Subject: Re: matching lists
Posted by John-David T. Smith on Mon, 13 Mar 2000 08:00:00 GMT
View Forum Message <> Reply to Message

Mark Fardal wrote:
>
> Hi,
>
> Somehow I knew that if I mentioned HISTOGRAM, that would stir up some
> real IDL programmers. :->
>
> J.D.'s sort method seems like the winner. The modest speed advantage
> of the histogram method in certain cases is not important. If you are
> in a situation where just matching up the elements is the limitation,
> you are probably going to be in trouble doing any analysis with them
> (let alone reading them in).
>
> The problem of repeated elements, which is the only advantage of
> WHERE_ARRAY, is not of any concern, at least to me. The point of
> the key variables a and b is that they are supposed to be unique
> identifiers. I would just like the routine not to break completely
> in case the same element was copied into the arrays twice. The
> sort method does fine in that respect (finds the last of the
> duplicate elements in a and the first in b).

By the way, I found an implementation I had mentioned a while back on the news group but had forgotten about from the nasa lib called "match" which does pretty much the same thing. It's probably less efficient, since it uses an auxiliary list of indices in addition to the flag vector, instead of just using the sort() results directly, and performs a few more "where" tests as a result. But a similar idea, written first in 1986! Match() is also more immune to changes in sort() than my routine, as a result of carrying around these additional index arrays.

Take a look.

JD

--
J.D. Smith |*| WORK: (607) 255-5842
Cornell University Dept. of Astronomy |*| (607) 255-6263
304 Space Sciences Bldg. |*| FAX: (607) 255-5875
Ithaca, NY 14853 |*|