Subject: 32 bit wchar merged
Posted by mirek on Thu, 02 Dec 2021 11:09:12 GMT
View Forum Message <> Reply to Message

sizeof(wchar) is changed to 4 (32 bits) to support non BMP unicode characters

This might bring some incompatibilities in the code that expects wchar to be 16 bit, which escpecially involves dealing with Win32 (and to lesser extend MacOS) APIs, so if your application
is doing that, please check all instances of WCHAR (UniChar on MacOS) or even wchar especially type casts.

To support host APIs, char16 is introduced (but there is no 16-bit String varian).

Use ToSystemCharsetW, FromSystemCharsetW to convert texts to Win32 API.

- Support of drawing non-BMP characters in GUI
- Vastly improved character font replacement code (when drawing characters missing with requested font, replacement font is used)
- Last instances of Win32 ANSI calls (those ending with A) are removed
- UTF handling routines are refactored and their's naming is unified
- RTF is now being able to handle non-BMP characters (RTF is used as clipboard format for RichText)

Other minor changes:

- fixed TryRealloc issue
- improved MemoryCheck
- Removed MemoryAlloc48/MemoryFree48
- In theide Background parsing should less often cause delays in the main thread

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Thu, 02 Dec 2021 11:47:25 GMT
View Forum Message <> Reply to Message

Mirek,

Thanks for your great effort!

Best regards,

Tom

Subject: Re: 32 bit wchar merged

Posted by mirek on Thu, 02 Dec 2021 13:49:13 GMT
View Forum Message <> Reply to Message

Tom1 wrote on Thu, 02 December 2021 12:47Mirek,

Thanks for your great effort!

Best regards,

Tom

Thanks for testing it!

Subject: Re: 32 bit wchar merged
Posted by Novo on Thu, 02 Dec 2021 18:44:05 GMT
View Forum Message <> Reply to Message

Linux.
./umk reference GLDrawDemo CLANG -bus
/home/buildbot/worker/l-upp/build/uppsrc/GLDraw/Text.cpp:72:7: error: use of undeclared identifier 'wstrlen'; did you mean 'strlen'?
```
        n = wstrlen(text);
            ^~~~~~~
            strlen
```

Broken in rev. 34ff69130826ac72f1481f82ef7f0fda0cce0a3e
Windows is fine. Actually, I didn't test this demo on Windows ...

Subject: Re: 32 bit wchar merged
Posted by Klugier on Sat, 04 Dec 2021 18:49:29 GMT
View Forum Message <> Reply to Message

Hello Mirek,

I just did quick test with Emojies. It seems that most of it 90% are not rendered. Should it work like this at the moment? Example:

Platform: Linux/Manjaro - GTK. We could use https://openmoji.org/ emojies library.

Klugier

File Attachments
1) Emojies.png, downloaded 664 times

Subject: Re: 32 bit wchar merged
Posted by mirek on Sat, 04 Dec 2021 19:16:00 GMT
View Forum Message <> Reply to Message

Klugier wrote on Sat, 04 December 2021 19:49Hello Mirek,

I just did quick test with Emojies. It seems that most of it 90% are not rendered. Should it work like this at the moment? Example:

Platform: Linux/Manjaro - GTK. We could use https://openmoji.org/ emojies library.

Klugier

Worth investigating. Simple explanation is that you do not have proper emoji font on your system.

Please try same emojis with OpenOffice, then we will investigate further.

Mirek

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Sat, 04 Dec 2021 19:45:39 GMT
View Forum Message <> Reply to Message

Hi,

I seem to have similar results on Linux Mint, but on Windows I can see quite a selection of emojis in UWord. (In black'n'white in Windows though.)

Libre Office in the same Linux Mint installation can show a whole lot of emojis for the same set.

Best regards,

Tom

Subject: Re: 32 bit wchar merged
Posted by Klugier on Sat, 04 Dec 2021 21:09:14 GMT
View Forum Message <> Reply to Message

Hello,

I could confirm what Tom said. Emojies works within LibreOffice:

For some reason basic :) emoij is displayed differently.

