

Re: Speeding up an application – general rules

Source: <http://coding.derkeiler.com/Archive/Perl/comp.lang.perl.misc/2006-12/msg01410.html>

- *From:* "Petyr David" <phynkel@xxxxxxxx>
 - *Date:* 22 Dec 2006 04:43:12 -0800
-

I will also review those URLs. Creating an app that did indexing of the files did not come up as this script came from a far simpler one that merely found files matching the single pattern and printed a link to the file. I also don't have the time to make this a full time job. Something was needed quick and dirty and that's what they got : -)

TX

On Dec 22, 4:28 am, "Todd W" <t...@xxxxxxxxxxxxxx> wrote:

"Petyr David" <phyn...@xxxxxxxx> wrote in
messagenews:1166757223.858558.144370@xx

I have a small Perl application that searches through a series of directories chosen by the user for files containing a pattern or group of patterns. The file names and matching patterns are returned to the user sorted by the file's modification time. The user also has the choice of how far back in time to search and how many lines of output he wants to see for each file.

With an expected and current increase of files and file sizes, the application is bogging down a bit. I didn't design it with performance in mind and I will be reviewing what I've done, but are there general rules or specific suggestions you could offer to enhance performance?

Basically: the script uses perl's system command to run a long winded "find" command which is piped to sed to correct patterns that match HTML markers. The matching lines are then shoved into an array. The elements of the array are moved into a hash for the purpose of sorting the file names. Then file names and matching lines are printed.

Re: Speeding up an application – general rules

Q: Can I speed things by eliminating the sed command and letting Perl filter and modify the matching patterns? If so, how much of a performance gain?

Is using Perl's grep to search through every file for the pattern faster than using the find command? The find command has the advantage that I can search for files of a certain date rather easily. Again: could that be done more rapidly by Perl's looking at the file's mod time?

Any thoughts or suggestions would be appreciated. The conventional way of doing what you are proposing is some how building an

index of the files. Your index interface then gives you pointers to results when a search is performed. If the data changes regularly, you also have to regularly reindex your files.

I've been using htdig in some form or another to accomplish what you suggest.

Your post, though, caused me to take another look on CPAN for relevant modules as I was sure the state of this technology has improved since I decided to use htdig (several years ago). The following module looks very promising:

<http://search.cpan.org/~dpavlin/Search-Estraiier-0.08/>

I think I'm going to give it a try as my next search engine. Heres another one that looks interesting:

<http://search.cpan.org/~snkwatt/Search-FreeText-0.05/>

I found these modules by going to:

<http://search.cpan.org/search?query=search&mode=all>

Enjoy,

Todd W.