

Re: Non-ascii characters in VS.NET service

Source:

<http://www.tech-archive.net/Archive/DotNet/microsoft.public.dotnet.languages.csharp/2007-02/msg01214.html>

- *From:* "Mike Schilling" <ap@xxxxxxxxxxxxxxxxxxxx>
 - *Date:* Fri, 9 Feb 2007 08:19:58 -0800
-

"Jon Skeet [C# MVP]" <skeet@xxxxxxxx> wrote in message
<news:MPG.203630616d52d59b98d83c@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>

Mike Schilling <msscottschilling@xxxxxxxxxxxx> wrote:

I've created a simple .NET 1.1 web service using VS.NET 2003: it has one method that takes a string parameter. It iterates through the input string, turning each character into hex and appending it to an output string, and returns the result.

How is it turning the character into hex?

```
ret += "0x" + ((int)s[i]).ToString("X");
```

I now send this service SOAP messages containing non-ASCII characters in the field that becomes the input string. Each SOAP message has an XML header that correctly describes the format of the non-ASCII characters. (I've tried both iso-8859-1 and utf-8).

What do you mean by "an XML header"? It should just be in the XML declaration.

Exactly. Each SOAP message specifies the correct encoding in its XML declaration, as shown below.

Re: Non-ascii characters in VS.NET service

For some reason, each XML character that's non-ASCII has been turned into a question mark "?" in the input string. Actually, a UTF-8 character that contains two bytes becomes two question marks.

That suggests that whatever's producing the XML file is wrong, *or* that you're looking at the XML in an inappropriate editor. How are you looking at the XML?

```
<?xml version="1.0" encoding="iso-8859-1"?>
<soap:Envelope xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance
xmlns:xsd="
http://www.w3.org/2001/XMLSchema
xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/>
<soap:Body>
<ToHex xmlns="http://tempuri.org/>
<s>aéiæf</s>
</ToHex>
</soap:Body>
</soap:Envelope>
```

NNTP is likely to garble the non-ASCII characters, but in hex the string inside the <s> tags is

141 351 354 346 146

verified by using od -b.

In iso-8859-1, these are respectively

a, e with an acute accent, i with a grave accent, ae, f

And I can parse the same file locally and observe that it's correct (e.g. with the program below). It only acts oddly when processed by the web service.

```
using System;
using System.Xml;
```

```
namespace XMLParser
{
class ParseXML
{
static void Main(string[] args)
{
XmlDocument doc = new XmlDocument();
doc.Load("c:\\java\\toHex.xml");
dumpStrings(doc);
Console.WriteLine("<Done>");
```

```
    }  
  
    private static void dumpStrings(XmlNode node)  
    {  
        if (node is XmlCharacterData)  
        {  
            Console.Out.WriteLine(node.Value);  
        }  
        else  
        {  
            for (XmlNode child = node.FirstChild;  
                child != null;  
                child = child.NextSibling)  
            {  
                dumpStrings(child);  
            }  
        }  
    }  
}
```

.