

Re: Format of string output of a socket server

Source:

<http://www.tech-archive.net/Archive/Development/microsoft.public.win32.programmer.networks/2006-10/msg00026>

- *From:* "Alexander Nickolov" <agnickolov@xxxxxxxx>
 - *Date:* Tue, 3 Oct 2006 15:37:37 -0700
-

ASCII is the same no matter what byte encoding is used.
It's characters beyond ASCII you should worry about.
By definition any ASCII string is in UTF-8 encoding.
UNICODE code points 128-2047 are encoded in 2 bytes,
code points 2048-65535 (excluding the invalid code point
range of 55296-57343) are encoded in 3 bytes and the
rest (e.g. 65536-1114111) are encoded in 4 bytes.

For quick reference on UTF-8 here's the wikipedia page:

<http://en.wikipedia.org/wiki/UTF-8>

--

=====
Alexander Nickolov
Microsoft MVP [VC], MCSD
email: agnickolov@xxxxxxxx
MVP VC FAQ: <http://www.mvps.org/vcfaq>
=====

"Angus" <nospam@xxxxxxxx> wrote in message
<news:%23TKRV8y5GHA.4484@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>

Well... think I am more confused now than when I asked the question ;)

I am talking about what the server will send. I am getting from these
comments that sending bytes (char) is OK. Basically a string response I
would send would be eg "my response\r\n" - ie byte chars followed by
carriage return line feed. I am supposing this is OK.

What for example does your standard POP3 server send? ASCII text just
like
I am saying here?

The client program can then convert to Unicode or whatever they see fit?

Angus

Re: Format of string output of a socket server

"Alexander Nickolov" <agnickolov@xxxxxxxx> wrote in message
news:OZSOk9x5GHA.3452@xxxxxxxxxxxxxxxxxxxxxxxxxxxx

UNICODE is the only sane choice of course. However, don't confuse with the Windows meaning of UNICODE which is really only the UTF-16 representation of UNICODE. I'd suggest you use UTF-8 representation of UNICODE to avoid byte-ordering issues on the network. What you return to your clients is up to you – you just need to do the appropriate format conversion (e.g. MultiByteToWideChar to get UTF-16 for example).

--

=====
Alexander Nickolov
Microsoft MVP [VC], MCS D
email: agnickolov@xxxxxxxx
MVP VC FAQ: <http://www.mvps.org/vcfaq>
=====

"Angus" <nospam@xxxxxxxx> wrote in message
news:eXxFh8m5GHA.400@xxxxxxxxxxxxxxxxxxxxxxxxxxxx

Hello

I am writing a socket server to deliver telephony events to clients on a network. For example the telephony server might send out text to connected clients. Clients might be written in C/C++, Java, Visual Basic,

anything

in fact which can talk to a socket.

My socket server is currently sending out char* . Do I have to worry about the format of string output? Should I be outputting Unicode? Some

other

format? Or would a C/C++ char* be OK? Will eg Java understand it? Do they

Re: Format of string output of a socket server
use UTF-8 or something?

Angus