
Subject: file_search problem

Posted by [greg.addr](#) on Wed, 25 Nov 2015 09:09:55 GMT

[View Forum Message](#) <> [Reply to Message](#)

I have a directory (on Windows7) containing some filenames with Romanian special characters, e.g.

în.txt

The OS 'dir' command shows them...

```
25/11/2015 09:41 <DIR>      .
25/11/2015 09:41 <DIR>      ..
25/11/2015 09:31          0 în.txt

                2 File(s)          0 bytes
                2 Dir(s) 1,043,905,544,192 bytes free
```

cheers,
Greg

Subject: Re: file_search problem

Posted by [Jim Pendleton](#) on Wed, 25 Nov 2015 15:45:34 GMT

[View Forum Message](#) <> [Reply to Message](#)

On Wednesday, November 25, 2015 at 2:09:56 AM UTC-7, greg...@gmail.com wrote:

> I have a directory (on Windows7) containing some filenames with Romanian special characters, e.g.

>

> în.txt

>

>

> The OS 'dir' command shows them...

>

> 25/11/2015 09:41 <DIR> .

> 25/11/2015 09:41 <DIR> ..

> 25/11/2015 09:31 0 în.txt

> 2 File(s) 0 bytes

> 2 Dir(s) 1,043,905,544,192 bytes free

>
>

>
>
> cheers,
> Greg

A work-around is to use the old FINDFILE function, but you should report this to support at exelisvis.com.

A substantial amount of work was put into I18N a few releases ago, but it looks like this is a special case.

Interesting...

```
IDL> print, byte('î')  
238  
IDL> print, string(238b)  
î
```

...however,

```
97  
IDL> print, string(97b)  
a
```

Jim P.

Subject: Re: file_search problem
Posted by [Lajos Foldy](#) on Wed, 25 Nov 2015 15:58:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Wednesday, November 25, 2015 at 4:45:37 PM UTC+1, Jim P wrote:

>
> A substantial amount of work was put into I18N a few releases ago, but it looks like this is a special case.
>
> Interesting...
>
> IDL> print, byte('î')
> 238
> IDL> print, string(238b)
> î
>
> ...however,

>

> 97

> IDL> print, string(97b)

> a

>

> Jim P.

I think the first one is in extended ASCII (0-255) and the second one is a true Unicode character.

Are there any Unicode string support plans for IDL?

regards,

Lajos

Subject: Re: file_search problem

Posted by [Heinz Stege](#) on Wed, 25 Nov 2015 17:05:56 GMT

[View Forum Message](#) <> [Reply to Message](#)

On Wed, 25 Nov 2015 07:58:07 -0800 (PST), fawltylanguage@gmail.com
wrote:

> On Wednesday, November 25, 2015 at 4:45:37 PM UTC+1, Jim P wrote:

>>

>> A substantial amount of work was put into I18N a few releases ago, but it looks like this is a special case.

>>

>> Interesting...

>>

>> IDL> print, byte('î')

>> 238

>> IDL> print, string(238b)

>> î

>>

>> ...however,

>>

>> IDL> print, byte('a')

>> 97

>> IDL> print, string(97b)

>> a

>>

>> Jim P.

>

> I think the first one is in extended ASCII (0-255) and the second one is a true Unicode character.

Let me add, that this conversion seems to take place during string input. No conversion happens, if the string "is really UTF-8":

```
IDL> a=['C3'xb,'AE'xb]
IDL> print,a
195 174
IDL> print,string(a)
î
IDL> print,byte(string(a))
195 174
IDL> b=['C4'xb,'83'xb]
IDL> print,string(b)
a
IDL> print,byte(string(b))
196 131
```

I hope, my news agent will choose the correct charset (UTF-8)! I'm not sure.

Cheers, Heinz

Subject: Re: file_search problem
Posted by [greg.addr](#) on Wed, 25 Nov 2015 20:38:25 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Wednesday, November 25, 2015 at 6:05:54 PM UTC+1, Heinz Stege wrote:

> On Wed, 25 Nov 2015 07:58:07 -0800 (PST), fawltlanguage
> wrote:

>

>> On Wednesday, November 25, 2015 at 4:45:37 PM UTC+1, Jim P wrote:

>>>

>>> A substantial amount of work was put into I18N a few releases ago, but it looks like this is a special case.

>>>

>>> Interesting...

>>>

>>> IDL> print, byte('î')

>>> 238

>>> IDL> print, string(238b)

>>> î

>>>

>>> ...however,

>>>

>>> IDL> print, byte('a')

>>> 97

>>> IDL> print, string(97b)

>>> a

>>>

>>> Jim P.

```

>>
>> I think the first one is in extended ASCII (0-255) and the second one is a true Unicode
character.
>
> Let me add, that this conversion seems to take place during string
> input. No conversion happens, if the string "is really UTF-8":
>
> IDL> a=['C3'xb,'AE'xb]
> IDL> print,a
> 195 174
> IDL> print,string(a)
> î
> IDL> print,byte(string(a))
> 195 174
> IDL> b=['C4'xb,'83'xb]
> IDL> print,string(b)
> a
> IDL> print,byte(string(b))
> 196 131
>
> I hope, my news agent will choose the correct charset (UTF-8)! I'm not
> sure.
>
> Cheers, Heinz

```

Thanks, everyone, for the comments. I've found that the same does happen for other non-ascii characters (Polish, this time):

Directory of D:\tmp\test

```

25/11/2015 21:05 <DIR>      .
25/11/2015 21:05 <DIR>      ..
25/11/2015 09:31          0 î.n.txt

```

4 File(s) 0 bytes

file_search gives:

```

IDL> file_search("d:\tmp\test\","*.*)
D:\tmp\test\î.n.txt

```

and findfile gives:

```
IDL> findfile("d:\tmp\test\*.txt")
d:\tmp\test\in.txt
d:\tmp\test\a.txt
d:\tmp\test\a.txt
d:\tmp\test\l.txt
```

So findfile sees the files, although the extended characters are simplified. However, the files are not identifiable through the simplified names:

```
IDL> a=findfile("d:\tmp\test\*.txt")
IDL> print,(file_info(a[3])).exists
0
```

This doesn't work either...

```
IDL> spawn,"dir /b /s d:\tmp\test\*.txt",res,err
IDL> res
d:\tmp\test\ En.txt
d:\tmp\test\a.txt
d:\tmp\test\a.txt
d:\tmp\test\l.txt
```

And surprisingly (to me!), even this fails:

```
IDL> spawn,"dir /b /s *.txt >dir.txt",res,err
```

...with dir.txt containing the same

```
d:\tmp\test\ En.txt
d:\tmp\test\a.txt
d:\tmp\test\a.txt
d:\tmp\test\l.txt
```

which is not the fault of IDL.

cheers,
Greg
