

2.6 字符串忽略大小写的搜索替换

问题

你需要以忽略大小写的方式搜索与替换文本字符串

解决方案

为了在文本操作时忽略大小写，你需要在使用 `re` 模块的时候给这些操作提供 `re.IGNORECASE` 标志参数。比如：

```
>>> text = 'UPPER PYTHON, lower python, Mixed Python'
>>> re.findall('python', text, flags=re.IGNORECASE)
['PYTHON', 'python', 'Python']
>>> re.sub('python', 'snake', text, flags=re.IGNORECASE)
'UPPER snake, lower snake, Mixed snake'
>>>
```

最后的那个例子揭示了一个小缺陷，替换字符串并不会自动跟被匹配字符串的大小写保持一致。为了修复这个，你可能需要一个辅助函数，就像下面的这样：

```
def matchcase(word):
    def replace(m):
        text = m.group()
        if text.isupper():
            return word.upper()
        elif text.islower():
            return word.lower()
        elif text[0].isupper():
            return word.capitalize()
        else:
            return word
    return replace
```

下面是使用上述函数的方法：

```
>>> re.sub('python', matchcase('snake'), text, flags=re.IGNORECASE)
'UPPER SNAKE, lower snake, Mixed Snake'
>>>
```

译者注：`matchcase('snake')` 返回了一个回调函数(参数必须是 `match` 对象)，前面一节提到过，`sub()` 函数除了接受替换字符串外，还能接受一个回调函数。

讨论

对于一般的忽略大小写的匹配操作，简单的传递一个 `re.IGNORECASE` 标志参数就已经足够了。但是需要注意的是，这个对于某些需要大小写转换的Unicode匹配可能还不够，参考2.10小节了解更多细节。