Klugier

File Attachments
1) EmojiesLibre.png, downloaded 574 times

Subject: Re: 32 bit wchar merged
Posted by Oblivion on Sat, 04 Dec 2021 21:30:46 GMT
View Forum Message <> Reply to Message

Hi,

Emojis work on Linux/Gnome 41 (Arch Linux, up-to-date)

I've tried TerminalCtrl with Noto Color Emoji Font (For some reason, the fallback list didn't work so I've passed the aforementioned font as the main font for TerminalCtrl):

Test file: https://unicode.org/Public/emoji/5.0/emoji-test.txt

As you can see, the font/line height seems incorrectly calculated - too high (it should be consistent with 12 pt). This is also a problem with TheIDE (and TheIDE's code-editor too displays the emojis if I directly pass Noto Color Emoji Font)

I'll do some more testing..

Best regards,
Oblivion

File Attachments
1) terminalctrl-emojis.png, downloaded 607 times

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Sun, 05 Dec 2021 10:55:18 GMT
View Forum Message <> Reply to Message

Hi,

I tried with the test file Oblivion pointed to:

https://unicode.org/Public/emoji/5.0/emoji-test.txt

UWord on Linux Mint could show some of the emojis, but most of them were not rendered. I

guess the font used just lacks the missing emojis. (?)

Best regards,

Tom

Subject: Re: 32 bit wchar merged
Posted by mirek on Sun, 05 Dec 2021 12:01:41 GMT
View Forum Message <> Reply to Message

Can you try Draw/FontCR.cpp:429

```
    if(rface[1].Find(i) < 0 && !(fi & Font::SPECIAL) && (fi & Font::SCALEABLE)) {
```

to

```
    if(rface[1].Find(i) < 0 && (fi & Font::SCALEABLE)) {
```

?

Explanation: By the time I was doing Linux code, I decided to exlude color fonts (I have reverted this decision for MacOS as there are no normal glyphs on MacOS, but never ported this decision back).

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Sun, 05 Dec 2021 14:11:18 GMT
View Forum Message <> Reply to Message

Hi Mirek,

The change (in Draw/FontCR.cpp) does not seem to change anything on UWord + Linux Mint for the test file.

Best regards,

Tom

Subject: Re: 32 bit wchar merged
Posted by Oblivion on Sun, 05 Dec 2021 15:05:58 GMT

I've changed the line, and unfortunately nothing has changed.

It seems that the fallback list and the new font rendering routines have some issues. Because upp wchar32 update can definitely render emojis on Linux, as it can be seen from the screenshot in my previous message.
In my test case, it displays all the emojis correctly. The problem is I have to pass the emoji font as the active font. (in TheIDE's code editor, for example)

P.s I have recreated the fallback font list, using the FontMap2. Noto Color Emoji is correctly listed in it, but It doesn't take it into account (or maybe it does bot not scaling correctly, because I am getting a single pixel instead of a emoji. (This may also have to do the incorrect line height issue I have mentioned in my previous post.)

Best regards,
Obilvion

Subject: Re: 32 bit wchar merged
Posted by mirek on Sun, 05 Dec 2021 15:24:05 GMT

Oblivion wrote on Sun, 05 December 2021 16:05I've changed the line, and unfortunately nothing has changed.

It seems that the fallback list and the new font rendering routines have some issues. Because upp wchar32 update can definitely render emojis on Linux, as it can be seen from the screenshot in my previous message.
In my test case, it displays all the emojis correctly. The problem is I have to pass the emoji font as the active font. (in TheIDE's code editor, for example)

P.s I have recreated the fallback font list, using the FontMap2. Noto Color Emoji is correctly listed in it, but It doesn't take it into account (or maybe it does bot not scaling correctly, because I am getting a single pixel instead of a emoji. (This may also have to do the incorrect line height issue I have mentioned in my previous post.)

Best regards,
Obilvion

Well, if the font is not found in the fixed list, all remaining fonts are loaded and tested (this takes time, the purpose of fixed list is to speedup the process in "normal" cases).

I think this is more likely related to color font rendering / handling.

Mirek

Subject: Re: 32 bit wchar merged
Posted by mirek on Tue, 07 Dec 2021 23:46:45 GMT
View Forum Message <> Reply to Message

mirek wrote on Sun, 05 December 2021 16:24
I think this is more likely related to color font rendering / handling.

Mirek

Confirmed, problem seems to be that cairo is unable to scale color font bitmaps. I will need to figure out a way how to extract the color bitmap from the font, then scale manually... (use completely different path for these glyphs).

Subject: Re: 32 bit wchar merged
Posted by mirek on Sun, 12 Dec 2021 15:41:13 GMT
View Forum Message <> Reply to Message

After wrestling a bit with FreeType, colored emoji are now hopefully supported in Linux. (Please test).

