Subject: Re: byte/unicode mismatch
Posted by R.Bauer on Fri, 21 Nov 2008 21:45:32 GMT
View Forum Message <> Reply to Message

Allan Whiteford schrieb:
> Reimar Bauer wrote:
>> That is all orthogonal.
>>
>> How can I decode and how can I encode?
>>
>> cheers
>> Reimar
>>
>
> Reimar,
>
> The question (and answer) isn't all that straightforward, byte values
> over 127 aren't well defined without an encoding system or a codepage.
>
> However, the answer you're probably looking for is:
>
> b=byte('ï¿½')   ; assumption 2
> print,b[1]+(b[0] eq 195)*64   ; assumption 1
>
> which is assuming:
>
> 1) you want byte values from (two byte) UTF-8 to ISO-8859-1
>
> and
>
> 2) that the u-umlaut character has entered the intepreter from a UTF-8
> environment.
>
> Please don't just cut and paste the above assuming all will be well.
>
> Thanks,
>
> Allan
>

Hmm this does confuse me more. Lets see if an other examples helps me.

If I write an output file using the ide e.g.

openw, 10, 'testfile.txt'
printf, 10, 'Jï¿½lich'
close, 10

If I run this program with iso encoding isn't the result different to utf-8?

Or how can I write it iso encoded independent from the user setting?
In python I have several methods for that.
http://effbot.org/zone/unicode-objects.htm

cheers
Reimar