

ALIBABA CLOUD

# 阿里云

开放搜索  
搜索引导

文档版本：20220527

 阿里云

## 法律声明

阿里云提醒您阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

# 通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
<b>粗体</b>	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[ ] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

# 目录

1. 下拉提示	05
2. 热搜和底纹	15
3. 黑白名单	25

# 1. 下拉提示

## 功能介绍

下拉提示是搜索服务的基础功能，在用户输入查询词的过程中，智能推荐候选query，提高用户输入效率，帮助用户尽快找到想要的内容。

下拉提示实现了基于用户文档内容的query智能抽取，可以通过中文前缀，拼音全拼，拼音首字母简拼查询以及汉字加拼音，分词后前缀，中文同音别字等查询下拉提示的候选query。

例如：连衣裙长款 这个query，可以通过如下方式查询得到：

- 中文前缀： 连, 连衣, ...
- 全拼前缀： l, li, lian, lianyi, lianyiqun, ...
- 简拼前缀： l, ly, lyq, ...
- 汉字加拼音： 连yi, 连衣qun, ...
- 分词后前缀： 长款, 长款连衣, 连衣长, ...
- 中文同音别字： 连衣群, 联谊群, ...

此外，用户可以通过干预词条对下拉提示的数据进行干预。可以看到下拉提示的多项业务指标，详见[下拉提示报表](#)。

## 使用须知

### 数据来源

下拉提示数据主要来源于应用文档。每个下拉提示最多可以从一个应用中选择3个字段作为数据来源。处理时，系统会选取应用中的部分文档（百万级别），对这些文档中被选中的字段按照指定的规则进行处理，生成下拉提示的候选数据。之后再按一定规则保留一定量的数据作为下拉提示候选query。

### query生成规则

对于候选query的生成规则，我们针对过去N(默认7)天的历史query，结合该query的词权重，召回结果数，历史搜索次数，近一天查询是否有结果等条件，选出一些热门历史查询词，作为下拉提示的候选词。系统支持两种规则生成候选query：**抽取生成**和**原值保留**。

- **抽取生成**：使用阿里nlp团队基于海量自然语言训练的分析器，对字段内容进行分词，抽取有意义的term进行组合，得到候选query，这种方式尽量保证生成的候选query能召回对应的文档。
- **原值保留**：该规则对字段内容不做分词处理，直接将其作为下拉提示的候选query。但当字段内容超过一定限制（30个字符）时，将截断字段，保留前30个字符的内容作为候选query。这种方式适用于不需要分词的应用字段（店铺名，用户名，歌曲名等）或者用户自己生成候选query，直接原样展示。建议使用内容较短且含义明确的字段。

## 行业模板

根据不同行业数据特点，下拉提示功能提供了相应优化模板。目前已支持通用、电商和内容行业模板。[行业模板介绍](#)：

- **通用行业模板**：使用中文通用分析器进行分词，通过通用分析器模板生成候选query。

- **电商行业模板**：使用阿里在电商领域多年积累的电商分析器进行分词，该分词模型用于识别query、title等电商文本中的产品词、品牌词、产品修饰词、风格、材质等标签以及每个词的词权重。然后根据电商行业常见标签组合得到一些候选query，并根据每个query的词权重，召回文档数，历史搜索次数等指标，综合计算query分数作为在线查询时的排序依据。
- **内容行业模板**：使用IT-内容分析器进行分词，根据内容行业常见标签组合得到候选query。

注意：

- 建议在搜索请求中设置raw\_query参数，便于数据处理时识别终端用户输入的原始查询词，[详见搜索处理文档](#)。
- 内容行业模板仍在优化中，如使用后存在疑问，请及时联系我们。

## 人工干预

下拉提示支持的干预项包括：

- 对候选query结果进行干预，包括黑名单和白名单。
- 对数据源应用文档设置过滤条件，配置过滤条件后，只会使用符合该过滤条件的应用文档生成下拉提示候选词。

参数名称	说明
过滤条件	<p>需填写OpenSearch应用表结构字段，该过滤条件作用于当前服务中的应用全量文档数据。<b>注意：</b></p> <ul style="list-style-type: none"> <li>● <b>比较符支持</b>：&lt;、&gt;、&lt;=、&gt;=、=、!=</li> <li>● <b>支持的字段类型</b>：数值、字符型，不支持array类型</li> <li>● <b>连接符</b>：多个过滤条件之间为AND关系，必须要使用英文逗号(,)分隔，表示且关系（暂不支持或关系）。</li> </ul> <p><b>举例</b>：当过滤条件为 status=1,level=1 时，则只能选取符合该条件的记录。</p>

控制台配置如下图所示：



## 候选query结果干预

**黑名单**：黑名单中的关键词支持包含匹配，包含黑名单关键词的query将不会出现在下拉提示结果中。因此，当下拉提示结果中出现一些业务上不想要的结果时，可将关键词添加到下拉提示黑名单中，实现对这些结果的屏蔽。

**白名单**：当白名单中的query满足推荐条件时，将优先在下拉提示中显示。因此，当某些优质query没有被识别出来，或者排序靠后时，可将其加到白名单中，实现更好的下拉提示效果。**\*\*黑名单和白名单的具体配置，可[点击此处](#)进行查看。**

## 配额说明

目前下拉提示为免费功能，因此计算资源和存储资源为系统统一设置。每个下拉提示的计算资源约为100QPS，存储资源约200W个候选query；

## 注意事项

- 标准版应用类型，不支持下拉提示。仅高级版应用支持；
- 单个应用最多创建10个下拉提示模型；
- 下拉提示的名称限制：同一用户不可设置名称相同的模型（包括下拉提示、人气模型、类目预测、热搜、底纹）；
- 下拉提示仅支持从一个应用中选择创建了索引的TEXT、SHORT\_TEXT、LITERAL、INT类型的字段作为数据来源；
- 同一个模型最多可选择3个训练字段；
- 修改应用结构时，不能修改下拉提示数据来源的字段；
- 删除应用时，对应的下拉提示也会被删掉；
- 下拉提示搜索时，query参数长度最大支持18字节，否则会报错，返回无结果；
- 下拉提示搜索时，hit参数的取值范围为(0, 10]的整数，否则按默认10处理，且会返回一个错误（例：当hit设置了0或-1或11时，默认按10处理并返回一个错误。）；
- 黑名单中的关键词条数不能超过500；
- 白名单中的query条数不能超过500；
- 黑名单和白名单数据有冲突时，黑名单的屏蔽优先级更高；
- 黑名单和白名单的修改均可实时生效；
- 下拉提示模型创建之后系统默认开启定时模型训练，下拉提示的数据随模型的定时训练周期性更新；
- 下拉提示的训练的时间和应用的数据量、系统负载有关；如果长时间（大于半个小时）没有训练结束，请联系我们；
- “中文同音别字”补全召回功能已默认生效，可以在请求中添加参数 `re_search="disable"` 关闭该功能。

