

8-es keyword normalizer

Created by 杨超 on 2019 Jun 26

目的：为了实现搜索的时候区分大小写的功能 比如搜索foo的时候 能把Foo这种也搜索出来

调研的结果: 1 对于默认的text 已经做了大小写归一化的工作 也就是统一把大小写的英文都转成了小写 所以搜索这种字段foo是能够搜出Foo 但是排序就不好说了

2 对于keyword 是一种“封闭”的结构 不分词 在导入索引的时候就已经决定了 是“原封不动”的导入的样子 所以搜索的时候foo只能搜出foo, Foo只能搜出Foo

如果要实现大小写无关这种特性，必须使用normalizer 也就是统一把keyword导入索引的时候转成小写 这样搜索的时候foo就可以搜出Foo

下面是个例子

```
PUT lowercase_test
{
  "settings": {
    "analysis": {
      "normalizer": {
        "my_normalizer": {
          "type": "custom",
          "char_filter": [],
          "filter": ["lowercase", "asciifolding"]
        }
      }
    }
  },
  "mappings": {
    "_doc": {
      "properties": {
        "foo": {
          "type": "keyword",
          "normalizer": "my_normalizer"
        }
      }
    }
  }
}
```

```
}
PUT lowercase_test/_doc/1
{"foo":"BAR"}
PUT lowercase_test/_doc/2
{"foo":"bar"}
PUT lowercase_test/_doc/3
{"foo":"baz"}
PUT lowercase_test/_doc/4
{"foo":"米ui"}
PUT lowercase_test/_doc/5
{"foo":"米uI"}
PUT lowercase_test/_doc/6
{"foo":"米UI"}
PUT lowercase_test/_doc/7
{"foo":"米ui水"}
GET lowercase_test/_search
{
  "query": {
    "term": {
      "foo": "BAR"
    }
  }
}
GET lowercase_test/_search
GET lowercase_test/_search
{
  "query": {
    "match": {
      "foo": "BAR"
    }
  }
}
```

Like Be the first to like this

No labels

地址：北京市朝阳区建国路86号佳兆业广场北塔6层梦想加空间601室

以太资本由艾普拉斯投资顾问(北京)有限公司运营，提供早期互联网项目的投融资对接服务

©2014-2017 以太资本 京ICP备14028208号