

# 393. UTF-8 Validation

题目描述: <https://leetcode.com/problems/utf-8-validation/>

给定数组, 判断是否为合法的UTF-8串。  
规则如下:

Char. number range (hexadecimal)	UTF-8 octet sequence (binary)
0000 0000-0000 007F	0xxxxxxx
0000 0080-0000 07FF	110xxxxx 10xxxxxx
0000 0800-0000 FFFF	1110xxxx 10xxxxxx 10xxxxxx
0001 0000-0010 FFFF	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

解题思路:

按照规则

代码1:

```

class Solution {
public:
    bool validUtf8(vector<int>& data) {
        int len = data.size();
        for(int i = 0; i < len;) {
            int tag = data[i];
            int num = 0;
            if(tag >= 0 && tag <= 0x7f) {
                num = 0;
            }
            else if(tag >= 0xc0 && tag <= 0xdf) {
                num = 1;
            }
            else if(tag >= 0xe0 && tag <= 0xef) {
                num = 2;
            }
            else if(tag >= 0xf0 && tag <= 0xf7) {
                num = 3;
            }
            else {
                return false;
            }
            if(len - i - 1 < num) {
                return false;
            }
            cout << num << endl;
            for(int j = i+1; j - i <= num; j++) {
                if((data[j] & 0xc0) != 0x80) {
                    return false;
                }
            }
            i = i + num + 1;
        }
        return true;
    }
};

```

代码2:

```
class Solution {
public:
    bool validUtf8(vector<int>& data) {
        int len = data.size();
        int count = 0;
        for(int i = 0; i < len; i++) {
            unsigned int k = data[i];
            if(count == 0) {
                if((k >> 7) == 0) count = 0;
                else if((k >> 5) == 0x6) count = 1;
                else if((k >> 4) == 0xe) count = 2;
                else if((k >> 3) == 0x1e) count = 3;
                else {
                    return false;
                }
            }
            else {
                if((k >> 6) != 0x2) return false;
                count--;
            }
        }
        return count == 0;
    }
};
```