**将 RTF 转换为纯文本（C# 编程指南）**

RTF 格式是 Microsoft 于 20 世纪 80 年代后期开发的一种文件格式，旨在实现不同操作系统之间的文档交换。Microsoft Word 和 WordPad 都可以读写 RTF 文件。在 .NET Framework 中，可以使用 [RichTextBox](https://msdn.microsoft.com/zh-cn/library/system.windows.forms.richtextbox.aspx) 控件创建支持 RTF 且支持用户以 WYSIWIG 方式将格式应用于文本的字处理器。

也可以使用 [RichTextBox](https://msdn.microsoft.com/zh-cn/library/system.windows.forms.richtextbox.aspx) 以编程方式将 RTF 格式代码从文档中移除，从而将该文档转换为纯文本。执行这种类型的操作无需在 Windows 窗体中嵌入该控件。

**在项目中使用 RichTextBox 控件**

1. 添加对 System.Windows.Forms.dll 的引用。
2. 为 **System.Windows.Forms** 命名空间添加 using 指令（可选）。

下面的示例将示例 RTF 文件转换为纯文本。文件包含 RTF 格式设置 (如字体信息)，四个 Unicode 字符和四个扩展的 ASCII 字符。代码示例打开文件，将其内容设置为 [RichTextBox](https://msdn.microsoft.com/zh-cn/library/system.windows.forms.richtextbox.aspx)，当 RTF，检索该内容作为文本，并在 [MessageBox](https://msdn.microsoft.com/zh-cn/library/system.windows.forms.messagebox.aspx)的文本和输出该文本到一个 UTF-8 格式的文件。

**MessageBox** 和输出文件包含以下文本：

The Greek word for "psyche" is spelled ψυχή. The Greek letters are encoded in Unicode.

These characters are from the extended ASCII character set (Windows code page 1252): âäӑå

// Use NotePad to save the following RTF code to a text file in the same folder as

// your .exe file for this project. Name the file test.rtf.

/\*

{\rtf1\ansi\ansicpg1252\deff0\deflang1033{\fonttbl{\f0\fswiss\fcharset0 Arial;}

{\f1\fnil\fprq1\fcharset0 Courier New;}{\f2\fswiss\fprq2\fcharset0 Arial;}}

{\colortbl ;\red0\green128\blue0;\red0\green0\blue0;}

{\\*\generator Msftedit 5.41.21.2508;}

\viewkind4\uc1\pard\f0\fs20 The \i Greek \i0 word for "psyche" is spelled \cf1\f1\u968?\u965?\u967?\u942?\cf2\f2 . The Greek letters are encoded in Unicode.\par

These characters are from the extended \b ASCII \b0 character set (Windows code page 1252): \'e2\'e4\u1233?\'e5\cf0\par }

\*/

class ConvertFromRTF

{

static void Main()

{

// If your RTF file isn't in the same folder as the .exe file for the project,

// specify the path to the file in the following assignment statement.

string path = @"test.rtf";

//Create the RichTextBox. (Requires a reference to System.Windows.Forms.)

System.Windows.Forms.RichTextBox rtBox = new System.Windows.Forms.RichTextBox();

// Get the contents of the RTF file. When the contents of the file are

// stored in the string (rtfText), the contents are encoded as UTF-16\.

string rtfText = System.IO.File.ReadAllText(path);

// Display the RTF text. This should look like the contents of your file.

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

// Use the RichTextBox to convert the RTF code to plain text.

rtBox.Rtf = rtfText;

string plainText = rtBox.Text;

// Display the plain text in a MessageBox because the console can't

// display the Greek letters. You should see the following result:

// The Greek word for "psyche" is spelled ψυχή. The Greek letters are

// encoded in Unicode.

// These characters are from the extended ASCII character set (Windows

// code page 1252): âäӑå

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

// Output the plain text to a file, encoded as UTF-8\.

System.IO.File.WriteAllText(@"output.txt", plainText);

}

}

RTF 字符以八位编码。但是，除了从指定的代码页，扩展 ASCII 字符之外用户可以指定 Unicode 字符。由于 [RichTextBox.Text](https://msdn.microsoft.com/zh-cn/library/system.windows.forms.richtextbox.text.aspx) 属性为 [string](https://msdn.microsoft.com/zh-cn/library/362314fe.aspx) 类型，因此字符按 Unicode UTF-16 进行编码。源 RTF 文档中的所有扩展的 ASCII 字符和 Unicode 字符都在文本输出中进行了正确的编码。

如果使用 [File.WriteAllText](https://msdn.microsoft.com/zh-cn/library/system.io.file.writealltext.aspx) 方法将文本写入磁盘，则文本将按 UTF-8（没有字节顺序标记）进行编码。