

Re: POSTing Chinese characters

Source:

<http://www.tech-archive.net/Archive/Internet-Server/microsoft.public.inetserver.iis/2004-10/1744.html>

From: David Wang [Msft] (*someone_at_online.microsoft.com*)

Date: 10/21/04

Date: Thu, 21 Oct 2004 15:54:00 -0700

There are specs that describe all of this (how URL is encoded and how forms parameters are encoded). I don't have them handy on me right now. If you are responsible for assembling the raw data, then you want to find, read, and understand those specs. They all have the same basic idea but different in implementation (due to design/scenarios). Feel free to web search for them.

The basic idea of %-encoding is to treat character encoding as a sequence of bytes (256 possible values), which can be %-encoded into another sequence of bytes (17 possible alphanumeric values) which can safely transport between systems since those 17 values remain idempotent between them. How you do this depends on who is encoding/decoding. If you have custom client/server, they can communicate however they want. Otherwise, your implementation will have to follow rules stated in server/client specs that govern their behavior.

What implementation are you actually having problems with? You were claiming that Pocket IE is having problems on the last post, and now you are saying you don't know how to do this with WinInet. Which is it?

Personally, I suggest that you get more familiar with how HTTP works and how character sets are transported between two machines regardless of their locale settings --- or else you will keep running into data truncation problems that look like "bugs" when in fact it isn't. No API can do this automatically for you --- you must know what you are doing if you want things to work. This is beyond the scope of a newsgroup discussion.

For example, your claims of POST'd data experiencing encoding problems does NOT make sense to me. POSTed form data is treated as opaque by the transport system (otherwise, binary file downloads won't work) --- but in your case, your POSTed data will be interpreted as forms --- meaning it has to follow the same set of encoding rules as if you sent it on the url via a FORM GET.

So, you are really talking about whether your input follows those rules --- and I suspect that data truncation happens because it does not. Please clarify the following:

microsoft.public.inetsrvr.iis: Re: POSTing Chinese characters

1. Client locale
2. Character encoding of data on the client side
3. Encoding used on transport from #2 to #4
4. Server locale

%-encoding is a good way to avoid lots of transport/encoding problems between the client/server, at the cost of extra encoding/decoding CPU costs, because it uses 17 characters that are idempotent between various character encodings.

```
--
//David
IIS
This posting is provided "AS IS" with no warranties, and confers no rights.
//
"Brian Burgess" <bburgess66@hotmail.com> wrote in message
news:OZ8vQPxtEHA.3152@TK2MSFTNGP14.phx.gbl...
Well I was meaning how to assemble the raw data ;-) ... %-encoding a
chinese character would mean %-encoding two bytes ... is that right? .. For
instance '??' would be encoded to '%9E4E1062%' .. is this correct?
Just a little background .. I'm not actually using the browser .. it is the
WinInet API instead .. (probably a little easier this way)
thx
-BB
"David Wang [Msft]" <someone@online.microsoft.com> wrote in message
news:upZrI8vtEHA.2116@TK2MSFTNGP14.phx.gbl...
> I don't do a lot of client-side web development, so I can't say how to do
> it.
>
> I'd imagine that with client side scripting in the web page you can use
the
> DOM in IE to access the values on the web page and %-encode them PRIOR to
> hitting submit (because after hitting submit, the web browser will take
the
> values from DOM and POST it to the web server -- it is at that moment that
> data corruption happens).
>
> --
> //David
> IIS
> This posting is provided "AS IS" with no warranties, and confers no
rights.
> //
> "Brian Burgess" <bburgess66@hotmail.com> wrote in message
> news:ewoFufktEHA.3364@TK2MSFTNGP10.phx.gbl...
> I did something similar to a net trace .. Finding your suspicion correct.
>
> How should %-encode the data to POST? .. can I get the byte values of
the
> chinese string somehow?
>
> thx
>
> -BB
>
> "David Wang [Msft]" <someone@online.microsoft.com> wrote in message
> news:OmtsvmVtEHA.1308@tk2msftngp13.phx.gbl...
> > Can you get the network trace coming into the IIS server, so that we can
> > conclusively state whether the issue is on the client or server.
> >
```

microsoft.public.inetsrv.iis: Re: POSTing Chinese characters

> > Right now, it seems that your issue is a Pocket IE issue.
> >
> > If it turns out to be something with the Pocket IE client, as with
> > mis-encoding issues, you are helpless on the server to "fix" anything.
> > Thus, I suggest you try to have client-side code which %-encodes
> everything
> > on the POST back to IIS so that you can work-around any client-side
> issues.
> > You can't do this on the server-side because the data loss would have
> > happened on the client prior to data transmission.
> >
> > --
> > //David
> > IIS
> > This posting is provided "AS IS" with no warranties, and confers no
> rights.
> > //
> > "Brian Burgess" <bburgess66@hotmail.com> wrote in message
> > news:e8z4FYosEHA.2124@TK2MSFTNGP11.phx.gbl...
> > Ok thx
> >
> > There is no Pocket IE groups (that I know of), only more generic WinCE
> and
> > Pocket PC groups.
> >
> > Thanks for the insight ..
> >
> > -BB
> >
> >
> > "Bernard" <qbernard@hotmail.com.discuss> wrote in message
> > news:%23cKFFfmsEHA.2804@TK2MSFTNGP14.phx.gbl...
> > > Well, what you can do is set the encoding and charsets at the targeted
> > page.
> > > but I doubt this would help you much, as you explained this only
happen
> > on
> > > pocketIE.
> > > and doesn't happen all the time.
> > >
> > > have you try pocket IE group ?
> > >
> > > --
> > > Regards,
> > > Bernard Cheah
> > > <http://www.tryiis.com/>
> > > <http://support.microsoft.com/>
> > > <http://www.msmvps.com/bernard/>
> > >
> > >
> > > "Brian Burgess" <bburgess66@hotmail.com> wrote in message
> > > news:#UIdlamsEHA.3200@TK2MSFTNGP14.phx.gbl...
> > > > Yessir ... but I'm not sure if the language settings are set to
> > > Simplified
> > > > all the time ... the language settings seem to change frequently and
> > > on
> > > > their own.. From Western European, to UTF-8, to Chinese Traditional,
> > > to
> > > ...
> > > >

