Subject: CJK woes
Posted by mobilehunter on Fri, 12 Jan 2007 02:36:37 GMT
View Forum Message <> Reply to Message

I just try the UWord with some japanese text,
but when i export it to PDF, then open it but no japanese character showed, only boxes.

Do you guys know how to get proper result?

Thanks

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Fri, 12 Jan 2007 10:12:05 GMT
View Forum Message <> Reply to Message

mobilehunter wrote on Thu, 11 January 2007 21:36I just try the UWord with some japanese text,
but when i export it to PDF, then open it but no japanese character showed, only boxes.

Do you guys know how to get proper result?

Thanks

This is most likely a bug in PdfDraw. Maybe a problem in TTF handling code.

Well, first things first - do you see text in UWord correct? What is the platform? If platform is Win32, does it print OK? Can you save and load the text?

If the last answer is yes, please save&zip&upload here the text so that we can fix the issue.

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Fri, 12 Jan 2007 12:09:11 GMT
View Forum Message <> Reply to Message

luzr wrote on Fri, 12 January 2007 19:12
This is most likely a bug in PdfDraw. Maybe a problem in TTF handling code.

Well, first things first - do you see text in UWord correct? What is the platform? If platform is Win32, does it print OK? Can you save and load the text?

If the last answer is yes, please save&zip&upload here the text so that we can fix the issue.

Mirek

My OS is Win XP Japanese.
I can see the text while editing.
I can open and see the content of qtf file properly, but not the pdf file.
Sorry i cannot print right now, i don't have one at home

File Attachments

1) japtest.rar, downloaded 569 times

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Fri, 12 Jan 2007 16:20:57 GMT
View Forum Message <> Reply to Message

Interesting, if I load text into UWord, I see boxes.

Looks like some weird font problem. What font is used? Is this font different from normal English XP?

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sat, 13 Jan 2007 00:50:10 GMT
View Forum Message <> Reply to Message

luzr wrote on Sat, 13 January 2007 01:20Interesting, if I load text into UWord, I see boxes.

Looks like some weird font problem. What font is used? Is this font different from normal English XP?

Mirek

Ah forgot to say. First i tried UWord from office computer with normal English XP(but able to show japanese characters), and the problem appeared. Then i tried at home with pure japanese, still have the problem.

If you mean font from UWord, i didn't use any special font.
My uword just has Arial and En-US.
How to find out which font i use?

Btw can you see japanese character from a japanese website?

Thanks

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Sat, 13 Jan 2007 08:16:59 GMT
View Forum Message <> Reply to Message

mobilehunter wrote on Fri, 12 January 2007 19:50luzr wrote on Sat, 13 January 2007 01:20Interesting, if I load text into UWord, I see boxes.

Looks like some weird font problem. What font is used? Is this font different from normal English XP?

Mirek

Ah forgot to say. First i tried UWord from office computer with normal English XP(but able to show japanese characters), and the problem appeared. Then i tried at home with pure japanese, still have the problem.

Hm, could be that, now thinking about it, I do not have raw English version, but east-european...

Quote:
If you mean font from UWord, i didn't use any special font.
My uword just has Arial and En-US.
How to find out which font i use?

Arial then

Quote:
Btw can you see japanese character from a japanese website?

Thanks

I think I can. Any website to test?

Somebody with English XP, could you test whether you see japanesse in UWord?

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sat, 13 Jan 2007 10:51:01 GMT
View Forum Message <> Reply to Message

Please try Google japan

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Sat, 13 Jan 2007 11:44:04 GMT
View Forum Message <> Reply to Message

mobilehunter wrote on Sat, 13 January 2007 05:51Please try Google japan

No boxes, jananesse glyphs everywhere

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sat, 13 Jan 2007 12:47:09 GMT
View Forum Message <> Reply to Message

luzr wrote on Sat, 13 January 2007 20:44mobilehunter wrote on Sat, 13 January 2007 05:51Please try Google japan

No boxes, jananesse glyphs everywhere

Mirek

Can you also write japanese character at UWord or notepad?
I think if japanese character shown, it would also possible to write one.

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Sat, 13 Jan 2007 22:43:20 GMT
View Forum Message <> Reply to Message

No, I cannot - I do not have appropriate keyboard methods installed, and even if I would, I do not know how to use it...

For now, I would concentrate on fact that you can save and load the file, see japanesse, but I do not....

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sun, 14 Jan 2007 01:06:58 GMT
View Forum Message <> Reply to Message

luzr wrote on Sun, 14 January 2007 07:43No, I cannot - I do not have appropriate keyboard methods installed, and even if I would, I do not know how to use it...

