Subject: Re: read multiple files with varying names
Posted by David Fanning on Mon, 25 May 2009 15:01:58 GMT
View Forum Message <> Reply to Message

sophie.hoss@gmail.com writes:

> probably a simple syntax question:
>
> i need to read multiple ascii files (all in one folder), with
> different names.
> e.g.
> ci0201.004
> ci0202.004
> ...
> do0201.004
>
> i want to read all of them that start with "ci", store them in one
> file, then read all of them that start with "do", store them and so
> on. (in order to extract data from them later on)
>
> I used to read in files with readf in older programms, like
>
> openr, lun, 'filepath/*.dat', /GET_LUN
> header = strarr(3)
> READF, lun, header
>
> but there it wasnt necessary to "group" them, means to read only the
> ones starting with some specific letters...
>
>
>
> the other idea I had is simply to include an IF statement somewhere to
> exclude all files that DON'T start with the specific letter.
> But I'm SURE there's a more elegant solution to this.. I just don't
> know it :)

You are looking for FILE_SEARCH.

```
CD, mydatadir
ci_files = FILE_SEARCH('ci*', COUNT=num_ci_files_found)
do_files = FILE_SEARCH('do*', COUNT=num_do_files_found)
```

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.

Subject: Re: read multiple files with varying names
Posted by Jean H. on Mon, 25 May 2009 15:04:32 GMT
View Forum Message <> Reply to Message

sophie.hoss@gmail.com wrote:
> hey guys
>
> probably a simple syntax question:
>
> i need to read multiple ascii files (all in one folder), with
> different names.
> e.g.
> ci0201.004
> ci0202.004
> ...
> do0201.004
>
> i want to read all of them that start with "ci", store them in one
> file, then read all of them that start with "do", store them and so
> on. (in order to extract data from them later on)

one way is to use file_search

print, file_search('ci*.004')

you can save the result to a new file.

Jean

>
> I used to read in files with readf in older programms, like
>
> openr, lun, 'filepath/*.dat', /GET_LUN
> header = strarr(3)
> READF, lun, header
>
> but there it wasnt necessary to "group" them, means to read only the
> ones starting with some specific letters...
>
>
>
> the other idea I had is simply to include an IF statement somewhere to
> exclude all files that DON'T start with the specific letter.

> But I'm SURE there's a more elegant solution to this.. I just don't
> know it :)
>
> thanks for your help
> sophie

Subject: Re: read multiple files with varying names
Posted by sophie.hoss on Thu, 28 May 2009 14:44:05 GMT
View Forum Message <> Reply to Message

yup, thanks.. i deleted the post like 10min after i wrote it, cause i found the solution myself. you guys were obviously quicker.

however, the next problem came right along..

i'm trying to read in a fair big amount of radiation ascii-files (12 colums, roughly 5000 rows each file), which have as i wrote in my last, deleted post, all different names. I guess the name problem is solved.
below you find the code i'm using. and i cannot find out why it gives me the error "READF: End of file encountered" (comes up in the line marked with ****). i'm very aware that this is a common error and i went over and over it again. with moderate success...

here's part of the code:

```
CD, 'filepath'
ci_files = FILE_SEARCH('ci*', COUNT=num_ci_files_found)

for i = 0,(num_ci_files_found-1) do begin

 file = ci_files(i)

  ; Get number of rows in file
    rows = File_Lines(file)

  ; Open file, read header
    OpenR, lun, file, /Get_Lun
    header = StrArr(34)
     ReadF, lun, header
     Point_Lun, -lun, currentLocation

  ; Read the first line and get number of columns
    line = ""
    ReadF, lun, line
    cols = N_Elements(StrSplit(line, /RegEx, /Extract))
```

```
  ; variable to hold the data
     data = strarr(cols, (rows(i)-N_Elements(header)))


  rows_data = rows - n_elements(header)
stop
  s=strarr(cols)
  n=0
  while (~ eof(file) and (n lt rows_data -1 )) do begin
   ; Read a line of data
   readf,lun,s
   ; Store it in data
    data[*,n]=s
    n=n+1       (*****)
  end
  data=data[*,0:n-1]
```

