



链滴

老司机带你 elasticsearch 安装中文分词器

作者: [gwy](#)

原文链接: <https://ld246.com/article/1571927136929>

来源网站: 链滴

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

发车

为什么要在elasticsearch中使用ik这样的中文分词，那是因为es提供的分词是英文分词，对于中文的分词就做的非常不好了，因此我们需要一个中文分词器来用于搜索和使用。今天我们就尝试安装下IK分词。

上车

1、去github 下载对应的分词插件

<https://github.com/medcl/elasticsearch-analysis-ik/releases>

根据不同版本下载不同的分词插件

2、到es的plugins 目录创建文件夹

```
cd your-es-root/plugins/ && mkdir ik
```

3、解压ik分词插件到ik文件夹

```
unzip elasticsearch-analysis-ik-6.4.3.zip
```

第二种安装方法

还有一种方式 直接通过es的命令进行安装， es版本需要大于5.5.1

```
./bin/elasticsearch-plugin install https://github.com/medcl/elasticsearch-analysis-ik/releases/download/v6.3.0/elasticsearch-analysis-ik-6.3.0.zip
```

安装好后 重启es

会看到 加载了 ik分词了

```
[2019-10-22T14:01:04,322][INFO ][o.e.p.PluginsService ] [node-1] loaded module [x-pack-security]
[2019-10-22T14:01:04,322][INFO ][o.e.p.PluginsService ] [node-1] loaded module [x-pack-sql]
[2019-10-22T14:01:04,322][INFO ][o.e.p.PluginsService ] [node-1] loaded module [x-pack-upgrade]
[2019-10-22T14:01:04,322][INFO ][o.e.p.PluginsService ] [node-1] loaded module [x-pack-watcher]
[2019-10-22T14:01:04,323][INFO ][o.e.p.PluginsService ] [node-1] loaded plugin [analysis-ik]
[2019-10-22T14:01:08,930][INFO ][o.e.x.s.a.s.FileRolesStore] [node-1] parsed [0] roles from file [/wwwroot/soft/elasticsearch-6.4.3/config/roles.yml]
[2019-10-22T14:01:09,455][INFO ][o.e.x.m.j.p.l.CppLogMessageHandler] [controller/20687] [Main.cc@109] controller (64 bit): Version 6.4.3 (Build 7a0781676dd492) Copyright (c) 2018 Elasticsearch BV
[2019-10-22T14:01:09,837][DEBUG][o.e.a.ActionModule ] Using REST wrapper from plugin org.elasticsearch.xpack.security.Security
```

到站

IK 分词器简介

- 1, Elasticsearch中文分词我们采用IK分词， ik有两种分词模式， ik_max_word和ik_smart模式;
 - ik_max_word 和 ik_smart 什么区别?

ik_max_word: 会将文本做最细粒度的拆分，比如会将“中华人民共和国国歌”拆分为“中华人民共和国,中华人民,中华,华人,人民共和国,人民,人,民,共和国,共和,和,国国,国歌”，会穷尽各种可能的组合；

ik_smart: 会做最粗粒度的拆分，比如会将“中华人民共和国国歌”拆分为“中华人民共和国,国歌”。

索引时，为了提供索引的覆盖范围，通常会采用ik_max_word分析器，会以最细粒度分词索引，搜索为了提高搜索准确度，会采用ik_smart分析器，会以粗粒度分词

实测

创建index

```
curl -XPUT http://localhost:9200/index
```

创建 mapping

```
curl -XPOST http://localhost:9200/index/index_mapping -H 'Content-Type:application/json' -d '{
  "properties": {
    "content": {
      "type": "text",
      "analyzer": "ik_max_word",
      "search_analyzer": "ik_smart"
    }
  }
}'
```

添加几个数据

```
curl -XPOST http://localhost:9200/index/index/1 -H 'Content-Type:application/json' -d' {"content":"美国留给伊拉克的是个烂摊子吗"} '
```

```
curl -XPOST http://localhost:9200/index/index/3 -H 'Content-Type:application/json' -d' {"content":"中韩渔警冲突调查：韩警平均每天扣1艘中国渔船"} '
```

```
curl -XPOST http://localhost:9200/index/index/4 -H 'Content-Type:application/json' -d' {"content":"中国驻洛杉矶领事馆遭亚裔男子枪击 嫌犯已自首"} '
```

进行查询

```
curl -XPOST http://localhost:9200/index/index/_search -H 'Content-Type:application/json' -d'
{
  "query": {
    "match": {
      "content": "中国"
    }
  },
  "highlight": {
    "pre_tags": ["<tag1>", "<tag2>"],
    "post_tags": ["</tag1>", "</tag2>"],
    "fields": {
      "content": {}
    }
  }
}
```

查看效果

历史记录

▼ 查询

http://10.180.18.38:9201/

_search

POST

```
{
  "query": {
    "match": {
      "content": "中国"
    }
  },
  "highlight": {
    "pre_tags": [
      "<tag1>",
      "<tag2>"
    ],
    "post_tags": [
      "</tag1>",
      "</tag2>"
    ],
    "fields": {
      "content": {}
    }
  }
}
```

提交请求 验证 JSON 易读

▶ 结果转换器

▶ 重复请求

▶ 显示选项

```
{
  "took": 9,
  "timed_out": false,
  "_shards": {
    "total": 6,
    "successful": 6,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 2,
    "max_score": 0.6489038,
    "hits": [
      {
        "_index": "index",
        "_type": "index",
        "_id": "4",
        "_score": 0.6489038,
        "_source": {
          "content": "中国驻洛杉矶领事馆遭亚裔男子枪击 嫌犯已自首"
        },
        "highlight": {
          "content": [
            "<tag1>中国</tag1>驻洛杉矶领事馆遭亚裔男子枪击 嫌犯已自首"
          ]
        }
      },
      {
        "_index": "index",
        "_type": "index",
        "_id": "3",
        "_score": 0.2876821,
        "_source": {
          "content": "中韩渔警冲突调查：韩警平均每天扣1艘中国渔船"
        },
        "highlight": {
          "content": [
            "中韩渔警冲突调查：韩警平均每天扣1艘<tag1>中国</tag1>渔船"
          ]
        }
      }
    ]
  }
}
```