
Subject: Re: Faster way to search/split a string?
Posted by [rjp23](#) on Thu, 24 Nov 2011 14:16:03 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Nov 24, 9:15 am, wlandsman <wlands...@gmail.com> wrote:

> Some thoughts (though I am not certain I understand the situation):
>
> 1. If you don't need regular expressions then I believe that using STRPOS() is quicker than using STREGEX().
> 2. If the ID is always in the first 20 characters of ALL_ROWS then I would created a new vector row_id = strmid(all_rows,0,20) and search on that.
> 3. If the ID is always exactly 20 characters, you could use a program like <http://idlastro.gsfc.nasa.gov/ftp/pro/misc/match.proto> then find the matching indices in row_id and id. This is similar to David's suggestion of sorting and using VALUE_LOCATE.
> 4. You want to make only a single call to STRSPLIT() for all 10,000 rows. Since IDL V8.0, STRSPLIT returns a list data type when supplied with an array (since in principle each string element could have a different number of "columns"). If you have an earlier version of IDL -- or if this capability is not available in GDL, then I would use the vector capability of STRMID (http://www.idlcoyote.com/code_tips/strmidvec.html)
>
> --Wayne

Thanks for that. Doing #2 made a huge difference along with David's suggestion.

On a related note but a different problem (and I know people aren't GDL experts here so apologies) I'm finding that strsplit in GDL is much slower than in IDL. Has anyone come across this before? I've tried searching the GDL help but it's obviously not as extensive as here.

I also just wanted to check that I'm doing this the optimum way for IDL.

If I have a string array like: array=[abc_001, abc_002, abc_003, bcdef_001, bcdef_002, cdf_001_001, cdf_001_002]

I've assumed the fastest way to get the start section characters is to do:

```
for loop=0, n_elements(array)-1 do begin
start_section[loop]=(strsplit(array[loop], '_', /extract))[0]
endfor
```

or is there a vectorised way to use strsplit in IDL? (limited to IDL 7 and/or GDL)