I did the while-loop because I learned that IDL might not read every
line separately. Don't know if it's the most elegant version but at
least it is one.

for the sake of my sanity, any help is appreciated!
cheers,
sophie

Subject: Re: read multiple files with varying names
Posted by Jean H. on Thu, 28 May 2009 15:25:56 GMT
View Forum Message <> Reply to Message

sophie.hoss@gmail.com wrote:

```
>    s=strarr(cols)
>    n=0
>    while (~ eof(file) and (n lt rows_data -1 )) do begin
>     ; Read a line of data
>     readf,lun,s
>     ; Store it in data
>      data[*,n]=s
>      n=n+1       (*****)
>    end
>    data=data[*,0:n-1]
>
>
>  I did the while-loop because I learned that IDL might not read every
>  line separately. Don't know if it's the most elegant version but at
>  least it is one.
```

>
> for the sake of my sanity, any help is appreciated!
> cheers,
> sophie

Sophie,
when reading strings, the whole line is read. So when you read in S, you are reading nb_cols LINES.. s[0] = line0, s[1] = line1 etc. You would have to read the whole line then separate the columns.

You can do it in one step

data = strarr(nbLines)
readf,lun,data

Jean

Subject: Re: read multiple files with varying names
Posted by sophie.hoss on Fri, 29 May 2009 12:51:50 GMT
View Forum Message <> Reply to Message

On 28 Mai, 17:25, "Jean H." <jghas...@DELTHIS.ucalgary.ANDTHIS.ca>
wrote:
> sophie.h...@gmail.com wrote:
>> s=strarr(cols)
>> n=0
>> while (~ eof(file) and (n lt rows_data -1)) do begin
>> ; Read a line of data
>> readf,lun,s
>> ; Store it in data
>> data[*,n]=s
>> n=n+1 (*****)
>> end
>> data=data[*,0:n-1]
>
>> I did the while-loop because I learned that IDL might not read every
>> line separately. Don't know if it's the most elegant version but at
>> least it is one.
>
>> for the sake of my sanity, any help is appreciated!
>> cheers,
>> sophie
>
> Sophie,
> when reading strings, the whole line is read. So when you read in S, you
> are reading nb_cols LINES.. s[0] = line0, s[1] = line1 etc. You would
> have to read the whole line then separate the columns.

>
> You can do it in one step
>
> data = strarr(nbLines)
> readf,lun,data
>
> Jean- Zitierten Text ausblenden -
>
> - Zitierten Text anzeigen -

Jean, thanks for that, but I'm still stuck with the same problem no matter how I do it. I cut the code short to

```
CD, 'filepath'
ci_files = FILE_SEARCH('ci*', COUNT=num_ci_files_found)

for i = 0,(num_ci_files_found-1) do begin

 file = ci_files(i)
     rows = File_Lines(file)
    OpenR, lun, file, /Get_Lun
    header = StrArr(34)
     ReadF, lun, header
     Point_Lun, -lun, currentLocation
  data = strarr(12, rows-(n_elements(header)))  ; i know there's 12
columns
                readF, lun, data

                ; from here on I want to extract one column only and
save this column (radiation data from one station for one month) in a
file where I later add values from all other stations for this
specific month, then take the average from all stations over the whole
month)
```

--> READF: End of file encountered.

