OTTAWA PC NEWS

PRODUCT REVIEW

On-Screen OCR   *by Alan German*

Some people seem to delight in making life hard. For example, there are those who protect PDF files so that, in theory, one can't copy the text. Of course, there are work-arounds for almost every such eventuality. In this case, one can simply print a page of the file, scan it, and run it through optical character recognition (OCR) software. Recently, a colleague asked if there was a way to shortcut this process by using OCR software directly from the computer screen.

A little research on the Internet turned up several candidate programs, including one in my particular favourite category of free and open-source software. But, how well do such programs perform? The acid test was running Capture2Text – an open-source program from SourceForge – on a page from Chris Taylor's article on disk image backups in the March issue of OP-CUG's newsletter.

Now, this wasn't because Chris is intent on stopping people from reproducing his ideas, nor is the newsletter's PDF file protected from copying, it was just that our newsletter provided a readily-available, and relatively-simple page layout, with both text and graphics, for me to use as a test case.

Having downloaded the OCR utility as a ZIP file, the embedded files can easily be extracted to a Capture2Text folder. The software doesn't need to be installed, and running the executable (Capture2Text.exe) establishes the program in a terminate-and-stay-resident mode. Right-clicking on the program's icon in the task bar provides a number of options. Checking "Show Popup Window" is recommended as this will subsequently provide direct confirmation of the text capture process (see below).

I found the program's operation to be both non-intuitive, and somewhat sensitive to mouse clicks. Checking out the instructions in the readme.txt file is highly recommended. You will find that the critical keyboard shortcut is Windows-Key-Q to both start the capture process and to end it. However, the subsequent part of the instructions aren't at all clear. They indicate: "Now, using your mouse, resize the capture box over the area of the screen that you want to OCR." The obvious choice would be to use the left mouse key to drag some sort of frame over the text to be selected. However, this is doomed to failure since, by default, a left-mouse click actually ends the capture process!

In fact, what you seem to need to do is press Windows-Key-Q, wait a few seconds, and then move the right-mouse button. A blue rectangle expands to select an area of text. Holding the right-mouse button down, and dragging the mouse, allows the rectangle to be repositioned on the screen so that, for example, the top-left corner of a section of text can be selected. Subsequently releasing the right-mouse button, and dragging the mouse across the screen once again expands the rectangle so that a specific block of text can be defined. When the desired text has been highlighted in this manner, pressing Windows-Key-Q a second time causes the text capture process to be completed. The selected text is now displayed in a pop-up window (entitled Capture2Text Popup Result) confirming that the desired text has been captured.

I found this mouse-key/select-drag sequence to be very non-intuitive. I was forever pressing the left-mouse key only to find that nothing further happened – because – of course – I had actually aborted the text-capture process!

Next Meeting: WEDNESDAY, June 11th, 2014

June Raffle

Courtesy of the Microsoft MVP program, we have a copy of **Windows 7 Ultimate** for the June raffle. This is the full version (not the "upgrade only" version) that can be installed on any PC. Valued at $350.

If you are still using Windows XP (for which security patches were discontinued after March 14, 2014) and you don't want to move to Windows 8, this is the perfect raffle prize for you!

Tickets are, as always, a good deal at $1 for one, a great deal at $2 for three or the unbelievable bargain of $5 for ten!

May Prize Winner

The winner of the raffle at our May meeting was **Gerald Larocque**.

He took home a Ubislate Tablet Computer!

Coming Up…

Wednesday, June 11, 2014

Pizza Night + eWaste Event, AND a guest speaker !

(see details next page)

Our next meeting will be in September.
Check our website at http://opcug.ca later this summer for our Fall line-up.

2014 CALENDAR

Meetings	Date	Time and Venue
OPCUG General Meeting	Wednesday, June 11th	7:30 p.m. Auditorium of the **Canada Science and Technology Museum**, 1867 St. Laurent Blvd. http://www.sciencetech.technomuses.ca/english/index.cfm
Beginners' SIG	Wednesday, June 11th	Immediately following the OPCUG General Meeting.
Linux / Open Source SIG	Wednesday, June 11th	Immediately following the OPCUG General Meeting.
Beer BOF (Wing SIG East)	Wednesday, June 11th	10:00 p.m. (after all other SIGs) at Liam Maguire's, St. Laurent Blvd. at Innes Rd.