microsoft.public.inetsrvr.iis: Re: POSTing Chinese characters

> > > > Any way to ensure the language is set properly ? ...
> programmatically?
> > > >
> > > > Many Thx
> > > >
> > > > -BB
> > > >
> > > > "Bernard" <qbernard@hotmail.com.discuss> wrote in message
> > > > news:uuuzh0ZseEHA.2660@TK2MSFTNGP12.phx.gbl...
> > > > > from PocketIE (on a PocketPC), this problem starts occuring
> > > > >> looks like client side issue to me
> > > > >
> > > > > 'But not all the time'?
> > > > >> it's even harder to trace then :(
> > > > >
> > > > > does the Pocket IE support chinese simplified ?
> > > > >
> > > > >
> > > > > --
> > > > > Regards,
> > > > > Bernard Cheah
> > > > > <http://www.tryiis.com/>
> > > > > <http://support.microsoft.com/>
> > > > > <http://www.msmvps.com/bernard/>
> > > > >
> > > > >
> > > > > "Brian Burgess" <bburgess66@hotmail.com> wrote in message
> > > > > news:u6PNDdRsEHA.636@TK2MSFTNGP09.phx.gbl...
> > > > > > Yes odd .. and yes I have tried all this too.
> > > > > >
> > > > > > To be more specific, if there is only one chinese char then the
> > > > > > submit
> > > > > > variable becomes = 'Submi', three chars submit = 'Sub'
> > > > > >
> > > > > > When calling from something like the desktop IE, then all works
> > > > > > properly,
> > > > > > but from PocketIE (on a PocketPC), this problem starts occuring.
> > > > > >
> > > > > > But not all the time: In this case the data originating from a
> > > > > > FoxPro
> > > > > > data
> > > > > > base on a English Win2k Advanced Server. IIS is on a
Simplified
> > > > > > Chinese
> > > > > > Win2K advanced server.
> > > > > >
> > > > > > Any thoughts?
> > > > > >
> > > > > > thx
> > > > > >
> > > > > > -BB
> > > > > > "Bernard" <qbernard@hotmail.com.discuss> wrote in message
> > > > > > news:Ooh82VQsEHA.832@TK2MSFTNGP10.phx.gbl...
> > > > > > > so with chinese chars (say 'Submit') the submit variable become
> > > > > > 'Subm'
> > > > > > and
> > > > > > not
> > > > > > > 'Submit' ?
> > > > > > > and you get this when do a response.write
request.form('submit')
> ?

microsoft.public.inetsrvr.iis: Re: POSTing Chinese characters

> > > > > >
> > > > > > ya, it's odd. can you just try a simple page, and a new target
> > form
> > to
> > > > > test
> > > > > > this again ?
> > > > > > it could be some 'part' in the existing script to misbehave.
> > > > > >
> > > > > > --
> > > > > > Regards,
> > > > > > Bernard Cheah
> > > > > > <http://www.tryiis.com/>
> > > > > > <http://support.microsoft.com/>
> > > > > > <http://www.msmvps.com/bernard/>
> > > > > >
> > > > > >
> > > > > > "Brian Burgess" <bburgess66@hotmail.com> wrote in message
> > > > > > news:eQGfmpPsEHA.1520@TK2MSFTNGP11.phx.gbl...
> > > > > > > Well the POSTed string looks like this:
> > > > > > > Login=;LE|;L&Passwd=123456&submit=Submit
> > > > > > > I have verified this, and it works in English
> > > > > > >
> > > > > > > When I look at the individual fields (with Request.Form)
after
> > the
> > > > > data
> > > > > > is
> > > > > > > POSTed, the values are:
> > > > > > > Login=;LE|;L
> > > > > > > Passwd=123456
> > > > > > > submit=Subm
> > > > > > >
> > > > > > > When English is used for both Login and Password, then
> 'submit'
> > > > always
> > > > =
> > > > > > 'Submit'
> > > > > > >
> > > > > > > Seem odd?
> > > > > > >
> > > > > > > thx
> > > > > > >
> > > > > > > -BB
> > > > > > >
> > > > > > > "Bernard" <qbernard@hotmail.com.discuss> wrote in message
> > > > > > > news:%23UcO\$uNsEHA.2688@TK2MSFTNGP14.phx.gbl...
> > > > > > > > Haven't seen this before, how do you post the data ?
> > > > > > > > and how do your verify the data length?
> > > > > > > >
> > > > > > > > --
> > > > > > > > Regards,
> > > > > > > > Bernard Cheah
> > > > > > > > <http://www.tryiis.com/>
> > > > > > > > <http://support.microsoft.com/>
> > > > > > > > <http://www.msmvps.com/bernard/>
> > > > > > > >
> > > > > > > >
> > > > > > > >
> > > > > > > > "Brian Burgess" <bburgess66@hotmail.com> wrote in message
> > > > > > > > news:emDYM2MsEHA.2300@TK2MSFTNGP09.phx.gbl...
> > > > > > > > > Hi all,

