Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by guillermo.castilla.ca on Sun, 03 May 2009 00:47:51 GMT
View Forum Message <> Reply to Message

On May 1, 12:36 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
> On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>
>
>> On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>
>>> Hello, everyone!
>
>>> Is there anyone knows a routine in IDL that be capable to remove
>>> duplicate elements from a multi-dimensional array efficiently? I 'm
>>> now working with huge arrays, and I have written one by myself, it
>>> works but is with low efficiency.
>
>>> example of my problem:
>>> the input array:
>>> 1,10,9,100,200
>>> 2,11,8,101,201
>>> 2,11,8,101,201
>>> 3,10,9,100,200
>>> 4,7,12,99,199
>>> 2,11,8,101,201
>
>>> goal:
>>> remove the duplicate elements with the same values for the second and
>>> the third column.
>
>>> expected output:
>>> 1,10,9,100,200
>>> 2,11,8,101,201
>>> 4,7,12,99,199
>
>>> Thanks for your help!
>
>>> Bo
>

If you don't have handy that ORD function Jeremy pointed out (I didn't know of it), and assuming your array is of byte type, you can do the following:

```
input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
      [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
```

```
keep = Where(Histogram(1000L*input[1,*]+input[2,*], rev=r) GT 0)
keep = r[r[keep]]
print, input[*,keep[sort(keep)]]
      1      10       9     100     200
      2      11       8     101     201
      4       7      12      99     199
```

Cheers

Guillermo