Please note that unless otherwise noted, SIGs meet at 9:00 p.m. (immediately following the OPCUG General Meeting).

Pizza Night and eWaste Event

Wednesday, June 11, 2014

5:30-7:00 pm: eWaste event (http://opcug.ca/public/history/eWaste/2014/eWaste.html)
6:30-7:30 pm: PIZZA PARTY (free)
7:30 pm: Regular meeting in the auditorium

At **5:30 pm**, the OPCUG will have an eWaste collection event, in partnership with Evolu-TIC of Gatineau, and in association with Ontario Electronic Stewardship (OES). Items that can be reused will be retained and refurbished, while all other older or not-useful electronic items will be taken to the processing plant and properly disposed of safely and responsibly with respect to OES guidelines. Everyone is invited to rid their homes of items no longer required. The eWaste event will be from 5:30pm to 7pm near the big tent in the front yard of the Canada Science and Technology Museum. For more information about our meeting please check our MEETINGS web page.

And at **6:30**, the club will have its annual Pizza night to thank its members for their continued support. Please come early to ensure you have a choice of several popular varieties of pizzas.

As usual, we will also supply several varieties of soft drinks, plus water and dessert.

The food will arrive, at 6:30 pm, under the big tent in the front yard of the Canada Science and Technology Museum.
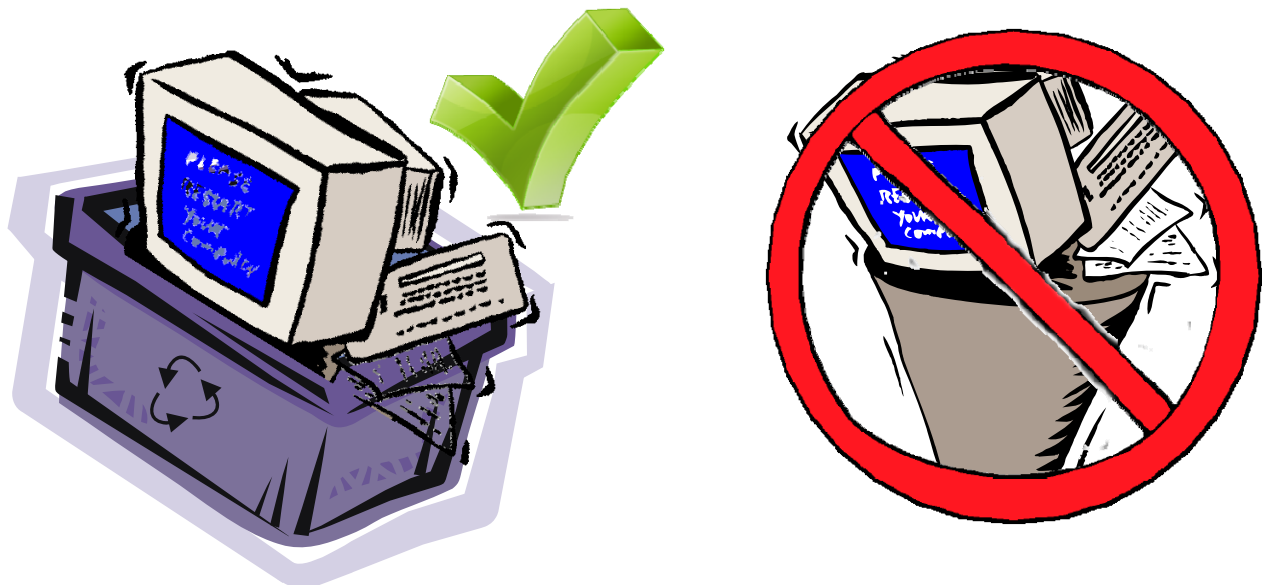
Guests are welcome.

