Subject: Re: Locating sequence of bytes within binary file
Posted by JDS on Fri, 18 Jun 2010 22:12:52 GMT
View Forum Message <> Reply to Message

On Jun 18, 12:47 pm, Axel Martínez <axe...@gmail.com> wrote:
>> As usual, we can (over-)use IDL's array strengths to brute-force this
>> using REBIN and dimensional TOTAL:
>
> The problem of this solution might be the size of the variable quill
> for a large "needle" (not my case).

Memory usage is actually fairly stable with needle size, and
comparable to the other solution. The larger the needle, the fewer
the rows. Put another way, "quill" is always of order the size of the
search array, and no larger.

> And also the speed, as you acknowledge...

This is where a large needle hurts you; the need to wrap the search
array around in n_needle steps.

>> That motivates another deeply unsatisfying, but resoundingly faster
>> (20-50x) option: simply replace all 0b's with 1b's in both input byte
>> array and search array, and just double-check for spurious matches as
>> you go:
>
> This is good, but it just uses double the amount of memory to preserve
> the original and the modified array. Works for me, but could be bad
> for others. And, honestly, it is really not so elegant.

Agreed. That said, if elegance is on your short list of programming
language features, I fear you shouldn't go anywhere near IDL. I think
the other message here is that finding where a given value occurs in
an array is, although seemingly just a short extrapolation away,
actually a completely different problem from finding where a sequence
of values appears. The latter has considerable room for optimization
over the former. For example, you might notice that STRPOS runs in a
time that is almost independent of needle length. Some algorithms run
faster the longer the needle: http://en.wikipedia.org/wiki/BoyerMoore_string_search_algori thm.
Brains always triumph over brute-force.

>> We might also lobby ITT to let STRPOS and its sort accept byte arrays
>> (since frankly there is very little difference between them
>> internally).
>
> This would be by far the best solution. I believe a VIP like you have
> more options to lobby ITT than I do :)

I don't know about that. Most new features start with a simple
request to ITT support; give it a try.

JD