## 使用建议

- 为进一步提升下拉提示效果（提升下拉提示引导搜索量、搜索结果点击率等），强烈建议将下拉提示请求与搜索请求关联。关联方法见文尾：下拉提示请求与搜索请求关联部分；
- 选择内容简洁，和文档主题相关的字段；
- 合理使用抽取生成和原值保留规则；

- 查询结果中，suggestions是本次查询的结果，errors表示本次查询是否有错误发生。errors不为空，不代表suggestions为空。因此，解析结果时，需要通过suggestions是否为空来判断是否无数据展示。

### 流程演示

1.搜索算法中心-->搜索引导-->下拉提示，点击“创建”：



2.输入“模型名称”、选择“行业类型”、选择“训练字段”及抽取方式、输入“过滤条件”【可省略】，点击“完成”：



3.在下拉提示列表页，点击“开始训练”，训练已创建的模型：

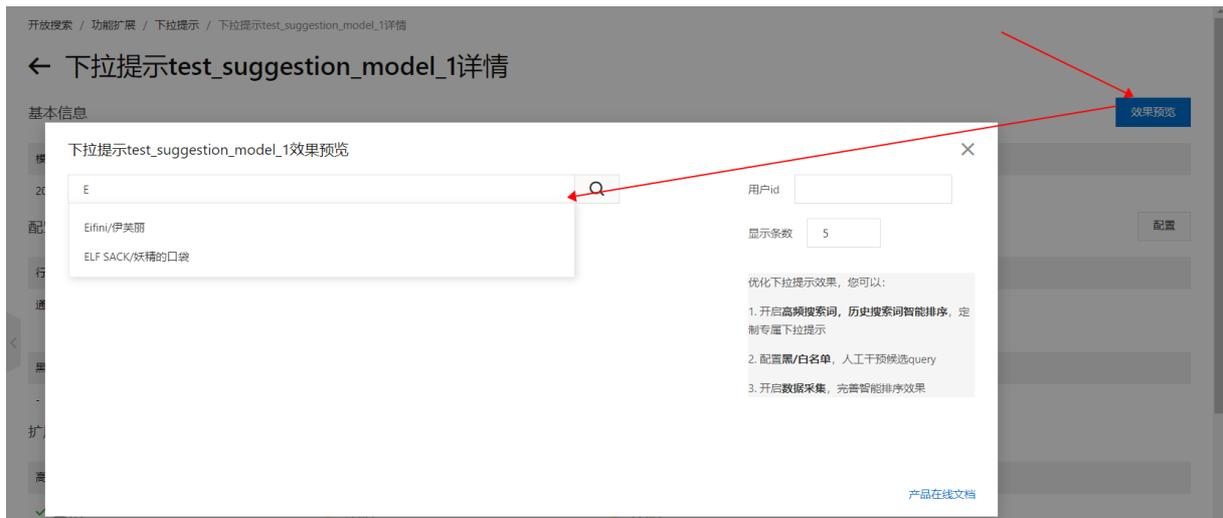


4.下拉提示模型开启训练后，需要20-30分钟左右的训练时间：



5.下拉提示效果测试，图中分别为“抽取”以及“原值保留”的两种抽取方式的效果测试：

原值保留：



抽取：



6.在线查询候选词，见下面Demo。接口详细说明见[下拉提示开发指南](#)。

## 下拉提示页面介绍

## 下拉提示列表页介绍

当用户通过开放搜索控制台：搜索算法中心—>搜索引导—>下拉提示，进入下拉提示列表页（如下图）：



列表页中显示下拉提示模型的“模型名称”、“是否已有可用模型”、“最后训练时间”、“近一次训练状态”（包括待训练、训练中、已训练并通过、数据异常四种状态）、“昨日PV”等信息。操作栏中可对模型进行模型详细查看、开始训练/重新训练模型、查看模型数据指标、删除模型等操作。

### 下拉提示模型详情页介绍

#### 基本信息和配置信息：



基本信息中主要展示了模型创建时间、是否已有可用模型、最后训练时间、近一次训练状态（如果为数据异常状态，则会在表格后展示“异常报告”悬浮窗）。配置信息中主要展示了模型配置的行业类型、定时训练的周期、训练字段、过滤条件以及黑/白名单。

### 扩展功能

#### 功能介绍：

- **高频搜索词**：将用户搜索频率作为候选词排序依据，优先将满足推荐条件的用户搜索频率较高的词作为下拉提示候选query，包含“生效中”和“未开启”两种状态；
- **历史搜索词**：优先展示用户之前搜索过的query，包含“生效中”和“未开启”两种状态(暂不支持)；
- **智能排序**：根据点击、购买等用户行为信息智能排序候选query，包含“生效中”和“未开启”两种状态(暂不支持)；

#### 开启方式：

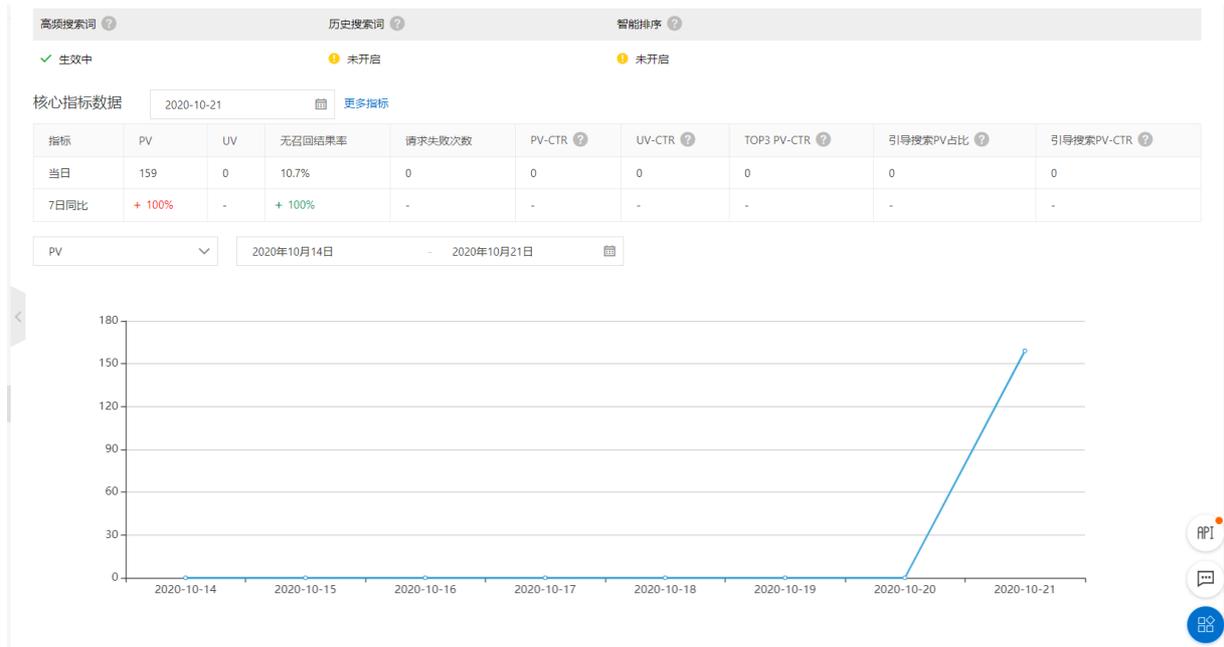
- **高频搜索词**：需要在搜索请求中上传raw\_query参数。(default索引也会开启，如需关闭，请提售后工单转人工处理)
- **历史搜索词**：需要在搜索请求中上传raw\_query、在下拉提示请求中上传参数user\_id参数(暂不支持)。
- **智能排序**：需要在搜索请求中上传raw\_query、user\_id、from\_request\_id参数，开启**数据采集**后效果更佳(暂不支持)。

