Subject: [solved] TheIDE "Run Options..." bug
Posted by cbpporter on Mon, 15 Sep 2014 14:01:40 GMT
View Forum Message <> Reply to Message

I spent the better part of the day trying to find out why when selecting Standard output: File from TheIDE with an application that sets UTF8 (UTF16 does not work for me in the console) as a console output I got gibberish.

At first I blamed Windows API, but ultimately I discovered that TheIDE introduces those extra characters.

If the code that invokes the .exe is able to detect that the output is UTF8 it should interpret the data as such. If this is not possible, we need a new option Standard Output: UTF8 File.

Subject: Re: TheIDE "Run Options..." bug
Posted by mirek on Mon, 15 Sep 2014 17:08:16 GMT
View Forum Message <> Reply to Message

After further investigation, it is unfortunately much more complicated than that.

The primary source of problems is Cout(), which puts data into output stream without conversion. It should either convert to local 8-bit code (e.g. Win-1252) or, much better, use WriteConsoleW.

Then LocalProcess should probably rather use W API as well...

Mirek

Subject: Re: TheIDE "Run Options..." bug
Posted by cbpporter on Mon, 15 Sep 2014 17:16:40 GMT
View Forum Message <> Reply to Message

During my day of testing, I found that it is near impossible to write Utf16 to the console and maintain the ability to reroute the output.

Also what I found on Google suggests that the only way it works (and it does; I tested) is to set Utf8 console output and write raw Utf8 without BOM.

My API is almost identical to Cout().

So while Cout may be changed, just adding an option for TheIDE to not convert that output to whatever encoding it uses should help.

Currently, I can reproduce with these test cases:
1. Cout() like stream, Utf8 console, Utf8 write, Run from TheIDE on cmd.exe. Works.
2. Cout() like stream, Utf8 console, Utf8 write, Run from cmd.exe with redirection (i.e. text.exe >

out.txt). Works.
3. Cout() like stream, Utf8 console, Utf8 write, Run from with "File" option. Does not work.

AFAIK WriteConsole does not redirect. It did not for me.

Subject: Re: TheIDE "Run Options..." bug
Posted by mirek on Mon, 15 Sep 2014 17:48:16 GMT
View Forum Message <> Reply to Message

cbpporter wrote on Mon, 15 September 2014 19:16During my day of testing, I found that it is near impossible to write Utf16 to the console and maintain the ability to reroute the output.

Yeah, it is a mess...

Anyway, I have tried to solve the issue converting to (in Cout) and from (in LocalProcess) OEM encoding, which is default 8-bit encoding for console. Seems to work for me.

Mirek

Subject: Re: TheIDE "Run Options..." bug
Posted by cbpporter on Wed, 17 Sep 2014 13:12:04 GMT
View Forum Message <> Reply to Message

Tested with latest nightly, does not work.

I'm writing 6 charters to the console, two of them being <128 and I get an output file of 29 bytes which interpreted as UTF8 shows 16 glyph.

TheIDE is still doing some conversion on the output.

As a test try to write 0xC0 in the console using Utf8 and you should get À (A with an accent) both in the opened up console and when you are redirecting to a file.

PS: The latest nightly really can't debug at all. Everything is 0.

Subject: Re: TheIDE "Run Options..." bug
Posted by mirek on Thu, 18 Sep 2014 05:47:03 GMT
View Forum Message <> Reply to Message

cbpporter wrote on Wed, 17 September 2014 15:12Tested with latest nightly, does not work.

I'm writing 6 charters to the console, two of them being <128 and I get an output file of 29 bytes

which interpreted as UTF8 shows 16 glyph.

TheIDE is still doing some conversion on the output.

As a test try to write 0xC0 in the console using Utf8 and you should get À (A with an accent) both in the opened up console and when you are redirecting to a file.

Can you show me the code please? I know it sounds trivial, but there are more interpretations to this...

Quote:
PS: The latest nightly really can't debug at all. Everything is 0.

Well, it works for me most of time, but I have to admit sometimes something weird happens here too. But never for long enough to actually debug it.

