

自然语言处理

产品介绍

文档版本 01
发布日期 2024-07-02



版权所有 © 华为技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编： 518129

网址： <https://www.huawei.com>

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 什么是自然语言处理	1
2 功能介绍	2
3 应用场景	3
4 基本概念	4
4.1 自然语言处理基础	4
4.2 语言生成	4
4.3 语言理解	5
4.4 机器翻译	5
5 计费说明	6
6 与其他服务之间的关系	10
7 约束与限制	12
8 如何访问自然语言处理服务	14
9 监控	15
9.1 自然语言处理的监控指标	15
9.2 查看监控指标	16

1 什么是自然语言处理

自然语言处理（Natural Language Processing，简称NLP）是一款基于人工智能技术，针对各类企业及开发者提供的用于文本分析及挖掘的云服务，旨在帮助用户高效地处理文本，常用于智能问答系统、文本分析、内容推荐、翻译等场景。

使用前必读

用户需要具备编程能力，熟悉Java、Python编程语言。

NLP服务需要用户通过调用API接口，将需要文件识别成可编辑的文本，然后返回JSON格式的识别结果，用户需要通过编码将识别结果对接到业务系统或保存为TXT、Excel等格式。

首次使用 NLP

如果您是首次使用NLP的用户，建议您学习并了解如下信息：

- 功能介绍
通过[功能介绍](#)章节的内容，了解NLP不同功能的具体介绍，主要包括自然语言处理基础（Natural Language Processing Fundamentals）、语言生成（Language Generation，简称LG）、语言理解（Language Understanding，简称LU）、机器翻译（Machine Translation，简称MT）功能。
- 入门使用
NLP以开放API的方式提供给用户，您可以参考[《快速入门》](#)学习并使用NLP服务。
- 使用方式
如果您是一个开发工程师，熟悉代码编写，想要直接调用NLP的API或SDK使用服务，您可以参考[《API参考》](#)或[《SDK参考》](#)获取详情。
- 由浅入深学习
您可以参考[成长地图](#)，由浅入深学习使用NLP。

2 功能介绍

自然语言处理包含如下子服务。

- 自然语言处理基础（Natural Language Processing Fundamentals），为用户提供包括分词、命名实体识别、关键词提取、短文本相似度等自然语言相关的API，可用于智能问答、对话机器人、内容推荐、电商评价分析等场景中。
- 语言生成（Language Generation，简称LG），为用户提供包括文本摘要等语言生成相关的API，可用于新闻摘要生成、文献摘要生成、搜索结果片段生成、商品评论摘要等场景中。
- 语言理解（Language Understanding，简称LU），为用户提供包括文本分类、情感分析等语言理解相关的API，可用于情感分析、内容检测、广告识别等场景中。
- 机器翻译（Machine Translation，简称MT），为用户提供快速准确的翻译服务，帮助用户跨语言沟通，可用于文档翻译等场景中。

优势

- 功能全面
提供多种常用自然语言类的算法模型及解决方案，可覆盖不同行业的各类需求。
- 高效准确
可快速分析大数据量的文本，深度理解文本语义，更加准确的挖掘出文本中的关键信息。
- 简单易用
简单易用的API接口。无需下载SDK、购买服务器，支持跨平台调用。丰富的产品种类，可一站式开通、部署。
- 稳定可靠
故障自动迁移，服务可用性达99.95%，保障业务连续。数据多副本，数据持久性99.99995%，保障数据不丢失。

3 应用场景

自然语言处理适用于智能问答系统、文本分析、内容推荐、翻译等场景。

- **智能问答系统**
通过中文分词、短文本相似度、命名实体识别等相关技术计算两个问题对的相似度，可解决问答、对话、语料挖掘、知识库构建等问题。
- **文本分析**
通过关键词提取、文本聚类、主题挖掘等算法模型，挖掘突发事件、公众话题导向，进行话题发现、趋势发现等。多维度分析公众情绪、热点、趋势、传播途径等，及时全面的掌握话题动态。
- **内容推荐**
通过关键词提取、短文本相似度等技术，提取关键语义信息，准确匹配出语义相似的内容，从而快速构建内容推荐场景。
- **翻译**
通过文本语言分析，准确翻译语句内容，从而帮助用户跨语言沟通。

