Subject: Re: multiple delimiters
Posted by Martin Schultz on Fri, 15 Sep 2000 07:00:00 GMT
View Forum Message <> Reply to Message

nrh@imag.wsahs.nsw.gov.au wrote:

>

- > Well, its actually a whole heap of strings, most fields separated by
- > blanks, and some fields, where there is more than one word, are
- > encased by quotation marks. The fields inside the quotes have spaces as
- > well, but we want them to be all one field, if you know what I mean.
- > Right now we pull out the strings within the quotes, replace all the
- > spaces with '\_', put it back in, remove the quotes and then we can use
- > the strsplit function to remove the extra white spaces created by
- > replacing the quotes.
- > so, in a nutshell, we have:
- > ....PROJECTION-R OTYP DP EXTN img PROC "CM CARDIAC MIBI" .....
- > and make it to be(through many painful string ops it is a huge
- > database file)
- > PROJECTION-R OTYP DP EXTN img PROC CM CARDIAC MIBI ........
- > and then we have to arrange it in a struct as every second field is the
- > info we actually need. Odd fields are the descriptors.
- > Clear as mud?

Your life could be a LOT easier if you had a formatted output, i.e. if all columns are aligned (I am sure the database that you are using should be able to produce this). Then you could simply use a formatted read statement

```
readf, lun, proj, otyp, dp, extn, img, proc, label, ..., format='(i4,i6,...,A20,...)'
or (even more elegantly) read into a structure temp = { proj:1.0, otyp:0L, ..., label:", ... }
readf, lun, temp, format='...'
```

It probably boils down to workflow optimization: If you have to do it once, don't bother and just use a simple code. If you have to do it several times - always with the same data set - convert the data set once into a better format (something binary to speed up reading, optimally a scientific data format for they allow better usability and are self-describing). If you need to do this several times with changing data sets, make sure the original data set producer change their format;-)

Cheers, Martin

--