

Getting data from the Web for Excel

Updated by Debbie Wolfe
Technology Training Editor
St. Petersburg Times
800-333-7505, ext. 2919

March. 9, 2004 2 pages
dpwolfe@peoplepc.com
http://canyonwolfephoto.com
FREE tip sheets, TECH TIPS
archives and consulting info

Text Files:

1. Check the URL and make sure that you are dealing with a text file (.txt).
Examples: www.census.gov/epcd/cbp/download/01_data/cbp01us.txt (delimited)
 www.census.gov/population/socdemo/race/indian/ailang1.txt ("fixed width" text)
2. Go to the **File** menu and select **Save As**. Save the file to your hard drive.
3. Minimize the browser window. Open Excel.
4. From the **File** menu, choose **Open**. Make sure the "Files of Type" box says Text Files. If it doesn't, click the down arrow and select Text Files. Select your saved file. Click **Open**.
5. The **Import Wizard** will launch. The Wizard will walk you through the import in three steps.
6. Determine if the file is "Delimited" or "Fixed Width." If there are commas or semi-colons or spaces separating the columns, you probably have a *delimited* file. If your columns appear to be lined up, the data may be *fixed width*. Click **Next**.
7. If you selected "Delimited," select the appropriate delimiter (comma, semicolon, space, etc.) Click **Next**.
8. If you selected "Fixed Width," insert the column lines where appropriate. Click **Next**.
9. In Step 3 of the Import Wizard, the column data format will default to *General*. You may want to change numbers like Zip Codes, addresses, and social security numbers to Text by selecting each column. Click **Finish**.
10. **Very Important!** You're file is still a text (.txt) file, you **must** save it as a spreadsheet (.xls). Select **File** from the menu, click on **Save As**. Name the file and from the "Save as Type" drop-down select Microsoft Excel (.xls).

HTML Files:

1. Check the URL and make sure that you are dealing with an "html" file (.htm/.html).
Example: ferret.bls.census.gov/macro/032003/health/h06_000.htm
(U.S. Census Bureau: Health Insurance Coverage Status by State for All People; last revised: September 11, 2003)
If you're lucky:
2. Select the data that you want in the spreadsheet by highlighting that information. From the **Edit** menu, select **Copy**.
3. Open Excel. From the **Edit** menu, select **Paste**. Data should be in the appropriate cells. Save the file as a spreadsheet (.xls).

If you're not so lucky:

1. Select the data that you want to use. From the **Edit** menu, choose **Copy**.
2. Open Excel. Click in a cell and from the **Edit** menu, click **Paste**. Each row of data will appear in one cell – the first one in each row (a.k.a. column A). You want all of that data to appear in individual cells. Click on "A" to select the entire column.
3. From the **Data** menu, select **Text to Columns**.
4. The **Import Wizard** will launch. The Wizard will walk you through the import in three steps.
5. Determine if the file is "Delimited" or "Fixed Width." Walk through the appropriate steps. Click **Finish**. Save the file as a spreadsheet (.xls).

If that doesn't work:

1. Select the data that you want to use by highlighting the information. From the **Edit** menu, choose **Copy**. (**Note:** If you want the full page, save the page as a Text (.txt) file and jump to Step 4.)
2. Open Notepad. Paste the data into Notepad.
3. Save the file as a Text (.txt) file. Close Notepad.
4. Open Excel. From the **File** menu select **Open**. Open your file. (If you can't find it, remember to change the "Files of Type" box to Text files.)

The **Import Wizard** will launch and you're home free. **Remember** to save the file as a spreadsheet (.xls).

PDF Files:

And you thought HTML files were tricky

U.S. Census Bureau: State Population Estimates by Selected Age Categories and Sex: July 1, 2002

Example: eire.census.gov/popest/data/states/tables/ST-EST2002-ASRO-01.pdf

(release date: Sept. 18, 2003)

1. To work with PDF files, you **must** be viewing the PDF file **independent** from the browser. To guarantee that you are not using Adobe's Acrobat Reader as a "plug-in" RIGHT-mouse click on the file that you want. LEFT-mouse click on **Save Target As** and download the file to your local desktop.
2. Open Acrobat Reader. From the **File** menu select **Open**, then open the file you just downloaded. You can now take advantage of all the tools that are in Acrobat Reader.
3. To select portions of the document, click on the Text Select Tool (this is the button that either says "abc" or has the capital "T" on it.) Highlight the text you want to copy. (**Note:** To clean the data effectively, you may need to select just portions of the pdf file. See Step 4.)
4. If you want to select a *portion* of a table rather than the entire table, press and hold down the **CTRL** key while you highlight the text you want with your mouse (TIP: draw a rectangle with your mouse). Once you have defined the correct area, on the menu bar click on **Edit → Copy** to place the data on the Windows clipboard.
5. Open Notepad and paste via **Edit → Paste**. Save the file as a text (.txt) file, then close it.
6. Open Excel. Find and open the text file you just created using Notepad. Excel's **Import Wizard** will automatically launch a text (.txt) file. Your data will probably be "delimited" and the delimiter is often a **space**. Follow the steps in the wizard. **Remember** to save the file as an Excel workbook file (.xls) after you finish the wizard steps.

**If you do not have Adobe's FREE Acrobat Reader on your computer or your version is old, you can download it from: www.adobe.com/products/acrobat/readstep2.html*

Spreadsheet/Database/Executable Files:

Just a few notes about these guys.

Sometimes you will luck out and find something that is already in a spreadsheet or database format. If your file has an extension like .xls or .dbf you can open the file directly into a spreadsheet by just choosing **File** from the menu and selecting **Open**. At all times you will want to save an .xls or .dbf file (with a RIGHT mouse click) to your hard drive first, then open it from within Excel or Access independent of your Web browser so you'll have access to all of the tools you may need. Opening one of these files inside your Web browser limits what you can do with technology.

If your file is an executable file (.exe), you will need to first download the file and then open it. This process typically inflates the physical size of a file that has been compressed.

Reminders:

- **Be careful.** Using stats can be tricky. Make sure you're comparing apples to apples.
- **Check your work.** Especially if you're sorting or using formulas. Save your work after each change.
- **Compare data.** If there is more than one source for the same data, check both to see if there are inconsistencies.

Have fun!

Common File Extensions:

.txt	.csv	.prn	.asc	=	text or ascii
.htm	.html			=	HTML or Hypertext Markup Language (Web pages)
.xls	.wk*			=	Excel, Lotus (spreadsheet files)
.mdb				=	Access
.dbf				=	typical database program, FoxPro
.doc				=	Word or WordPad
.wp*				=	WordPerfect
.pdf				=	Adobe Acrobat Reader
.zip				=	ZIP (compressed file)
.exe				=	executable (often, a program file; sometimes a self-extracting ZIP file)
.jpg	.tif	.gif		=	graphics (photos are most often .jpg)
.mpg	.mov			=	video
.ppt	.pps			=	PowerPoint presentation or slide show (if you do not have PowerPoint on your PC and you would like to view these types of files, you can download a n
					FREE viewer from Microsoft caution total download time = 17 minutes @ 28.8k
					office.microsoft.com/downloads/2000/Ppview97.aspx)

Good for these PowerPoint versions:
95, 97, 2000 and 2002

Special thanks to:

Jennifer Small Evert (Lexis-Nexis, formerly of the *San Antonio Express-News*) who authored the original document.