Subject: Re: Read ASCII
Posted by davidf on Wed, 07 Jan 1998 08:00:00 GMT
View Forum Message <> Reply to Message

David Mottershead (DMottershead@mhl.nsw.gov.au) writes
from down under:

> I am trying to read in some ASCII data in a column form with a format
> defined in my main procedure like this:
>
> fileData = {WBD_File, $
> yr:0, $
> mn:0, $
> day:0, $
> hour:0, $
> minute:0, $
> value:0.0}

[much code clipped]

I've read this post over quite a few times, and I am still
not *totally* sure I understand what you are trying to do.
It looks to me like you have column data in the form of
6 columns by 35136 rows. Going on that assumption, I would
say you are trying to do too much all at once. Take it a
little bit slower.

I might try it like this. I would make my info structure,
which is where I store *all* of the information I need to
make my program work, with a field for the data template
and another field for the data itself. It might be defined
like this:

```
fileData = {WBD_File, yr:0, mn:0, day:0, hour:0, $
          minute:0, value:0.0}
info = { template:fileData, $  ; The file template.
       header:StrArr(5), $   ; Five line data file header.
       rows:35136L, $        ; Number of rows in the data files.
       data:Ptr_New() }      ; A null pointer, for now.
```

And just from this one code example, I'm guessing this program
could use a little modularity. I would probably define the
Open File button like this, so I could work with a smaller
event handler:

```
openfileID = Widget_Button(menubase, Value='Open File...', $
   Event_Pro='Open_File_Button_Event')
```

So, then, my event handler might look something like this:

```
   PRO Open_File_Button_Event, event

      ; Get the data file name from the user.

   filename = Dialog_Pickfile(/Read, Filter='*.dat')
   IF filename EQ '' THEN RETURN

      ; Go read the file. First, get the info structure.

   Widget_Control, event.top, Get_UValue=info, /No_Copy

      ; Create an array of structures. Read the data.

   dataArray = Replicate(info.template, info.rows)
   OpenR, lun, filename, /Get_Lun
   Readf, lun, info.header, dataArray
   Free_Lun, lun

      ; Make the data vectors.

   yr = dataArray(*).(0)
   mn = dataArray(*).(1)
   day = dataArray(*).(2)
   hour = dataArray(*).(3)
   minute = dataArray(*).(4)
   value = dataArray(*).(5)

       ; Store data as structure in the pointer location.

   info.data = Ptr_New({yr:yr, mn:mn, day:day, hour:hour, $
      minute:minute, value:value}, /No_Copy)

   Widget_Control, event.top, Set_UValue=info, /No_Copy
   END
```

Now, any module that needs to do something with the data
can access it like this:

```
   Plot, (*info.data).day, (*info.data).value
```

Hope that gives you some ideas.

Cheers,

David

P.S. You might also look at the two new functions ASCII_TEMPLATE
and READ_ASCII. They are made specifically to read this kind
of column data.

David Fanning, Ph.D.
Fanning Software Consulting
E-Mail: davidf@dfanning.com
Phone: 970-221-0438
Coyote's Guide to IDL Programming: http://www.dfanning.com/

Subject: Read ASCII
Posted by DMottershead on Thu, 08 Jan 1998 08:00:00 GMT
View Forum Message <> Reply to Message

I am trying to read in some ASCII data in a column form with a format
defined in my main procedure like this:

```
fileData = {WBD_File, $
            yr:0, $
            mn:0, $
            day:0, $
            hour:0, $
            minute:0, $
            value:0.0}

info={obj:obj, $
     filedata:filedata}
ptr=ptr_new(info,/no_copy)
```

I am opening and reading the data in my event handler like this:

```
'open': begin
       filename='d:\idl\mhl\syd9712.wbd'

         WIDGET_CONTROL, (*ptr).infowbd.fileText, SET_VALUE = filename

x=35136
arr=fltarr(6,x)
hdr=strarr(5)
count=0

;make a the same structure as fileData

a=(*ptr).fileData

data=replicate(a,x)
```

```
openr, lun, filename,/get_lun

;read header in

readf,lun,hdr

WHILE NOT EOF(lun) DO BEGIN
   READF, lun, a
      data(count)=a
      count=count+1
endwhile

data=data(0:count-1)

; tell user abour file header

WIDGET_CONTROL, (*ptr).infowbd.fileHeader, SET_VALUE = hdr

; get the data values
```

