Subject: Re: Faster way to search/split a string?
Posted by rjp23 on Thu, 24 Nov 2011 14:41:31 GMT
View Forum Message <> Reply to Message

On Nov 24, 2:16 pm, Rob <rj...@le.ac.uk> wrote:
> 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 likehttp://idlastro.gsfc.nasa.gov/ftp/pro/misc/match.prototh en 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

Ah after re-reading #4 a few times I realised it's what I was asking

strmid(variable_names, 0, transpose(strpos(variable_names, '_')))

Thanks :-)