
Subject: Re: Download files from the web

Posted by on Mon, 16 Dec 2013 14:46:05 GMT

[View Forum Message](#) <> [Reply to Message](#)

Den måndagen den 16:e december 2013 kl. 15:14:10 UTC+1 skrev Mats Löfdahl:

> Den måndagen den 16:e december 2013 kl. 14:41:08 UTC+1 skrev Helder:

>

>>>> if it helps, i used the IDLnetUrl object and then use the getProperty method to get the Response_code value. Not sure if it helps, but I have been downloading files successfully with https. I also use the callback_function to make a progress bar.

>

>>>> Not sure if it helps, but it might be a place to start...

>

> Thanks. But it seems it has the same problem as the wget function, in that it can't tell the difference between a proper download and a 404 error web page.

>

> I simplified your code a bit (because I don't need the progress bar) and came up with this:

>

> function downloadurl, url, file

>

> url_scheme = (strsplit(url, ':',/extract))[0]

>

> url_hostname = strjoin((strsplit(url, '/',/extract))[1:*,'/'])

>

> oUrl = OBJ_NEW('IDLnetUrl', URL_SCHEME = url_scheme, URL_HOSTNAME = url_hostname)

>

> retrievedFilePath = oUrl->Get(FILENAME=file)

>

> oUrl->GetProperty, RESPONSE_CODE=RespCode ; 200 = OK

>

> oUrl->CloseConnections

>

> OBJ_DESTROY, oUrl

>

> return, RespCode eq 200 ; True if OK

>

> end

>

> I tried it both with a url pointing to an existing web page and to a non-existing page. In both cases I get RespCode eq 200. With the non-existing page I again had downloaded a 404 error page.

>

> I got the value 200 for OK from the list at

http://www.exelisvis.com/docs/IDLnetURL.html#objects_network_1009015_1417867

Thought it might work better to use spawn and wget and read its exit status. But that seems to have the same problem: 404 error page downloaded in case the remote file doesn't exist, but exit

status 0 (=OK) regardless.

So this does not seem to be an IDL problem. It is just hard to get the information I want from the download process.

The web server obviously knows the requested file does not exist but isn't there a way to make it tell the downloading process this in a more condensed way than constructing a web page with a 404 error?
