

Re: STRANGE QUESTION : search engine query and web site dowload

Source: http://coding.derkeiler.com/Archive/C_CPP/comp.lang.cpp/2003-10/4084.html

From: Moonlit (*alt.spam_at_jupiter.universe*)

Date: 10/27/03

Date: Mon, 27 Oct 2003 18:28:02 +0100

Hi,

"igthibau" <igthibau@wanadoo.fr> wrote in message
news:bnheu6\$1tg\$1@news-reader2.wanadoo.fr...

> *Hmmmmm not sure I understood all that, but it looks as though I am
embarking*

> *on network programming...*

You sure do ;-)

About sockets

<http://www.developerweb.net/sock-faq/flatfaq.php>

Don't know where but if you google for RFC and HTTP you probably find lots
of info on the hypertext protocol.

The first thing you might try if you have written the basics (a simple
network connection) is sending the following 4 lines (2 empty returns at the
end) Host should be the host you connect. Use port 80. (You could actually
try this with telnet <host> 80). This should get you back the main page of
that site.

GET / HTTP1/1

host: <fill in the server to connect to>

Regards, Ron AF Greve.

> *on which I have not a clue.*

> *any book / reference I might find useful ?*

>

> *thanks*

> *G.*

>

> *Moonlit wrote:*

>

> > *Hi,*

> >

>>
>> *"igthibau" <igthibau@wanadoo.fr> wrote in message*
>> *news:bnfug9\$ihe\$1@news-reader3.wanadoo.fr...*
>>> *Hello everyone,*
>>> *well, the title says pretty much all.*
>>> *First off though I'd like to say I have no idea what ng I should post*
>>> *this to. I chose this one because I am more likely to understand*
>>> *c-based answers than other language's.*
>>> *So, my problem is simple : to submit a query to a search engine, one*
goes
>> *to*
>>> *, say www.google.com, enters keywords and a long list appears. Then one*
>>> *clicks onto the appropriate hyperlink to go to the chosen site.*
>>> *Question : how does one do that using scripts / fortran / C / html /*
>>> *whatnot. As in : using command line interfacing to as to enable, say,*
>> *loops*
>>> *or non-graphical interactivity?*
>>> *The idea is to be able to, within a program, call up a search engine,*
run
>> *it*
>>> *using keywords and then download the first X sites.*
>>
>> *Well would you should do is create a connection to the server (port 80)*
>> *using a socket. Then you would send information to it and read from it.*
>> *The protocol used is described in the http rfc (google for it). You*
could
>> *then create a query and scan the pages for the results. Of course you*
have
>> *to look at the pages to see how the results are returned.*
>>
>> *There is also a library available somewhere that let you build for*
>> *instance a browser, It has al kinds of http function which might save*
you
>> *a lot of time.*
>>
>> *I actually wrote an app that talks to our local SMS http server in the*
>> *company to send sms to mobile phones. I just intercepted all trafic from*
>> *my webbrower to the site and back to see how they communicated, then I*
>> *wrote a class that implemented a very small subset of the http*
protocol..
>> *Such an approach might work for you too.*
>>
>>
>>>
>>> *There, I said it. I have no idea where to start from let alone what*
>> *langage*
>>> *to use and would hugely appreciate any help whatsoever !!!*
>>>
>>> *thanks to all.*
>>> *G.*
>>

comp.lang.c++: Re: STRANGE QUESTION : search engine query and web site download

> > *Regards, Ron AF Greve.*

>