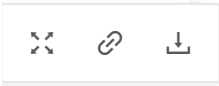
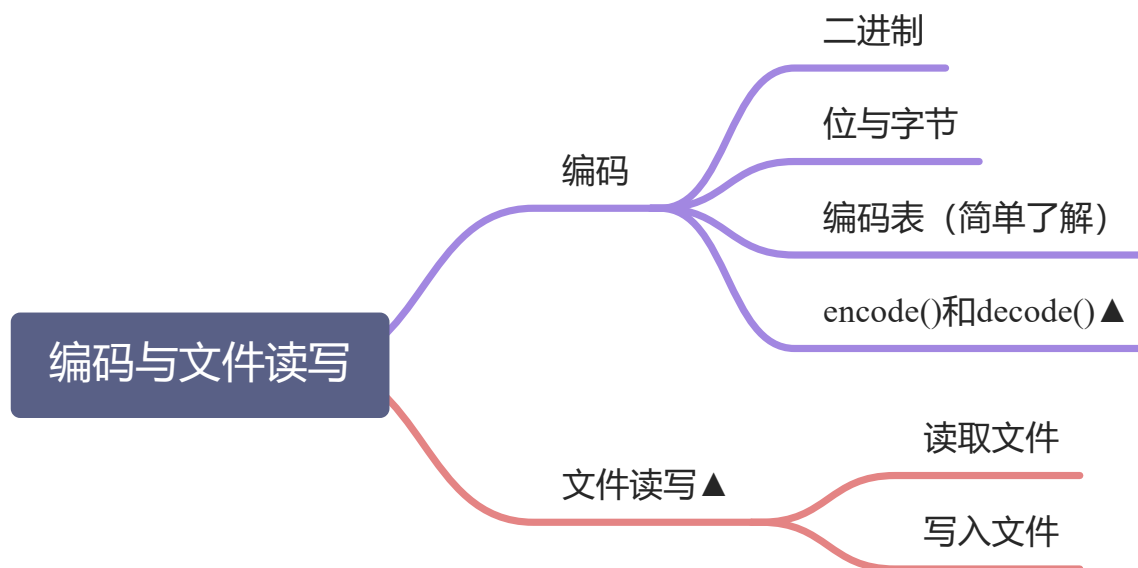


第15课 编码与文件读写

一、课程结构导图





注：▲为重点知识点。

二、知识点讲解

2.1 编码

概念：编码是用文字、符号或者数码来表示某种信息（数值、语言、操作指令、状态等）的过程，组合0和1数码就是二进制编码¹。与之相反，将数码还原成原始信息的过程就是解码。

2.1.1 二进制

概念：二进制使用0和1两个数字来表达信息，是计算机存储数据的方式。

2.1.2 位与字节

概念：位与字节是计算机内存中常用的数据存储单位。【位】也称【比特】(bit)是最小的存储单位，用以存放一个二进制的0或1。【字节】(byte)是最常用的存储单位，包含有8位。

2.1.3 编码表

ASCII: 美国国家标准学会（ANSI）制定的一种使用7个比特位来表示128个字符的编码方式，包含有大小写英文字母、数字、一些常见标点符号和一些特殊的计算机控制符号。

GBK: 为了解决ASCII编码对于中文没有支持的问题，我国制定了GBK编码，GBK编码采用了双字节编码，收录了21003个汉字。

Unicode: Unicode为世界上所有的字符都分配了唯一一个数字编码。现在Unicode可以容纳100多万个符号，所有语言都可以互通，一个网页上也可以显示多国语言。

UTF-8: UTF-8规定了Unicode编码表的二进制存储方式，简而言之，它是Unicode的一种实现方式。

编码表	适用性	特点
ASCII码	英文大小写,字符,不支持中文	美国人的发明,占用空间小
GB2312码、GBK码	支持了中文	GBK码是GB2312的升级
Unicode码	支持国际语言	占用空间大,适用性强。在ASCII码前面补8个就成了Unicode码
UTF-8码	支持国际语言	是Unicode的升级,两者可以非常容易地互相转化,占用空间小。ASCII码被UTF-8码包含

2.1.4 encode()和decode()

用法：在Python中，使用encode()函数和decode()函数来实现数据的编码与解码。

```
'你想编码的内容'.encode('你使用的编码表')  
  
'你想解码的内容'.decode('你使用的编码表')
```

示例:

```
1 print('吴枫'.encode('utf-8')) # 打印b'\xe5\x90\xb4\xe6\x9e\xab'  
2 print('吴枫'.encode('gbk'))   # 打印b'\xce\xe2\xb7\xe3'  
3 print(b'\xe5\x90\xb4\xe6\x9e\xab'.decode('utf-8')) # 打印吴枫  
4 print(b'\xce\xe2\xb7\xe3'.decode('gbk'))   # 打印吴枫  
5  
6 print(type('吴枫')) # 打印<class 'str'>  
7 print(type(b'\xce\xe2\xb7\xe3')) # 打印<class 'bytes'>
```

代码解析: 第1, 2行代码的打印结果和第3, 4行代码中的字母'b'表示是bytes(字节)数据, 可以使用type()函数来查看数据类型, 如第6, 7行代码。