#### 注意：

- raw\_query、user\_id、from\_request\_id参数介绍可[点击此处](#)进行查看。
- 开启智能排序需要一定点击量，建议每天有超过1w次的下拉提示点击。
- 高频搜索词和历史搜索词在上传的参数满足条件后会自动开启，智能排序功能在上传的参数满足条件后需提工单后台开启(暂不支持)。

### 核心指标数据

可选择不同的时间区间，以表格和折线图的形式展示下拉提示模型的核心指标数据：



注意：具体指标含义，可参考[下拉提示报表](#)。

### 下拉提示 SDK 演示

API:

```
GET v3/openapi/suggestions/{suggestion_name}/actions/search?hit=10&query={your_query}&re_search=homonym&user_id=xxx
```

Java SDK的maven依赖：

```
<dependency>
  <groupId>com.aliyun.opensearch</groupId>
  <artifactId>aliyun-sdk-opensearch</artifactId>
  <version>4.0.0</version>
</dependency>
```

相关链接：[版本说明](#)

代码演示：

```
package com.example.opensearch;

import com.aliyun.opensearch.OpenSearchClient;
import com.aliyun.opensearch.SuggestionClient;
```

```

import com.aliyun.opensearch.sdk.generated.OpenSearch;
import com.aliyun.opensearch.sdk.generated.commons.OpenSearchClientException;
import com.aliyun.opensearch.sdk.generated.commons.OpenSearchException;
import org.junit.After;
import org.junit.Before;
import org.junit.Test;

import java.nio.charset.Charset;

public class SuggestDemo {
    static private final String accesskey = "填入accesskey信息";
    static private final String secret = "填入secret信息";
    static private final String host = "填入下拉提示关联应用所在区域的host";
    OpenSearch openSearch;
    OpenSearchClient openSearchClient;
    static private final byte hits = 8; //最大返回下拉提示的条数
    static private final String suggestionName = "填入下拉提示名称"; //填入下拉提示名称

    @Before
    public void setUp() {
        //初始化openSearch对象
        openSearch = new OpenSearch(accesskey, secret, host);
        openSearchClient = new OpenSearchClient(openSearch);
    }

    @Test
    public void TestEnv() {
        //查看文件和默认编码格式
        System.out.println(String.format("file.encoding: %s", System.getProperty("file.encoding")));
        System.out.println(String.format("defaultCharset: %s", Charset.defaultCharset().name()));

        //创建下拉提示对象
        SuggestionClient suggestionClient = new SuggestionClient(suggestionName, openSearchClient);
        String query = "填入查询词";
        try {
            SuggestParams suggestParams = new SuggestParams();
            suggestParams.setQuery(query); //设置查询词
            suggestParams.setHits(10); //设置最大返回下拉提示的条数
            suggestParams.setUserId("12345678"); //设置user_id
            //“中文同音别字”补全召回功能已默认生效，可以在请求中添加参数re_search进行调整
            //ReSearch.findByValue(1)表示禁用；ReSearch.findByValue(0)或者不传该参数表示开启
            suggestParams.setReSearch(ReSearch.findByValue(1));
            SearchResult result = suggestionClient.execute(suggestParams);
            System.out.println(result); //打印召回的结果
        } catch (OpenSearchException e) {
            e.printStackTrace();
        } catch (OpenSearchClientException e) {
            e.printStackTrace();
        }
    }
}

```

```
@Override
public void clean() {
    openSearch.clear();
}
}
```

下拉提示的Java SDK的详情可参考：[下拉提示demo](#)。

召回结果展示：

```
{
  "request_id": "159851481919726888064081",
  "searchtime": 0.006246,
  "suggestions": [
    {
      "suggestion": "裙子潮"
    },
    {
      "suggestion": "裙子小个子 连衣裙"
    },
    {
      "suggestion": "裙子 波点连衣裙"
    },
    {
      "suggestion": "裙子 减龄"
    },
    {
      "suggestion": "裙子 波点"
    },
    {
      "suggestion": "裙子小个子"
    },
    {
      "suggestion": "裙子小个子波点"
    }
  ]
}
```

注意：召回结果返回的requestID，可以用于搜索请求的关联。

## 下拉提示请求与搜索请求关联

将下拉提示请求和搜索请求关联起来后，对业务有如下帮助：

1. 可以统计出一些指标，用于衡量下拉提示对搜索的效果。指标包括：下拉提示引导搜索pv，引导搜索点击率，引导搜索无/少结果率等，更多详细指标见[下拉提示报表](#)。
2. 基于关联请求数据，得到下拉提示点击结果等数据，优化下拉提示排序模型，进而提升下拉提示引导搜索的效果。

**关联方法：**

在opensearch app search请求中，如果此次请求是由下拉提示引导的，则在search请求中带上参数

```
from_request_id={from_request_id}
```

 其中from\_request\_id表示：本搜索请求从哪里引导而来，如

果当前的query来自下拉提示、热词、底纹等功能的推荐列表，那么请求这个推荐列表的request\_id可以赋给这个参数，通过关联这个引导事件，可以计算上游功能的各项指标，衡量使用效果，为优化功能提供依据。[搜索处理文档](#)中也有此参数说明。

## 举例：

通过下拉提示API查询出来的结果中的requestID是：159851481919726888064081，如下如果想关联搜索请求，可参考：

```
SearchParams searchParams = new SearchParams(config);
searchParams.setQuery("title:'裙子小个子'"); //通过下拉提示引导的

//添加from_request_id参数
Map<String, String> customParam =new HashMap<>();
customParam.put("from_request_id","159851481919726888064081");
searchParams.setCustomParam(customParam);

//执行查询语句返回数据对象
SearchResult execute = searcherClient.execute(searchParams);
//以字符串返回查询数据
String result = execute.getResult();
System.out.println(result);
```

## 2.热搜和底纹

### 什么是热搜和底纹

热搜和底纹是一个完整搜索引擎必备的基本功能，通常占据着搜索框入口的重要位置，提供不可或缺的业务价值。下图是看一个典型电商平台的热搜和底纹的位置。



热搜和底纹处于搜索引擎整个工作流程的最上游，为搜索优化起铺垫作用，可以大大降低query理解、排序、运营干预等环节的调优难度，并且结合运营策略可以有比较大的发挥空间。

从用户的角度来看，热搜底纹一般可以满足如下的需求：

② 说明

1. 我想随便逛逛，不知道搜什么好，能不能给我推荐一些优质的查询词？
2. 我想知道大家都搜了些什么，随大流不会错
3. 最好能结合我的兴趣推荐query，也要有多样性，我既想看感兴趣的内容，又想探索一些兴趣之外的内容