4 基本概念

4.1 自然语言处理基础

分词

将文本切分成以独立的词为单位的序列，且在该过程中，对切分得到的词汇进行词性的标注。

依存句法分析

分析句子中词汇和词汇之间的相互依存关系，得到句子的句法结构。例如中文中使用依存句法分析，将一句话分析出主谓宾结构，将宾语定义为谓语动词的支配对象等。

文本相似度

对文本对进行相似度计算，详细内容请参见[文本相似度接口参考](#)中的示例。

句向量

将语句映射至固定维度的实数向量。详细内容请参见[句向量接口参考](#)中的示例。

实体链接

针对通用领域的文本进行实体链接分析，识别出其中的实体，并返回实体相关信息。

4.2 语言生成

文本摘要

在不改变文档语义的情况下总结出文档的主要内容。可应用于新闻摘要生成、文献摘要生成、搜索结果片段生成、商品评论摘要生成等场景。

诗歌生成

根据用户输入的诗歌主题，自动赋诗。生成的诗歌种类多样，言简意深，对仗工整，平仄协调。用户不仅可以指定诗歌类型（五言绝句、七言绝句，五言律诗，七言律诗），而且还可以指定是否生成藏头诗，能够满足用户在不同节日、场景的需求。

4.3 语言理解

文本分类

对文本按照一定的分类体系或标准进行自动分类标记。目前支持对给定文本自动判断是否是广告。

情感分析

情感分析是文本分类的一种，是指针对带有主观描述的文本，自动判断该文本的情感极性类别。情感极性分为正面和负面。

意图理解

针对天气类、报时、新闻类、笑话类、翻译类、提醒类、闹钟类、音乐类共8个领域进行意图理解，对用户的问题进行领域识别并提取领域内的参数。

4.4 机器翻译

语种识别

语种识别是为了识别文本所属的语种。对于用户输入的文本，返回识别出的所属语种。

5 计费说明

计费项

自然语言处理接口调用次数：调用自然语言处理的接口收费，提供按调用次数按需计费的方式。

计费模式

- 免费试用**

部分API支持首次免费试用，需要购买免费试用套餐包后使用。免费试用套餐包只允许每个用户免费购买一次，一经退订无法再购买，可以以原价续费。
- 按需计费**

按需计费是指按照API调用次数阶梯价格计费，计费价格请参见[自然语言处理价格详情](#)。
- 折扣套餐包**

折扣套餐包方式是用户可以购买套餐包，扣费时调用次数会先在套餐包内进行抵扣，抵扣完后的剩余调用量默认转回按需计费方式。

折扣套餐包方式相对按需付费方式提供了更大的折扣，对于长期使用者，推荐该购买方式。

📖 说明

- 调用自然语言处理API的时间需在套餐包生效期内，该次调用才可在套餐包内抵扣。
 - 折扣套餐包从购买之日起生效，有效期1年，到期自动结束。例如您1月1日购买了1个套餐包，则该套餐在次年1月1日会自动结束，即使您在该有效期内未调用自然语言处理服务，该套餐也不会延期，且无法退还费用。
- API支持计费模式

接口类型	接口	支持计费模式
自然语言处理基础接口	分词	<ul style="list-style-type: none">免费试用：免费试用套餐包包年包月：基础套餐包按需计费：自然语言处理基础版API
	多粒度分词	免费调用

接口类型	接口	支持计费模式
	依存句法分析	免费调用
	命名实体识别（基础版）	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	命名实体识别（领域版）	<ul style="list-style-type: none"> ● 包年包月：领域套餐包 ● 按需计费：自然语言处理领域版API
	文本相似度（基础版）	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	文本相似度（高级版）	免费调用
	句向量	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	实体链接	按需计费：自然语言处理定制版API
	关键词抽取	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	事件抽取	免费调用
语言生成接口	文本摘要（基础版）	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	文本摘要（领域版）	免费调用
	诗歌生成	按需计费：自然语言处理定制版API
语言理解接口	情感分析（基础版）	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API