2.2 文件读写

2.2.1 读取文件

完整代码

```
1 file1 = open(r'C:\Users\Ted\Desktop\test\abc.txt', 'r', encoding='utf-8') # 文件的绝对路径  
2 filecontent = file1.read() # 读取文件内容  
3 print(filecontent) # 打印读取的内容  
4 file1.close() # 关闭文件
```

- **open()函数打开文件**

用法：open函数最常用的参数有三个，文件路径、打开文件模式、编码格式。

第一个参数是文件的路径。文件的路径含有绝对路径和相对路径两种，绝对路径就是你电脑上的文件或目录在硬盘上真正的路径，相对路径是相对与某个基准目录的路径。

示例：

```
1 # 绝对路径
2 open('C:\\Users\\Ted\\Desktop\\test\\abc.txt') # 写法一：将'\''替换成'\\'
3 open(r'C:\Users\Ted\Desktop\test\abc.txt') # 写法二：在路径前加上字母r
4
5 # 相对路径
6 open('abc.txt') # 相对于test文件夹的相对路径
```

注意：windoes电脑的绝对路径需要在字符串前加上r，或者是将字符串里的'\''替换成'\\'。

第二个参数是打开文件的模式。'r'代表阅读模式，'w'是写入模式（写入会覆盖掉之前的内容），'a'是追加写入（不会覆盖掉之前的内容），'rb'和'wb'分别表示对二进制文件（视频、音频、图片等）的读取和写入。

示例：

open()函数				
		b(bytes,字节)	+	b +
r(read,读)	r只读，指针在开头 文件不存在则报错	rb二进制只读， 其余同左	r+读写， 其余同左	rb+二进制读写， 其余同左
w(write,写)	w只写，文件不存在 则新建，存在则覆盖	wb二进制只写， 其余同左	w+读写， 其余同左	wb+二进制读写， 其余同左
a(append,追加)	a追加，文件存在指 针放在末尾，文件 不存在则新建	ab二进制追加， 其余同左	a+追加且可读， 其余同左	ab+二进制追加， 且可读，其余同左

第三个参数是编码格式。表示的是返回的数据采用何种编码，一般采用utf-8编码或者gbk编码。

示例：

```
1 file1 = open(r'C:\Users\Ted\Desktop\test\abc.txt','r',encoding='utf-8')
```

注意：file1为open()函数调用后返回的文件对象，我们之后对于文件的操作都是基于对file1对象的操作。

• read()函数读取文件

用法：我们可以直接用open函数返回的对象.read()来读取文件。

示例：

```
1 filecontent = file1.read()
2 print(filecontent) # 打印读取的内容
```

• close()函数关闭文件

用法：对文件进行操作之后我们要记得使用.close()关闭文件，这样主要有三个目的：1. 释放程序所申请的系统资源。2. 保存文件。3. 增强对文件操作的安全性。

示例：

```
1 file1.close()
```

2.2.2 写入文件

完整代码

```
1 file2 = open(r'C:\Users\Ted\Desktop\test\abc.txt', 'w', encoding='utf-8')
2 file2.write('张无忌\n')
3 file2.write('宋青书\n') # \n表示换行
4 file2.close()
```

- **open()函数打开文件**

用法：

写入文件用的open()函数也需要三个参数，与读取文件不同，其中第二个参数需要写为'w'或者是'a'，'w'表示覆盖掉原内容写入，'a'表示在原文件的基础上追加写入。file2为调用open()函数返回的文件对象。

示例：

```
1 file2 = open(r'C:\Users\Ted\Desktop\test\abc.txt', 'w', encoding='utf-8')
```

- **write()函数打开文件**

用法：我们需要用到open函数返回的对象.write()来往文件中写入内容。

```
1 file2.write('张无忌\n')
2 file2.write('宋青书\n') # \n表示换行
```

- **close()函数关闭文件**

用法：此处同读取文件中的close()函数。

示例：

```
1 file2.close()
```

- **with关键字**

用法：在with关键字之后使用open()函数来读写文件，可以不用.close()关闭文件。

示例：

```
1 # 普通写法
2 file1 = open('abc.txt', 'a')
3 file1.write('张无忌')
4 file1.close()
5
6 # 使用with关键字的写法
7 with open('abc.txt', 'a') as file1: # 格式: with open('文件地址', '读写模式') as 变量名:
8     file1.write('张无忌') # 格式: 对文件的操作要缩进, 无需close()关闭文件
```

三、巩固练习

1. (实操题) 请把你的名字按照'utf-8'的编码方式来编码, 并在终端中打印结果。
2. (单项选择) 小明有一个文本文件, 命名为robots.txt, 文本中原有一行内容, 现在小明想在其后继续追加写入一些东西, 请问下列选项目中哪一种写法是正确 () ?

A、

```
1 file = open(r'C:\user\robots.txt', 'w', encoding='utf-8')
2 file.write('the second line...\n')
3 file.close()
```

B、

```
1 file = open('C:\user\robots.txt', 'a', encoding='utf-8')
2 file.write('the second line...\n')
3 file.close()
```

C、

```
1 file = open(r'C:\user\robots.txt', 'a', encoding='utf-8')
2 file.write('the second line...\n')
3 file.close()
```

D、

```
1 file = open(r'C:\user\robots.txt', 'r', encoding='utf-8')
2 file.write('the second line...\n')
3 file.close()
```

参考文献:

1. 钱晓捷著, 《计算机系统: 基础概念及编程实践》, 第33页。

