PPCUG*NEWS*

A PUBLICATION OF THE PRINCETON PC USERS GROUP

Douglas Dixon

Podcasting, Blogs, Vlogs, & RSS -- Personal Mass Media Phenomenon
Monday, November 14, 2005

New Internet technologies are changing how people communicate. It is estimated that there are over 5 million Podcast listeners and that this will grow to 60 million by 2010. Doug Dixon will give an introduction to Podcasting (radio shows on your computer or media player). Blogs (online journals), Vlogs (Video Blogs), and RSS (distributed news headlines on the Web). Doug will discuss how they are used, how you can tap in and learn, and how to get started.

Douglas Dixon is an independent technology consultant, author, and speaker specializing in digital media. He makes his articles and technical references freely available on his Manifest Technology website (www.manifest-tech.com).

Lawrence Library
Meeting Rooms 1 & 2
US Alternate Route 1 South & Darrah Lane, Lawrenceville, NJ

*Meetings of an organization at any of the facilities of the Mercer County Library System
in no way imply endorsement of its programs.*

In this issue:

Coming Events:

December 12—Membership Party
January 9—TBA
February 13—TBA
March 13—TBA

About PPCUG

MINUTES OF THE OCTOBER MEETING

Clarke started the meeting at 7:36 p.m.

Martin Mosho queried as to where would be a good place to host a web site that he plans to develop. On this site, he plans to collect money for a book that he'll be selling. Yours truly, Joe Budelis, stated that domain registration, web hosting and web design services are available on my web site from reputable, dependable and reliable companies. Others in the audience suggested LoHost, 1&1, and especially for eCommerce, GoDaddy and Yahoo. MSN and Verizon also offer web space; however, these might not be the best options for eCommerce.

There was a question about a machine running WinXP that was rebuilt by a friend of the questioner. There were some problems in the summer – it would continually reboot itself. There were lots of virus problems when she was simply using a cable modem; now she has a router. She mentioned that she was using Panda antivirus. Can not uninstall Panda. Now that she is attempting to use the machine again, she finds that she can not go onto the web. If she tries to go into Control Panel, it reboots. John Lemasney said that it sounded like a virus and/or processor overheating. He recommended that Windows updates be installed. Kim pointed out that a memory problem could also cause reboots. It was recommended that the PC be opened and all dust be blown out. Also, all boards should be checked that they are plugged in tight. It could be a problem with the hard drive so scandisk should be run to do a surface scan. The problem could certainly be a combination of virus infection and hardware. She should make sure all fans are running. A software firewall, such as Zonealarm which is free, should be installed and the MS XP firewall turned off. MS Antispyware which is currently free in beta testing may be worth installing although it isn't that good. Ad-Aware and Spybot are free antivirus programs; many of us run both of them.

Sol introduced John Lemasney's presentation on Open Source Software.

John stated that he was presenting as himself and as a representative of the Linux Users Group of Princeton (http://www.lugip.org/) which meets in the same building albeit in another room. John's presentation is at http://ghost.rider.edu/cii/presen/oss.

John Lemasney has audio recorded the meeting and has made it available at
http://orion.rider.edu/podcasts/open_source_software_28474323.mp3.

- *Joe Budelis*

In next month's newsletter there will be a synopsis of John's talk. *- Editor*

◄◄◄◄►►►►

Help Wanted
Treasurer

The group is in need of a Treasurer to track membership dues, pay bills and make reimbursements.　See Paul Kurivchack for more information.

Help Wanted
Secretary

The group is in need of a Secretary to take minutes of the meetings.　Also to read and response to club email. See Joe Budelis for more information.

President's Message

John Lemasney in his talk listed several Open Source applications that sounded very interesting. If you agree, maybe you can explore one of these applications and give a talk or write an article.

At the November meeting we will be accepting nominees for the various Club Officers. Then in December we will vote on next years Officers. If you are interested in any of the positions, especially Treasurer and/or Secretary, please contact me or any of the Board members.

We are seeking presenters for our Winter meetings. Please contact Sol Libes (sol@libes.com) if you wish to make a presentation or have a topic and/or speaker to suggest.

- Clarke Walker

◄◄◄◄◄►►►►

Link of the Month

Don Arrowsmith, former Secretary of the Club, has submitted this answer to the question "What *could* Microsoft do with the error reports we send them?":

http://msexp.streamnavig.com/msexp/player.asp?
e=E9999&s=9999_en_w&f=9999_en&uid=0003BFFD82DDFD4E&lng=nl&cou=du#

Have you discovered a useful link? Then share it with the members of the P PC UG.

