**在旧式编码与 Unicode 之间转换（C# 编程指南）**

在 C# 中，内存中的所有字符串都是按 Unicode (UTF-16) 编码的。将数据从存储器移动到 **string** 对象中后，数据将自动转换为 UTF-16。如果数据仅包含从 0 到 127 的 ASCII 值，则此转换无需您执行任何额外的工作。但若源文本包含扩展的 ASCII 字节值（128 到 255），则默认情况下，将根据当前代码页解释扩展字符。若要指定应该根据其他某个代码页解释源文本，请使用 [System.Text.Encoding](https://msdn.microsoft.com/zh-cn/library/system.text.encoding.aspx) 类，如下面的示例所示。

下面的示例演示如何转换按 8 位 ASCII 编码的文本文件，此转换根据 Windows 代码页 737 解释源文本。

class ANSIToUnicode

{

static void Main()

{

// Create a file that contains the Greek work ψυχή (psyche) when interpreted by using

// code page 737 ((DOS) Greek). You can also create the file by using Character Map

// to paste the characters into Microsoft Word and then "Save As" by using the DOS

// (Greek) encoding. (Word will actually create a six-byte file by appending "\r\n" at the end.)

System.IO.File.WriteAllBytes(@"greek.txt", new byte[] { 0xAF, 0xAC, 0xAE, 0x9E });

// Specify the code page to correctly interpret byte values

Encoding encoding = Encoding.GetEncoding(737); //(DOS) Greek code page

byte[] codePageValues = System.IO.File.ReadAllBytes(@"greek.txt");

// Same content is now encoded as UTF-16

string unicodeValues = encoding.GetString(codePageValues);

// Show that the text content is still intact in Unicode string

// (Add a reference to System.Windows.Forms.dll)

System.Windows.Forms.MessageBox.Show(unicodeValues);

// Same content "ψυχή" is stored as UTF-8

System.IO.File.WriteAllText(@"greek\_unicode.txt", unicodeValues);

// Conversion is complete. Show the bytes to prove the conversion.

Console.WriteLine("8-bit encoding byte values:");

foreach(byte b in codePageValues)

Console.Write("{0:X}-", b);

Console.WriteLine();

Console.WriteLine("Unicode values:");

string unicodeString = System.IO.File.ReadAllText("greek\_unicode.txt");

System.Globalization.TextElementEnumerator enumerator =

System.Globalization.StringInfo.GetTextElementEnumerator(unicodeString);

while(enumerator.MoveNext())

{

string s = enumerator.GetTextElement();

int i = Char.ConvertToUtf32(s, 0);

Console.Write("{0:X}-", i);

}

Console.WriteLine();

// Keep the console window open in debug mode.

Console.Write("Press any key to exit.");

Console.ReadKey();

}

/\*

\* Output:

8-bit encoding byte values:

AF-AC-AE-9E

Unicode values:

3C8-3C5-3C7-3B7

\*/

}