Do I need to include a format statement? I read some posts saying it is sometimes (in easy cases) better to forget about any format statements...

thanks for your input...
sophie, going crazy

Subject: Re: read multiple files with varying names
Posted by David Fanning on Fri, 29 May 2009 13:15:30 GMT
View Forum Message <> Reply to Message

sophie.hoss@gmail.com writes:

> Jean, thanks for that, but I'm still stuck with the same problem no
> matter how I do it. I cut the code short to
>
> CD, 'filepath'
> ci_files = FILE_SEARCH('ci*', COUNT=3Dnum_ci_files_found)
>
> for i = 0,(num_ci_files_found-1) do begin
>
> file = ci_files(i)
> rows = File_Lines(file)
> OpenR, lun, file, /Get_Lun
> header = StrArr(34)
> ReadF, lun, header
> Point_Lun, -lun, currentLocation
> data = strarr(12, rows-(n_elements(header))); i know there's 12 cols
> readF, lun, data
>
> ; from here on I want to extract one column only and
> save this column (radiation data from one station for one month) in a
> file where I later add values from all other stations for this
> specific month, then take the average from all stations over the whole
> month)
>
> --> READF: End of file encountered.

What this means is that you are trying to read data past the
end of the file. In other words, you are trying to read more
data than is actually in the file. In practical terms, it is
likely the number of rows is wrong, as you seem sure about the
number of columns.

Programs that count the number of rows in a file, such as FILE_LINES,
actually just counts the number of "end of line" characters found in
the file. This is an inexact science. Your data file may not have
an end-of-line character on the last line of the file. In which case,
FILE_LINES will report one less line that you actually have in the
file (not your problem here). Or, there may be extra, blank lines,
in the file (usually at the end of the file). Then, FILE_LINES would
report *more* lines than there are lines with actual data on them.
In which case you would try to read more data that exists and you
would get this error. This seems the most likely explanation to me. :-)

It may be you have to do some "conditioning" of your data files before
you try to read them.

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 multiple files with varying names
Posted by David Fanning on Fri, 29 May 2009 13:22:18 GMT
View Forum Message <> Reply to Message

David Fanning writes:

> It may be you have to do some "conditioning" of your data files before
> you try to read them.

A simple test would be to read all the lines in your file
as a string array, and then see if any of the strings were
blank:

```
file = 'datafile.dat'
rows = File_Lines(file)
data = StrArr(rows)
OpenR, lun, file, /Get_Lun
ReadF, lun, data
Free_lun, lun
I = Where( data EQ "", count)
IF count GT 0 THEN Print, "I got trouble!!"
```

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 multiple files with varying names
Posted by Jean H. on Fri, 29 May 2009 14:01:19 GMT
View Forum Message <> Reply to Message

sophie.hoss@gmail.com wrote:

> On 28 Mai, 17:25, "Jean H." <jghas...@DELTHIS.ucalgary.ANDTHIS.ca>
> wrote:
>> sophie.h...@gmail.com wrote:
>>> s=strarr(cols)
>>> n=0
>>> while (~ eof(file) and (n lt rows_data -1)) do begin
>>> ; Read a line of data
>>> readf,lun,s
>>> ; Store it in data
>>> data[*,n]=s
>>> n=n+1 (*****)
>>> end
>>> data=data[*,0:n-1]
>>> I did the while-loop because I learned that IDL might not read every
>>> line separately. Don't know if it's the most elegant version but at
>>> least it is one.
>>> for the sake of my sanity, any help is appreciated!
>>> cheers,
>>> sophie
>> Sophie,
>> when reading strings, the whole line is read. So when you read in S, you
>> are reading nb_cols LINES.. s[0] = line0, s[1] = line1 etc. You would
>> have to read the whole line then separate the columns.
>>
>> You can do it in one step
>>
>> data = strarr(nbLines)
>> readf,lun,data
>>
>> Jean- Zitierten Text ausblenden -
>>
>> - Zitierten Text anzeigen -
>
> Jean, thanks for that, but I'm still stuck with the same problem no
> matter how I do it. I cut the code short to
>
> CD, 'filepath'
> ci_files = FILE_SEARCH('ci*', COUNT=num_ci_files_found)
>
> for i = 0,(num_ci_files_found-1) do begin
>
> file = ci_files(i)
> rows = File_Lines(file)
> OpenR, lun, file, /Get_Lun
> header = StrArr(34)
> ReadF, lun, header
> Point_Lun, -lun, currentLocation
> data = strarr(12, rows-(n_elements(header))) ; i know there's 12