The question now remains: When doing replacement, often some replacement font with black&white emoji glyph is closer in appearance than Color Emoji font. That results in mixing colored and B&W emojis. While technically probably more correct, should in case I am looking for emoji glyph always prioritize Noto Color Emoji (or maybe any color emoji font) over matching appearance of font?

Mirek

Subject: Re: 32 bit wchar merged
Posted by Oblivion on Sun, 12 Dec 2021 16:55:18 GMT
View Forum Message <> Reply to Message

Hello Mirek,

Thanks!
Initial test: it works, and rather looks nice! (taken fron TerminalCtrl, test code):

As for your other question: In my experience, the general preference and rule seems to be mixing

a set of consistent color emoji with font, where possible, though it is not set in stone.

Best regards,
Oblivion

File Attachments

1) emojis.png, downloaded 528 times

Subject: Re: 32 bit wchar merged
Posted by pvictor on Mon, 13 Dec 2021 11:42:02 GMT
View Forum Message <> Reply to Message

Hello!

I've found an error in displaying Cyrillic characters in latest U++ releases.
Here's a testcase:

```
#include <CtrlLib/CtrlLib.h>

using namespace Upp;

struct MyApp : TopWindow {
 void Paint(Draw& w) {
  ImagePainter ip(GetSize());
  ip.DrawRect(GetSize(), White());


  ip.DrawText(10, 10, s);    // OK

  String s2 = ToCharset(CHARSET_WIN1251, s);
  ip.DrawText(10, 50, s2, CHARSET_WIN1251); // draws rectangles instead of letters

  Image im = ip;
  w.DrawImage(0, 0, im);
 }
};

GUI_APP_MAIN
{
 MyApp().Run();
}
```

It works fine under U++ 15947.
Platform: Linux Mint.

Best regards.

Victor

Subject: Re: 32 bit wchar merged
Posted by mirek on Mon, 13 Dec 2021 22:06:23 GMT
View Forum Message <> Reply to Message

pvictor wrote on Mon, 13 December 2021 12:42Hello!

I've found an error in displaying Cyrillic characters in latest U++ releases.
Here's a testcase:

```
#include <CtrlLib/CtrlLib.h>

using namespace Upp;

struct MyApp : TopWindow {
 void Paint(Draw& w) {
  ImagePainter ip(GetSize());
  ip.DrawRect(GetSize(), White());


  ip.DrawText(10, 10, s);    // OK

  String s2 = ToCharset(CHARSET_WIN1251, s);
  ip.DrawText(10, 50, s2, CHARSET_WIN1251); // draws rectangles instead of letters

  Image im = ip;
  w.DrawImage(0, 0, im);
 }
};

GUI_APP_MAIN
{
 MyApp().Run();
}
```

It works fine under U++ 15947.
Platform: Linux Mint.

Best regards.
Victor

Well, I was planning to remove that DrawText variant with charset, looks like I will just deprecate it then. Fixed.

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Wed, 15 Dec 2021 07:29:12 GMT
View Forum Message <> Reply to Message

mirek wrote on Sun, 12 December 2021 17:41The question now remains: When doing replacement, often some replacement font with black&white emoji glyph is closer in appearance than Color Emoji font. That results in mixing colored and B&W emojis. While technically probably more correct, should in case I am looking for emoji glyph always prioritize Noto Color Emoji (or maybe any color emoji font) over matching appearance of font?

Mirek

Hi,

Prioritizing color emoji would give the most consistent user experience on any desktop environment.

Best regards,

Tom

Subject: Re: 32 bit wchar merged
Posted by mirek on Wed, 15 Dec 2021 09:53:57 GMT
View Forum Message <> Reply to Message

Tom1 wrote on Wed, 15 December 2021 08:29mirek wrote on Sun, 12 December 2021 17:41The question now remains: When doing replacement, often some replacement font with black&white emoji glyph is closer in appearance than Color Emoji font. That results in mixing colored and B&W emojis. While technically probably more correct, should in case I am looking for emoji glyph always prioritize Noto Color Emoji (or maybe any color emoji font) over matching appearance of font?

Mirek

Hi,

Prioritizing color emoji would give the most consistent user experience on any desktop environment.

Best regards,

Tom

A bit of issue is that Win32 does not support color emoji in GDI as of now. We would need to go Direct2D or something to do that. At least use Direct2D to extract those pictures.

