

## 2-15 附：分词与内置分词器

### 什么是分词？

把文本转换为一个个的单词，分词称之为analysis。es默认只对英文语句做分词，中文不支持，每个中文字都会被拆分为独立的个体。

- 英文分词：I study in [imoooc.com](https://www.imoooc.com)
- 中文分词：我在慕课网学习

```
POST /_analyze
{
  "analyzer": "standard",
  "text": "text文本"
}
```

```
POST /my_doc/_analyze
{
  "analyzer": "standard",
  "field": "name",
  "text": "text文本"
}
```

### es内置分词器

- standard：默认分词，单词会被拆分，大小会转换为小写。
- simple：按照非字母分词。大写转为小写。
- whitespace：按照空格分词。忽略大小写。
- stop：去除无意义单词，比如 the / a / an / is ...
- keyword：不做分词。把整个文本作为一个单独的关键词。

```
{
  "analyzer": "standard",
  "text": "My name is Peter Parker, I am a Super Hero. I don't like the Criminals."
}
```

检测到您还没有关注慕课网服务号，无法接收课程更新通知。请扫描二维码即可绑定



下一节

2-15 附：分词与内置分词器

播放下一节

重新观看