

Re: Large text file to a CSV file...

Source:

<http://www.tech-archive.net/Archive/Access/microsoft.public.access.externaldata/2004-05/0582.html>

From: Ted Allen (*anonymous_at_discussions.microsoft.com*)

Date: 05/21/04

Date: Fri, 21 May 2004 14:32:04 -0700

Hi Louie,

Well, even after reading your repost I agree with Joe's initial comment that it isn't clear why the variation in data has anything to do with the space delimiter. From the repost it sounds like sometimes the user name (Admin in the sample) would be a dash instead, but that wouldn't create a problem as long as a space was still inserted after. Maybe I misread your post.

It does seem like you may have problems with embedded spaces if all text fields are not qualified with quotes.

In any case, you can write code to open the text file and read the contents (line by line) into a variable, work with the contents of the line and then output it.

I have done this in cases where I received csv files from third parties that had embedded quotes and commas that were preventing the data from importing correctly. I used code to go through line by line to identify problem lines and replace the embedded quotes and commas. Basically I first counted the quotes and commas and checked for extras, if there were extras then it would go through to identify the positions of the embedded ones and replace them using some defined logic that worked in all cases in that particular file.

In order to do this though, you have to be able to identify some logic that will be able to correct the problems in all cases. It sounds like you may be able to do this since your data is fairly structured.

If you are interested, the code would look something like the following (without the dim statements, file path variable assignments and error handler). strFileName (In/Out) would contain the path and file name to the

appropriate file.

```
intOutputHandle = FreeFile
Open FileNameOut For Output As #intOutputHandle
```

```
intInputHandle = FreeFile
Open strFileNameIn For Input As #intInputHandle
```

```
Do While Not EOF(intInputHandle)
  Line Input #intInputHandle, strBuffer
  (put your code here to analyze/manipulate strBuffer)
  Print #intOutputHandle, strBuffer
Loop
```

```
Close #intInputHandle
Close #intOutputHandle
```

And, actually on mine I also output some statistics to a second output file (such as line number, comma count, and quote count). That way I could see which lines had been manipulated and check some of them to see if the code appeared to work correctly.

HTH, Ted Allen

```
>-----Original Message-----
>I have a WebsitePro Access Log Combined (W3C/NCSA)
Format
>file; and yes, I know there is an ODBC routine out there
>to do this, but my company doesn't want to spend the
money
>to get it because we are abandoning WebSite Pro in the
>next few months. The data is on a classified system so
I
>will attempt to recreate the problem with generic data.
I
>already see problems with the blank delimiters...
>
>An example from the book (string these together with a
>space between fields.
>
>172.18.96.157 (IP Address)
>
>gryphonaire.com (Hostname of server
>receiving request)
>
>Admin (Authentication user
name)
>
>[02/Jan/2001:09:44:56 - 0800] (Date timestamp UTC
offset)
```

Re: Large text file to a CSV file...

>
>"GET /members.html HTTP/1.1" (Complete HTTP request)
>
>200 (HTTP Response Code)
>
>178 (Number of bytes
transferred)
>
>"<http://gryphonaire.com/index.html>" (Complete referring
>URL)
>
>"Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)" (ID
of
>browser)
>
>"Member=WebSiteRules" (Cookie)
>
>The problem in the Access import is if the
Authentication
>username isn't available. WSP puts in a – This is
>common for most of the requests. When an actual ID is
>present, it doesn't fit the pattern with the others.
The
>data is consistantly like the above example (" [, and
>all).
>
>Hopefully that explains it.
>
>L
>
>.
>