Subject: Faster way to search/split a string?
Posted by rjp23 on Wed, 23 Nov 2011 12:13:39 GMT
View Forum Message <> Reply to Message

I was hoping that someone might have a cleverer way of approaching this problem.

The following command is the bootleneck in my code:

row_of_data=strsplit(all_rows[(where(stregex(all_rows, id, /boolean) eq 1))[0]],' ', /extract)

I have a large text file with lots of columns of data (which I don't know exactly what the columns are until I've read them in). There are then say 10000 rows of this data.

This is all read into one large string array (all_rows) which contains each row as a single very long string.
The first 20 characters of the row contain a unique id which I need to search the rows for and then extract the entire matching row. This row then needs to be split up into it's columns (space delimited).

Hopefully that all makes sense.

The problem is having to do this 10000 times, (once for for each id) is very slow and the time to do all of the other stuff done in the code, reading, writing, some maths, etc is being dominated by this one line.

Any thoughts or suggestions?

Cheers

Rob

p.s. This needs to be GDL compatible as well which I think most solutions would be anyway.

Subject: Re: Faster way to search/split a string?
Posted by Vincent Sarago on Wed, 23 Nov 2011 18:15:56 GMT
View Forum Message <> Reply to Message

I'm not sure to understand, maybe with examples it will be easier to me.

tmp = stregex(all_rows, '^[a-zA-Z0-9]{20}',/extract) ; Array of all IDs

```
test = uniq[sort(tmp)] ; Array determining witch ID is uniq

for ii = 0, n_elements(test) - 1 do begin

  id = tmp[test[ii]]

  test = where(tmp eq id, count)
  if count ne 0 then begin
    tmp2 = strsplit(all_rows[test], ",/extract) ; Array of ID + Other but split

   void = all_rows[test] ; all row for uniq ID
subset = stregex(b[ii], string(id[ii], format='("[^",a,"].+")'), /extract)
    ;do what you need to do here



  endif

endfor
```