

Re: parsing xml

Source: <http://www.tech-archive.net/Archive/Data/microsoft.public.data.ado/2006-08/msg00270.html>

- *From:* hgeron <hgeron@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Wed, 23 Aug 2006 21:21:01 -0700
-

I messaged back to soon. The fix works some times. Many records that should have a node.text don't. sometimes one record got several node.text's. I took out the movelast and only record 1 had a node.text.
--
hgeron

"Mark J. McGinty" wrote:

"hgeron" <hgeron@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx> wrote in message <news:58D00426-E245-4779-8005-6B1F2C2FDCC5@xxxxxxxxxxxxxxxxxxxxxx>

Mark,
I've been struggling with MSxml for weeks, I would certainly like to know more about you so I can include it with my documentation. I understand if you can't reveal more about yourself.

My display name on these NG posts is my actual name... There is a short bio blurb on www.codeguru.com to which I have contributed a few things... What else did you want to know?

Again, many thanks ---hgeron

You're welcome.

-Mark

www.myspace.com/hgeron

Re: parsing xml

"Mark J. McGinty" wrote:

"hgeron" <hgeron@xxxxxxxxxxxxxxxxxxxxxxxxxxxx> wrote
in message
news:E3233D51-C3FF-4190-AE39-DB1D61E71AC9@xxxxxxxxxxxxxxxxxxxx

Well, that certainly listed everything.

I have never got Debug.print to work for
me... maybe it does not work
in an Access vb module.

I thought VBA supported it... docs say it does. <shrug> It
only works
when
running in debugging mode; it dumps output to the
immediate window.

I wrote the output to a file. The node
"catalog" thru the end of the
XML
is
output
several times. I cut nodes off at 20 chars
max, and the output was
16kb.
The whole books.xml was only 5kb. The
extra size is repeating nodes.
I
didn't see any attributes repeated. I know
you have spent a lot of
time
on
this, but do you see a way of avoiding
repeating nodes? I thought
about
capturing the node output and
comparing it to the last, but that wouldn't
work because when nodes are
within other nodes, the repeating node is not
the last node.

It's repeating the same output because you are dumping Node
inside a loop

Re: parsing xml

that doesn't change it. ChildNode is the one used by For Each.

The other duplication you are seeing is because you are dumping the xml property. Try .name and .value for attributes, and .nodeName, .nodeValue for nodes. That still effectively dumps 2 copies of each NODE_TEXT node, the parent nodes of which are apparently named #text... making the NODE_TEXT output redundant.

-Mark

Here's my code now.

```
Option Compare Database
Dim Opened As Boolean
Sub mark()
Dim xmlDoc As New
MSXML2.DOMDocument50
Dim Node As IXMLDOMNode
Dim NodeList As IXMLDOMNodeList
Open "c:\temp\output.txt" For Output As #1:
Reset
xmlDoc.Load ("m:\xml\books.xml")
'xmlDoc.Load ("m:\xml\bittner2.xml")
RecurseXML xmlDoc.documentElement, 0
Reset
End Sub
Function RecurseXML(Node As
IXMLDOMNode, Level As Long) As
Boolean
Dim NodeList As IXMLDOMNodeList
Dim ChildNode As IXMLDOMNode
If Not Opened Then Open
"c:\temp\output.txt" For Append As #1:
Opened =
True
'MsgBox ("Begin Recurse Level: " & Level)
Print #1, "Begin Recurse Level: "; Level
Set NodeList = Node.childNodes
Dim Att As IXMLDOMAttribute
If Not Node.Attributes Is Nothing Then
```

Re: parsing xml

```
attnum = 0
For Each Att In Node.Attributes
attnum = attnum + 1
'MsgBox ("ATTR:" & vbCr & Att.Value)
Print #1, "ATTR:"; attnum, Att.XML,
Att.Value
Next
End If
If Node.nodeType = NODE_TEXT Then
'MsgBox "NODE_TEXT:" & vbCr &
Node.XML
Print #1, "NODE_TEXT:"; nde, Node.XML
Else
If Not NodeList Is Nothing Then
For Each ChildNode In NodeList
nde = nde + 1
'MsgBox ("NODE: " & vbCr & Node.XML)
Print #1, "NODE:(left 20 of";
Len(Node.XML); " chars)";
Left(Node.XML, 20)
If RecurseXML(ChildNode, Level + 1) =
False Then Close
1:
Opened = False: Exit Function
Next
End If
End If
RecurseXML = True
End Function
```