Name: 陳柏瑜

Student ID: 41047054S

Department: Computer Science & Information Engineering 114

HW0106 wchart

Question

In this class, I said that I prefer to keep char as a character type. As you know, it is impossible to encode all Chinese words in char. Actually, in C standard library, there is a type called wchar_t, which implies Wide Character. It was first defined in C90 as an integral type whose range of values can represent distinct codes for all members of the largest extended character set specified among the supported locales.

It sounds great, right? However, when considering portability, you should not use wchar_t. Why? Please write down your reason and provide some references.

Answer

First of all, we see how SPEC says about wchar_t.

wchar_t, which is an integer type whose range of values can represent distinct codes for all members of the largest extended character set specified among the supported locales; the null character shall have the code value zero. Each member of the basic character set shall have a code value equal to its value when used as the lone character in an integer character constant if an implementation does not define.

Hence we know that wchar_t is a type that is larger than type char

and is used for storing (Unicode) characters not included in ASCII, especially Chinese and Arabic. However, the type wchar_t must be compatible with type char, namely ASCII.

The size of wchar_t is 2 for windows but 4 for the other, which means the size may vary from different platform. That is the reason why you should not use wchar_t when considering portability.

Reference

- ITHome: https://www.ithome.com.tw/voice/135711
- SPEC