> columns
> readF, lun, data
>
> ; from here on I want to extract one column only and
> save this column (radiation data from one station for one month) in a
> file where I later add values from all other stations for this
> specific month, then take the average from all stations over the whole
> month)
>
> --> READF: End of file encountered.
>
> Do I need to include a format statement? I read some posts saying it
> is sometimes (in easy cases) better to forget about any format
> statements...
>
> thanks for your input...
> sophie, going crazy

Sophie,

you are still reading a text line, so without the format, the whole line
is stored where you believe one column value would be stored. In other
words, you are reading 12 * nb_lines LINES, which is obviously too much.

So, if possible, you would have to play with the formats...

For example, reading a simple file that contains:
a b c
d e f

can be read as:
data = strarr(3,2)
readf,lun,data, format = '(3A2)' ;3 times a text (A) of length 2 (the space)

but if you do
readf,lun,data
then you get the same error as you are know. And if you print data[0,0],
you get "a b c"

Jean

Subject: Re: read multiple files with varying names
Posted by sophie.hoss on Fri, 29 May 2009 15:54:50 GMT
View Forum Message <> Reply to Message

On 29 Mai, 15:22, David Fanning <n...@dfanning.com> wrote:
> David Fanning writes:

>> It may be you have to do some "conditioning" of your data files before
>> you try to read them.
>
> A simple test would be to read all the lines in your file
> as a string array, and then see if any of the strings were
> blank:
>
> file = 'datafile.dat'
> rows = File_Lines(file)
> data = StrArr(rows)
> OpenR, lun, file, /Get_Lun
> ReadF, lun, data
> Free_lun, lun
> I = Where(data EQ "", count)
> IF count GT 0 THEN Print, "I got trouble!!"
>
> 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.")

David, that was also my first guess, so i took a look at the .dat-file
(opened it in an editor) and it seems to me as if the last line was
indeed empty (cursor can be placed into the last line which is empty).
then I did the conditioning, but it says there is no "empty" lines
(count = 0).
however, i think i made it work now - at least that's what my tired
brain thinks.
I'll take a closer look into this after the weekend.
bye everybody

Subject: Re: read multiple files with varying names
Posted by David Fanning on Fri, 29 May 2009 16:39:38 GMT
View Forum Message <> Reply to Message

sophie.hoss@gmail.com writes:

> David, that was also my first guess, so i took a look at the .dat-file
> (opened it in an editor) and it seems to me as if the last line was
> indeed empty (cursor can be placed into the last line which is empty).
> then I did the conditioning, but it says there is no "empty" lines
> (count =3D 0).

> however, i think i made it work now - at least that's what my tired
> brain thinks.
> I'll take a closer look into this after the weekend.
> bye everybody

Even more pernicious, sometimes those empty lines will have
space characters on them. Aaauuuggghhh! Try doing a string
compress on the string array before checking for empty
strings.

```
file = 'datafile.dat'
rows = File_Lines(file)
data = StrArr(rows   OpenR, lun, file, /Get_Lun
ReadF, lun, data
Free_lun, lun
data = StrCompress(data, /REMOVE_ALL)
I = Where( data EQ "", count)
IF count GT 0 THEN Print, "I got trouble!!"
```

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 multiple files with varying names
Posted by David Fanning on Fri, 29 May 2009 16:45:58 GMT
View Forum Message <> Reply to Message

David Fanning writes:

> Even more pernicious, sometimes those empty lines will have
> space characters on them. Aaauuuggghhh!

By the way, it is *extremely* important that you not work
with these kinds of data files if you have firearms nearby.
Homicidal rage directed at the creator of such data files
is the rule, rather than the exception. :-(

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.")