◄◄◄◄◄►►►►

Product Review: Brother MFC-7820N

by Lewis A. Edge, Jr., lewis@edgeassoc.com

Member of the Princeton PC Users Group

After giving me more than a decade of service and processing tens of thousands of FAX transmissions, my heavy-duty, expensive Canon laser L780 fax machine finally quit working. Since my machine was so old, the local dealer told me that it would be a couple of weeks before the needed repair parts could be obtained. I could not do without a FAX machine that long.

This sent me looking for a replacement. The many choices at the local office supply stores were mind-boggling, but the Brother MFC-7820 caught my eye for several reasons. It was selling for less than $300 after rebate, the 2,500 page replacement toner cartridges sell at retail for less than $60 (less than 2½ ¢ per page) and this machine has additional features that I could not have imagined when I bought my Canon FAX machine more than ten years ago.

Aside from being able to scan a 20 page FAX in about a minute and then transmit the pages automatically from its memory, the Brother MFC-7820N has a built-in CAT-5 network interface enabling it to serve as a network laser printer. Any user on the network can send a FAX directly through this machine without printing it first. This machine also performs well as a digital 20 ppm auto-feed photocopy machine. The machine's automatic document feeder, which holds about 35 pages, also works with its scanner, enabling anyone on the network to

scan multiple page documents into any application that supports scanning. I use it to automatically scan multiple pages into Adobe Acrobat which is much more convenient than hand-feeding the pages into my flat-bed scanner.

In addition to its automatic document feeder, this machine also has a flat platen to scan, copy or FAX bound or bulky documents. This multi-function machine's built-in Website enables the network administrator to set up speed-dial and one-touch dialing from a computer, monitor the status of the machine's consumables and obtain other statistics.

Things to like:
1. This machine has the lowest price that I've found for a machine with so many features and it is simple to set up.
2. Scanning a multi-page FAX is fast and reliable and it automatically sends documents at the fastest speed possible to other FAX machines from memory; trying again if the receiving machine's line is busy or doesn't answer.
3. From a cold start this machine can print or copy faster than other machines and its FAX, print and copy image quality is excellent.
4. The cost of consumables is reasonable and dramatically less than inkjet.
5. This multi-function machine is surprisingly compact, lightweight and takes up much less space than a separate scanner, copier, network printer and FAX machine would take. FAX pages come out face down in the same order as they were received, so you don't need to reshuffle the pages into the right order.

Things not to like:
1. The paper tray holds only a half-ream of paper.
2. Scan quality is adequate but not good enough for critical work. For serious graphics work you would need to use a dedicated scanner.
3. Since this machine was just introduced earlier in this calendar year, its long-term reliability is unknown.
4. It copies and prints only on letter-size paper, despite specifications that say otherwise.
5. If you depend upon this machine, or any other multi-function machine, for all of its capabilities you will lose them all when it needs repair or replacement.

In summary, the Brother MFC-7820N takes up less space and costs less than buying a separate printer, scanner, FAX machine and copier. It also costs less than other laser multi-function machines. Yet it does all of those jobs capably and quickly. This makes the Brother MFC-7820N ideal for someone with a tight budget and a small office or home application where all of those features would be useful.

◄◄◄◄►►►►

Indexing the Web: Spiders, Web Crawlers & Bots

by Brian K. Lewis, Ph.D., www.spcug.org
Member of the Sarasota, FL PC Users Group

Have you ever wondered how search engines such as Google manage to get the answers to your queries so rapidly? How could they search the web that fast, I mean usually less than one second to find the words you ask for? Well --- They don't. Actually the searching goes on constantly, 24/7. And, the mechanism they use is just a modification of what you use for browsing the web.

Although you may have heard about spiders, web crawlers and web bots, they don't actually traverse the web any more than does your web browser (Internet Explorer or Firefox or whatever browser you use). Instead they download web pages that are then scanned and the significant words added to an index.

To simplify the terminology, I will refer to all the web searching programs as "spiders". (It takes less space and is easier to type.) These spiders are programs designed to find web addresses (URL's) and to

download the pages. Some also do the indexing of the words on the page. However, Google uses a separate indexing program and stores the downloaded pages for future reference. Now if a single spider were being used to locate and download pages, the task would really be impossible. Sergey Brin and Lawrence Page, the originators of Google, published a paper while they were graduate students at Stanford that utilized three spiders. Each spider kept about 300 connections open simultaneously. With four spiders they could download about 600 pages per second. This paper referred to the prototype that became the commercial Google enterprise. Even with the prototype system they were able to download and index 24 million pages in a week. Their current methodology is proprietary and so is not public, but it is probably a significant improvement over their prototype system.

