

---

Subject: Re: URL Parsing for wget in IDL

Posted by [Allan Whiteford](#) on Tue, 17 Jul 2007 16:34:44 GMT

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

---

Ben Panter wrote:

> Hi Folks,  
>  
> Does anyone know of a handy way of parsing URLs in IDL? Or even a nice  
> perl script to do it? The situation is that we have a server that can  
> accept SQL queries via URL encoding of a get command, which works in the  
> main part but falls over with '&' and '+' type syntax.  
>  
> I'll write a regexp version myself if necessary, but it's not trivial  
> - you have to ensure the ordering is correct as there are several  
> encodings which use characters which would other wise be replaced.  
>  
> cheers,  
>  
> Ben  
>

Ben,

Maybe I'm misunderstanding your problem but if you're just spawning wget  
then adding the /sh keyword might fix most of your problems.

e.g.:

url=' http://www.google.co.uk/webhp?ie=UTF-8&oe=UTF-8&hl=en&q=&tab=iw'

spawn,'wget '+url ; doesn't work

spawn,'wget '+url,/sh ; works

Thanks,

Allan

---

---

Subject: Re: URL Parsing for wget in IDL

Posted by [Ben Panter](#) on Tue, 17 Jul 2007 17:06:41 GMT

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

---

Allan Whiteford wrote:

> Maybe I'm misunderstanding your problem but if you're just spawning wget  
> then adding the /sh keyword might fix most of your problems.  
>  
> e.g.:

```
>
> url=' http://www.google.co.uk/webhp?ie=UTF-8&oe=UTF-8&hl=en&q=&tab=iw'
>
> spawn,'wget '+url ; doesn't work
> spawn,'wget '+url,/sh ; works
```

Hi Allan,

Thanks, but I'm afraid that doesn't help - the /sh keyword isn't necessary as I use an option on the wget (-o) that pipes it to stdout. There is no problem with the wget operation as long as there is no \_ in the query string, the problem is when I want to send the query itself as a URL - hence all spaces turn into %20, and + and ? confuse the cgi script at the other end. If the URL is encoded properly, this is not a problem - the other end can decode it fine... but the encoding is an issue.

Thanks for the suggestion though!

Ben

--

Ben Panter, Edinburgh, UK.  
Email false, <http://www.benpanter.co.uk>  
or you could try ben at ^^^^^^^^^^^

---

Subject: Re: URL Parsing for wget in IDL  
Posted by [Dick Jackson](#) on Tue, 17 Jul 2007 17:34:53 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

"Ben Panter" <[me@privacy.net](mailto:me@privacy.net)> wrote in message  
news:f7ions\$b7h\$1@scotsman.ed.ac.uk...

```
> Hi Folks,
>
> Does anyone know of a handy way of parsing URLs in IDL? Or even a nice perl
> script to do it? The situation is that we have a server that can accept SQL
> queries via URL encoding of a get command, which works in the main part but
> falls over with '&' and '+' type syntax.
>
> I'll write a regexp version myself if necessary, but it's not trivial - you
> have to ensure the ordering is correct as there are several encodings which
> use characters which would otherwise be replaced.
```

Hi Ben,

I just stumbled on this yesterday: IDL 6.4 comes with a new function called Parse\_URL. Quoting Online Help:

PARSE\_URL returns an anonymous structure containing the disassembled segments of the URL. The fields in the structure are:

SCHEME  
USERNAME  
PASSWORD  
HOST  
PORT  
PATH  
QUERY

Does that help?

Cheers,  
-Dick

--

Dick Jackson Software Consulting                      <http://www.d-jackson.com>  
Victoria, BC, Canada                      +1-250-220-6117                      [dick@d-jackson.com](mailto:dick@d-jackson.com)

---

---

Subject: Re: URL Parsing for wget in IDL  
Posted by [Marshall Perrin](#) on Tue, 17 Jul 2007 19:25:26 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Ben Panter <[me@privacy.net](mailto:me@privacy.net)> wrote:

> Thanks, but I'm afraid that doesn't help - the /sh keyword isn't  
> necessary as I use an option on the wget (-o) that pipes it to stdout.  
> There is no problem with the wget operation as long as there is no \_ in  
> the query string, the problem is when I want to send the query itself as  
> a URL - hence all spaces turn into %20, and + and ? confuse the cgi  
> script at the other end. If the URL is encoded properly, this is not a  
> problem - the other end can decode it fine... but the encoding is an issue.

Hi Ben,

I looked a little while ago and couldn't find any pre-existing IDL code to encode URL strings. But it's a pretty straightforward search-and-replace operation that shouldn't take too long to code up. In case you need it, there's a list of the required codes on Wikipedia: <http://en.wikipedia.org/wiki/Percent-encoding>

Incidentally, there are some scripts that access web pages directly from within IDL, such as webquery from the Goddard library. These may be slightly faster for repeated use, since they don't involve spawning external processes. I believe there's also a web access object in IDL 6.4 now, but I'm stuck back at 6.2 right now.

- Marshall

---

---

Subject: Re: URL Parsing for wget in IDL  
Posted by [Ben Panter](#) on Tue, 17 Jul 2007 21:45:11 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dick Jackson wrote:

> I just stumbled on this yesterday: IDL 6.4 comes with a new function called  
> Parse\_URL. Quoting Online Help:  
>  
> PARSE\_URL returns an anonymous structure containing the disassembled segments of  
> the URL. The fields in the structure are:

Thanks Dick - I want to go the other way, but a related object I found on the page is IDLnetURL, which might do some of the things that I would like. Will investigate further!

Thanks,

Ben

--  
Ben Panter, Edinburgh, UK.  
Email false, <http://www.benpanter.co.uk>  
or you could try ben at ^^^^^^^^^^^

---

---

Subject: Re: URL Parsing for wget in IDL  
Posted by [Ben Panter](#) on Tue, 17 Jul 2007 21:48:06 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thanks Marshall, I hadn't seen the wget function before and that could be very useful for the project. The wiki page is what I will have to base my encoding script if the object approach doesn't work.

Thanks!

Ben

--  
Ben Panter, Edinburgh, UK.  
Email false, <http://www.benpanter.co.uk>  
or you could try ben at ^^^^^^^^^^^

---