概率
nonsensitive	Double	不敏感的概率

4. 示例

输入

```
https://wenzhi.api.qcloud.com/v2/index.php?
    Action=TextSensitivity
    &Nonce=345122
    &Region=sz
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
```

版权所有：腾讯云计算（北京）有限责任公司

```
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &content=六四事件是一个历史问题
    &type=2
```

输出

```
{
  "code": 0,
  "message": "",
  "sensitive": 0.7310585786300049,
  "nonsensitive": 0.2689414213699951
}
```

下载抽取API相关接口
下载抽取API

最近更新时间：2017-11-08 11:53:53

下载抽取API

1. 接口描述

域名：wenzhi.api.qcloud.com
接口名: ContentGrab

下载接口基于分布式爬虫系统，用户只需提供一个url即可轻松完成数据抓取, 也可与下载团队合作打造专有的定向抓取服务。分布式爬虫系统通过对全网 url 进行精准调度、智能压力挖掘、自适应页面更新周期预测，可以实现自动路由、url 作弊识别、智能主题抓取等功能。扁平的架构设计使得系统可以进行任意的扩展，同时结合公司海量运营的经验，在系统监控、运营告警等方面都不断进行完善使得系统可以稳定高效运行。
抽取接口支持 Html 网页的标题、正文等十数种重要字段的抽取，以及特殊类型网页的定制化抽取服务。抽取后台完成网页内容的归一化、结构化处理工作，用户只需要调用抽取 API 即可高效完成从指定页面获得丰富的结构化信息。目前已经接入业务包括网页转码、广点通等。

2.输入参数

参数名称	必选	类型	描述
url	是	String	待抓取的 url

3. 输出参数

参数名称	必选	描述
code	Int32	错误码, 0: 成功, 其他值: 失败
message	String	错误信息
title	String	返回输入 uri 对应网页的标题
content	String	返回输入 uri 对应网页的正文

下载抽取API错误码详细信息如下：

错误码	含义说明
400	HTTP Method不正确
401	HTTP请求参数不符合要求

503	调用额度已超出限制
504	服务故障

4. 详细示例

示例业务详细信息如下表：

	参数名称	参数描述	必选	参数值示例
腾讯云公共参数	Action	方法名	是	ContentGrab
	SecretId	SecretId	是	AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
	Timestamp	当前时间戳	是	1408704141
	Nonce	随机正整数	是	345122
业务参数	url	待抓取的url	是	http://www.qq.com

下面以上述业务为例，详细说明"下载抽取 API"接口的使用方法。

4.1 接口鉴权

示例要调用服务的数据为：{"url":" http://www.qq.com "}
则上述业务的参数列表如下：

```
{
        'Action' : 'ContentGrab',
        'Nonce' : 345122,
        'Region' : 'sz',
        'SecretId' : 'AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA',
        'Timestamp' : 1408704141,
        'url': 'http://www.qq.com'
    }
```

根据上述参数列表进行签名，得出的数字签名为：HgIYOPcx5lN6gz8JsCFBNAWp2oQ（示例），详细的数字签名的生成方法请参照：《腾讯云接口鉴权》。

注意：

1）在生成签名的过程中，需要将加密字符串中包含的"_"改写成"."，从而加密产生签名；

2）鉴权时，需要将参数列表按key进行排序：字典序，同时大写在前。

4.2 API调用

根据上一步（8.4.1）中得到的数字签名，以POST请求为例构造请求URL，将数字签名加入到参数Signature中。

```
https://wenzhi.api.qcloud.com/v2/index.php?
  Action=ContentGrab
  &Nonce=345122
  &Region=sz
```

```
    &SecretId=AKIDz8krbsJ5yKBZQpn74WFkmLPx3gnPhESA
    &Timestamp=1408704141
    &Signature=HgIYOPcx5lN6gz8JsCFBNAWp2oQ
    &url=http://www.qq.com
```

执行上述操作之后，会将数据{"url":" http://www.qq.com "} 发送给API接口，进行相应分析。

注意：在发送请求过程中，不能将参数字符串中包含的“”改写成“.”。