We can use the original Google system as a model of what could be used by search engines to prepare the index and database of web pages that you access when you send a query. The first step is to send a list of URLs to the spider to download. This is done by a server that maintains a list of URLs. The spider will download pages and also follow any hyperlinks to other pages. Addresses of pages that were linked to the original search list are also sent back to the server to be checked to see if they were already on the list. If not, they are added to the URL lists. Not every spider uses a URL server. The spider will continue crawling the web until it reaches a dead end or a page with no further links.

As I mentioned earlier, a spider isn't just working with one page, but has hundreds of connections open to different pages. Given that there are billions of pages on the Web, even with thousands of spiders collecting information, only a small fraction of the entire web is scanned. Some web sites, such as those with news or rapidly changing information are visited hourly. Every spider has a re-visitation policy that determines how frequently a page will be revisited and checked for changes.

There is another general policy that is usually programmed into these spiders. That is called the "politeness" policy. This is used to prevent the overloading of web sites. After all, there is a finite limit to bandwidth and it would be possible to overwhelm a web site with visits from multiple spiders in a short period of time. This policy provides for an interval of time to elapse between accesses by a spider. This time interval seems to vary from 20 seconds to 3-4 minutes. This would be the case where multiple pages need to be downloaded from a single server. Revisiting indexed and stored web sites occurs at less frequent intervals.

However, even this politeness policy is sometimes inadequate. Frequent visits by spiders may result in complaints being sent back to the owner of the spider. So it is also possible to enter code on a web page which asks the spider to not access or download a page or pages. This can be done by the addition of meta tags in the page header or by a robots.txt file placed in the root directory for the web site. This is especially appropriate for game pages. These pages use a dynamic format that changes when pages are viewed or links are followed. When a spider downloads these pages the game program may respond as if a very high-speed player were logged on. This can create problems for the program and may result in crashing the game server. So we now have the robot exclusion protocol being used by owners of web pages that do not want their pages included in the search engine indexing.

In the original Google system the web pages were sent to another program referred to as the indexer. This program sorts through every word on the page and stores them in a database. The exceptions are the simple words such as a, an, the. However, simply entering the words into a database is not sufficient. They have to be identified to the particular page from which they came, the location on that page and a relative ranking in importance. The frequency with which they appear on the page as well as the position on the page may be used in determining the weight or relative rank. Words in the title or near the top of the page may be ranked as more important. So the storage of the words include the URL, and a

calculated weight in an encoded format.

The word database is then indexed to speed the retrieval of the information. This is usually done by the building of a Hash Table. Hashing evens out the alphabetical sections so that it takes no longer to find a "z" than it does a more popular letter like "m". It also separates the index from the actual entry for the word. This improves the efficiency of the storage of this information. The indexing and the Hash Table also speed the overall retrieval of the information. The complete web page is also stored in a separate location. Once the indexing process is completed, the information is available for your query.

Given the size of the web and the continuing changes to web pages, the spider's search is never ending. It may also be one where we will never have every page indexed. One other aspect of the size of the web and the time required for the crawling process is that broken links will always occur. If a page is not re-visited frequently, it may still be in the index and the database long after it has been removed from its server. Another situation may be where the URL has changed and the new location has not yet been crawled. So, the process is not perfect by any means.

The other aspect of searching the web is the design of the query you want to submit to a search engine. As I'm sure you know, you can simply list a few keywords in the search engine and hope you will get a useful result. Many times you will also get results that have no relationship to the information you are seeking. In some of these cases, you need to try the advanced search or learn to use Boolean operators. Those most frequently used are:

AND	– all the terms joined by "AND" must appear in the pages or documents.
OR	– at least one of the terms joined by "OR" must appear in the pages or documents.
NOT	– the term or terms following "NOT" must appear.
Quotation marks	– Words between quotation marks must appear as a phrase.
Followed By	– one of the terms must be followed by the other.
Near	– one of the terms must be within a specified number of words of the other.

Generally, search engines can use these Boolean operators to provide results more closely aligned to the topic you are trying to locate.

Like everything else related to computers, web indexing and searching are not static technologies. The search engine companies are researching "natural language" queries such as those handled by "Ask Jeeves". Currently, these queries can accommodate only relatively simple phrases. However, there is heavy competition to develop an engine that can work with much more complex queries. Another area that is being pursued is "concept-based" searching. This would use a form of statistical analysis to deter-mine if the page fit your query. And, as you may have read, Google has plans to put the content of the world's libraries on the web.

Just imagine what it would be like if we didn't have these search engines to help us find information on the web. So good searching and I hope you find what you are looking for.

Dr. Lewis is a former university & medical school professor. He has been working with personal com-puters for more than thirty years. He can be reached via e-mail: bwsail@yahoo.com.