从运营者的角度，热搜和底纹可以提供这样的价值：

② 说明

1. 我想知道哪些query被搜得最多，热门query是用户兴趣的风向标，通过分析热门query可以把握用户的兴趣走向，对制定运营策略提供决策依据
2. 我想给用户推荐一些优质query，在用户有输入的情况下，下拉提示引导用户意图，但是在没有任何输入的时候，如何推荐优质query呢？
3. 如果给用户推荐热门query，不能总是固定给出最热的那几个query，需要考虑到多样性，一方面兼顾用户体验，另一方面需要给部分次热门query曝光机会
4. 通过分析用户的行为，结合用户的兴趣来推荐query，既兼顾用户体验，又可以有的放矢的提升业务目标

【控制台热搜榜】和【热搜推荐接口】有以下几点区别：

② 说明

1. 【热搜推荐接口】只统计最近14天的，【控制台热搜榜】可以选择时间区间；
2. 【热搜推荐接口】只统计raw\_query，【控制台热搜榜】除了统计raw\_query还会统计default索引上的查询词；
3. 【热搜推荐接口】会过滤掉搜索无结果的query，【控制台热搜榜】则不会；

### 温馨提示

- 目前热搜与底纹算法模型需要通过统计历史搜索请求中设为 raw\_query 的查询词来进行训练，若历史搜索请求没有出现过raw\_query，模型可能会创建失败。
- 最多各创建3个热搜、底纹模型；
- 默认显示热搜榜数据条数是10000条，可根据实际情况进行修改，但最低不少于100条。
- 目前的算法模型只依赖搜索日志，按照一定算法做统计，暂不支持设定其他训练目标，因此仅需创建模型一次。
- 系统接收到模型创建请求后，会立刻开始训练模型。
- 一旦创建模型成功，系统会每天自动重新训练模型，不再需要人为干预。
- 热搜/底纹模型开启并且已经训练后，可查询半年内任意一个时间段的榜单数据，默认为近14天的榜单数据。

### 创建热搜或底纹模型

#### API/SDK创建热搜和底纹模型：

### 管控SDK依赖：

```
<dependency>
  <groupId>com.aliyun</groupId>
  <artifactId>aliyun-java-sdk-opensearch</artifactId>
  <version>0.7.0</version>
</dependency>
<dependency>
  <groupId>com.aliyun</groupId>
  <artifactId>aliyun-java-sdk-core</artifactId>
  <version>4.5.0</version>
</dependency>
```

## Java SDK Demo

```
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.exceptions.ServerException;
import com.aliyuncs.http.FormatType;
import com.aliyuncs.http.HttpResponse;
import com.aliyuncs.opensearch.model.v20171225.*;
import com.aliyuncs.profile.DefaultProfile;

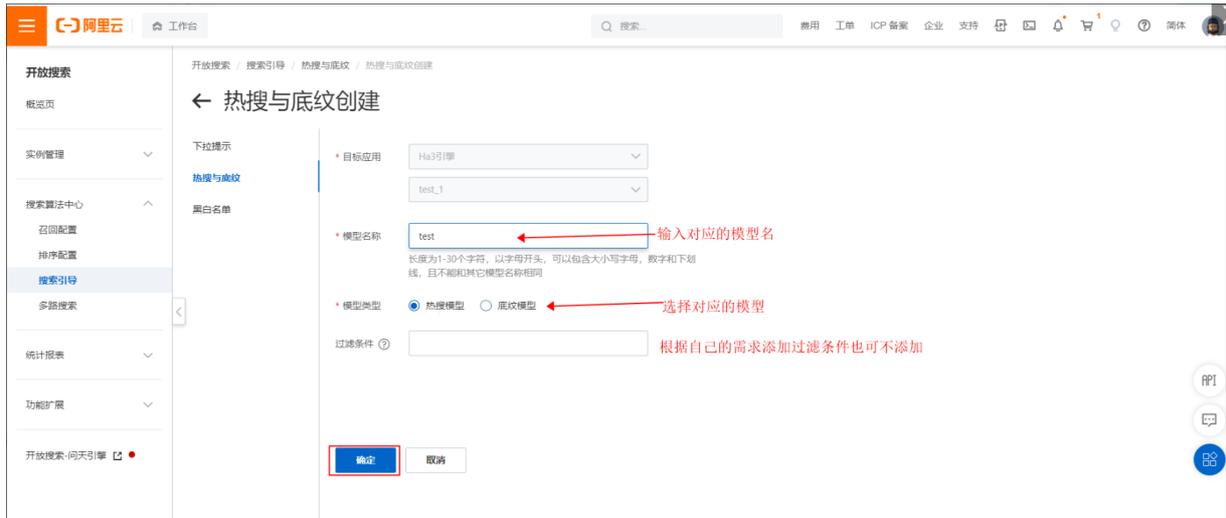
public class CreateModel {
    public static void main(String[] args) {
        DefaultProfile profile = DefaultProfile.getProfile("cn-beijing", "<accessKeyId>", "<accessSecret>");
        IAcsClient client = new DefaultAcsClient(profile);
        CreateModelRequest request = new CreateModelRequest();
        request.setAppGroupIdentity("app_name");
        //type为算法类型：热搜设置为hot，底纹设置为hint；modelName为模型名称：自定义，由英文数字、字母、下划线组成，非纯数字，不超过30个字符
        String content = "{\"type\": \"hot\", \"name\": \"model_name\"}";
        request.setHttpContent(content.getBytes(), "UTF-8", FormatType.JSON);
        try {
            HttpResponse response = client.doAction(request);
            System.out.println(response.getHttpContentString());
        } catch (ServerException e) {
            e.printStackTrace();
        } catch (ClientException e) {
            e.printStackTrace();
        }
    }
}
```