Perhaps you can join (and test) in this thread:

http://www.ultimatepp.org/forums/index.php?t=msg&th=9043 &start=0&

I believe we are very close to catching the problem now... (really, the source of problem is that Win32 symbol API is not well documented).

Mirek

Subject: Re: TheIDE "Run Options..." bug
Posted by cbpporter on Thu, 18 Sep 2014 06:18:58 GMT
View Forum Message <> Reply to Message

EDIT: Wrong thread.

Subject: Re: TheIDE "Run Options..." bug
Posted by cbpporter on Fri, 19 Sep 2014 09:02:52 GMT
View Forum Message <> Reply to Message

[quote title=mirek wrote on Thu, 18 September 2014 08:47]
Can you show me the code please? I know it sounds trivial, but there are more interpretations to this...

[quote title=Quote:]
That's not that easy since technically I am not written C++ code :).

Anyway, here is a bare bones WinAPI sequence that prints thing as expected:

```
SetConsoleOutputCP(65001);

uint8 s[] = { 0xC3, 0x80 };
uint8* h = GetStdHandle(4294967285u);
uint32 dummy = 0u;
WriteFile(h, s, 2, &(dummy), 0);
```

The console is set to UTF8. The À (0xC0) codepoint is encoded with two Utf8 code units 0xC3, 0x80. Now, if the console is set to use a non-bitmap font, it will work very well with a wide range of characters. If the character is not supported, one or more empty little rectangles are rendered. The good news is that you can copy&paste them and the correct information is preserved.

So technically if the String is Utf8, it should work. It works in my code and even with Utf8 U++ Strings prior to your changes. But TheIDE/LocalProcess messes up these values.

I did not manage to encode À as 0x00C0 (or even 0xC000) in Utf16 and send it to the console. Google seems to suggest that the console is inherently 8bit.

Subject: Re: TheIDE "Run Options..." bug
Posted by mirek on Fri, 19 Sep 2014 11:47:40 GMT
View Forum Message <> Reply to Message

What a mess...

The problem is that you change encoding in console of your app, but unfortunately there seems to be no way how to tell what code-page is program using for its output, which is something we clearly need in LocalProcess (and ide console capture).

That is why my change converts all Cout data into _default_ console codepage and expects the same in LocalProcess.

Mirek

Subject: Re: TheIDE "Run Options..." bug
Posted by cbpporter on Fri, 19 Sep 2014 12:33:37 GMT
View Forum Message <> Reply to Message

Maybe there should be an option at least to have no conversion, taking the data as is and giving the responsibility of interpreting the data to the caller?

Subject: Re: TheIDE "Run Options..." bug

Posted by mirek on Sat, 20 Sep 2014 07:33:23 GMT
View Forum Message <> Reply to Message

cbpporter wrote on Fri, 19 September 2014 14:33Maybe there should be an option at least to have no conversion, taking the data as is and giving the responsibility of interpreting the data to the caller?

Perhaps... but that will have to be on both sides.

I guess:

- ability to set Cout encoding (defaults to default console code page)
- ability to set LocalProcess encoding (again defaults to default console code page)

I guess that would do the trick, right?

Mirek

Subject: Re: TheIDE "Run Options..." bug
Posted by cbpporter on Sat, 20 Sep 2014 07:50:40 GMT
View Forum Message <> Reply to Message

And ability to just handle data without any conversions? Would not work for WString...

Subject: Re: TheIDE "Run Options..." bug
Posted by mirek on Sat, 20 Sep 2014 16:05:55 GMT
View Forum Message <> Reply to Message

cbpporter wrote on Sat, 20 September 2014 09:50And ability to just handle data without any conversions? Would not work for WString...

That equals to setting both encodings to utf-8, does not it?

Subject: Re: TheIDE "Run Options..." bug
Posted by mirek on Mon, 22 Sep 2014 12:44:22 GMT
View Forum Message <> Reply to Message

Hopefully resolved:

http://www.ultimatepp.org/forums/index.php?t=msg&goto=43 694&#msg_43694