My problem exists when I try to "copy" the "data" back to the fileData structure.
How do I make the fileData structure the same size as the "data". I have tried to replicate (*ptr).fileData without any success. Also, how do I copy the values from "data" to "(*ptr).fileData"? In advance, thanks for your help.

--
Regards

David

 ** *******
David Mottershead Phone: +61 2 9949 0234
Manly Hydraulics Laboratory Fax: +61 2 9948 6185
110b King St, Manly Vale, 2093 email: dmottershead@mhl.nsw.gov.au
SYDNEY, AUSTRALIA WWW: http://www.mhl.nsw.gov.au
 ** *******

Subject: Re: read ascii
Posted by greg michael on Mon, 17 Nov 2008 10:55:08 GMT
View Forum Message <> Reply to Message

file_lines() is very useful in this situation - then you can avoid all that looping and EOF stuff. The next problem is that when you read a

string it takes the whole line - not just the part that looks like a
string to you. So the error comes when you try to read the time and
it's at the start of a new line. I'd start like this:

```
n=file_lines('test.dat')
s=strarr(n)
openr,1,'test.dat'
readf,1,s
close,1
```

and then chop up the lines with strmid() and convert to the types you
need.

cheers,
Greg

Subject: Re: read ascii
Posted by julia.walterspiel on Mon, 17 Nov 2008 11:08:43 GMT
View Forum Message <> Reply to Message

good idea, thanks Greg, I'll go with this

Subject: Re: read ascii
Posted by R.Bauer on Mon, 17 Nov 2008 12:32:31 GMT
View Forum Message <> Reply to Message

julia.walterspiel@gmail.com schrieb:
> hi
>
> i've been fiddling about this problem the whole morning and it seems I
> can't get it done properly:
>
> I'm trying to read in a huge (!) ascii-file ('test.dat') where the
> first column is a string (name of station), followed by 2 columns with
> integer values (second row date in the format yymmddhourminute, third
> row integer numbers between 0 and 10).
>
> example 'test.dat':
> SMA 200001010010 0
> SMA 200001010020 0
> SMA 200001010030 1
> SMA 200001010040 0
> SMA 200001010050 3
> SMA 200001010100 4
> SMA 200001010110 5

> SMA 200001010120 0
> SMA 200001010130 0
> SMA 200001010140 0
>
> no header.
>
> I don't know how many rows I got, but I'm sure it's a LOT, since excel
> crashes when trying to copy the file :)
>
> I'm using David's Code like this:
>
> OPENR, lun, '/filepath/test.dat', / GET_LUN
>
> station = strarr (1000000)
> time = fltarr (1000000)
> sunshine_duration = fltarr(1000000)
>
> s = '???' --> not sure what to insert here
> t = 0.0
> sd = 0.0
> count = 0
>
> WHILE (NOT EOF(lun)) DO BEGIN
> READF, lun, s, t, sd
> station(count) = s
> time (count) = t
> sunshine_duration(count) = sd
> count = count+1
> ENDWHILE
>
> station = station(0:count-1)
> time = time(0:count-1)
> sunshine_duration = sunshine_duration(0:count-1)
>
> FREE_LUN, lun
>
> then I get the error
> "READF: Input conversion error. Unit: 101"
> I tried to google this error but couldnt find any useful
> information...
>
> I figure, I don't understand 100% what I'm doing here, that's why I
> don't see where I make the mistake...
> furthermore, I'm not sure how the file is delimited (tab or white
> space) if this is crucial to anything..?
>
> any help appreciated!
> cheers,

