Subject: Re: readcol procedure
Posted by R.Bauer on Sun, 15 Sep 2002 15:16:14 GMT
View Forum Message <> Reply to Message

Reimar Bauer wrote:

> Pavel A. Romashkin wrote:
>> Reimar Bauer wrote:
>>
>>> if you use the eof method you have to read line by line. As you know idl
>>> is an array orientated language so reading in an array is much faster.
>>> It's really fast. If you have only 10 lines it doesn't matter but
>>> sometimes we got datafiles of nearly 100.000 lines. In this case it is
>>> very important.
>>
>>
>> I am sorry to disagree.
>> I routinely read large (60k-200k rows) ASCII files with unknown number
>> of lines. I always use large arrays to read into and never ever use EOF
>> with line by line reading.
>> All I have to do is to catch I/O error in case my buffer array is too
>> big as my reading approaches the end of file, then look up what size it
>> should have actually been, resize the buffer, then read the last portion
>> of file only. Reading a file with 80k lines using this method takes
>> about 0.1 s.
>> Take a look:
>> http://www.ainaco.com/idl/idl_library/read_ascii_columns.pro
>> Cheers,
>> Pavel
>
> I don't understand where you are disagree.
> I will try a comparison with the usb device and no file cache or
> how should comparisons be done?

I did a test today of both routines on my usb 1.1 device which could probaly have a max speed of 1MByte/sec.
I learned that's umounting und remounting the device clears the cache.

To test only the reading speed I have both routines compiled by my compile routine into a sav file which will loaded if the routine is called.
The testfile of 100000 lines by sindgen was altered in the first line with a column name which is useable as a structure name for read_ascii_columns.

Result is:

read_ascii_columns: 2.048 seconds
read_data_file: 4.418 seconds

read_ascii_columns speed goes linear with the speed of the device.
by read_data_file the most time is used for the interpretation of the data
from bytearray to data.
I believe it is possible to improve the routine a bit but at the moment it's
for us fast enough.
It would be fine to see read_ascii_columns with an autodetection of headers
and columns and a translation of header description in useful tagnames.
e.g. H2(ppm) isn't possible to set as a tagname.

regards

Reimar

--
Forschungszentrum Juelich
email: R.Bauer@fz-juelich.de
http://www.fz-juelich.de/icg/icg-i/
 ========== ======
a IDL library at ForschungsZentrum Juelich
 http://www.fz-juelich.de/icg/icg-i/idl_icglib/idl_lib_intro. html