

云搜

操作指南

产品文档



腾讯云

【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】

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

操作指南

应用管理

基本信息

字段配置

数据处理

选购方案

体验测试

高级组件

分词方式

智能联想

按域检索

LBS

排序定制

运营管理

模版管理

操作指南

应用管理

基本信息

最近更新时间：2017-03-08 22:30:51

此页面包含应用的基本信息如：应用名称、应用ID、状态、文档量及检索量等数据。也有应用操作的三个快捷入口。

应用操作快捷入口：

- 1) 数据上传：页面提供了文件上传数据的接口，可手工上传文件；
- 2) 体验测试：用户上传数据后，即可进行检索效果测试，测试检索效果；
- 3) 应用删除：删除对应应用。

其他：

4) 应用ID：系统为每个应用分配一个ID，称为APP_ID；应用ID是全局唯一的；

5) 应用状态：

状态1：尚未开通。表示您还没有激活搜索服务，您可以通过试用或者购买激活搜索服务。

状态2：免费试用中。您正在试用搜索服务，期间您可以购买搜索服务或者暂停搜索服务。暂停后，云搜将停止检索服务。

状态3：试用已结束。试用已到期，您可以购买搜索服务。

状态4：待支付。该状态表明您的账单还未支付&支付失败，或者支付后，支付状态还未同步。您可以查看账单，看是否已支付成功，如果支付成功2小时后，还未开通，请联系客服。

状态5：运行中。表示云搜正在提供的搜索服务，请注意文档量和搜索峰值是都超标。期间您可以暂停或者终止服务。终止服务后，服务立即停止，数据清除。

状态6：已过期-运行中。表示您的云搜服务已到期，为了保障您的服务，系统给予了1天的续费缓冲期。24小时后，搜索服务将被冻结。请您注意续费。

状态7：服务已停用-冻结中。表示您的云搜服务已到期，搜索服务被冻结，且将在到期7天后清除数据，下架服务。请及时续费。

状态8：服务已下线。表示您的云搜服务已终止服务，请数据被清除。如需使用，可重新上传数据并购买服务。

及其他状态，如有疑问，可加入云搜讨论群433367976。

字段配置

最近更新时间：2017-03-09 10:54:34

开发者可以根据业务情况对应用结构进行修改，修改方式同新建应用时的定义应用结构。修改后请注意以下情况

1. 删除字段：删除的字段如果在智能联想、按域检索和高级排序中使用，会影响到检索服务，请慎重操作并立即调整对应服务；
2. 以下情况需要重建索引才能生效：
 - 增加、删除、修改应用字段
 - 修改文本字段的数据长度；
 - 修改数值字段的数据类型

数据处理

最近更新时间：2017-03-09 04:14:38

数据处理操作主要包括：数据上传、数据清空以及重建索引三块。包括在其他页面修改配置引发的重建索引等操作，也可以在此处查询进度。

1. 数据上传：用户通过页面提供的接口上传文件格式的数据，数据将用于检索；

2) 数据清空：用户将当前上传的数据彻底删除，数据清空操作不可逆，由于数据清空涉及到较多的后台流程，约需要5分钟左右（警告：清空数据后，未上传新数据的情况下如果重建索引，您的搜索服务将无法使用）；

