1、默认的分词器

standard

standard tokenizer：以单词边界进行切分
standard token filter：什么都不做。。。。。
lowercase token filter：将所有字母转换成小写
stop token filter（默认被禁用）：移除停用词，比如：a the it 等等

2、修改分词器的设置

启用 english 停用词 token filter

```
PUT /my_index1                    // 启用停用词 token filter
{
  "settings": {
    "analysis": {
      "analyzer": {
        "es_std":{
          "type":"standard",
          "stopwords":"_english_"
        }
      }
    }
  }
}

GET /my_index1/_analyze
{
  "analyzer": "es_std",
  "text": ["i have a dog,and i love my dog"]
}
```

3、定制自己的分词器

```
PUT /my_index2
{
  "settings": {
    "analysis": {
      "char_filter": {
        "&_to_and":{
          "type":"mapping",
          "mappings":["&=> and "]
        }
      },
      "filter": {
        "my_stopwords":{
          "type":"stop",
          "stopwords":["the", "a"]
        }
      },
      "analyzer": {
        "my_analyer":{
          "type":"custom",
          "char_filter":["html_strip","&_to_and"],
          "tokenizer":"standard",
          "filter":["lowercase","my_stopwords"]
        }
      }
    }
  }
}

GET /my_index2/_analyze
{
  "analyzer": "my_analyer",
  "text": ["i love my dad&mom,<a>pilipala</a>, the dog is so CUTE"]
}
```