

# 基础正则表达式介绍与练习

## 一、实验介绍

### 1.1 实验内容

本次实验将介绍正则表达式的一些基本概念和用法。

### 1.2 实验知识点

- 正则表达式的概念
- 特殊符号

### 1.3 实验环境

课程使用的实验环境为 Ubuntu Linux 14.04 64 位。实验中会用到程序：

- Xfce终端

## 二、实验步骤

接下来开始实验。

### 2.1 什么是正则表达式

在做文字处理或编写程序时，用到查找、替换等功能，使用正则表达式能够简单快捷的完成目标。简单而言，正则表达式通过一些特殊符号的帮助，使用户可以轻松快捷的完成查找、删除、替换等处理程序。例如 grep, expr, sed, awk. 或 Vi 中经常会使用到正则表达式，为了充分发挥 shell 编程的威力,需要精通正则表达式。正规表示法基本上是一种『表示法』，只要工具程序支持这种表示法，那么该工具程序就可以用来作为正规表示法的字符串处理之用。也就是说，例如 vi, grep, awk, sed 等等工具，因为她们有支持正规表示法，所以，这些工具就可以使用正规表示法的特殊字符来进行字符串的处理。

### 2.2 正则表达式特殊符号

首先是正则表达式的特殊符号：

## 正则表达式基础 (/courses/90)

[:alpha:]代表英文大小写字母  
 [:blank:]代表空格和 tab 键  
 [:cntrl:]键盘上的控制按键, 如 CR,LF,TAB,DEL  
 [:digit:]代表数字  
 [:graph:]代表空白字符以外的其他  
 [:lower:]小写字母  
 [:print:]可以被打印出来的任何字符  
 [:punct:]代表标点符号  
 [:upper:]代表大写字母  
 [:space:]任何会产生空白的字符如空格, tab,CR 等  
 [:xdigit:]代表 16 进位的数字类型

### 2.2.1 特殊符号实例

任意建立一个文本并编辑。或者以原有的文本做实验皆可。

```
$ touch regular_express.txt
$ vi regular_express.txt
//或者 sudo gedit regular_express.txt
```

操作截图：

```
shiyancelou:~/ $ cat regular_express.txt [17:10:06]
"Open Source" is a good mechanism to develop programs.
apple is my favorite food.
Football game is not use feet only.
this dress doesn't fit me.
However, this dress is about $ 3183 dollars.
GNU is free air not free beer.
Her hair is very beauty.
I can't finish the test.
Oh! The soup taste good.
motorcycle is cheap than car.
This window is clear.
the symbol '*' is represented as start.
Oh!My god!
The gd software is a library for drafting programs.
You are the best is mean you are the no. 1.
The world <Happy> is the same with "glad".
I like dog.
google is the best tools for search keyword.
goooooogle yes!
go! go! Let's go.
# I am VBird
shiyancelou:~/ $
```



[17:10:22]

文本内容为：

正则表达式基础 (regular expression) to develop programs.  
 apple is my favorite food.  
 Football game is not use feet only.  
 this dress doesn't fit me.  
 However, this dress is about \$ 3183 dollars.  
 GNU is free air not free beer.  
 Her hair is very beauty.  
 I can't finish the test.  
 Oh! The soup taste good.  
 motorcycle is cheap than car.  
 This window is clear.  
 the symbol '\*' is represented as start.  
 Oh!My god!  
 The gd software is a library for drafting programs.  
 You are the best is mean you are the no. 1.  
 The world <Happy> is the same with "glad".  
 I like dog.  
 google is the best tools for search keyword.  
 gooooooogle yes!  
 go! go! Let's go.  
 # I am VBird

查找小写字母：

```
$ grep -n '[:lower:]' regular_express.txt
```

操作截图：


蓝色为匹配成功的字符

```
shiyanolou:~/ $ grep -n '[:lower:]' regular_express.txt
1:"Open Source" is a good mechanism to develop programs.
2:apple is my favorite food.
3:Football game is not use feet only.
4:this dress doesn't fit me.
5:However, this dress is about $ 3183 dollars.
6:GNU is free air not free beer.
7:Her hair is very beauty.
8:I can't finish the test.
9:Oh! The soup taste good.
10:motorcycle is cheap than car.
11:This window is clear.
12:the symbol '*' is represented as start.
13:Oh!My god!
14:The gd software is a library for drafting programs.
15:You are the best is mean you are the no. 1.
16:The world <Happy> is the same with "glad".
17:I like dog.
18:google is the best tools for search keyword.
19:gooooooogle yes!
20:go! go! Let's go
21:# I am VBird
```

[11:58:34]



查找数字：

 # 正则表达式基础 (/courses/90) regular\_express.txt

操作截图：

```
shiyanolou:~/ $ grep -n '[[[:digit:]]]' regular_express.txt [12:58:30]
5:However, this dress is about $ 3183 dollars.
15:You are the best is mean you are the no. 1.
```

## 2.3 语系对正则表达式的影响

由于不同语系的编码数据不同，所以造成不同语系的数据选取结果有所差异。以英文大小写为例，zh\_CN.big5 及 C 这两种语系差异如下：

- LANG=C 时：0 1 2 3 4...ABCDE...Zabcde...z
- LANG=zh\_CN 时：0 1 2 3 4...aAbBcCdD.....zZ

在使用正则表达式[A-Z]时，LANG=C 的情况下，找到的仅仅是大写字符 ABCD..Z。而在 LANG=zh\_CN 情况下，会选取到 AbBcCdD.....zZ 字符。因此在使用正则表达式时要特别留意语系。

由于我们一般使用的兼容与 POSIX 的标准，因此使用 C 语系。

## 三、实验总结

本节实验介绍了正则表达式的一些基本概念以及特殊符号。下一节我们将进入 grep 命令与正则表达式的学习。

\*本课程内容，由作者授权实验楼发布，未经允许，禁止转载、下载及非法传播。

下一节：grep命令与正则表达式 (/courses/90/labs/780/document)