3) 重建索引：当应用结构发生变化时，需要重建索引才能生效，重建根据数据量一般需要十几分钟到几个小时不等。仅仅上传数据，无需重建索引。

主要有以下场景，具体会在对应场景中解释，这里可做了解：

修改应用结构：包括添加、删除、修改应用结构字段，数据长度、数据类型；

智能联想：设置或修改了智能联想的字段；

按域检索：添加或修改了检索用的域；

分词方式：修改了文本域字段的分词方式；

高级排序：修改了初级排序或者自定义了高级排序；

选购方案

最近更新时间：2018-05-25 16:41:35

购买流程

未购买时您可选择购买体验版、专业版或极速版。

体验版购买：体验版实行一价制，30元每月，首月按照剩余天数进行收费。

专业版&极速版：专业版和极速版，您可以选择合适的文档和搜索请求峰值进行设置。设置完成后，将为您计算所需费用与价格，单击即可进行购买。

购买成功后，页面展示如下：

建议：建议您首先通过体验版体验云搜，并开发完成SDK后，再进行购买，这样可以节省费用。

更改配置

购买完成后，您可以选择升级或者降级现有配置。单击更改配置后，根据您的需要，调整版本及对应的文档量与QPS即可进行更改。

注意：

1、您更改后的文档量不能小于您当前已上传的文档量大小，如果需要调整，需要清除掉多余的数据后再进行调整

其他

1、当您申请的文档量和QPS超过一定数额时，需要提交需求，待我方工作人员与您联系后，才可以进行购买，开通服务。

2、特殊情况下，您购买的服务可能需要两个工作日后，才能开通，请耐心等待。

3、购买和更改配置产生的费用详见购买指导部分。

体验测试

最近更新时间：2017-03-09 03:36:51

当您上传文档后，即可进入此页面，进行搜索体验了。同时，当修改了网站的配置时，也可以立即前来体验。

备注：测试操作详见API使用说明，输入框中填写内容与您编码时所传参数相同。

高级组件

分词方式

最近更新时间：2018-05-25 16:43:14

包括：正常分词、单字分词、混合分词。

1. 正常分词是最常规的分词方式，按照正常的词组对用户的数据和检索词进行切分。以"腾讯云搜"为例，分词结果为"腾讯"、"云搜"。这样搜到的结果与检索词强相关，效果好；
2. 字分词是将用户输入的检索串切分成一个一个的单字，仍以"腾讯云搜"为例，分词结果为"腾"、"讯"、"云"、"搜"。适用对象为内容较少，对搜索结果要求不高的用户。这样可以使搜索结果增多，但某条数据包含任一单字，就会被展示，搜索结果会较差；
3. 混合分词则是上面两种分词的组合，将正常分词和单字分词结合在一起使用，得到的结果是两者的和。这种搜索可以在保证一定质量的基础上，减少无结果和少结果的搜索。

如何选择分词方式：

用户可以直接在界面中通过拖动的方式为各个字段分配分词方式。

在分配完成后单击"保存"按钮，重建索引（详见初级排序介绍）后新分配的分词方式生效。

智能联想

最近更新时间：2018-05-25 16:44:58

智能联想：当用户在输入检索串的过程中，可以提示用户可能的检索串，为用户补全检索串。如下图：

用户可以开启或者关闭智能联想。开启方式为，用户在下拉菜单中选择用户最可能搜索的文本字段，单击提交即可。提交后同其他高级组件，需要建索引后才能生效。

作用：假设开发者选择了字段“应用名称”为智能联想字段，当用户输入“zhiwu”时，如果应用名称中包含“植物大战僵尸”、“植物XXX”等应用名时，系统可以自动补全相关的应用名称，减少用户的输入。

使用：开启功能后，开发者可以配置当检索时是否开启智能联想，智能联想的字段为：`is_smartbox`；设置为1则代表开启智能联想检索。 详见API/SDK使用。

按域检索

最近更新时间：2017-03-06 15:31:12

通用检索在用户输入检索词搜索时，会默认在所有文本字段中查找与检索词相关的内容，这样不仅降低了检索效率，同时也会有一定几率将无关内容呈现在检索结果中。

按域检索：一种将检索范围限制在部分字段中的搜索方式。云搜的数据是由文档组成，而文档又是由字段组成，一个或者多个字段可以组成一个域。

以小说为例：用户在搜小说时，如果网站认为用户都是搜标题或者作者来搜书。那么可以设置域A（IA）包含字段title、域B（IB）包含字段author。然后可限制搜索仅在IA和IB两个域内检索。或者将title和author包含在一个域中。

保存后，重建索引完成，则新建或者修改的域生效。

按域检索的使用方法如下：

检索时检索串中带上相应的域名即可在指定域内检索，如"index:IA:凡人修仙传"（其中IA是域名，凡人修仙传是用户输入的检索串），则检索结果中只出现书名与"凡人修仙传"相关的结果。具体见API和SDK使用说明。

LBS

最近更新时间：2017-03-06 22:45:54

1 简介

LBS定制可以提供给基于地理位置距离进行优先的搜索业务，例如查找距离用户（经度：113.959633，纬度：22.54138）最近的川菜馆，这样即要考虑商户的经纬度信息和用户自身经纬度的球面距离（单位为米）作为返回结果的一个重要权重信息，同时，云搜提供distance字段作为LBS值的排序字段（精细排序）。

2 配置

第一步：在“应用结构修改”数值域字段添加“经纬度”两个字段（字段类型为浮点型），用于存储商户的经纬度信息。如下图

第二步：在高级组件定制“LBS”里设置经纬度字段，并保存开启LBS距离优先策略。

3 使用

在检索过程中，需要传入用户所在的经（参数名为longitude）纬（参数名为latitude）度值，用来计算两点的距离，如下图所示：

排序定制

最近更新时间：2018-05-25 16:49:20

排序定制允许用户自定义搜索结果排序方式，通过设置初排字段和精细排序表达式，来对最终展示结果的排序进行干预。

1 初级排序定制

效果：配置该字段，即可在用户检索前，预先把数据按照该字段值的优劣进行排序。因此请选择最能代表一条数据优劣的数值字段。

作用：因为用户搜索时，对搜索结果进行倒排索引和求交以及相关性计算是一个比较耗时的过程，特别是遇到数据规模很巨大的业务。配置该字段后，可以保证在一定的时间内，尽量召回优质的文档展现给用户。

