Subject: Re: matching lists
Posted by Craig Markwardt on Fri, 10 Mar 2000 08:00:00 GMT
View Forum Message <> Reply to Message

"J.D. Smith" <jdsmith@astro.cornell.edu> writes:
>
> Mark Fardal wrote:
>>
>> Hi,
>>
>> I have been looking at properties of particles in a simulation, and
>> sometimes need to match up the particles in two different subsets. I
>> typically have (quantity A, index #) for one set of particles, and
>> (quantity B, index #) for another set, and want to compare quantities
>> A and B for the particles that are in both sets.
>>
>> As of late last night I could not think of a good way to do this;
>> WHERE inside a for-loop would be very slow. Maybe I'm missing
>> something easy, but in any case here's a solution inspired by the
>> recently submitted SETINTERSECTION function. Hope somebody finds
>> it useful.
>>
>
> The standard where_array, as posted a few years back, and modified
> slightly for the case of the null intersection, is attached. It
> will work with floating point and other data types also. It works
> by inflating the vectors input to 2-d and testing for equality in
> one go. It will also handle the case of repeated entries.
> ...

I also submit CMSET_OP, a function I recently posted on my web page.
(Actually, I'm not sure if Mark is referring to that by
SETINTERSECTION).

Advantages are:
 * works on any numeric or string data type
 * works in order (n1+n2)*log(n1+n2) time or better, rather than n1*n2
 * uses the histogram technique for short integer lists as JD suggests
 * also does "union" and "exclusive or"
 * also does A and NOT B or vice versa

Disadvantages:
 * it removes duplicates, treating the two lists strictly as sets.
 * returns values, not indices

Craig

http://cow.physics.wisc.edu/~craigm/idl/idl.html

--
---------- ----------
Craig B. Markwardt, Ph.D. EMAIL: craigmnet@cow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
---------- ----------