At **7:30** the regular meeting will take place in the auditorium.
Topic: "Web Comics"
Speaker: Mark Shainblum (http://www.shainblum.com/), creator of Northguard (http://en.wikipedia.org/wiki/Northguard)

REMINDER: Parking is a flat fee of $3 after 5 pm. Payment methods includes coins, VISA, and MasterCard, and the proof of payment must be left in the car and be visible in the front windshield. For those who don't mind a couple of minutes' easy stroll, there is free parking just before the museum along Gladwin Crescent.

PRODUCT REVIEW

Resizing Windows *by Alan German*

I seem to be running into all sorts of odd problems with my recent projects. The latest involves viewing proprietary files with a reader that insists on always opening in full-screen mode. Now, this doesn't work for me since the display of this particular file type is merely one of a number of windows that I wish to have open concurrently. This means that I have to resize the window so that it only occupies part of the screen.

Now, this isn't difficult, but it does mean clicking on the window control to turn off full-screen mode, dragging the window over to one side of the screen and, finally, grabbing corners and sides of the window to resize it so that I can see the other relevant programs. This means lots of mouse activity. There has to be a better way, and indeed there is...

Sizer is a tiny app (created by the aptly-named Brian Apps) that provides the required functionality. The utility runs as a terminate-and-stay-resident program so that right-clicking on any window pops-up a resizing menu (Figure 1).
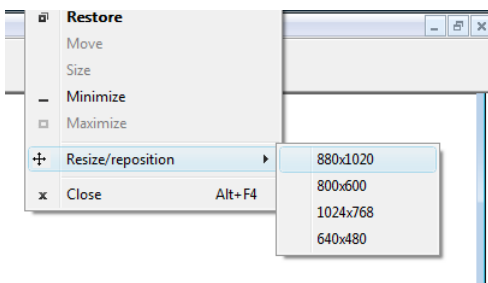

Figure 1. Resizer's pop-up menu

Several default window sizes are available but it is also possible to configure Sizer to use custom sizes. This latter feature was just what I needed for my specific application since 880x1020 was perfect for the proprietary file viewer on my monitor.

Now, with just a couple of clicks the window can be easily fixed at the desired size, and the file contents can be viewed in the context of a number of other data sources (Figure 2).
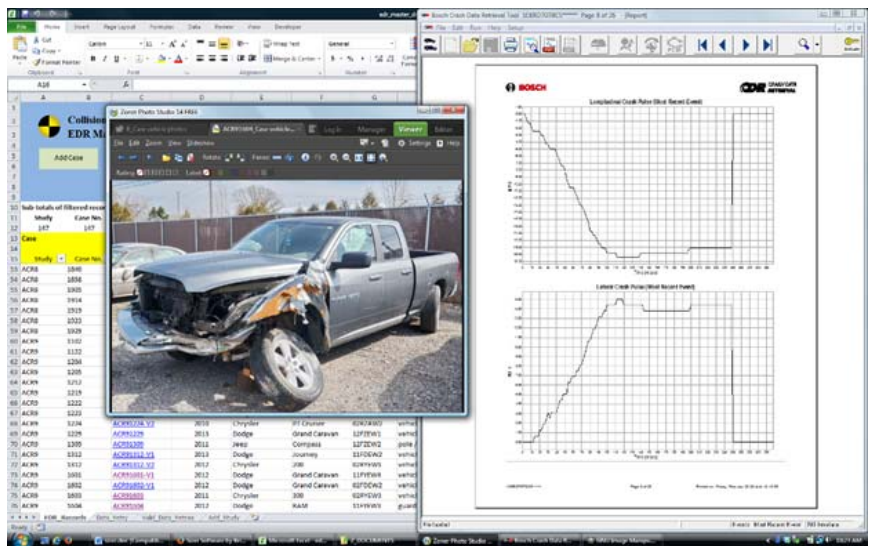

Figure 2. Multiple applications on screen with re-sized window at right

Most applications are well behaved and when set for windowed operation they retain the previous window size and its location when the program is rerun. However, clearly, some programmers don't adhere to this highly-desirable doctrine. In such situations, Sizer can come to the rescue.