## 页面创建模型

1、在搜索引导->热搜与底纹页面，点击创建按钮创建热搜或底纹模型：



2、填写名称，选择模型类型（热搜/底纹模型），填写过滤条件（非必选）：



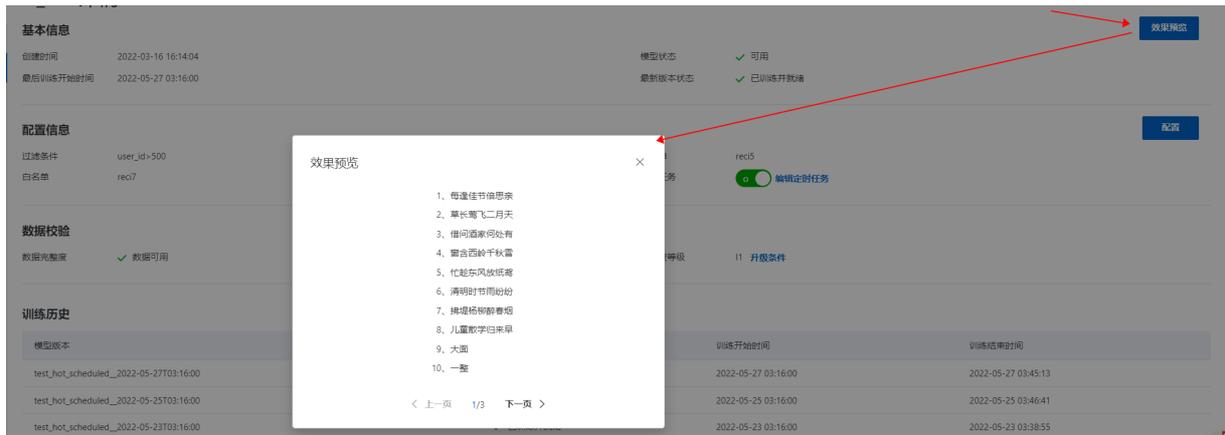
参数名称	说明
过滤条件	需填写OpenSearch搜索请求参数的判断条件，该过滤条件用于筛选符合条件的历史搜索词。支持如下格式：OpenSearch请求参数 biz_type (<、>、<=、>=、=、!=) 数值，暂仅支持biz_type和user_id参数。多个过滤条件之间为AND关系，必须要使用英文逗号(,)分隔，表示且关系（暂不支持或关系）。例如当过滤条件为 user_id=1,biz_type=1 时，则只能选取符合该条件的记录。

3、在热搜与底纹页面的列表页中找到刚创建的状态为“待训练”的模型，点击训练：

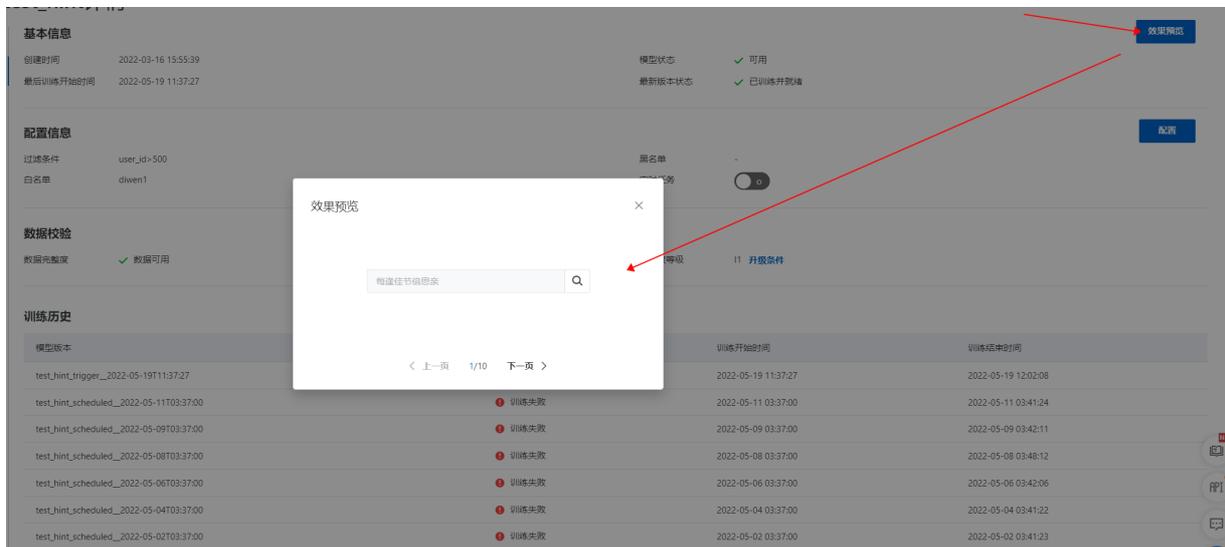


4、训练完成后可通过热搜效果预览查看效果：

### 热搜效果预览：



### 底纹效果预览：



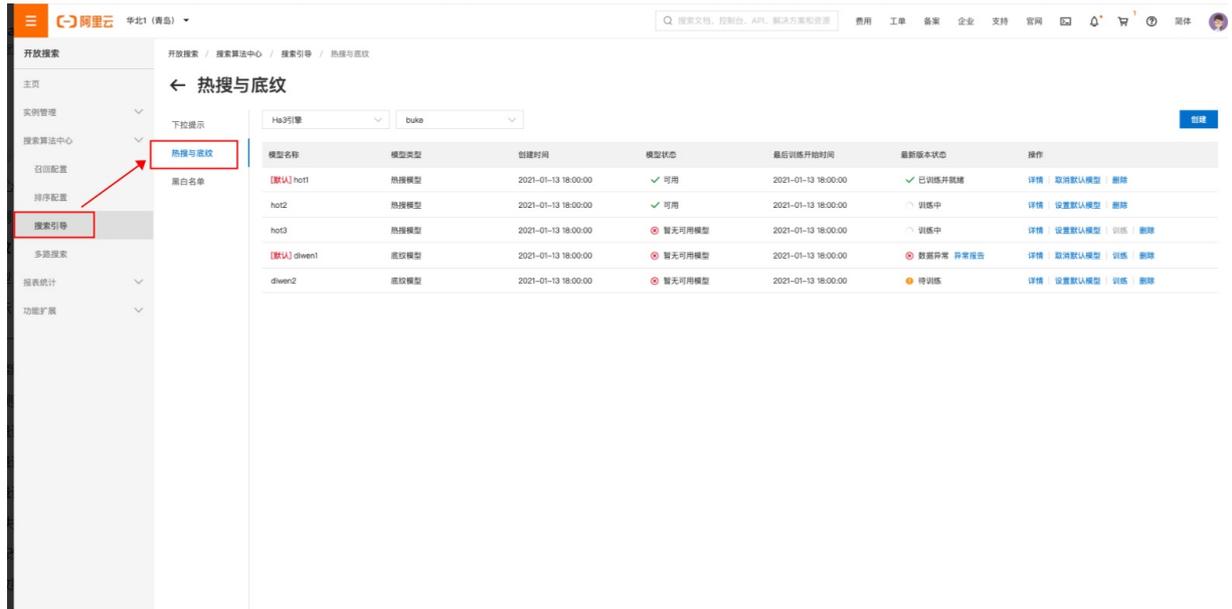
### 注意：

- 热搜的效果预览，弹框每次返回的10个推荐的热搜词，最多20组推荐的热搜词。
- 底纹的效果预览，弹框每次返回1个底纹词，最多30组推荐的底纹词。

## 热搜与底纹页面介绍

## 热搜与底纹列表介绍

可通过搜索引导->热搜与底纹下查看，模型名称、模型类型、创建时间、模型状态、最后训练开始时间、最新版本状态和操作信息。用户可以根据自身需求设置默认模型，如果参数中没有设定模型名称，则访问默认模型。



### 热搜与底纹 详情页介绍

其基本信息包含创建时间、模型状态、最后训练开始时间、最新版本状态。也可在训练历史下查看模型版本、版本状态、训练开始时间、训练结束时间等信息。模型训练后为两种状态训练成功和训练失败