(And yes, in general, I think it might be the time to move from GDI, but I do not want to do that

now.)

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Wed, 15 Dec 2021 10:12:56 GMT
View Forum Message <> Reply to Message

Mirek,

OK, I understand. But please keep me posted if you decide to go ahead with Direct2D or any other GDI replacement option. The 2D graphics functionality and performance on Windows is critically important for me -- and has always been -- so I wish to participate already in early stages to ensure a smooth transition.

Best regards,

Tom

Subject: Re: 32 bit wchar merged
Posted by mirek on Sun, 26 Dec 2021 19:12:41 GMT
View Forum Message <> Reply to Message

Tom1 wrote on Wed, 15 December 2021 08:29mirek wrote on Sun, 12 December 2021 17:41The question now remains: When doing replacement, often some replacement font with black&white emoji glyph is closer in appearance than Color Emoji font. That results in mixing colored and B&W emojis. While technically probably more correct, should in case I am looking for emoji glyph always prioritize Noto Color Emoji (or maybe any color emoji font) over matching appearance of font?

Mirek

Hi,

Prioritizing color emoji would give the most consistent user experience on any desktop environment.

Best regards,

Tom

I made U++ prioritize color emoji in Linux for now. Or most of them, it is not easy to decide which codepoints should prefer color; e.g. copyright sign is considered emoji, but I do not think it would be a good idea to prefer color font for it.

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Sun, 26 Dec 2021 21:09:50 GMT
View Forum Message <> Reply to Message

mirek wrote on Sun, 26 December 2021 21:12Tom1 wrote on Wed, 15 December 2021 08:29mirek wrote on Sun, 12 December 2021 17:41The question now remains: When doing replacement, often some replacement font with black&white emoji glyph is closer in appearance than Color Emoji font. That results in mixing colored and B&W emojis. While technically probably more correct, should in case I am looking for emoji glyph always prioritize Noto Color Emoji (or maybe any color emoji font) over matching appearance of font?

Mirek

Hi,

Prioritizing color emoji would give the most consistent user experience on any desktop environment.

Best regards,

Tom

I made U++ prioritize color emoji in Linux for now. Or most of them, it is not easy to decide which codepoints should prefer color; e.g. copyright sign is considered emoji, but I do not think it would be a good idea to prefer color font for it.

Yes, I see your point. It is indeed difficult to decide which way to go with each of them, but certainly (R) and (C) must be b/w.

This is not a perfectly clear pattern, but it seems that Firefox prioritizes b/w emojis for 16-bit codes and color emojis for higher codes. Also, please note that in Firefox there is a combined gender+skintone+profession in one image, whereas UWord shows a sequence of emojis to identify each property separately. Please open emoji-test.txt (https://unicode.org/Public/emoji/14.0/emoji-test.txt) in Firefox (and maybe Chromium too) to see how they render it. Then compare with UWord.

Best regards,

Tom

Edit: Added link to latest emoji-test.txt.

Subject: Re: 32 bit wchar merged
Posted by mirek on Mon, 27 Dec 2021 07:33:27 GMT
View Forum Message <> Reply to Message

[quote title=Tom1 wrote on Sun, 26 December 2021 22:09]mirek wrote on Sun, 26 December 2021 21:12Also, please note that in Firefox there is a combined gender+skintone+profession in one image, whereas UWord shows a sequence of emojis to identify each property separately.

I am well aware of that, but it (unicode combining characters) is a task for another day....

Mirek

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Tue, 28 Dec 2021 14:48:53 GMT
View Forum Message <> Reply to Message

mirek wrote on Mon, 27 December 2021 09:33
I am well aware of that, but it (unicode combining characters) is a task for another day....

Mirek

Hi Mirek,

No worries. Another year is equally acceptable. (Just mentioned it as I came across it while testing.)

Another thing I found while testing is that on Linux Mint exporting a PDF from UWord using emoji-test.txt results in an empty file. Without emojis the PDF export works OK.

On Windows, exporting a PDF from UWord using emoji-test.txt results in a file with mostly weird symbols instead of the expected emojis shown on UWord.

(Just let me know if you wish me to stop testing here... :))

Best regards,

Tom

Subject: Re: 32 bit wchar merged
Posted by mirek on Wed, 05 Jan 2022 08:33:46 GMT
View Forum Message <> Reply to Message

Tom1 wrote on Tue, 28 December 2021 15:48mirek wrote on Mon, 27 December 2021 09:33
I am well aware of that, but it (unicode combining characters) is a task for another day....

