1. 16进制字符流D981DF92

(1)采用UTF16编码, 包含1个字符.

UTF16编码若小于0xD800或大于0xDFFF, 则可使用2字节在BMP平面中寻找对应字符; 否则需要再读2字节, 使用4个字节进行相应处理后在ucs寻找对应字符.

(2)采用UTF8编码, 包含2个字符.

0xD981DF92转换成2进制表示为11011001 10000001 11011111 10010010.

第1个字节高3位是110, 第2个字节高2位是10, 这两个字节在UTF8中代表一个字符; 第3个和第4个字节同样的方法判断, 也在UTF8中代表一个字符.

2. UTF8如何判断用几个字节编码?

读取一个字节, 若最高位是0, 则是使用1个字节进行编码; 若最高3位是110, 则是使用2字节进行编码; 若最高4位是1110, 则是使用3字节进行编码; 若最高5位是11110, 则是使用4字节进行编码.

3. (1) UTF16是可变长编码, 使用2字节或者4字节进行编码.

(2) UTF16不能表示UCS中的所有字符, UTF16最多能表示UCS中0x00组0x10平面的字符, 即UTF16只能表示UCS中17个平面内的字符, 无法表示UCS所有的字符.