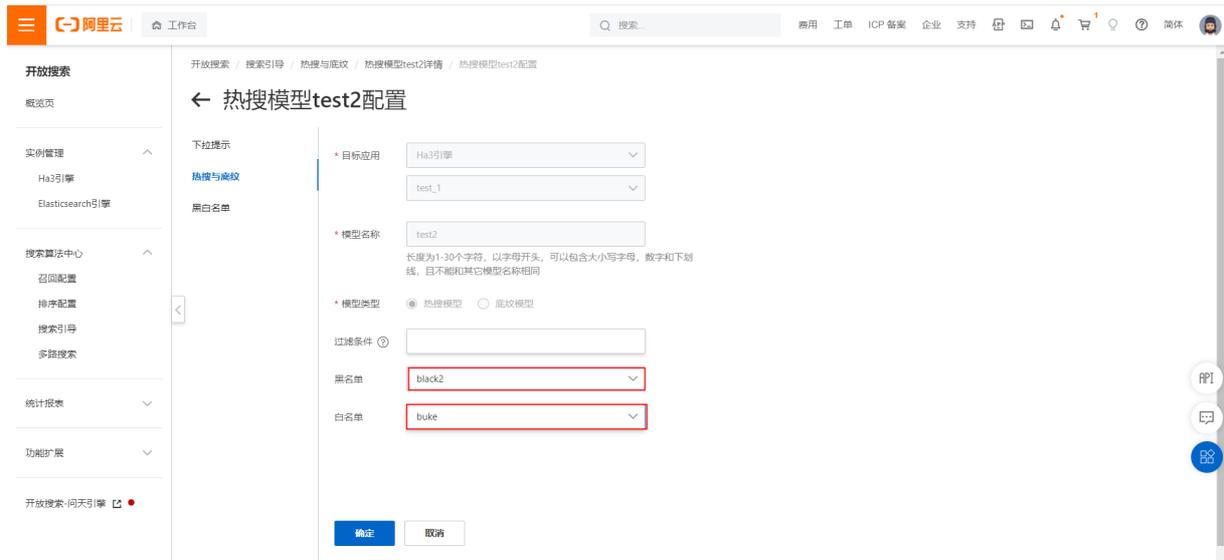
#### 训练失败问题查看：

可点击详情进入详情页查看：可以通过异常报告查看失败的原因，在升级条件下查看所需条件。



#### 个性化配置：

黑白名单配置：用户可根据自身业务的需求对模型进行黑白名单配置在模型详情下配置信息处点击配置



### SDKSDK获取热搜/底纹结果

在产品中集成热搜和底纹功能，可以通过API/SDK来调用热搜底纹的推荐结果，[热搜和底纹api文档](#)。SDK Demo请参考下文。

热搜功能默认返回10个query，底纹功能默认返回3个query，用于在搜索框中轮播。

#### 温馨提示：

- 目前没有获取算法任务执行状态的接口，需要用户调用API/SDK根据是否有结果来判断是否可使用。
- 一旦调用创建模型成功后，后续每日系统内部会自动执行算法训练任务，用户每日通过API/SDK获取更新的热搜词即可。
- 目前通过SDK调用热搜和底纹接口时，hit参数（召回个数）热搜最大是30，底纹最大是10。
- 热词榜单api: [ListStatisticLogs](#)

#### 流量SDK依赖：

```
<dependency>
  <groupId>com.aliyun.opensearch</groupId>
  <artifactId>aliyun-sdk-opensearch</artifactId>
  <version>3.5.0</version>
</dependency>
```

### Java SDK Demo

```
import java.util.HashMap;
import java.util.Map;
import com.alibaba.fastjson.JSON;
import com.aliyun.opensearch.OpenSearchClient;
import com.aliyun.opensearch.sdk.generated.OpenSearch;
import com.aliyun.opensearch.sdk.generated.common.OpenSearchClientException;
public class Hot {
    private static final String accesskey = "替换为accesskey";
    private static final String secret = "替换为secret";
    private static final String host = "替换为endpoint, 例如: http://opensearch-cn-hangzhou.aliyuncs.com";
    private static final String appName = "替换为opensearch应用名";
    private static final String HOT_API_PATH = "/apps/{app_name}/actions/hot";
    public static void main(String[] args) {
        OpenSearch openSearch = new OpenSearch(accesskey, secret, host);
        // Create OpenSearchClient
        OpenSearchClient client = new OpenSearchClient(openSearch);
        String requestPath = HOT_API_PATH.replaceAll("\\{app_name\\}", appName);
        Map<String, String> params = new HashMap<>();
        params.put("model_name", "your_model_name");//设置模型名称
        params.put("hit", "20");//设置查询的热搜词数据条数
        try {
            String response = client.call(requestPath, params, OpenSearchClient.METHOD_GET);
            ;
            System.out.println(JSON.toJSONString(response));
        } catch (OpenSearchClientException e) {
            e.printStackTrace();
        }
    }
}
```

## 效果优化

### 行为数据

基于统计搜索日志的热搜/底纹结果可以满足冷启动阶段的需求，在这个功能开始发挥作用后，建议关联热搜/底纹的用户点击事件，系统内部通过采集行为数据（`from_request_id`和`user_id`）来进一步优化效果，采集了行为数据有如下几个收益：

#### ② 说明

1. 可以统计得到各项指标，诸如引导搜索的pv、uv、无结果率等，用来衡量这个功能的使用效果，为后续改进提供依据。
2. 可以分析用户群的兴趣走向，为制定运营策略提供依据。
3. 可以采用智能化的手段来推荐query，通过用户的点击行为对数据进行标注，可能根据不同的优化目标来训练模型（默认按点击率优化），通过模型来推荐query，具备较强的泛化能力。

### 热搜/底纹引导搜索流量

API/SDK获取到的热搜/底纹结果中会返回一个request\_id，用于唯一标识这一次请求。通常，如果产品中集成了热搜/底纹结果，当终端用户点击了热搜词或底纹词后，便会以该热搜词或底纹词作为搜索词触发一次搜索请求，通过在搜索请求中将from\_request\_id参数设置为热搜/底纹结果返回的request\_id，就可以关联这一次用户点击行为，后台通过搜索日志分析和统计了这些点击行为，就可以在业务运营报表中看到由热搜/底纹引导的搜索流量指标。如果在搜索请求中设置了user\_id参数，业务运营报表中还可以看到热搜/底纹引导的搜索UV指标。

温馨提示：

- 由于搜索的翻页请求也会被认为是独立的搜索PV，所以对于热搜/底纹引导的搜索请求翻页时候，也同样要设置from\_request\_id参数。
- 搜索请求的参数详情，请查阅from\_request\_id和user\_id的参数说明。

