**Subject: Re: byte/unicode mismatch**
Posted by R.Bauer on Fri, 21 Nov 2008 10:10:22 GMT
View Forum Message <> Reply to Message

That is all orthogonal.

How can I decode and how can I encode?

cheers
Reimar

Allan Whiteford schrieb:
> Heinz Stege wrote:
>> On Thu, 20 Nov 2008 09:23:52 -0800 (PST), mgalloy@gmail.com wrote:
>>
>>
>>> On Nov 20, 3:19 am, Reimar Bauer <R.Ba...@fz-juelich.de> wrote:
>>>
>>>> Hi
>>>>
>>>> the ascii table is gone.
>>>>
>>>> IDL> print,byte('ï¿½')
>>>> 195 188
>>>>
>
>> The string entered in the workbench command line is encoded in UTF8.
>
> Picking up on this point (and the one made by Mike) - it's mostly to do
> with your editor. The workbench seems to be unicode aware so it really
> is passing a two byte representation of ï¿½ into the interpreter.
>
> If I use the simple command line interface running through an xterm
> (X.Org 6.8.99.903) which I guess isn't unicode aware then I get 252 with
> the same version of IDL:
>
> IDL> print,!version
> { x86 linux unix linux 7.0 Oct 25 2007    32    64}
> IDL> print,byte('ï¿½')
>  252
>
> but with the workbench:
>
> IDL> print,!version
> { x86 linux unix linux 7.0 Oct 25 2007    32    64}
> IDL> print,byte('ï¿½')
>  195 188
>

> I would expect that if you read the character from a file (either as
> data or in a .pro file) it depends on the program which wrote the file
> and whether your editor was unicode-aware.
>
> In saying all this, I don't understand unicode properly (does anyone?!?)
> - I'm just reporting on the fact that it isn't just the IDL interpreter
> which is the issue, it's to do with the editor which sends the character
> to the interpreter.
>
> This has already been said - I've just rephrased it using more
> (unnecessary?) words. I hope it's helpful.
>
> Thanks,
>
> Allan

---