

Re: Strange ' ' character output when using simplexml load string

Re: Strange ' ' character output when using simplexml load string

Source: <http://coding.derkeiler.com/Archive/PHP/comp.lang.php/2008-04/msg00399.html>

- *From:* AnrDaemon <anrdaemon@xxxxxxxxxxx>
 - *Date:* Tue, 8 Apr 2008 19:00:00 +0400
-

Greetings, bizt.

In reply to Your message dated Monday, April 7, 2008, 17:19:28,

I converting an XML string using
simplexml_load_string function. It is
giving me a character for some reason
dotted around the text.

simplexml always outputs in UTF-8. Is your page's
encoding UTF-8?

At a guess, ISO-8859-1 or perhaps ISO-8859-15.

In UTF-8, a "prefix" of an 0xC2 byte is used to access the top half of the "Latin-1 Supplement" block which includes a lot of juicy characters such as currency symbols, fractions, superscript 2 and 3, the copyright and registered trademark symbols, and the non-breaking space.

However in ISO-8859-1 and -15, the byte 0xC2 represents an , so if UTF-8 is misinterpreted as one of those, then you get followed by some other nonsense character.

Probably the easiest solution would be to take the output from SimpleXML and pass it through iconv():

```
$xmlout = iconv('UTF-8', 'ISO-8859-15//TRANSLIT', $xmlout);
```

Note that UTF-8 is capable of representing a far greater range of characters than ISO-8859-1/-15 are, so certain characters may not properly survive conversion. (Using the '//TRANSLIT' option tells iconv to do its best, and if, say, a particular accented character is not available in ISO-8859-1, then to substitute an unaccented one in its place.)

Re: Strange ' ' character output when using simplexml load string

Hi, ive tried what you said which worked for one of my pages but when i tried it on another i got the following:

```
Notice: iconv() [function.iconv]: Detected an illegal character in input string in /home/public_html/search_apartments.php on line 67
```

Im using the following to convert my XML string which is fetched via cUrl:

```
$result = iconv('UTF-8', 'ISO-8859-15//TRANSLIT', $result);
```

Would it be the case that my \$result string, im not providing the iconv() with the correct input encoding? If so, is there a way for me to detect the input encoding?

As a guess, Your "B" probably followed by space and represent a non-breaking space.

To Your trouble with iconv on \$result, I think You should take care of the SOURCE BEFORE using simplexml_load_string.
And see what the encoding it use. Because if Your source in, say, ISO-8859-15, You can't have any untranslatable characters in UTF-8 what You can't convert back to ISO-8859-15.

—
Sincerely Yours, AnrDaemon <anrdaemon@xxxxxxxxxxxx>

.