

Re: simple (I hope!) screen scraping script – pointers?

Re: simple (I hope!) screen scraping script – pointers?

Source:

<http://www.tech-archive.net/Archive/Scripting/microsoft.public.scripting.vbscript/2005-08/msg00989.html>

- *From:* "McKirahan" <News@xxxxxxxxxxxxxx>
 - *Date:* Wed, 24 Aug 2005 23:40:33 -0500
-

"Sylvia" <Puget4753@xxxxxxxxxx> wrote in message
news:1124820081.399551.41380@xx

> Hello – is vbscript a good way to go for this project?
>
> Basically, on this page:
>
> <http://www.cityofbellevue.org/page.asp?view=38806>
>
> There's a set of links to docs. I need to:
> 1. get the text from each doc file (can perl work with doc files?)
> 2. then put all the crime reports into one text file, ordered first by
>
> District, and then by Date.
>
> Thoughts, help? I've used vbscript before, but I'm not sure it can
> take a text file and order it the way I want.
>
> thanks much!
> Sylvia
>

I've created a script that created the following sample CSV file
using the documents from 2005_08_23; specifically, the documents:

Press_2005_08_23_0600.doc
Press_2005_08_23_1400.doc
Press_2005_08_23_2000.doc

which are located under:

<http://www.cityofbellevue.org/departments/Police/files/>

Save it as "20050823.csv" then double-click on it to open it up in MS-Excel.
In MS-Excel you can format, sort, filter, and print it to suit your needs.

It contains 13 lines including the header row; watch for word-wrap.

```
DATE,TIME,COMPLETED BY,Incident,District  
#,Location,Address,Occurred,Description,Information  
08/23/05,0600,T172,ROBBERY,District #1-1,05B-9571,00 blk NE 1,08/22 at 2010
```

Re: simple (I hope!) screen scraping script – pointers?

Re: simple (I hope!) screen scraping script – pointers?

hrs,Victims reports that 3 subjects approached him and demanded his wallet at gunpoint.,
08/23/05,0600,T172,THEFT,District #1-1,05B-9570,e 2nd & 111 Ave NE,08/22 at 2016 hrs,Victims plates were both taken from vehicle.,Vehicle:
08/23/05,0600,T172,FORGERY,District #1-1,05B-9564,0100 blk NE 8th St; Bartell's,08/22 at 1645 hrs,Suspect attempted to fill forged prescription for oxycontin.,
08/23/05,0600,T172,MV PROWL,District #2-8,05B-9557,500 blk 134th Ave SE; Kelsey Ridge Apts,"08/21 to 08/22, 2100 – 0800 hrs",Unk susp(s) entered unlocked vehicle and took items.,Total amount of loss: more than \$250; Vehicle: Whi 90 Jeep Cher 4D
08/23/05,0600,T172,MALICIOUS MISCHIEF,District #2-8,05B-9566,500 blk 145th Pl SE; Highcroft,"08/21 to 08/22, 1900 – 0700 hrs",Neighbors had a dispute about parking and this morning victim discovered vehicle had been keyed.,Vehicle: Blk 99 Ford Taur 4D
08/23/05,0600,T172,MALICIOUS MISCHIEF,District #3-1,05B-9548,200 blk 140th Ave NE; Construction site,"08/19 to 08/22, 1500 – 1030 hrs",Unk susp(s) broke water pipes in new house construction causing water damage.,"Total amount of loss: over \$10,000"
08/23/05,0600,T172,THEFT,District #3-4,05B-9508,3000 blk Bel Red Rd; Circuit Services,08/21 at 1400 hrs,Unk susp(s) stole three copper rods from outside of the listed address.,Total amount of loss: more than \$250
08/23/05,1400,T153,MALICIOUS MISCHIEF,District #2-1,05B-9576,600 blk Bellevue Way SE; Barney/ Als Chevron,8/23 at Unknown Time,Unk susp(s) broke vic's veh window.,Vehicle: Gray 92 Chev
08/23/05,1400,T153,MALICIOUS MISCHIEF,District #3-4,05B-9530,3900 blk NE 20th ST; Chevron,"8/19 to 8/21, Unknown Time",Unk susp(s) pried lock on air machine.,Total amount of loss: less than \$250; Related: 05B9531
08/23/05,1400,T153,MALICIOUS MISCHIEF,District #4-1,05B-9531,6200 blk NE 8th St; Chevron,"8/19 to 8/21, Unknown Time",Unk susp(s) damaged locking mechanism to coin operated air machine.,Total amount of loss: less than \$250; Related: 05B9530
08/23/05,1400,T153,MV THEFT,District #7-2,05B-9573,000 blk 119th AV SE,"8/22 to 8/23, 2330 – 0141 hrs",Unk susp(s) stole vic/s vehicle during times listed,Vehicle: Blk 94 Hond Acc 2D
08/23/05,2000,T150,MALICIOUS MISCHIEF,District #4-1,05B-9561,5600 blk NE 8th St; Crossroads Mall,08/22 at 1445 – 1530 hrs,Unkn person(s) knocked over motorcycle causing damage.,Vehicle: Red/whi 98 Hond Shad MC

Before I post the script, I have a few questions:

- 1) Is this what the result your looking for?
- 2) How can it be improved?
- 3) What is your interest is in this data?
- 4) Is your interest one-time or ongoing?
- 5) Is your interest for a certain period or incident?

Re: simple (I hope!) screen scraping script – pointers?

Re: simple (I hope!) screen scraping script – pointers?

6) Do you work for the City of Bellevue, Washington?

Here is an overview of the script:

This Visual Basic Script (VBS) program does the following:

- 1) downloads the Web page source and identifies all ".doc" files mentioned.
- 2) downloads each "doc" file (if new) and converts it to a ".txt" file.
- 3) parses each ".txt" file and writes a ".csv" file.

The CSV file may then be opened up in MS–Excel for formatting and sorting.

Optionally, the last step can be rerun by itself and takes about 10 seconds.

This is desirable when the "parse" process is modified and needs to be rerun.

This is possible because the ".txt" versions of all the documents are saved.

The script (at this time) ignores the ARRESTS section.

I've used it to process 683 reports in 2005 through "2005_08_24_0600".

The CSV file generated contains 4,094 lines plus a header row and is 843KB.

All 2005 reports are available via:

<http://www.cityofbellevue.org/page.asp?view=7863>

BELLEVUE POLICE DAILY RECAP OF ACTIVITIES

Each day at 6:00 AM, 2:00 PM and 8:00PM,

the Records Unit of the Bellevue Police

Department generates a activity summary

for police employees and the news media.

These reports are now online.

These reports contain Public Information

and is generated in accordance with Public

Records Laws of the State of Washington (RCW 42.17).

If you have any questions reference the information

on these recaps, please contact the Bellevue Police

Department Public Information Officer at 425–452–4129.

August, 2005

July, 2005

June, 2005

May, 2005

April, 2005

March, 2005

February, 2005

January, 2005

.

Re: simple (I hope!) screen scraping script – pointers?

- **Follow-Ups:**
 - ◆ **Re: simple (I hope!) screen scraping script – pointers?**
 - ◇ From: Sylvia

- **References:**
 - ◆ **simple (I hope!) screen scraping script – pointers?**
 - ◇ From: Sylvia

- Prev by Date: **Re: use vbscript to "click" a taskbar button**
- Next by Date: **Re: RTF Files, How to Read?**
- Previous by thread: **Re: simple (I hope!) screen scraping script – pointers?**
- Next by thread: **Re: simple (I hope!) screen scraping script – pointers?**
- Index(es):
 - ◆ **Date**
 - ◆ **Thread**