

SQL Tools for Bizweaver

Samples of tools that can be used with SQL Server in Bizweaver.

Function to Strip HTML Tags

This is a user-defined function that strips out various HTML tags from data being inserted into a SQL table. This is very helpful when the input comes from a web-based API.

To use this function

- First create the function in the database where you will be doing the table insert. The code below creates the function. Additional tags can be added as needed.

```
/****** Object:  UserDefinedFunction [dbo].[udf_StripHTML]      Script Date:
10/28/2019 2:16:13 PM *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE FUNCTION [dbo].[udf_StripHTML]
(
    @HTMLText VARCHAR(MAX)
)
RETURNS VARCHAR(MAX)
AS
BEGIN
    DECLARE @START INT
    DECLARE @END INT
    DECLARE @LENGTH INT

    -- Replace the HTML entity &#amp; with the '&' character (this needs to be
    -- done first, as
    -- '&' might be double encoded as '&#amp;#amp;')
    SET @START = CHARINDEX('&#amp;', @HTMLText)
    SET @END = @START + 4
    SET @LENGTH = (@END - @START) + 1

    WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
        SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, '&')
        SET @START = CHARINDEX('&#amp;', @HTMLText)
        SET @END = @START + 4
        SET @LENGTH = (@END - @START) + 1
    END

    -- Replace the HTML entity &lt; with the '<' character
```

```
SET @START = CHARINDEX('<', @HTMLText)
SET @END = @START + 3
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, '<')
SET @START = CHARINDEX('<', @HTMLText)
SET @END = @START + 3
SET @LENGTH = (@END - @START) + 1
END

-- Replace the HTML entity '<'; with the '<' character
/*
SET @Start = CHARINDEX('<', @HTMLText)
SET @End = @Start
SET @Length = (@End - @Start)

WHILE (@Start > 0 AND @End > 0 AND @Length > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @Start, @Length, '<')
SET @Start = CHARINDEX('<', @HTMLText)
SET @End = @Start + 1
SET @Length = (@End - @Start)
END
*/

-- Replace the HTML entity > with the '>' character
SET @START = CHARINDEX('>', @HTMLText)
SET @END = @START + 3
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, '>')
SET @START = CHARINDEX('>', @HTMLText)
SET @END = @START + 3
SET @LENGTH = (@END - @START) + 1
END

-- Replace the HTML entity & with the '&' character
SET @START = CHARINDEX('&', @HTMLText)
SET @END = @START + 4
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, '&')
SET @START = CHARINDEX('&', @HTMLText)
SET @END = @START + 4
SET @LENGTH = (@END - @START) + 1
END

-- Replace the HTML entity &nbsp; with the ' ' character
SET @START = CHARINDEX('&nbsp;', @HTMLText)
```

```
SET @END = @START + 5
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, ' ')
SET @START = CHARINDEX('&nbsp;', @HTMLText)
SET @END = @START + 5
SET @LENGTH = (@END - @START) + 1
END

-- Replace any <br> tags with a newline
SET @START = CHARINDEX('<br>', @HTMLText)
SET @END = @START + 3
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, CHAR(13) + CHAR(10))
SET @START = CHARINDEX('<br>', @HTMLText)
SET @END = @START + 3
SET @LENGTH = (@END - @START) + 1
END

-- Replace any <br/> tags with a newline
SET @START = CHARINDEX('<br/>', @HTMLText)
SET @END = @START + 4
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, CHAR(13) + CHAR(10))
SET @START = CHARINDEX('<br/>', @HTMLText)
SET @END = @START + 4
SET @LENGTH = (@END - @START) + 1
END

-- Replace any <br /> tags with a newline
SET @START = CHARINDEX('<br />', @HTMLText)
SET @END = @START + 5
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, CHAR(13) + CHAR(10))
SET @START = CHARINDEX('<br />', @HTMLText)
SET @END = @START + 5
SET @LENGTH = (@END - @START) + 1
END

-- Remove anything between <STYLE> tags
SET @START = CHARINDEX('<STYLE', @HTMLText)
SET @END = CHARINDEX('</STYLE>', @HTMLText, CHARINDEX('<', @HTMLText)) + 7
SET @LENGTH = (@END - @START) + 1
```

```
WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, '')
SET @START = CHARINDEX('<STYLE', @HTMLText)
SET @END = CHARINDEX('</STYLE>', @HTMLText, CHARINDEX('</STYLE>',
@HTMLText)) + 7
SET @LENGTH = (@END - @START) + 1
END

-- Remove anything between <whatever> tags
SET @START = CHARINDEX('<', @HTMLText)
SET @END = CHARINDEX('>', @HTMLText, CHARINDEX('<', @HTMLText))
SET @LENGTH = (@END - @START) + 1

WHILE (@START > 0 AND @END > 0 AND @LENGTH > 0) BEGIN
SET @HTMLText = STUFF(@HTMLText, @START, @LENGTH, '')
SET @START = CHARINDEX('<', @HTMLText)
SET @END = CHARINDEX('>', @HTMLText, CHARINDEX('<', @HTMLText))
SET @LENGTH = (@END - @START) + 1
END

RETURN LTRIM(RTRIM(@HTMLText))

END
GO
```

- Use the function during the insert process as follows.

In this example, assume your input string is coming from a FileReader step. both the FileReader and SQL table column name is "TextData". Your insert statement will look like:

```
INSERT INTO TableA
(TextData)
VALUES
(dbo.udf_StripHTML('Workflow(FileReader_readList).Get({FIELD_TextData})'))
```

From:

<https://wiki.twbs.com/> - Wiki

Permanent link:

https://wiki.twbs.com/doku.php/bw/sql_tools

Last update: **2019/10/28 14:36**

