vincent@ubuntu:~$ **awk '$0~/Brown/ {print}' grade.txt**

意思是：将所有匹配 Brown 的行打印出来。其中，$0~/Brown/是一个条件，表示所指定的域（这里是$0）要匹配的规则（这里是 Brown），也就是说，grade.txt 中的一行只要包含有单词 Brown，就会被选出来然后打印出来。

这里的单词 Brown 被包含在两个正斜杠之间，事实上这两个正斜杠之间写的就是大名鼎鼎的“正则表达式”（即 Regular Expression 有时简称 RE），RE 是脚本编程当中的高富帅，一般人还真不容易搞懂，但他是一切文本处理的基础，没有 RE 很多工具基本上就变得毫无用处，甚至是像 awk 和 sed 这样的神器，不客气地说，也都只是一堆废铁而已。这么玄乎的东东，当然要来一瞧究竟：

正则表达式 RE 说到底就是一些约定的规则，用这些规则来匹配字符串，从而达到过滤文本信息的目的。RE 使用各种字符和符号，来筛选你要的信息。假设文件 exmaple.txt 里面的内容是：

vincent@ubuntu:~$ **cat example.txt -n**

```
1  this is a testing string.
2  $22 apple, $33 banana, strawberry $88
3  $99
```

如果要找到 apple 这个单词，那么只要这么做就行了：

vincent@ubuntu:~$ **grep 'apple' example.txt -n**

2: $22 **apple**, $33 banana, strawberry $88

这里，我们使用了'apple'来直接完整地匹配了整一个单词，一个字符都不差。这里的'apple'其实就是一个 RE，是一个没有使用任何特殊字符的 RE。现在，假如想要找到所有以 ing 结尾的单词，该怎么做呢？请看：

vincent@ubuntu:~$ **grep '[^]*ing ' example.txt**

2: this is a **testing** **string**.

此时我们发现，testing 和 string 两个单词都被匹配了出来，其中的[^]*ing 是一个比上面的 apple 更为典型也更复杂的 RE，里面包含了几个特殊的模式符号。其实 RE 有一系列特殊的模式字符，来表示不同的含义，下表总结了正则表达式中最重要的特殊符号：

元字符	含义	示例	说明
.	匹配一个任意字符	a.b	匹配以 a 开头 b 结尾，中间有一个任意字符的单词
[ ]	匹配一个指定范围的字符	a[xyz]b	匹配以 a 开头 b 结尾，中间有一个 x 或 y 或 z 的单词
[^]	匹配一个不在指定范围的字符	a[^xyz]b	匹配以 a 开头 b 结尾，中间有一个不是 x 或 y 或 z 的字符的单词
^	匹配行首	^ab	匹配以 ab 为行首的单词
$	匹配行尾	ab$	匹配以 ab 为行尾的单词
\	转义符	\<ab\>	使<和>成为有特殊含义的符号
< >	匹配一个指定的单词	\<ab\>	精准匹配 ab 这个单词
|	逻辑或	ab|AB	匹配 ab 或者 AB

?	使前面的字符重复 0 次或 1 次	a?	匹配 a 重复了 0 次或 1 次的单词
*	使前面的字符重复 0 次或多次	a*	匹配 a 重复了 0 次或多次的单词
+	使前面的字符重复 1 次或多次	a+	匹配 a 重复了 1 次或多次的单词
{n}	使前面的字符重复 n 次	a{6}	匹配 a 重复了 6 次的单词
{n,}	使前面的字符重复 n 次或以上	a{6,}	匹配 a 重复了 6 次或以上的单词
{n,m}	使前面的字符重复 n 次到 m 次	a{6,9}	匹配 a 重复了 6 次到 9 次的单词

图 1-35 正则表达式

现在来解释一下刚才的命令：

```
vincent@ubuntu:~$ grep '[^ ]*ing ' example.txt -nE
    1:  this is a testing string.
```

[^]表示一个不为空格的任意字符，*表示前面的模式（即一个不为空格的任意字符）被重复 0 次或多次，ing 精准匹配 ing。因此整个 RE 的意思就是：找出不以空格开头且以 ing 结尾的单词。

针对表格中的样例，下面再举几个例子加以说明：

```
vincent@ubuntu:~$ grep '\<is\>' a.txt -nE
    1:  this is a testing string.
```

只精准匹配单词 is，如果不加尖括号，this 也会被匹配出来。

```
vincent@ubuntu:~$ grep '^\$.{2}$' example.txt -nE
    3:  $99
```

其中，^\$表示匹配以$为行首的单词，$前面必须加\进行转义，否则$就表示行尾了。接着，.{2}指的是一个重复了两遍的任意字符，最后一个$表示行尾。整个 RE 的意思是：找出以美元符号$为行首，且紧跟着两个任意字符为行尾的单词。