

How can I get the URL to the Web page the clipboard was copied from?

 devblogs.microsoft.com/oldnewthing/20140721-00

July 21, 2014



Raymond Chen

When you copy content from a Web page to the clipboard and then paste it into OneNote, OneNote pastes the content but also annotates it “Pasted from ...”. How does OneNote know where the content was copied from?

As noted in [the documentation for the HTML clipboard format](#), Web browsers can provide an optional `SourceURL` property to specify the Web page the HTML was copied from.

Let’s write a Little Program that mimics what OneNote does, but just in plain text, because I don’t want to try to parse HTML. This is much easier to do in C#, because the BCL provides most of the helper functions.

```
using System;
using System.IO;
using System.Windows;
class Program {
    [STAThread]
    public static void Main() {
        System.Console.WriteLine(Clipboard.GetText());
        using (var sr = new StringReader(
            Clipboard.GetText(TextDataFormat.Html))) {
            string s;
            while ((s = sr.ReadLine()) != null) {
                if (s.StartsWith("SourceURL:")) {
                    System.Console.WriteLine("Copied from {0}", s.Substring(10));
                    break;
                }
            }
        }
    }
}
```

First, we get the text from the clipboard and print it. That’s the easy part.

Next, we get the HTML text from the clipboard. This is a bunch of text in a [particular format](#). We look for an entry that specifies the `SourceURL` ; if we find it, then we print the URL.

This code is rather sloppy. For example, if the HTML itself contains the string `SourceURL:haha-fakeout`, we risk misdetecting it as the source. To do this properly, we would have to verify that the string appears in the header area of the HTML (before the first `StartFragment`).

But this is a Little Program, so I can skip all that stuff.

Here's a sketch of the equivalent C/C++ version:

```
int __cdecl main(int, char **)
{
    if (OpenClipboard(NULL)) {
        // Obtain the Unicode text and print it
        HANDLE h = GetClipboardData(CF_UNICODETEXT);
        if (h) {
            PCWSTR pszPlainText = GlobalLock(h);
            ... print pszPlainText ...
            GlobalUnlock(h);
        }
        // Obtain the HTML text and extract the SourceURL
        h = GetClipboardData(RegisterClipboardFormat(TEXT("HTML Format")));
        if (h) {
            PCSTR pszHtmlFormat = GlobalLock(h);
            ... break pszHtmlFormat into lines ...
            ... look for a line that begins with "SourceURL:" ...
            ... if found, print it ...
            GlobalUnlock(h);
        }
        CloseClipboard();
    }
    return 0;
}
```

[Raymond Chen](#)

Follow

