Chapter 1

文本

1.1　re: 正则表达式

1.1.1　模式语法

重复

模式中有 5 中表示重复的方法

1. 元字符 *，则表示重复 0 次或多次（运行一个模式重复 0 次是指这个模式即使不出现也可以出现匹配）

2. + ，那么模式必须至少出现 1 次才能匹配

3. ? 表示出现 0 次或 1 次

4. 如果需要指定出现次数，需要在模式后面使用 {m}，这里 m 是模式应重复的次数

5. 如果要允许一个可变但有限的重复次数，那么可以使用 {m,n}，这里 m 是最小重复次数，n 是最大重复次数。如果省略 n (m,) 则表示值必须知道出现 m 次，但没有最大限制

字符集

A character set is a group of characters, any one of which can match at that point in the pattern. For example, [ab] would match either a or b.

A character set can also be used to exclude specific characters. The caratˆmeans to look for characters that are not in the set following the carat.

As character sets grow larger, typing every character that should (or should not) match becomes tedious. A more compact format using **character ranges** can be used to define a character set to include all of the contiguous characters between the specified start and stop points.

As a special case of a character set, the meta-character dot, or period (.), indicates that the pattern should match any single character in that position.

转义码

Table 1.1: Regular Expression Escape Codes

Code	Meaning
d	A digit
D	A non-digit
s	Whitespace (tab, space, newline, etc.)
S	Non-whitespace
w	Alphanumeric
W	Non-alphanumeric