
Subject: Re: Intersection of 2 sets--Beginner IDL question
Posted by [John-David T. Smith](#) on Wed, 17 Oct 2001 05:00:34 GMT
[View Forum Message](#) <> [Reply to Message](#)

David Fanning wrote:

>
> Ted Cary (tedcary@yahoo.com) writes:
>
>> Is there any programming technique for finding the intersection of two sets
>> (arrays) of numbers without using WHERE in a loop to search the larger array
>> for every element in the smaller array? It seems like a very clumsy way to
>> find values shared by both arrays, especially with integer sets/arrays.
>
> How about the very tiny program, SetIntersection,
> which uses--what else--a Histogram! :-)
>
> http://www.dfanning.com/tips/set_operations.html

It's amazing how much recycled information flows through the newsgroup, if you watch it long enough. I remember just like it was 4 years ago the detailed discussions with which we whiled away our days, concerning value-based intersection vs index-based intersection, order N vs. unknown order operations, etc.

I do, however, still cringe when I read on that page of yours:

"These routines are faster than previously published functions, e.g. Contain..."

since I whipped up contain() to prevent the problem of an operation scaling out of control on sparse data sets (as set_intersection does -- a feature they warn you of). You can read all about a variety of intersection methods if you like:

<http://groups.google.com/groups?selm=38CBF8B6.5BF0AB50%40ast.ro.cornell.edu>

In a classic example posed by Mark Fardal, you are matching up social security numbers in two lists containing age and income. The set_intersection style solution fails miserably here, and to a lesser degree for any arrays which are somewhat sparse (where *somewhat* seems to be about 1 in 10, depending on lots of factors).

Ah, the glory days.

JD