接口类型	接口	支持计费模式
	情感分析（领域版）	按需计费：自然语言处理领域版API
	文本分类	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	属性级情感分析	<ul style="list-style-type: none"> ● 包年包月：领域套餐包 ● 按需计费：自然语言处理领域版API
	属性级情感分析（高级版）	按需计费：自然语言处理定制版API
	实体级情感分析	免费调用
	意图理解	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	文档分类	<ul style="list-style-type: none"> ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
机器翻译接口	文本翻译	按需计费，详情请见 价格详情页 > 文本翻译按需付费价格 。
	语种识别	<ul style="list-style-type: none"> ● 免费试用：免费试用套餐包 ● 包年包月：基础套餐包 ● 按需计费：自然语言处理基础版API
	文档翻译任务创建	按需计费：文档翻译
	文档翻译状态查询	免费试用：免费试用套餐包

变更配置

NLP默认使用按需计费。您也可以根据您的实际情况购买资源包，购买后对应区域的资源将自动转为包年包月计费方式。如果超出当前已购买资源包的额度，超出部分将自动转为按需计费。

欠费

按需购买的接口是按照API调用次数阶梯价格计费，当余额不足，无法对当前产生费用进行扣费，就会导致欠费，欠费后有**保留期**。您续费后可继续正常使用，请注意在保留期进行的续费，是以原到期时间作为生效时间，您应当支付从进入保留期开始到续费时的服务费用。

您账号欠费后，会导致部分操作受限，建议您尽快续费。具体受限操作如下所示：

- 按需方式的API接口不可调用
- 不可重置套餐包方式的API接口，例如“10万次基础套餐包”，在欠费后如果套餐包有剩余，可继续使用，但不可购买，也不可续期
- 每日或每月可重置套餐包方式的API接口，例如“每月20000次基础套餐包”，在调用超过每日或每月限额导致欠费，会进入保留期，保留期后会被冻结，无法使用该套餐包
- 无法开通服务

续费

资源包到期后，您可以进行续费以延长资源包的有效期，也可以设置到期自动续费。续费相关操作，请参见[续费管理](#)。

服务到期

包年包月资源包到期后，自动转为按需计费。转按需后如果账号欠费，会根据“客户等级”定义不同的保留期时长，保留期内资源处理和费用详见“[保留期](#)”。保留期满仍未续订或充值，数据将被删除且无法恢复。

6 与其他服务之间的关系

介绍了自然语言处理与其他服务的关系。

图 6-1 自然语言处理服务与其他服务关系示意图



统一身份认证服务

统一身份认证服务（Identity and Access Management，简称IAM），自然语言处理使用统一身份认证服务（IAM）进行鉴权。IAM更多信息请参见[统一身份认证服务IAM](#)。

云监控

云监控（Cloud Eye）可以监控自然语言处理的相关指标，具体如[表6-1](#)所示。用户可以通过指标及时了解自然语言处理的使用情况。Cloud Eye的更多信息请参见[云监控CES](#)。

表 6-1 NLP 的监控指标

指标名称	指标含义	取值范围	测量对象
调用服务成功次数	该指标用于统计用户请求服务成功次数，以“次每分钟”为单位。	≥0次每分钟	自然语言处理
调用服务失败次数	该指标用于统计用户请求服务失败次数，以“次每分钟”为单位。	≥0次每分钟	自然语言处理

说明

每个子服务都有调用成功次数和失败次数两个指标。

对象存储服务

对象存储服务（Object Storage Service，简称OBS）是稳定、安全、高效、易用的云存储服务。自然语言处理接口都涉及到对用户的数据处理，用户的大量数据采用OBS批量方式处理，可以提升云上处理的总体效率。

