

文智自然语言处理

产品简介

产品文档



腾讯云

【版权声明】

©2015-2016 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】



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

文档声明.....	2
产品概述	4
产品功能	5
产品优势	7
应用场景	8

产品概述

文智中文语义开放平台是基于并行计算系统和分布式爬虫平台，结合独特的语义分析技术，一站式满足用户NLP、转码、抽取、全网数据抓取等中文语义分析需求的开放平台，用户能够基于平台对外提供的OpenAPI实现搜索、推荐、舆情、挖掘等语义分析应用，也能够通过与我们合作定制产品特色的语义分析解决方案。

产品功能

文智提供的API功能包括词法分析、句法分析、篇章分析、下载四大类别。具体API功能如下：

1.分词/命名实体识别API

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接口即可获得到所需结果，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.情感分析API

为有情感分析需求的产品提供的服务。该服务能够对信息进行情感上的正向、负向及中性进行评价。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。

3.主题分类API

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。

4.关键词提取API

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。

5.敏感信息识别API

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。

6.句法分析API

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。

7.同义词API

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。

8.纠错API

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。

9.转码API

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。

10.下载抽取API

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。水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。

产品优势

1.分析精准

十年专注语义分析，千亿级互联网语料，众多腾讯产品应用经验，为文智API的精准效果打下了基础。

2.服务高效

服务支撑高效稳定，同时配备完善及时的开发者支持

3.接口全面

一站解决中文语义分析需求，集合词法、句法、篇章、下载模块，提供分词\命名实体识别、情感分析、主题分类、关键词提取、同义词、纠错、下载等多种API，且更多API也将不断推出。

应用场景

文智自然语言处理的应用场景颇为广泛。只要有大量文本数据的地方，都可使用文智API做分析。以下列举几个经典使用场景。

1.品牌口碑监控

使用情感分析API分析针对相关品牌的言论中带有的情感倾向，点评言论可来自下载抽取API抓取的社交网络或点评网站言论，也可来自市场调查等。

2.个性化推荐

以读物推荐为例（可为小说、新闻、图书推荐等等），使用文本分类、关键词提取、情感分析API分析推荐库内的读物，再基于文本类型，关键词，情感倾向，结合用户类型/兴趣将读物与用户匹配。

3.垃圾/敏感信息识别

使用文本分类、敏感信息识别API可以识别广告信息、色情信息、政治敏感信息。

4.下载抽取

需要爬取网站内容、结构化文本信息时，可使用下载抽取API。