Subject: Re: fast array comparison
Posted by David Fanning on Sun, 08 Dec 2002 17:45:00 GMT
View Forum Message <> Reply to Message

Sean Raffuse (sean@me.wustl.edu) writes:

> Let's say I have two arrays.
>
> requested_array = [5,6,7,8,9,10]
> available_array = [3,7,8,9,12,13,16]
>
> What is the absolute fastest way to determine the indices of available_array
> that contain values in requested_array? The indices need not match. i.e.,
> if the two arrays above were used, I would like to return index=[1,2,3]
> because the requested values 7, 8 and 9 are in the available array.

The absolute fastest way MUST involve histograms, so
I maintain with a great deal of confidence (say, in the 40-50
percent range) that this is the fastest possible algorithm:

```
 ********************************************************** ************
FUNCTION SetIntersection, a, b, Indices=indices
minab = Min(a, Max=maxa) > Min(b, Max=maxb) ;Only need intersection of
ranges
maxab = maxa < maxb

  ; If either set is empty, or ranges don't intersect: result = NULL.

IF maxab LT minab OR maxab LT 0 THEN RETURN, -1
r = Where((Histogram(a, Min=minab, Max=maxab) NE 0) AND  $
        (Histogram(b, Min=minab, Max=maxab) NE 0), count)
IF Arg_Present(indices) THEN $
   indices = Where((Histogram(a, Min=minab, Max=maxab) NE 0))
IF count EQ 0 THEN RETURN, -1 ELSE RETURN, r + minab
END
 ********************************************************** ************
```

Use it like this:

```
  IDL> request_array = [5,6,7,8,9,10]
  IDL> avail_array = [3,7,8,9,12,13,16]
  IDL> int = setintersection(avail_array, request_array, Indices=i)
  IDL> print, int, i
        7       8       9
        2       3       4
```

Cheers,

David

--
David W. Fanning, Ph.D.
Fanning Software Consulting, Inc.
Phone: 970-221-0438, E-mail: david@dfanning.com
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Toll-Free IDL Book Orders: 1-888-461-0155
