4.8 跳过可迭代对象的开始部分¶

问题¶

你想遍历一个可迭代对象，但是它开始的某些元素你并不感兴趣，想跳过它们。

解决方案¶

itertools 模块中有一些函数可以完成这个任务。 首先介绍的是 itertools.dropwhile() 函数。使用时，你给它传递一个函数对象和一个可迭代对象。 它会返回一个迭代器对象，丢弃原有序列中直到函数返回Flase之前的所有元素，然后返回后面所有元素。

为了演示，假定你在读取一个开始部分是几行注释的源文件。比如：

```
>>> with open('/etc/passwd') as f:
...     for line in f:
...         print(line, end='')
...
##
# User Database
#
# Note that this file is consulted directly only when the system is running
# in single-user mode. At other times, this information is provided by
# Open Directory.
...
##
nobody:*:-2:-2:Unprivileged User:/var/empty:/usr/bin/false
root:*:0:0:System Administrator:/var/root:/bin/sh
...
>>>
```

如果你想跳过开始部分的注释行的话，可以这样做：

```
>>> from itertools import dropwhile
>>> with open('/etc/passwd') as f:
...     for line in dropwhile(lambda line: not line.startswith('#'), f):
...         print(line, end='')
...
nobody:*:-2:-2:Unprivileged User:/var/empty:/usr/bin/false
root:*:0:0:System Administrator:/var/root:/bin/sh
...
>>>
```

这个例子是基于根据某个测试函数跳过开始的元素。 如果你已经明确知道了要跳过的元素的序号的话，那么可以使用 itertools.islice() 来代替。比如：

```
>>> from itertools import islice
>>> items = ['a', 'b', 'c', 1, 4, 10, 15]
>>> for x in islice(items, 3, None):
...     print(x)
...
4
10
15
>>>
```

在这个例子中， islice() 函数最后那个 None 参数指定了你要跳过前面3个元素，获取第4个到最后的所有元素， 如果 None 和3的位置对调，意思就是仅仅获取前三个元素恰恰相反， (这个跟切片的相反操作 [3:] 和 [:3] 原理是一样的)。

讨论¶

函数 dropwhile() 和 islice() 其实就是两个帮助函数，为的就是避免写出下面这种冗余代码：

```
with open('/etc/passwd') as f:
    # Skip over initial comments
    while True:
        line = next(f, '')
        if not line.startswith('#'):
            break

    # Process remaining lines
    while line:
        # Replace with useful processing
        print(line, end='')
        line = next(f, None)
```

跳过一个可迭代对象的开始部分跟通常的过滤是不同的。 比如，上述代码的第一个部分可能会这样重写：

```
with open('/etc/passwd') as f:
    lines = (line for line in f if not line.startswith('#'))
    for line in lines:
        print(line, end='')
```

这样写确实可以跳过开始部分的注释行，但是同样也会跳过文件中其他所有的注释行。 换句话讲，我们的解决方案是仅仅跳过开始部分满足测试条件的行，在那以后，所有的元素不再进行测试和过滤了。

最后需要着重强调的一点是，本节的方案适用于所有可迭代对象，包括那些事先不能确定大小的， 比如生成器，文件及其类似的对象。