Bottom Line

Sizer (Freeware)
Brian Apps
Version 3.34
http://www.brianapps.net/sizer/

ARTICLE

Installation Caution   *by Alan German*

If you are a fan of free and open-source software, have you noticed an insidious trend to bundle unwanted additions in the installers of various packages? Some of these are downright sneaky. They offer an express and/or recommended installation process simply by clicking the Install button at the bottom of a dialogue box. The trouble is we are all so used to clicking on multiple buttons like this that there is a distinct danger of installing something that we really don't want.

So, first we should scan that long, detailed, and thoroughly-boring End User Licence Agreement (EULA) to make sure that we aren't agreeing to something that we don't agree with! And, secondly, we are now also pretty well obligated to carefully read the "install now" screen to see what is actually going to happen – rather than we what we think is going to happen!

For example, note in the above dialogue box for FreeFileSync that, although the express option tells you that it is going to change your default home page and search engine, it is necessary to check the "Custom installation (advanced)" button in order to turn these options off (as shown) and only install FreeFileSync.

It's sad to see even the developers of some open-source products falling under the evil spell to make money by peddling such junk software as an "option". Various flame wars (e.g. FileZilla Forums) are testimony to the dissatisfaction of many users with this process where the "it's just an option" argument gets pretty short shrift.

In contrast, some developers (including the authors of GIMP) have taken down their pages on SourceForge because SourceForge has been bundling junk software with modified installers for the developers' programs.

So, what's a poor end user to do? Clearly, one needs to read the text that is associated with the installation of free software, especially that immediately preceding the option to press an "Install" button. However, even this isn't always sufficient. Sometimes, one needs to select "Custom Install" (or similar), rather than some pre-checked "Express Install" option, as only this secondary route will provide a series of boxes that must be unchecked in order to prevent unwanted software from being installed.
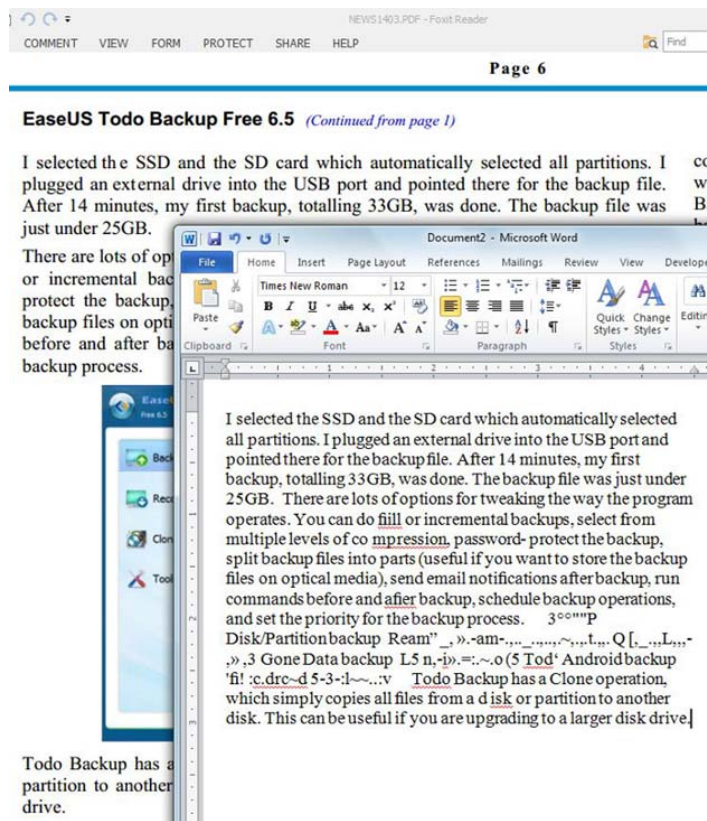
And, it's even worse than that. For at least one software offering, that used "OpenCandy" for its installer, I had to go through three or four junk software installation options – and decline each in turn! – before I got to the final screen that I needed to install the software of choice.

It's a software jungle out there so – be aware of what you are being offered – and be especially aware of what you accept!

♦♦♦

OCR   *(Continued from page 1)*

However, after a little trial and error, I successfully saw the text that I had selected displayed in the pop-up box. The same text had also been copied to the Windows clipboard and so could readily be pasted into a word-processing file. The following figure shows the results for one section of a page from OPCUG's newsletter.