> juls

hi

you can read the file once as byte data file and then do a conversion to
the types you want for each column. That tool can be brought to
perfection if you involve pointers ;)

cheers
Reimar

e.g.

```
file = 'test.dat'
lines = file_lines(file)
struct = replicate(create_struct('var1', bytarr(5),$
                                 'var2', bytarr(13),$
                                 'var3', bytarr(11)), lines)
openu,lun,file,/get_lun
readu,lun,struct
free_lun,lun

; example usage of the data
print, strtrim(struct.var1, 2)
print, long64(strtrim(struct.var2, 2))
print, long(strtrim(struct.var3, 2))

; or
result =  create_struct('var1', strtrim(struct.var1, 2),$
                        'var2', long64(strtrim(struct.var2, 2)),$
                        'var3', long(strtrim(struct.var3, 2)))

print, result.var1

end
```

Subject: Re: read ascii
Posted by David Fanning on Mon, 17 Nov 2008 14:35:59 GMT
View Forum Message <> Reply to Message

julia.walterspiel@gmail.com writes:

> I understand, that NaN has some problems with certain routines, like
> TOTAL, but obviously it also has problems with MEAN?! Whenever I have
> a "NaN" in a part of the data I want to get the mean of, the result is
> "NaN"... isn't the nice thing about working with NaN that IDL
> recognizes those "values" as bad values and ignores them in further

> calculations??
> would the other way to set all NA-data to e.g. 10000 and then continue
> working with the MAX_VALUE function the better solution here?

Have you tried setting the NAN keyword on the MEAN function?

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: read ascii
Posted by julia.walterspiel on Mon, 17 Nov 2008 15:11:48 GMT
View Forum Message <> Reply to Message

yeah sorry David, i got confused with the TOTAL routine.. i realized
what bulls*** i had asked and I tried to delete my last post right
after I uploaded it but I assume you were quicker :))

Subject: Re: read ascii
Posted by David Fanning on Mon, 17 Nov 2008 15:13:12 GMT
View Forum Message <> Reply to Message

julia.walterspiel@gmail.com writes:

> yeah sorry David, i got confused with the TOTAL routine.. i realized
> what bulls*** i had asked and I tried to delete my last post right
> after I uploaded it but I assume you were quicker :))

Yeah, at 6:00 AM I'm all over everything. ;-)

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: read ascii
Posted by R.Bauer on Mon, 17 Nov 2008 15:13:31 GMT
View Forum Message <> Reply to Message

julia.walterspiel@gmail.com schrieb:
> yeah sorry David, i got confused with the TOTAL routine.. i realized
> what bulls*** i had asked and I tried to delete my last post right
> after I uploaded it but I assume you were quicker :))

I see three posts

Reimar

Subject: Re: read ascii
Posted by Brian Larsen on Mon, 17 Nov 2008 15:14:17 GMT
View Forum Message <> Reply to Message

yikes, bytes, bits, structures, and splitting oh my. That certainly works but... yikes

This can be done with a cleaver call to format and a type conversion that is (in my opinion) the best and fastest way. Play with the type conversion as you want I didn't include it here.

```
lines=file_lines('test.dat')
dat = strarr(3, lines)
openr, lun, 'test.dat', /get_lun
readf, lun, dat, format = '(a4,a16,a)'
free_lun, lun
IDL> print, dat[0,*]
SMA
SMA
SMA
SMA
SMA
SMA
SMA
SMA
SMA
SMA
SMA
IDL> print, dat[1,*]
200001010010
200001010020
200001010030
200001010040
200001010050
```

200001010100
200001010110
200001010120
200001010130
200001010140
IDL> print, dat[2,*]
0
0
1
0
3
4
5
0
0
0

Cheers,

Brian

Brian Larsen
Boston University
Center for Space Physics
http://people.bu.edu/balarsen/Home/IDL

Subject: Re: read ascii
Posted by R.Bauer on Mon, 17 Nov 2008 15:18:55 GMT
View Forum Message <> Reply to Message

Brian Larsen schrieb:
> yikes, bytes, bits, structures, and splitting oh my. That certainly
> works but... yikes
>

yeah

can one make a speed comparision and the other question can format be
used with structures too?

