

10.8 读取位于包中的数据文件¶

问题¶

你的包中包含代码需要去读取的数据文件。你需要尽可能地用最便捷的方式来做这件事。

解决方案¶

假设你的包中的文件组织成如下：

```
mypackage/  
    __init__.py  
    somedata.dat  
    spam.py
```

现在假设spam.py文件需要读取somedata.dat文件中的内容。你可以用以下代码来完成：

```
# spam.py  
import pkgutil  
data = pkgutil.get_data(__package__, 'somedata.dat')
```

由此产生的变量是包含该文件的原始内容的字节字符串。

讨论¶

要读取数据文件，你可能会倾向于编写使用内置的I/O功能的代码，如open()。但是这种方法也有一些问题。

首先，一个包对解释器的当前工作目录几乎没有控制权。因此，编程时任何I/O操作都必须使用绝对文件名。由于每个模块都包含有完整路径的__file__变量，这弄清楚它的路径不是不可能，但它很凌乱。

第二，包通常安装作为.zip或.egg文件，这些文件并不像在文件系统上的一个普通目录里那样被保存。因此，你试图用open()对一个包含数据文件的归档文件进行操作，它根本不会工作。

pkgutil.get_data()函数是一个读取数据文件的高级工具，不用管包是如何安装以及安装在哪。它只是工作并将文件内容以字节字符串返回给你

get_data()的第一个参数是包含包名的字符串。你可以直接使用包名，也可以使用特殊的变量，比如__package__。第二个参数是包内文件的相对名称。如果有必要，可以使用标准的Unix命名规范到不同的目录，只要最后的目录仍然位于包中。