Character Filters（字符过滤器）

原文链接：https://www.elastic.co/guide/en/elasticsearch/reference/5.3/analysis-charfilters.html

译文链接：http://www.apache.wiki/pages/viewpage.action?pageId=9406443

贡献者：谢雄，ApacheCN，Apache中文网

Character filters（字符过滤器）用于字符流传递到分词器（tokenizer）之前对它进行预处理。

一个 Character filters（字符过滤器）接收原始文本作为字符流，通过 adding（添加），removing（删除）或 changing （更改）字符来转换流。例如，可以使用字符过滤器将 Arabic numerals（阿拉伯数字）（ ）转换为和它等价的 Latin（拉丁数字）（0123456789），也可以用于从字符流中剥离 <b> 等 HTML 元素。

Elasticsearch 内置了许多的 character filters（字符过滤器），可以用来构建 custom analyzers（自定义分词器）。

HTML Strip Character Filter

该 html_strip 字符串过滤器可以删除类似 <b> 的 HTML 元素和解码类似于 &amp 这样的 HTML 实体。

Mapping Character Filter

该 mapping 字符串过滤器可以将所有指定的字符串替换成特定的字符串。

Pattern Replace Character Filter

该 pattern_replace 字符串过滤器可以将满足正则表达式的所有字符串替换成特定的字符串