day16：二进制文件，编码规则

笔记本： Python基础
创建时间： 2018/6/27 10:35 **更新时间：** 2018/7/2 19:39
作者： liuchang_0412@163.com

1.bytes类型：以字节为单位存储数据，每个字节的值是0~255
①bytes常量的表示方法：
b'' 空字节串
b'ABCD' 有四个字节的字节串
b'\x41\x42' 有两个字节的字符串
②bytes运算：
+，+=，*，*=
<，<=，>，>=，==，!=
in，not in
③bytes相关函数
len(x) 求字节个数
max(x) 求最大值，返回十进制数
min(x) 求最大值，返回十进制数
sum(x) 求和，返回十进制数
any(x)
all(x)

2.创建字节串(bytes)的函数：
bytes( ) #创建空的字节串
bytes(整数)
bytes(整型可迭代对象)
bytes(字符串, encoding='utf-8')
#字节串可以看做是序列
#字节串是不可变的
>>>help(bytes)

3.bytes和str转换
str ---> bytes
str.encode(encoding='utf-8') 方法来转换
例：
b="英文abc".encoding('utf-8')

bytes ---> str
B.decode(encoding='utf-8')
#练习：写一个程序，从键盘输入一段字符串，将其字符串转换为字节串后，计算长度并打印此字节串，然后将此字节串再转换为字符串，比较此字符串是否与原来输入的字符串相同。

4.二进制文件的读写：

①什么是二进制文件：文件以字节（byte）为单位存储，不以换行符(\n)为单位分隔内容的文件。

②二进制文件操作的方法：

F=open(filename, 'rb'('wb'))

F.read(size=-1)　#从一个文件流中最多读取size个字符

F.write(字符串/字节串)写一些数据到文件流中，返回写入的字节数（字符数）

F.tell( )　返回当前文件流的绝对位置

F.seek(offset, whence=0)　改变数据流的位置，返回新的绝对位置

F.readable( )　判断这个文件是否可读，可读返回True

F.writable( )　判断这个文件是否可写，可写返回True

F.read( )函数返回类型：

文本文件，返回字符串

二进制文件，返回字节串（字节序列）

F.write(x)函数：

见write_file.py

F.seek(偏移量，相对位置)

偏移量：

大于0的数代表向文件尾方向移动

小于0的数代表向文件头方向移动

相对位置：

0 代表从文件头开始偏移

1 代表从前位置开始偏移

2 代表从文件尾开始偏移

示例：

F.seek(10,0)　#从头开始向后10个字节

F.seek(5,1)　#从当前位置向后5个字节

F.seek(-10,2)　#从末尾开始向前2个字节

5.bytearray类型：

#bytes类型为不可变类型

#bytearray类型为可变的数据类型

①bytearray运算：

+，+=，*，*=

<，<=，>，>=，==，!=

in，not in

②bytearray的方法：

B.clear( )　清空

B.append(n)　追加一个字节（n为0~255的整数）

B.remove(value)　删除第一次出现的字节，如果没有出现，则产生ValueError错误

B.reverse( )　字节的顺序进行反转

B.decode(encoding = 'utf-8')
B.find(sub [, start [, end]])　返回索引，查找失败返回-1

③创建bytearray的方法：
bytearray( )　#创建空的bytearray
bytearray(整数)
bytearray(整型可迭代对象)
bytearray(字符串, encoding='utf-8')
#练习：有一个bytearray字节序列：ba=bytearray(b'a1b2c3d4'):
1.得到字符串'1234'和'字符串'abcd'。
2.将上述bytearray改为(b'A1B2C3D4')

6.标准输入输出文件：
sys.stdin
sys.stdout
sys.stderr
模块sys

Linux 下 Ctrl+D 输入文件结束符

7.汉字编码（只讲两种）：
GB18030(GBK(GB2312))　#国标
UNICODE <-> UTF-8　　#

①GB2312-80编码：
1980年发布
用两个字节进行编码，编码范围（A1A1~FEFE）
包含汉字 6763个和682个其他字符
②GBK：
1995年发布
用两个字节进行编码，编码范围（9140~FEFE）（剔除XX7F）
收录文字21003个
③GB18030-2005编码
2005年发布
用两字节或四字节进行编码
收录了27533个汉字
④UNICODE-16(两字节)
0x0000 -- 0xFFFF
ASCII(0-127) --> (0x0000-0x007F)
⑤UNICODE-32(四字节)
0x00000000 -- 0xFFFFFFFF
ASCII(0-127) --> (0x00000000-0x0000007F)
⑥UTF-8(8bit Unicode Transformation Font)

UNICODE <<------>> UTF-8 可以互转

互转规则：

0x0000~0x007F　一字节

0x0080~0x07FF　二字节

0x8000~0xFFFF　三字节（中文落在此区域）

.....

#练习：在windows上用记事本编写一段中文，在Linux下能读出中文的内容