For now, I would concentrate on fact that you can save and load the file, see japanesse, but I do

not....

Mirek

Hope this link setup japanese font help.
From the link, one of the font used is MSPMincho.
But thanks anyway.

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Sun, 14 Jan 2007 23:09:13 GMT
View Forum Message <> Reply to Message

Interestingly, on my other machine, glyphs are visible in UWord.

So we are step further

So I have checked what is going on and glyphs are simply missing in the font data that Win returns as font used.

It feels like for the sake of screen painting, Win32 "redirects" paint routine to the font with CJK glyphs - the one I unfortunately am not able to detect yet.

Sounds familiar to anybody?

Relevant code for loading font data is in PdfDraw.cpp at line
744....

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Thu, 18 Jan 2007 00:10:46 GMT
View Forum Message <> Reply to Message

luzr wrote on Mon, 15 January 2007 08:09
Relevant code for loading font data is in PdfDraw.cpp at line
744....

changed to Arial,TimesNewRoman for example.

Then i add Font::Other to richedit code. Now the pdf can contain japanese character even the

result was bad.

So may be if UWord default font can be added with Ms PMincho or any other CJK fonts, we can have correct result for PDF.

Reza

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Thu, 18 Jan 2007 06:14:17 GMT
View Forum Message <> Reply to Message

[quote title=mobilehunter wrote on Wed, 17 January 2007 19:10

changed to Arial,TimesNewRoman for example.
[/quote]

I guess, this is relevant trace

Quote:
Then i add Font::Other to richedit code. Now the pdf can contain japanese character even the result was bad.

Not sure I quite understand.

Quote:
So may be if UWord default font can be added with Ms PMincho or any other CJK fonts, we can have correct result for PDF.

RichEdit::FontFaces(const Vector<int>& face);

allows you to set fonts to use with RichEdit. Can you try this with MS PMincho?

Anyway, the right solution still is to somewhat patch the system to work for Arial too...

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Thu, 18 Jan 2007 08:38:25 GMT
View Forum Message <> Reply to Message

ff.Add(Font::OTHER);
I added that line to RichEdit constructor.

About adding new font, i'm still not quite sure how to add font based on font name, sorry.

Is there any function to add font just by name?

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Thu, 18 Jan 2007 15:50:06 GMT
View Forum Message <> Reply to Message

Ops....

OTHER means first index of font that is not "standard"

Each font has an integer index ("face"). If you want to add specific font, you have to add its index.

You can enumerate fonts (getting index-name pairs), you can also find index for given face name.

See
http://www.ultimatepp.org/src$Draw$Font$en-us.html

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sat, 20 Jan 2007 15:41:33 GMT
View Forum Message <> Reply to Message

luzr wrote on Fri, 19 January 2007 00:50
OTHER means first index of font that is not "standard"

I have modified the UWord source to add japanese font

```
Vector<int> ff;
ff.Add(Font::ARIAL);
ff.Add(Font::ROMAN);
ff.Add(Font::COURIER);
for(int i = Font::COURIER + 1; i < Font::GetFaceCount(); i++)
    if(i==81) //This is MS PMincho for my system
    {
 if(Font::GetFaceInfo(i) & Font::SCALEABLE &&
    !(Font::GetFaceInfo(i) & Font::SYMBOLTYPE))
  ff.Add(i);
    }
editor.FontFaces(ff);
```

But the rendering result is ugly, next character overlapped with half of the previous character.
Then i changed DrawText.cpp, GetPage function.

```
ptr->GetMetrics(page, ptr->default_width);
//ptr->GetMetrics(140, ptr->default_width);
/*for(int i = 0; i < 256; i++)
 if(ptr->default_width[i].width) {
  for(int j = 0; j < 256; j++)
  ptr->default_width[j] = ptr->default_width[i];
  break;
      }*/
```

I don`t know my changes will break everything, but the result is better.

Which line that do text drawing for richedit control?

BTW the pdf exporting still error.
It stopped at this line from TTFReader.cpp

```
if(head.magicNumber != 0x5F0F3CF5)
 Error();
```

Thanks

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Sun, 21 Jan 2007 08:20:47 GMT
View Forum Message <> Reply to Message

Quote:
Then i changed DrawText.cpp, GetPage function.

```
ptr->GetMetrics(page, ptr->default_width);
//ptr->GetMetrics(140, ptr->default_width);
/*for(int i = 0; i < 256; i++)
 if(ptr->default_width[i].width) {
  for(int j = 0; j < 256; j++)
  ptr->default_width[j] = ptr->default_width[i];
  break;
      }*/
```

I don`t know my changes will break everything, but the result is better.