## 业务运营报表

### 界面介绍

#### 热搜：

可以通过【搜索算法中心-->搜索引导-->热搜与底纹】，“查看指标”进入热搜的运营报表；



也可以直接在“统计报表-->热搜报表”里查：

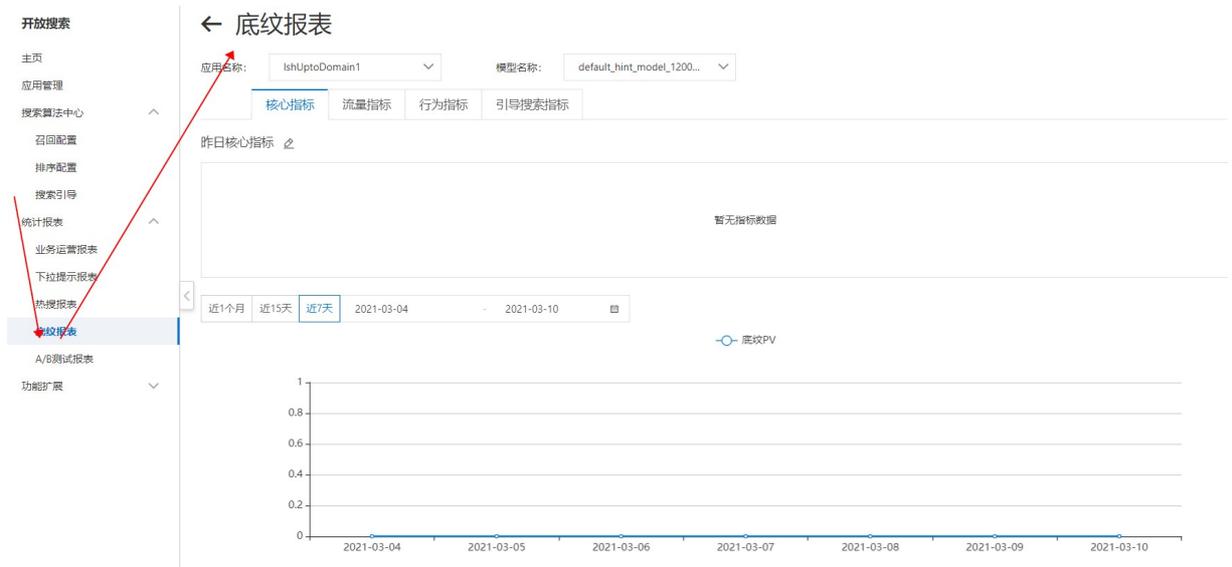


底纹：

可以通过【搜索算法中心-->搜索引导-->热搜与底纹】，“查看指标”进入底纹的运营报表；



也可以直接在“统计报表-->底纹报表”里查：



注意：

- 热搜报表对应指标介绍可[点击此处](#)进行查看。
- 底纹报表对应指标介绍可[点击此处](#)进行查看。

# 3.黑白名单

## 使用介绍

黑白名单主要用于对下拉提示、热搜、底纹候选query的干预，其中包含黑名单关键词的query将不会出现在候选query中；当白名单中的query满足推荐条件时，将优先出现在候选query中。

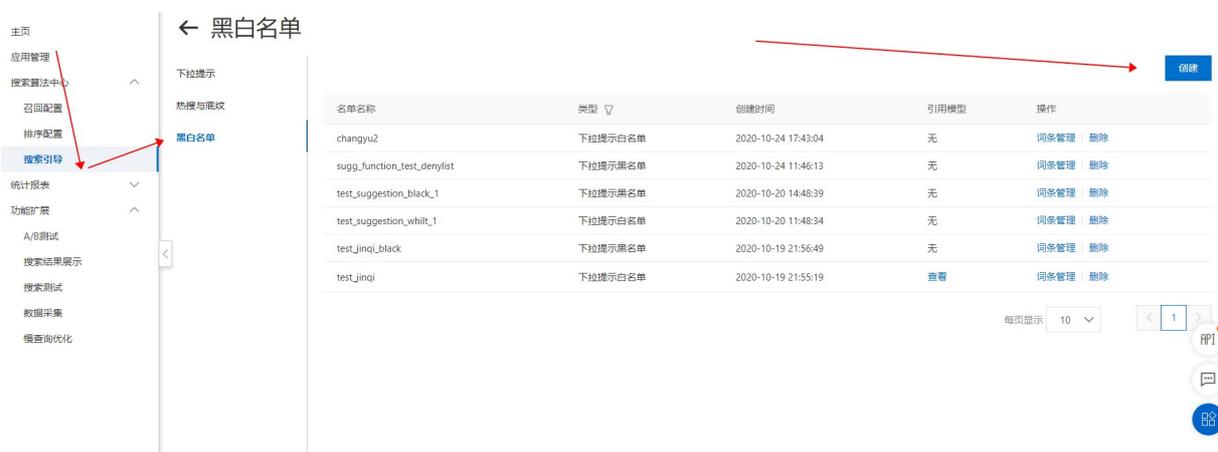
当候选query推荐结果中出现一些业务上不想要的结果时，可将关键词添加到黑名单中，实现对这些结果的屏蔽。反之，当某些优质query没有被推荐，或排序靠后时，可以将其加到白名单中，实现人工干预。