限制：初排字段只能设置为数值类型的字段，不支持其他类型的字段。

配置过程：请按照如下图进行配置

单击“提交”之后，系统会提示重建索引，如图示：

待索引重建完成，初级排序立即生效。

提示：如果您有多个配置要修改，建议选择稍后重建，等全部修改完成后，再手工在“数据处理”处单击重建索引，节约费用。

2 精细排序定制

精细排序是直接影响文档展示顺序的最后一级排序，允许开发者自主对搜索结果进行更个性化整合和混合排序。当前主要考虑了几个维度的混排因子，比如距离因子，精确匹配因子，城市因子，相关性因子，权威度因子，以及各数值域字段因子。

精细排序是二维排序方式，首先根据业务情况把文档界定出几个档位（即优先级，如下图优先级1、2、3），优先级高的数据整体排在优先级低的数据前面，以此类推。

每个优先级都对应一个数据范围表达式，该表达式最终结果为一个bool类型的值，计算结果为true或者>0的结果的所有文档都属于该优先级。

在同一个优先级范围内，通过指定排序因子（可以有多个因子，用下划线_隔开）的方式，决定内部的排序。排序方式为先按照第一个因子大小排序，前一个因子相等的情况下，再通过下一个因子进行排序，依次类推，默认是按照降序排序，如果要按照升序排序，则直接在因子前加一个减号“-”。

例如：

1.最优文档为优先级1的文档，我认为符合精确匹配结果的文档为最优质的文档，则该优先级对应的数据范围表达式为：`exactmatch >0`，则所有符合该条件的文档都一定会优先于不符合该条件的其他文

档，这是一个维度的排序。

2.对于以上符合精确匹配结果的文档，指定按照权威度优先，再按照距离优先进行排序，则排序因子为：

3 排序运算表

数据范围和排序因子支持以下四则运算规则：

分类	字段名	描述
内置特征字段	docWeight	文本相关性分数，用于衡量query与文档的匹配度和质量
	exactmatch	是否精确匹配，query和文档是否完全匹配，0或者1
	authority	权威度，业务制定的权威度值
	textRelLevel	文本相关性分档：1:bad 2:normal 3:good 4:perfect
业务特征字段	所有数值类型字段名	所有数值类型字段名
内置函数	distance(xxx,yyy,longitude,latitude)	计算两点之间的位置距离，xxx, yyy为文档经纬度值，longitude和latitude为检索query的经纬度值
	max(NA,NB,query_a,...)	计算最大值
	min(NA,NB,query_a,...)	计算最小值
	avg(NA,NB,query_a,...)	计算平均值
	sum(NA,NB,query_a,...)	计算平均值

	abs(NA):	计算绝对值
内置 函数	+ - * /	四则运算
	==(等于) !=(不等于) >(大于) <(小于) >=(大于等于) <=(小于等于):	关系运算符

运营管理

最近更新时间：2017-03-07 10:18:03

为了便于应用的日常运营工作，为用户提供了操作日志统计和检索日报。

1. 操作日志查询

用户对应用的所有重要操作会被记录到日志中，以便查看。记录日志的种类包括：数据上传、数据清除、索引重建、配置修改四类，具体提供操作的时间、结果和详情。

1. 检索日报

日报信息包括：检索量统计、检索耗时统计、检索热词统计以及零结果检索词统计。他们的用途分别为：

- 检索量统计：可以查看每日网站的检索量及分布情况
- 检索耗时统计：可以查看每次检索请求消耗的时间
- 检索热词：可以查看当前最热的检索词及被检索次数，方便运营
- 零结果检索词：可以查看用户检索较多，但应用中没有相应内容的检索词，方便内容增加或者引导

注意事项：统计具有一定的延时性，报表上不能立即展示数据变化。请稍后再做统计查询。

模版管理

最近更新时间：2018-05-25 16:40:07

1. 模板新建

单击“+”号，开始新建模板。

第一步：填写模板信息

第二步：填写模板信息

支持自主创建，或者选择已有模板进行修改，具体操作同新建应用。

第三步：模板信息确认

可确认模板相关信息是否有误

第四步：创建成功

2. 编辑模板

允许用户对模板进行修改。

3. 删除模板

将已新建完成的模板完全清除。

4. 导出模板

将模板导出，存为本地文件。