A number of points are worth noting:

1. The basic text is generally reproduced quite well by the OCR process.

2. The OCR process does not capture new paragraphs. When pasted from the clipboard into a word processor, the text appears as a single, continuous stream. For example, the text from the first and second paragraphs are run together as "…just under 25GB. There are lots…"

3. There are some errors in transcription, e.g. "fiill" should have been "full" and "afier" should have been "after" . [However, sometimes the OCR process is deadly accurate – too accurate perhaps. The display of "co mpression" and "d isk" (direct from my Foxit Reader software) were both faithfully reproduced in the OCR'd text!]

4. The program has an editable suggestions.txt file that can be used to "train" the OCR to recognize (or at least replace) mistakes such as those noted above.

5. If too many errors are noted, it may be worthwhile changing the on-screen zoom level for the original PDF file, to show larger text, and then re-running the OCR process.

6. The occurrence of a block of nonsensical text (e.g. _, ».-am-.,..._..,..,.~,.,.t.,,. Q [,_.,,L,,,-,») clearly indicates the location of the graphic on the page so it easy to identify and remove this material. [However, note that the OCR software does a creditable job of identifying snippets of text contained in the image (e.g. Disk/Partition backup)!]

7. Running a spell checker on the word-processing file is quite successful in identifying specific problems in the transcribed text (e.g. "co mpression")

8. Italic text didn't phase the OCR software; it correctly identified "Todo Backup has a *Clone* operation…"

When testing the software on another PDF file, it was found that this free software did a much better job on tabular information than a number of commercial applications. While the data elements from a sample table were captured as a single "paragraph" on the clipboard, the individual numeric values were clearly delimited by spaces and held the possibility of being transferred fairly easily into individual columns in a spreadsheet. Since this is my colleague's specific application, the program shows considerable promise.

But, even better, when backing up the recently-changed files on my hard drive, I discovered that the \Capture2Text\Output folder contained a file named ocr.txt that held these tabular values – in a simple ASCII format – in individual rows! While this is clearly, an interim file created by the program as part of the OCR process, if not overwritten by another screen capture, it is available for use outside of the program. And, the data format should make transferring the tabular values to a spreadsheet really simple.

Capture2Text does a pretty good job at capturing on-screen text, converting it to machine-readable form through an OCR process, and allowing the results to be saved to a file. If you have such a need, give this program a try. You may not be disappointed and, given the price, you certainly won't be out of pocket!

Bottom Line

Capture2Text (Open source)
Christopher Brochtrup
Version 3.1
http://capture2text.sourceforge.net/

OPCUG Free Software Guide – Part 49

Compiled by
Alan German

This guide features an annotated list of free computer programs. The software mentioned has not been reviewed (except where noted) nor have any tests necessarily been conducted. Consequently, no guarantees are provided that the individual programs will perform as described. Rather the list of available software is provided for the information of our members who may find one or more of the programs useful.

NetWorx
This simple yet powerful tool helps you to evaluate your bandwidth situation. You can collect usage data and measure the speed of your Internet connection. NetWorx can help identify the sources of network problems, ensure that you do not exceed the bandwidth limits specified by your ISP, or track down suspicious network activity.
Current Release: Version 5.2.12
Web Site: http://www.softperfect.com/products/networx/

JetClean
With a single click, JetClean gives you that fresh, new PC feeling again! Lightweight and easy to use. Removes system junk and improves your PC's performance.
Current Release: Version 1.5.0
Web Site: http://www.bluesprig.com/jetclean.html

Trillian
A modern instant messaging client for home and work that prioritizes interoperability and security. Start a chat on your iPad at the breakfast table; continue on your phone as you take the train to work; wrap things up when you sit down in front of your desktop. Trillian automatically keeps all of your devices synchronized throughout the day.
Web Site: https://www.trillian.im/

Auto-Mute
Do you wish your laptop wouldn't disturb those around you when you turn it on by playing a loud sound? Do screensavers and other applications cause you similar grief? Auto-Mute will keep your machine quiet – automatically – or you can use a keyboard shortcut.
Current Release: Version 4.5.0.71
Web Site: http://auto-mute.com/