目前已支持对下拉提示、热搜、底纹的候选Query进行黑白名单干预。

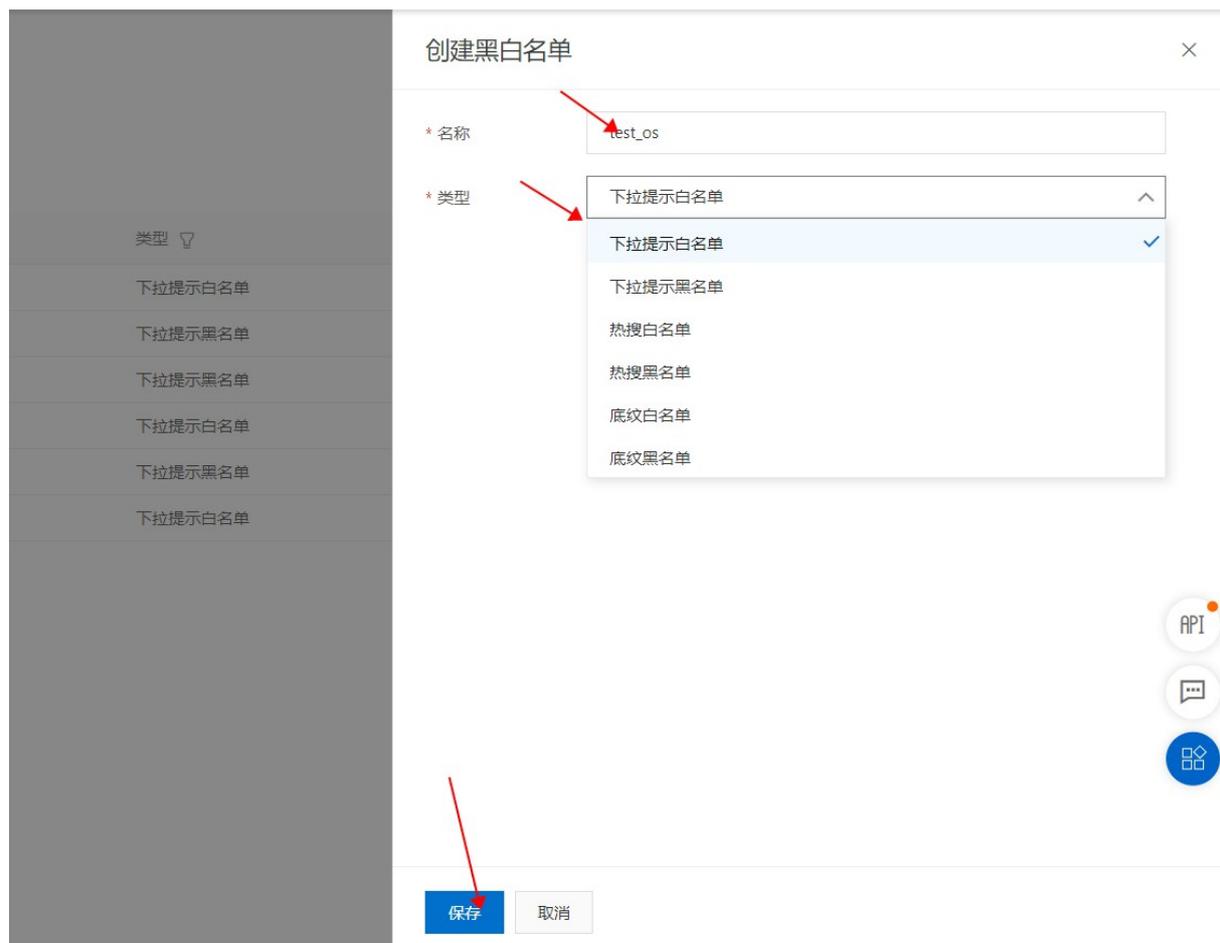
## 流程演示

当您的业务中需要优先展示，或者屏蔽一些候选query时可参考如下步骤：

1.创建词典，在搜索算法中心—>搜索引导—>黑白名单，点击“创建”按钮：



2.填写相关信息，输入“名称”，选择对应的“类型”【下拉提示白名单/下拉提示黑名单】，点击保存：



3.添加干预词条，在黑白名单列表页，找到对应词典，点击【操作】栏下的词条管理：

开放搜索 / 功能扩展 / 黑白名单

### ← 黑白名单

① 创建

词典名称	类型 ▾	创建时间	最近修改时间	引用模型	操作
test_suggestion_black_1	下拉提示黑名单	2020-10-20 14:48:39	2020-10-20 14:48:39	无	词条管理 删除
test_suggestion_whilt_1	下拉提示白名单	2020-10-20 11:48:34	2020-10-20 11:48:34	无	词条管理 删除
test_jinqi_black	下拉提示黑名单	2020-10-19 21:56:49	2020-10-19 21:56:49	查看	词条管理 删除
test_jinqi	下拉提示白名单	2020-10-19 21:55:19	2020-10-19 21:55:19	查看	词条管理 删除

每页显示 10 < 1 >

新增干预词条：

开放搜索 / 功能扩展 / 黑白名单 / 词条管理

### ← 词条管理

名称: test\_suggestion\_black\_1  
词典类型: 下拉提示黑名单  
引用应用: -  
创建时间: 2020-10-20 14:48:39  
词条数:

新增干预词条 输入Query

关键字 修改时间 没有数据

### 新增干预词条

下拉提示黑名单

夏季旧款老式修身裙

+ 继续添加

②

保存 取消

4.将配置好的干预词典，添加到对应的下拉提示模型中：

开放搜索

- 应用管理
- 搜索算法中心
- 召回配置
- 排序配置
- 搜索引导**
- 统计报表
- 业务运营报表
- 下拉提示报表
- 热搜报表
- 底纹报表
- A/B测试报表
- 功能扩展
- A/B测试
- 搜索结果展示
- 搜索测试
- 数据采集
- 查询优化

开放搜索 / 功能扩展 / 下拉提示

### ← 下拉提示

① 下拉提示是搜索服务的基础功能，在用户输入查询词的过程中，智能推荐候选query，提高用户输入效率，帮助用户尽快找到想要的内容。了解更多使用详情请访问产品在线文档

ishlptoDomain1

模型名称	是否已有可用模型	最后训练时间	近一次训练状态	昨日PV	操作
test1	否	①	训练中 2 %	-	详情 删除

每页显示 10 < 1 >

开放搜索 / 功能扩展 / 下拉提示 / 下拉提示test\_suggestion\_model\_1详情

### ← 下拉提示test\_suggestion\_model\_1详情

基本信息 效果预览

模型创建时间	是否有可用模型	最后训练时间	近一次训练状态
2020-10-20 10:42:21	是	2020-10-20 12:01:27	✓ 已训练并通过

配置信息 配置

行业类型	定时训练	训练字段	过滤条件
通用	开启	title 抽取生成 brand 原值保留 detail 原值保留	-

黑名单 白名单

开放搜索 / 功能扩展 / 下拉提示 / 下拉提示test\_suggestion\_model\_1详情 / 下拉提示test\_suggestion\_model\_1配置

### ← 下拉提示test\_suggestion\_model\_1配置

\* 定时训练  编辑定时任务

\* 行业类型  通用  电商  内容

\* 训练字段 选择训练字段

title  抽取生成  原值保留 × brand  抽取生成  原值保留 × detail  抽取生成  原值保留 ×

过滤条件

黑名单 test\_suggestion\_black\_1

白名单 test\_suggestion\_whit\_1

取消 完成

开放搜索 / 功能扩展 / 下拉提示 / 下拉提示test\_suggestion\_model\_1详情 / 下拉提示test\_suggestion\_model\_1配置

### ← 下拉提示test\_suggestion\_model\_1配置

\* 定时训练  编辑定时任务

\* 行业类型  通用  电商  内容

\* 训练字段 选择训练字段

title  抽取生成  原值保留 × brand  抽取生成  原值保留 × detail  抽取生成  原值保留 ×

过滤条件

黑名单 test\_suggestion\_black\_1

白名单 test\_suggestion\_whit\_1 词库管理

下拉提示配置

模型test\_suggestion\_model\_1配置保存成功! 黑白名单配置将实时生效, 无需重新训练模型。

确定

取消 完成

## 注意事项

- 黑名单为包含匹配，例：黑名单中配置查询词“垃圾”，那么候选query被分词的term中只要包含“垃圾”的term，该候选query都会被屏蔽；
- 当黑名单和白名单数据有冲突时，黑名单优先级更高；
- 黑名单支持正则表达式；
- 黑名单和白名单修改后，无需重新训练模型，实时生效；

- 
- 热搜白名单中已生效的Query将在热搜的指定位置展示，如果相应位置未设置推荐Query，则按热搜模型结果顺序展示；
  - 底纹白名单中已生效的Query将作为推荐词在底纹中轮播；
  - 已被模型引用的黑名单或者白名单无法直接删除；

## 功能限制

- 下拉提示黑名单中的query条数不能超过500；
- 下拉提示白名单中的query条数不能超过500；
- 热搜黑名单中的query条数不能超过500；
- 热搜白名单中的query条数不能超过10；
- 底纹黑名单中的query条数不能超过500；
- 底纹白名单中的query条数不能超过3；