**VC中Unicode字符串的处理**

* **在windows上编程，字符串使用： TCHAR 、\_tcscpy(), \_tcslen()、\_stscanf() 、\_T(“string”) 、\_tfopen()**

    1. 如何取得一个既包含单字节字符又包含双字节字符的字符串的字符个数？    
    可以调用Microsoft     Visual     C++的运行期库包含函数\_mbslen来操作多字节（既包括单字节也包括双字节）字符串。    
    调用strlen函数，无法真正了解字符串中究竟有多少字符，它只能告诉你到达结尾的0之前有多少个字节。    
    2. 如何对DBCS（双字节字符集）字符串进行操作？    
    函数 描述    
    PTSTR     CharNext     （     LPCTSTR     ）; 返回字符串中下一个字符的地址    
    PTSTR     CharPrev     （     LPCTSTR,     LPCTSTR     ）； 返回字符串中上一个字符的地址    
    BOOL     IsDBCSLeadByte(     BYTE     )； 如果该字节是DBCS字符的第一个字节，则返回非0值    
    3. 为什么要使用Unicode？    
    （1） 可以很容易地在不同语言之间进行数据交换。    
    （2） 使你能够分配支持所有语言的单个二进制.exe文件或DLL文件。    
    （3） 提高应用程序的运行效率。    
    Windows     2000是使用Unicode从头进行开发的，如果调用任何一个Windows函数并给它传递一个ANSI字符串，那么系统首先要将字符串转换成Unicode，然后将Unicode字符串传递给操作系统。如果希望函数返回ANSI字符串，系统就会首先将Unicode字符串转换成ANSI字符串，然后将结果返回给你的应用程序。进行这些字符串的转换需要占用系统的时间和内存。通过从头开始用Unicode来开发应用程序，就能够使你的应用程序更加有效地运行。    
    Windows     CE     本身就是使用Unicode的一种操作系统，完全不支持ANSI     Windows函数    
    Windows     98     只支持ANSI，只能为ANSI开发应用程序。    
    Microsoft公司将COM从16位Windows转换成Win32时，公司决定需要字符串的所有COM接口方法都只能接受Unicode字符串。    
    4. 如何编写Unicode源代码？    
    Microsoft公司为Unicode设计了WindowsAPI，这样，可以尽量减少代码的影响。实际上，可以编写单个源代码文件，以便使用或者不使用Unicode来对它进行编译。只需要定义两个宏（UNICODE和\_UNICODE），就可以修改然后重新编译该源文件。    
    \_UNICODE宏用于C运行期头文件，而UNICODE宏则用于Windows头文件。当编译源代码模块时，通常必须同时定义这两个宏。    
    5. Windows定义的Unicode数据类型有哪些？    
    数据类型 说明    
    WCHAR Unicode字符    
    PWSTR 指向Unicode字符串的指针    
    PCWSTR 指向一个恒定的Unicode字符串的指针    
    对应的ANSI数据类型为CHAR，LPSTR和LPCSTR。    
    ANSI/Unicode通用数据类型为TCHAR，PTSTR,LPCTSTR。    
    6. 如何对Unicode进行操作？    
    字符集 特性 实例    
    ANSI 操作函数以str开头 strcpy    
    Unicode 操作函数以wcs开头 wcscpy    
    MBCS 操作函数以\_mbs开头 \_mbscpy    
    ANSI/Unicode 操作函数以\_tcs开头 \_tcscpy（C运行期库）    
    ANSI/Unicode 操作函数以lstr开头 lstrcpy（Windows函数）    
    所有新的和未过时的函数在Windows2000中都同时拥有ANSI和Unicode两个版本。ANSI版本函数结尾以A表示；Unicode版本函数结尾以W表示。Windows会如下定义：    
    #ifdef       UNICODE    
    #define       CreateWindowEx       CreateWindowExW    
    #else    
    #define       CreateWindowEx       CreateWindowExA    
    #endif         //     !UNICODE    
    7. 如何表示Unicode字符串常量？    
    字符集 实例    
    ANSI “string”    
    Unicode L“string”    
    ANSI/Unicode T(“string”)或\_TEXT(“string”)if(szError[0] ==   \_TEXT(‘J’)){     }    
    8. 为什么应当尽量使用操作系统函数？    
    这将有助于稍稍提高应用程序的运行性能，因为操作系统字符串函数常常被大型应用程序比如操作系统的外壳进程Explorer.exe所使用。由于这些函数使用得很多，因此，在应用程序运行时，它们可能已经被装入RAM。    
    如：StrCat，StrChr，StrCmp和StrCpy等。    
    9. 如何编写符合ANSI和Unicode的应用程序？    
    （1） 将文本串视为字符数组，而不是chars数组或字节数组。    
    （2） 将通用数据类型（如TCHAR和PTSTR）用于文本字符和字符串。    
    （3） 将显式数据类型（如BYTE和PBYTE）用于字节、字节指针和数据缓存。    
    （4） 将TEXT宏用于原义字符和字符串。    
    （5） 执行全局性替换（例如用PTSTR替换PSTR）。    
    （6） 修改字符串运算问题。例如函数通常希望在字符中传递一个缓存的大小，而不是字节。这意味着不应该传递sizeof(szBuffer),而应该传递（sizeof(szBuffer)/sizeof(TCHAR)。另外，如果需要为字符串分配一个内存块，并且拥有该字符串中的字符数目，那么请记住要按字节来分配内存。这就是说，应该调用malloc(nCharacters     \*sizeof(TCHAR)),而不是调用malloc(nCharacters)。    
    10. 如何对字符串进行有选择的比较？    
    通过调用CompareString来实现。    
    标志 含义    
    NORM\_IGNORECASE 忽略字母的大小写    
    NORM\_IGNOREKANATYPE 不区分平假名与片假名字符    
    NORM\_IGNORENONSPACE 忽略无间隔字符    
    NORM\_IGNORESYMBOLS 忽略符号    
    NORM\_IGNOREWIDTH 不区分单字节字符与作为双字节字符的同一个字符    
    SORT\_STRINGSORT 将标点符号作为普通符号来处理    
    11. 如何判断一个文本文件是ANSI还是Unicode？    
    判断如果文本文件的开头两个字节是0xFF和0xFE，那么就是Unicode，否则是ANSI。    
    12. 如何判断一段字符串是ANSI还是Unicode？    
    用IsTextUnicode进行判断。IsTextUnicode使用一系列统计方法和定性方法，以便猜测缓存的内容。由于这不是一种确切的科学方法，因此     IsTextUnicode有可能返回不正确的结果。    
    13. 如何在Unicode与ANSI之间转换字符串？    
    Windows函数MultiByteToWideChar用于将多字节字符串转换成宽字符串；函数WideCharToMultiByte将宽字符串转换成等价的多字节字符串。