Well, the purpose of that code is to save the space needed for character widths for CJK fonts (as

all glyphs seem to be monospaced).

Unfortunately, the code was designed for chinesse glyphs (tries to read specific characters starting at unicode 140*256) and very likely these characters are missing in PMincho.

BTW, it is also interesting that widths are correct when the font is accessed as Arial - most likely in that case, some other font is used there to provide chinesse chars.

Anyway, I guess we should try to fix this for PMincho as well. What we need to have there is the correct monospaced width of CJK glyph. Any ideas?

Can you check whether width is gound (break is taken)?

Thinking about it, perhaps we should simply check the first 'page' used instead of 140:

```
FontInfo::CharMetrics *FontInfo::GetPage(int page) const
{
 if(page >= 46 && !ptr->default_width) {
  ptr->default_width = new CharMetrics[256];
  ptr->GetMetrics(page, ptr->default_width);
  for(int i = 0; i < 256; i++)
   if(ptr->default_width[i].width) {
    for(int j = 0; j < 256; j++)
     ptr->default_width[j] = ptr->default_width[i];
    break;
   }
 }
 CharMetrics *& cm = ptr->width[page];
 if(!cm) {
  cm = new CharMetrics[256];
  ptr->GetMetrics(page, cm);
  if(page == 1)
   ComposeMetrics(ptr->font, cm);
  if(page >= 46) {
   for(int i = 0; i < 256; i++) {
    if(!(cm[i] == ptr->default_width[i]) && cm[i].width)
     return cm;
   }
   delete[] cm;
   cm = ptr->default_width;
  }
 }
 return cm;
}
```

Can you check please?

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sun, 21 Jan 2007 08:45:17 GMT
View Forum Message <> Reply to Message

luzr wrote on Sun, 21 January 2007 17:20
Can you check whether width is gound (break is taken)?

Hmm sorry, i dont understand about gound, any example?

luzr wrote on Sun, 21 January 2007 17:20
Thinking about it, perhaps we should simply check the first 'page' used instead of 140:

I replaced 140 with page, and commented the for loop below it.

luzr wrote on Sun, 21 January 2007 17:20
Can you check please?

Sorry, you mean check againt codes you just posted? I think it7s the same code with original GetPage codes.

Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Sun, 21 Jan 2007 11:24:58 GMT
View Forum Message <> Reply to Message

gound -> found (typo)

In fact, what I am asking for is to replace 140 with page, but leave the loop there.

The purpose of the loop is to avoid any zero width in the array. Zero width means missing glyph.

What we really are trying to do here is to save memory for monospaced CJK glyphs. We need an array with all values equal to CJK "mono" width. Then we can check for CJK chars whether all characters have the same width (or zero for missing chars) and in that case we can avoid using separate array for the "page" and use single shared array with default widths.

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sun, 21 Jan 2007 13:54:36 GMT

luzr wrote on Sun, 21 January 2007 20:24gound -> found (typo)

In fact, what I am asking for is to replace 140 with page, but leave the loop there.

The purpose of the loop is to avoid any zero width in the array. Zero width means missing glyph.

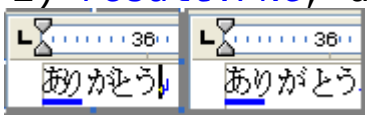
What we really are trying to do here is to save memory for monospaced CJK glyphs. We need an array with all values equal to CJK "mono" width. Then we can check for CJK chars whether all characters have the same width (or zero for missing chars) and in that case we can avoid using separate array for the "page" and use single shared array with default widths.

Mirek

I replaced the 140 with 'page', and test with loop and no loop.
Here i attached the image.
On the left is with loop, and the right is without loop.
I think the right is better, but may be just coincidence.

File Attachments

1) result.PNG, downloaded 1084 times



Subject: Re: Howto export japanese word to pdf?
Posted by mirek on Sun, 21 Jan 2007 23:21:35 GMT

Ops, found a bug causing CJK character widths to misbehave:

Draw.h, 227:

```
struct CharMetrics {
  int  width;
  int  lspc;
  int  rspc;

  bool operator==(const CharMetrics& b) const
      { return width == b.width && lspc == b.lspc && rspc == b.rspc; }
};
```

In GetPage, I have used 'page' instead of 140, that is better too.

Please test!

Which leaves us still with two problems: PDF and IME....

Moving this thread to "development area" now...

Mirek

Subject: Re: Howto export japanese word to pdf?
Posted by mobilehunter on Sun, 21 Jan 2007 23:48:29 GMT
View Forum Message <> Reply to Message

Tested, and it looks better now
Thanks
