UDH Search分词功能使用说明

目录

[一、概述 1](#_Toc424201403)

[二、环境准备 1](#_Toc424201404)

[2.1参照UDH部署手册，搭建Search集群 1](#_Toc424201405)

[2.2查看分词功能是否生效 2](#_Toc424201406)

[三、分词库修改 2](#_Toc424201407)

[四、创建Collection及构建索引 3](#_Toc424201408)

[4.1构建hbase表 3](#_Toc424201409)

[4.2 创建Collection 3](#_Toc424201410)

[4.3创建lily indexer配置文件 3](#_Toc424201411)

[4.4构建索引 4](#_Toc424201412)

# 一、概述

UDH1.0.3增加了在线上传下载分词库功能，使用者可以通过UDH Manager手动查看、修改Search集群的分词库，而不用单独到操作每个Search服务，简化了Search分词库管理工作。

# 二、环境准备

## 2.1参照UDH部署手册，搭建Search集群

通过UDH Manager安装Search服务的时候，同时会安装分词处理程序udh-search-ik，同时会在的schema.xml模板中添加应用分词功能的字段类型：text\_ik。

## 2.2查看分词功能是否生效

 打开Solr Web管理界面，选择已有的collection，在Analysis界面输入一段文字，看下分词情况，效果如下图：



# 三、分词库修改

3.1进入管理->分词管理下载已有的分词文件。

3.2根据需要修改分词文件（注意，词库中每个中文词汇独立占一行，文件编码需要UTF-8）。

3.3上传修改后的分词文件。

3.4手动重启Search服务。



# 四、创建Collection及构建索引

这里以批量构建hbase索引的方式创建基于分词器的Collection。

## 4.1构建hbase表



## 4.2 创建Collection

$ solrctl instancedir --generate /tmp/hbase-demo

$ vim /tmp/hbase-demo/conf/schema.xml

 在schema.xml的fields标签里添加：

<field name="info" type="text\_ik" indexed="true" stored="true" />

然后依次执行：

$ solrctl instancedir --create hbase-demo /tmp/hbase-demo

$ solrctl collection --create hbase-demo

## 4.3创建lily indexer配置文件

$ vim /tmp/morphline-hbase-mapper.xml



$vim /etc/hbase-solr/conf/morphlines-demo.conf



## 4.4构建索引

执行命令：



这个过程可能需要几分钟，出现“hadoop.ForkedMapReduceIndexerTool: Succeeded with job: jobName”语句表明执行成功。

*注：构建collection和构建索引阶段如果出现Direct buffer memory相关错误时，可以先关闭solr\_hdfs\_blockcache\_direct\_memory\_allocation特性，具体见search配置项页面。*