
Subject: Re: need a quicker way to read ascii file w/a structure

Posted by [m.hadfield](#) on Wed, 06 Jun 2001 00:40:26 GMT

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

From: "Lucas Miller" <differentiable@hotmail.com>

> OK. I've got an ascii file laid out in columns with a string in the
> first column, like this
>
> yyyy-ddd // hh:mm:ss.ms 0 12 -1.00 -1.00 -1.00 -1.00
> (lots and lots of rows)
>
> I defined a structure
> mystruct = {utctime: ' ', mpos: 0, sector: 0, \$
> arate: 0.0, brate: 0.0, crate: 0.0, drate: 0.0}
>
> I know the number of rows so I replicate,
> data = replicate(mystruct, num_of_rows)
>
> and I read,
> readf, myfilelun, data
>
> and find out that the entire row is read into my mystruct.utctime
> string variable.

Yes, because mystruct.utctime is a string and strings are "greedy" in read operations unless the field width is specified explicitly.

yyyy-ddd // hh:mm:ss.ms 0 12 -1.00 -1.00 -1.00 -1.00

> I don't want to use an
> explicit format statement because the format varies....

That's a pity. If the widths of the fields were fixed then an explicit format would work.

Have you considered going to the person who generated the text files and suggesting she produce them with fixed widths. While you're at it, ask her why she couldn't have done that in the first place.

> ...The only thing I can think of doing at
> this point is reading my file in as a string array, separating it into
> variables, and feeding it into my structure.

Well, you don't have to load the entire file into an array, you can read each line into a scalar string, get the data you want and load it into the output array, then discard the line.

> This would

> take a LOT of time...

Not all that much in my experience. Read each line, split off the time string, read the numbers out of the remainder with a reads statement--it wouldn't take much longer to write the code than to describe it.

> ...and besides isn't there a more elegant way to do
> it?

Not that I can think of.

I have written a lot of routines to process text data files and I have found that the formats are so variable and loosely defined that a general solution is not possible and not worth attempting.

Mark Hadfield
m.hadfield@niwa.cri.nz <http://katipo.niwa.cri.nz/~hadfield>
National Institute for Water and Atmospheric Research

--

Posted from clam.niwa.cri.nz [202.36.29.1]
via Mailgate.ORG Server - <http://www.Mailgate.ORG>

Subject: Re: need a quicker way to read ascii file w/a structure
Posted by [R.Bauer](#) on Wed, 06 Jun 2001 07:03:01 GMT
[View Forum Message](#) <> [Reply to Message](#)

Mark Hadfield wrote:

>
> From: "Lucas Miller" <differentiable@hotmail.com>
>> OK. I've got an ascii file laid out in columns with a string in the
>> first column, like this
>>
>> yyyy-ddd // hh:mm:ss.ms 0 12 -1.00 -1.00 -1.00 -1.00
>> (lots and lots of rows)
>>
>> I defined a structure
>> mystruct = {utctime: ' ', mpos: 0, sector: 0,\$
>> arate: 0.0, brate: 0.0, crate: 0.0, drate: 0.0}
>>
>> I know the number of rows so I replicate,
>> data = replicate(mystruct,num_of_rows)
>>
>> and I read,

```

>> readf, myfilelun, data
>>
>> and find out that the entire row is read into my mystruct.utctime
>> string variable.
>
> Yes, because mystruct.utctime is a string and strings are "greedy" in read
> operations unless the field width is specified explicitly.
>
> yyyy-ddd // hh:mm:ss.ms  0 12 -1.00 -1.00 -1.00 -1.00
>
>> I don't want to use an
>> explicit format statement because the format varies....
>
> That's a pity. If the widths of the fields were fixed then an explicit
> format would work.
>
> Have you considered going to the person who generated the text files and
> suggesting she produce them with fixed widths. While you're at it, ask her
> why she couldn't have done that in the first place.

```

The best is to ask about a non string date/time identifier.
 For example you can use seconds or julian seconds.
 julian seconds are defined by Ray Sterner
 as seconds since 2000-1-1 00:00:00 UTC.

For this timeformat are lots of routines available to format them
 in other time formats.

regards

Reimar

```

>
>> ...The only thing I can think of doing at
>> this point is reading my file in as a string array, separating it into
>> variables, and feeding it into my structure.
>
> Well, you don't have to load the entire file into an array, you can read
> each line into a scalar string, get the data you want and load it into the
> output array, then discard the line.
>
>> This would
>> take a LOT of time...
>
> Not all that much in my experience. Read each line, split off the time
> string, read the numbers out of the remainder with a reads statement--it
> wouldn't take much longer to write the code than to describe it.

```

>
>> ...and besides isn't there a more elegant way to do
>> it?
>
> Not that I can think of.
>
> I have written a lot of routines to process text data files and I have found
> that the formats are so variable and loosely defined that a general solution
> is not possible and not worth attempting.
>
> ---
> Mark Hadfield
> m.hadfield@niwa.cri.nz <http://katipo.niwa.cri.nz/~hadfield>
> National Institute for Water and Atmospheric Research
>
> --
> Posted from clam.niwa.cri.nz [202.36.29.1]
> via Mailgate.ORG Server - <http://www.Mailgate.ORG>

--
Reimar Bauer

Institut fuer Stratosphaerische Chemie (ICG-1)
Forschungszentrum Juelich
email: R.Bauer@fz-juelich.de
<http://www.fz-juelich.de/icg/icg1/>

=====

a IDL library at Forschungszentrum Jülich
http://www.fz-juelich.de/icg/icg1/idl_icglib/idl_lib_intro.html

<http://www.fz-juelich.de/zb/text/publikation/juel3786.html>

Subject: Re: need a quicker way to read ascii file w/a structure
Posted by [dirk](#) on Wed, 06 Jun 2001 19:22:09 GMT
[View Forum Message](#) <> [Reply to Message](#)

In article <003701c0ee21\$447c6b70\$d938a8c0@hadfield>,
Mark Hadfield <m.hadfield@niwa.cri.nz> wrote:
> From: "Lucas Miller" <differentiable@hotmail.com>
>> OK. I've got an ascii file laid out in columns with a string in the
>> first column, like this
>>
>> yyyy-ddd // hh:mm:ss.ms 0 12 -1.00 -1.00 -1.00 -1.00
>> (lots and lots of rows)
>>
>> I defined a structure
>> mystruct = {utctime: ' ', mpos: 0, sector: 0, \$

```
>>      arate: 0.0, brate: 0.0, crate: 0.0, drate: 0.0}
>>
>> I know the number of rows so I replicate,
>> data = replicate(mystruct,num_of_rows)
>>
[snip]
>
> Yes, because mystruct.utctime is a string and strings are "greedy" in read
> operations unless the field width is specified explicitly.
>
> yyyy-ddd // hh:mm:ss.ms  0 12 -1.00 -1.00 -1.00 -1.00
>
[snip]
> I have written a lot of routines to process text data files and I have found
> that the formats are so variable and loosely defined that a general solution
> is not possible and not worth attempting.
```

I have had a lot of success with the function readcol.pro found on the idlastro archives. <http://idlastro.gsfc.nasa.gov/contents.html>

Get all the supporting programs for this, it's worth it.

Basically, it will separate each line into tokens of various type and stick those values in arrays that you specify. It will be easy to put them in the structure from there. Since it is based on the token idea, it can deal with varying formats, as long as there are spaces between the tokens.

Good luck.

- Dirk
