1、什么是分词器

切分词语， normalization（提升 recall 召回率）

给你一段句子，然后将这段句子拆分成一个一个的单个的单词，同时对每个单词进行 normalization（时态转换，单复数转换），分词
提高 recall 召回率，搜索的时候，增加能够搜索到的结果数量

character filter，在一段文本进行分词之前，进行预处理，比如说最常见的就是，过滤 html 标签（hello ---> hello），& --> and（I&you --> I and you）
tokenizer, 分词，hello you and me --> hello, you, and, me
token filter (normalization 有关操作): lowercase 大小写转换, stop word 停用词转换, synonym 同义词转换;
例如 dog -- > dogs, liked -- > like, Tom -- > tom, a/the/to ---> 干掉， mother ---> mom, small ---> little

一个分词器很重要，将一段文本进行各种处理，最后将处理好的结果才会拿去建立倒排索引

2、 内置分词器的介绍

要处理的文本 Set the shape to semi-transparent by calling set_trans(5)

standard analyzer: set，the， shape， to， semi， transparent， by， calling，set_trans， 5（标准分词器）
simple analyzer：set， the， shape，to，semi，transparent，by，calling，set，trans (简单分词器)
whitespace analyzer：Set，the，shape，to，semi-transparent，by，calling，set_trans(5)（空格分词器）
language analyzer：set， shape，semi，transpar，call，set_tran，5 （特定语言分词器）