Mirek

Hi Mirek,

No worries. Another year is equally acceptable. (Just mentioned it as I came across it while testing.)

Another thing I found while testing is that on Linux Mint exporting a PDF from UWord using emoji-test.txt results in an empty file. Without emojis the PDF export works OK.

On Windows, exporting a PDF from UWord using emoji-test.txt results in a file with mostly weird symbols instead of the expected emojis shown on UWord.

(Just let me know if you wish me to stop testing here... :))

Best regards,

Tom

Windows situation should be now fixed (with black emojis).

Outputing color emoji in Linux needs some more work...

Mirek

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Wed, 05 Jan 2022 10:42:17 GMT
View Forum Message <> Reply to Message

Hi Mirek,

Thanks! I see the PDF from Windows is OK now.

It seems that the multi character issue is different in Windows compared to Linux:

On Windows I see an empty character (space) following a small BW smiling face for the fully-qualified version above. This is possibly caused by the FE0F -code, as similar behavior can be found in many places of the file (emoji-test.txt). However, on Linux there is just the smiling face without any additional spaces for both variants. Another place where this causes issues is the subgroup keycap:
subgroup: keycap

On Linux Mint all of the characters fall into their correct places pretty well. However, on Windows there are slight alignment issues with unqualified variants and fully-qualified variants are completely off having a space between the character and the keycap box.

BTW: Just noticed that you can actually copy and paste the above unicode sequences directly to the UWord for testing. :)

Best regards,

Tom

Subject: Re: 32 bit wchar merged
Posted by mirek on Wed, 05 Jan 2022 11:09:55 GMT
View Forum Message <> Reply to Message

Tom1 wrote on Wed, 05 January 2022 11:42Hi Mirek,

Thanks! I see the PDF from Windows is OK now.

It seems that the multi character issue is different in Windows compared to Linux:

On Windows I see an empty character (space) following a small BW smiling face for the fully-qualified version above. This is possibly caused by the FE0F -code, as similar behavior can be found in many places of the file (emoji-test.txt). However, on Linux there is just the smiling face without any additional spaces for both variants. Another place where this causes issues is the subgroup keycap:
subgroup: keycap

On Linux Mint all of the characters fall into their correct places pretty well. However, on Windows there are slight alignment issues with unqualified variants and fully-qualified variants are completely off having a space between the character and the keycap box.

BTW: Just noticed that you can actually copy and paste the above unicode sequences directly to the UWord for testing. :)

Best regards,

Tom

Frankly, this is at the moment way beyond what I plan to implement. I think combining characters for normal texts and RTL handling has much higher priority.

(I am at crossroads there - sensible approach would be to embed harfbuzz but it seems quite heavyweight and possibly breaking too much of current CtrlLib).

Subject: Re: 32 bit wchar merged
Posted by coolman on Wed, 12 Jan 2022 09:23:31 GMT
View Forum Message <> Reply to Message

Hi,

Have you checked this 32 bit wchar changes together with the sqlite plugin? I have some problems with that. It seems, that only one letter is read / inserted instead of string.

BR, Radek

Subject: Re: 32 bit wchar merged
Posted by mirek on Wed, 12 Jan 2022 13:24:00 GMT
View Forum Message <> Reply to Message

Thanks, should be now fixed.

If you could test other engines, would be great.

Subject: Re: 32 bit wchar merged
Posted by coolman on Wed, 12 Jan 2022 15:25:44 GMT
View Forum Message <> Reply to Message

Sqlite plugin now works (at least my problem disappeared). Thank you

---

Subject: Re: 32 bit wchar merged
Posted by mirek on Wed, 12 Jan 2022 23:20:15 GMT
View Forum Message <> Reply to Message

Tom1 wrote on Tue, 28 December 2021 15:48

Another thing I found while testing is that on Linux Mint exporting a PDF from UWord using emoji-test.txt results in an empty file. Without emojis the PDF export works OK.

Hopefully finally fixed.

Mirek

Subject: Re: 32 bit wchar merged
Posted by Tom1 on Thu, 13 Jan 2022 07:37:37 GMT
View Forum Message <> Reply to Message

Hi Mirek,

Thanks! It looks good now.

Best regards,

Tom