Skitch
Get your point across with fewer words using annotation. Skitch provides simple tools to let you mark up any image. You can add arrows, shapes and text to point out something important, or use a highlighter on existing text.
Current Release: Version 2.3.0.159
Web Site: http://evernote.com/skitch/

Macrium Reflect Free
A free disaster recovery solution for your home computer. Create disk images using intuitive wizards with no need to shut down Windows. Easily restore a complete image or mount the image to recover individual files.
Current Release: Version 5.2.6474
Previously reviewed: Reflections on Disk Imaging Software (Alan German)
http://opcug.ca/public/Reviews/MacriumReflectFree.htm
Web Site: http://www.macrium.com/reflectfree.aspx

Microsoft Flight
Packed with new features, including state of the art aircraft, new game play elements, new mission types, and gorgeous vistas, the free version includes everything you need to take-off and fly around the Big Island of Hawaii.
Web Site: https://microsoftflight.com/en-us/home/

Wise Registry Cleaner
Locate and correct problems in the Windows registry, such as missing references to shared DLLs, unused registration entries for file extensions, and missing referenced application paths. Wise Registry Cleaner can mark registry items which are not advisable to remove, and it supports manual backup for the whole registry before running any clean-up operations.
Current Release: Version 7.94.524
Web Site: http://preview.tinyurl.com/35cc8u4

OTTAWA PC NEWS

Ottawa PC News is the newsletter of the Ottawa PC Users' Group (OPCUG), and is published monthly except in July and August. The opinions expressed in this newsletter may not necessarily represent the views of the club or its members.

Member participation is encouraged. If you would like to contribute an article to Ottawa PC News, please submit it to the newsletter editor (contact info below). Deadline for submissions is three Sundays before the next General Meeting.

Group Meetings

OPCUG meets on the second Wednesday in the month, except July and August, at the Canada Science and Technology Museum, 1867 St. Laurent Blvd, Ottawa. Meetings are 7:30–9:00 p.m. and Special Interest Groups (SIGs) go until 10 p.m.

Parking is available at the museum for a flat fee of $3 after 5pm. Limited street parking is available for free. Details at http://opcug.ca/public/regmtg.htm

OPCUG Membership Fees:	$25 per year
Mailing Address:	3 Thatcher St., Nepean, Ontario, K2G 1S6
Web address:	**http://opcug.ca**
Bulletin board—The PUB (BBS)	**http://opcug.ca/default.htm**
Follow us on Twitter:	**http://twitter.com/opcug**

President and System Administrator
    Chris Taylor   **chris.taylor@opcug.ca**   613-727-5453
Meeting Coordinator
    (Mr.) Jocelyn Doire   **jocelyn.doire@opcug.ca**
Treasurer
    Alan German   **alan.german@opcug.ca**
Secretary
    Gail Eagen   **gail.eagen@opcug.ca**
Membership Chairman
    Mark Cayer   **mark.cayer@opcug.ca**   613-823-0354
Newsletter
    Brigitte Lord   **brigittelord@opcug.ca**
    (editor/layout)
    (Mr.) Jocelyn Doire   **jocelyn.doire@opcug.ca**
    (e-mail distribution)
Public Relations
    Jeff Dubois   **PR@opcug.ca**   613-366-7936
Facilities
    Bob Walker   613-489-2084
Webmaster
    Brigitte Lord   **opcug-webmaster2@opcug.ca**
Privacy Director
    Wayne Houston   **privacy2@opcug.ca**
Special Events Coordinator
    Bob Gowan   **bob.gowan@opcug.ca**
Beginners' SIG
    Chris Taylor   **chris.taylor@opcug.ca**   613-727-5453
Linux / Open-Source SIG
    (vacant)

Reduce, Reuse, Recycle

Bring your old computer books, software, hardware, and paraphernalia you want to GIVE AWAY to the General Meetings, and leave them at the table near the auditorium's entrance. Please limit magazines to publication dates of less than two years old.

You may TAKE AWAY any items of use to you.

Any items left over at the end of the meeting have to be taken back home by those who brought them in.