

## 6. 字符串类型

### 学习目标

---

1. 了解字符串在Python中的4种表示方式
2. 熟练使用字符串的替换、去除指定字符的方法
3. 熟练掌握字符串的三种格式化方式
4. 了解字符串切片原理
5. 掌握字符串拼接方式
6. 了解字符串转义和原始字符串的使用
7. 熟练掌握字符串的编码和解码，了解字符串常见的编码解码方式

### 字符串类型

---

#### 字符串的声明

在Python里面，凡是用英文引号引起来的字符就是字符串。引号可以是单引号，双引号和三引号。

```
s1 = 'hello'
s2 = "world"
s3 = '''python'''
s4 = """我爱学习"""
# 对于三引号，可以换行
s5 = '''我是第一行字符串。
然后我又重新写了一行
这里面可以包含单引号，'hello'，
和双引号 "world" 而不需要转义
如何换行就如何输出
'''
```

字符串的类型用 `str` 表示，如果想要让其它数据类型转换为字符串，那么直接使用 `str` 类的实例就可以了。

```
a = 123
str(a)
str(123.456)
```

#### 字符串转义

如果要在单引号声明的字符串里面使用单引号或者是在双引号声明的字符串里面使用双引号，那么就必须对里面的单引号或双引号进行转义才能使用，否则会报错，转义就是在需要使用的单引号前面用反斜杠进行转义。

```
s1 = 'i\'m 30 years old.' # 这里我们对里面的单引号进行了转义
s2 = "如何学习一门编程语言是个很值得深思的问题，比如\"python\"这门语言"
```

对于在单引号里面使用双引号，或在双引号字符串里面使用单引号则不需要进行转义。

## 原始字符串

如果字符串里面包含一些其它类型的转义字符（ASCII码表中的），比如换行符 `\n`、制表符 `\t` 这些那么这些字符会原本的进行转义出来。如果是换行，那么打印出来的时候也会换行。那么如果需要展示的是原始的样子，那么一种方法是进行转义，另一种方法就是使用原始字符串。

原始字符串的定义就是在字符串前面加上字母 `r` 或 `R`

```
s = r"hello \n world"
```

## 获取子字符串

如果我们想要获取到字符串里面的某一个子串，比如存在里面的数字，那么就需要用到切片的方式进行访问了。

可以把切片理解为两把刀，从你要截取的起点和终点的位置各砍一刀下去，然后就能拿到里面的子串了。

- `[:]` 提取整个字符串
- `[start:]` 从start提取到结尾
- `[:end]` 从头提取到end-1
- `[start: end]` 从start提取到end-1
- `[start: end: step]` 从start提取到end-1，每隔step提取一个

## 判断一个字符串是否包含某个字符串

可以通过成员运算符 `in` 或者是 `not in` 来判断一个字符串是否在另一个字符串里面。

想想看，这个有什么用？？？

## 字符串重复

假如我们需要一个重复N遍的字符串，比如hellohellohellohello这样的，那么应该怎么样操作呢？

## 字符串拼接

两个或多个字符串要合并在一起的话可以用加号 + 相连，或者使用后面更高级的用法 `join` 方法

## 字符串格式化

如果我们需要让一个字符串生成某种格式的话，那么就需要用到字符串格式化。比如我们生成网站的sitemap或者是构造URL链接等等。都是要生成指定格式的字符串，这时候字符串格式化就派上用场了。

目前Python3中支持的字符串格式化有3种，为了能够最大限度的看懂别人的代码。所以我们需要对3种字符串格式化都要有所了解才行。

### 百分号格式化 (%)

百分号在字符串里面具有特殊的作用，就是用来做格式化的，因此如果是要单纯的显示一个百分号还需要特定的写法才行。

常用的百分号格式化的符号和含义

符号	说明
%d	代表的是整数
%s	代表字符串
%f	代表浮点数
%x	代表的是十六进制数
%%	代表百分号本身

### format函数格式化

format函数格式化在我们的Python学习过程中是比较常用的，而且速度也比较快。使用起来非常的方便。

具体用法看实例演示。

### f-string格式化

f-string是Python3.6才新增的一种格式化，比起format更加的方便一些，但是灵活性还不怎么够。

具体用法看视频演示。

## 字符串常用函数

---

方法	描述
<code>string.count(str, beg=0, end=len(string))</code>	返回str在string里面出现的次数,如果指定beg和end则返回的是范围内的
<code>string.encode(encoding='utf-8', errors='strict')</code>	以encoding指定的编码格式编码string, 如果出错, 默认报一个ValueError异常, 除非errors指定的是'ignore'或者是'replace'
<code>string.endswith(obj, beg=0, end=len(string))</code>	检查字符串是否以obj结束, 如果是beg和end指定范围, 则检查范围内是否是以obj结束
<code>string.startswith(obj, beg=0, end=len(string))</code>	检查字符串是否以obj开始, 如果是beg和end指定范围, 则检查范围内是否是以obj开始
<code>string.find(str, beg=0, end=len(string))</code>	检测str是否包含在string中, 如果指定beg和end, 则检查是否在范围内,如果是, 则返回索引值, 否则返回-1
<code>string.index(str, beg=0, end=len(string))</code>	跟find一样, 只不过不存在的话会报异常
<code>string.lower()</code>	将字符串中的所有的大写字母转换成小写
<code>string.upper()</code>	将字符串中的所有的小写字母转换成大写
<code>string.join(seq)</code>	以string作为分隔符, 将seq中所有的元素合并为一个新的字符串
<code>string.replace(str1, str2, num=string.count(str1))</code>	把string中的str1替换为str2, num指定最多替换的次数
<code>string.lstrip()</code>	删除string前面的空字符
<code>string.rstrip()</code>	删除string末尾的空字符
<code>string.strip([obj])</code>	默认删除string前后的空字符,obj可以指定删除的字符
<code>string.split(str="",num=string.count(str))</code>	以str为分隔符切片string, 如果num有指定值, 则只分割num个字符

具体看实例演示。

# 字符串转义

要在单引号的字符串中使用单引号，或者是双引号中使用双引号，那么就需要进行转义才能正常使用。

字符串中如果包含一些ASCII转义字符的时候，往往也会被python转义成ASCII转义字符。

转义字符	含义
\n	换行
\r	回车
\t	横向制表符 tab
\'	单引号
\"	双引号
\\	反斜杠

## 作业

- 假如需要提取网页源码中h1标签里面的内容，使用字符串的方式应该如何提取？请写出来。
- 如果要生成网站所有页面的sitemap地图，那么如何利用字符串格式化的方法来制作sitemap的模板然后生成对应的sitemap模块呢？
- 要判断关键词里面是否包含违禁词应该用什么方法？
- 如果要去除掉字符串里面的所有换行符应该怎么做？
- 根据模板批量生成关键词应该怎么做？自己设计自己行业的关键词模板