cheers
Reimar

> This can be done with a cleaver call to format and a type conversion
> that is (in my opinion) the best and fastest way. Play with the type

> conversion as you want I didn't include it here.
>
> lines=file_lines('test.dat')
> dat = strarr(3, lines)
> openr, lun, 'test.dat', /get_lun
> readf, lun, dat, format = '(a4,a16,a)'
> free_lun, lun
> IDL> print, dat[0,*]
> SMA
> SMA
> SMA
> SMA
> SMA
> SMA
> SMA
> SMA
> SMA
> SMA
> IDL> print, dat[1,*]
> 200001010010
> 200001010020
> 200001010030
> 200001010040
> 200001010050
> 200001010100
> 200001010110
> 200001010120
> 200001010130
> 200001010140
> IDL> print, dat[2,*]
> 0
> 0
> 1
> 0
> 3
> 4
> 5
> 0
> 0
> 0
>
>
> Cheers,
>
> Brian
>
> -- --------------
> Brian Larsen

> Boston University
> Center for Space Physics
> http://people.bu.edu/balarsen/Home/IDL
>
>
>

Subject: Re: read ascii
Posted by David Fanning on Mon, 17 Nov 2008 15:24:24 GMT
View Forum Message <> Reply to Message

Reimar Bauer writes:

> julia.walterspiel@gmail.com schrieb:
>> yeah sorry David, i got confused with the TOTAL routine.. i realized
>> what bulls*** i had asked and I tried to delete my last post right
>> after I uploaded it but I assume you were quicker :))
>
>
> I see three posts

It's a user interface issue. They need to get that DELETE
button further away from the SEND button. :-)

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: read ascii
Posted by julia.walterspiel on Mon, 17 Nov 2008 15:43:01 GMT
View Forum Message <> Reply to Message

> I see three posts

i didn't see my post after 10min so I figured something went wrong and
added it again. I am not a very patient person.. :) (which makes
programming my absolute favourite :))

good, besides the posting-deleting-issue here's another simple

question which is obviously so simple that I cannot find a decent answer on the internet:

I put my 3 arrays (Station, Time, Sunshine_duration) in one structure because I'm trying to avoid having to index the sunshine_duration array (the data runs from 01 jan 2000 to 31 okt 2008 and I want to be able to plot selected months).

Am I right in my assumption, that every value of the "sunshine_duration" has automatically assigned the right date ("time") when it was measured? Means, if I want to let's say plot data from April 2007 i can do some magic like

index = where(structure.time EQ 2007)

and this gives me the corresponding Sunshine_duration-values? (I really hope it does, otherwise I would not understand the sense of building structures...)

... last question for today! promise

Subject: Re: read ascii
Posted by julia.walterspiel on Mon, 17 Nov 2008 16:09:22 GMT
View Forum Message <> Reply to Message

On 17 Nov., 16:43, julia.waltersp...@gmail.com wrote:
> > I see three posts
>
> i didn't see my post after 10min so I figured something went wrong and
> added it again. I am not a very patient person.. :) (which makes
> programming my absolute favourite :))
>
> good, besides the posting-deleting-issue here's another simple
> question which is obviously so simple that I cannot find a decent
> answer on the internet:
>
> I put my 3 arrays (Station, Time, Sunshine_duration) in one structure
> because I'm trying to avoid having to index the sunshine_duration
> array (the data runs from 01 jan 2000 to 31 okt 2008 and I want to be
> able to plot selected months).
>
> Am I right in my assumption, that every value of the
> "sunshine_duration" has automatically assigned the right date ("time")
> when it was measured? Means, if I want to let's say plot data from
> April 2007 i can do some magic like
>
> index = where(structure.time EQ 2007)

>
> and this gives me the corresponding Sunshine_duration-values? (I
> really hope it does, otherwise I would not understand the sense of
> building structures...)
>
> ... last question for today! promise

answer to myself: yes it does...I love structures!
good night and thanks to all of you