自然语言处理接口支持从OBS上采用临时授权或者匿名公开授权的方式获取数据并进行处理。OBS的更多信息请参见[对象存储服务OBS](#)。

7 约束与限制

使用区域限制

自然语言处理目前支持的区域为“华北-北京一”和“华北-北京四”，其中机器翻译不支持“华北-北京一”区域。

自然语言处理基础

表 7-1 自然语言处理基础的使用限制

规格项	规格限制
分词	文本长度，中文长度1~512字符，英文长度1~2000字符，文本编码为UTF-8。
命名实体识别（基础版）	文本长度，中文长度1~512字符，英文长度1~2000字符，文本编码为UTF-8。
命名实体识别（领域版）	文本长度，中文长度1~512字符，英文长度1~2000字符，文本编码为UTF-8。
句向量文本列表	列表长度为1~1000。
句向量文本	文本分词后数组，长度为1~1000，文本编码为UTF-8。
实体链接	长度为1~50字符，文本编码为UTF-8。
文本相似度	文本长度，中文长度1~512字符，英文长度1~2000字符，文本编码为UTF-8。

语言生成

表 7-2 语言生成的使用限制

规格项	规格限制
文本标题	文本长度为1~1000字符，文本编码为UTF-8。

规格项	规格限制
文本正文	文本长度为1~10000字符，文本编码为UTF-8。

语言理解

表 7-3 语言理解的使用限制

规格项	规格限制
情感分析（基础版）	文本编码为UTF-8。限定400个字符以内，文本长度超过400个字符时，只检测前400个字符。
情感分析（领域版）	文本编码为UTF-8。type字段支持0, 1和2。 <ul style="list-style-type: none"> • type为0, 1时，限定200个字符以内，文本长度超过200个字符时，只检测前200个字符。 • type为2时，限定400个字符以内，文本长度超过400个字符时，只检测前400个字符。
文本分类	文本编码为UTF-8。限定400个字符以内，文本长度超过400个字符时，只检测前400个字符。
意图理解	文本长度为1~32字符，文本编码为UTF-8。

机器翻译

表 7-4 机器翻译的使用限制

规格项	规格限制
文本翻译	文本长度为1~1000字符，文本编码为UTF-8。
语种识别	文本长度为1~1000字符，文本编码为UTF-8。

8 如何访问自然语言处理服务

公有云提供了Web化的服务管理平台，即管理控制台和基于HTTPS请求的API（Application programming interface）管理方式。

- API方式

如果用户需要将公有云平台上的自然语言处理服务集成到第三方系统，用于二次开发，请使用API方式访问自然语言处理服务，具体操作请参见《自然语言处理服务接口参考》。

- 管理控制台方式

其他相关操作，请使用管理控制台方式访问自然语言处理服务。如果用户已注册

公有云，可直接登录[管理控制台](#)，在页面左上角单击，选择“人工智能 > 自然语言处理 NLP”服务。

9 监控

9.1 自然语言处理的监控指标

功能说明

本节定义了自然语言处理服务上报云监控服务的监控指标的命名空间，监控指标列表和维度定义，用户可以通过云监控服务提供管理控制台或API接口来检索自然语言处理服务产生的监控指标和告警信息。

命名空间

SYS.NLP

监控指标

表 9-1 NLP 支持的监控指标

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
successful_calls_of_service	调用服务成功次数	统计周期内调用服务成功次数。 单位：次/分钟	≥ 0 times/min	接口	1分钟
failed_calls_of_service	调用服务失败次数	统计周期内调用服务失败次数。 单位：次/分钟	≥ 0 times/min	接口	1分钟

维度

表 9-2 NLP 监控维度

Key	Value
call_of_interface	接口

9.2 查看监控指标

1. 登录管理控制台。
2. 选择“管理与部署 > 云监控服务”。
3. 展开左侧导航树的“云服务监控”，单击待查看的云服务。
4. 单击操作列“查看监控指标”，进入指标监控页面。
5. 在监控区域，您可以通过选择时长，查看对应时间的监控数据。
当前支持查看云服务“近1小时”、“近3小时”、“近12小时”的监控数据。