◀◀◀◀▷▷▷▷

Network Your Home or Office Internet

by Ira Wilsker, iwilsker@apcug.net
APCUG Director; Columnist, The Examiner, Beaumont, TX

On my weekly radio show (KLVI 560AM 1-3pm Saturdays) I am frequently asked questions about connecting an internet source to a home network, such that several computers can simultaneously share the internet connection. This type of home networking is typically only practical if there is a broadband source, and generally impractical if the source is dialup, and a voice phone line is to be shared for internet access.

Basically, there are three good options available, and each has its relative advantages and disadvantages, and varies in terms of cost effectiveness based on individual circumstances. The three basic methods or technologies available for home or small office networking are wired, wireless, and "homeplug".

Wired networking is a viable choice if it is practical to run a network cable between computers. The cable, which can be relatively inexpensive if purchased in bulk from electronics or home/hardware stores, but somewhat pricey if purchased in pre-cut sections at retail stores. Cable can be run through an attic, suspended from cup hooks along a ceiling, run along baseboards, or other methods that comply with appropriate safety and electrical codes. Other than running the cable, the installation of the network is fairly simple, and requires readily available and generally inexpensive hardware, manufactured by countless companies. A device called a router is connected to the broadband input, and the computers are connected via network cable to the router. Most wired routers come with a floppy or CD with simple configuration utilities that allow the network to be setup and functioning within minutes of connection. Wired routers can be purchased from discount stores, electronics stores, office supply stores, and a variety of other sources. Since almost all contemporary computers, both desktop and notebook, come with an integral network connector, generally no other hardware is required other than the router and cable. In terms of security, wired is generally the most secure method of networking. Many of the inexpensive wired routers include some type of hardware firewall to protect the network from intrusion, but it is still generally a good idea to have a software firewall on each computer.

One of the most popular methods of home or small office networking is wireless, often referred to with its industry standard designation, 802.11, with a letter suffix, such as an "a", "b", or "g". The suffix indicates which version of the standard hardware complies with in terms of frequency, and bandwidth. Generally, the least expensive and most widely used type is the 802.11b, which shares the same 2.4 GHz digital frequencies of many cordless phones and other household electronics, which may possibly cause some interference with the wireless connection. Under ideal conditions, 802.11b can transfer data at a rate of up to11 megabytes per second (Mbps). The newer 802.11a and 802.11g standards offer higher transmission speeds, and will likely encounter less interference from household appliances, but some interference is always a risk.

To connect computers to a wireless network requires a wireless router, and a device called an "access point" for each computer. The main advantage to a wireless network is the ease of installation, and the flexibility of not being connected to a wire while surfing the net. The wireless router is connected to the broadband input, and an antenna on the router broadcasts the signal to the access points. Range is a function of building construction and interference, but at typical household and small office distances and construction, a single router is typically adequate. Users should be sure to follow the instructions included with the router in terms of placement in order to maximize the signal. Many wireless routers include some form of hardware firewall, but again a software firewall on each computer is generally recommended. "Access Points" are generally small devices that attach to a computer either through a USB port, or to the network input on the computer. As they are radio transceivers, placement will affect performance. It is imperative to note that unless the integral encryption and other security options offered by the router and access points are implemented, there can be some serious security issues with wireless connections, in that it is easy for unauthorized persons to access the network.

Wireless routers and access points are readily available wherever electronics are sold, including discount stores, warehouse clubs, office supply stores, and other retailers. As the technology has been improving, prices have been plunging. Some provisos are in order at this point; first, be sure that whatever wireless hardware is purchased, be sure that it is compatible, in that the suffix "a", "b", or "g" match, although much of the hardware available today will work with any standard, often listed as "a/b/g". A second proviso is the "weakest link" factor, in that a wireless (or any other network) connection is only as fast as the slowest segment. If a broadband connection is running at 2Mbps, a faster wireless connection may be wasted unless files or hardware (such as printers or video) are being shared among the network. It should also be noted that many of the new home theater devices now coming on the market utilize the same wireless interconnectivity and hardware as computers, and may require the newer and faster standards of connectivity.

There is a third technology available for home and small office networking that is still in its infancy, but holds a promising future, and most of the major networking hardware manufacturers such as Belkin, Linksys, GigaFast, IOGear, Netgear, and others are producing compatible hardware. This technology is called "Homeplug" (www.homeplug.org) that can safely utilize the existing household electrical wiring as an already wired network. Homeplug compatible hardware is priced about the same as wireless, but is not yet as readily available as traditional wired or wireless hardware, but it may be worth the effort to shop for it. The standard, which meets UL and other safety requirements, is possibly the easiest to configure, has a 14Mbps bandwidth, and is more secure than wireless. To connect broadband to Homeplug, the broadband input is plugged into a Homeplug input device (similar to a router), and that device is then simply plugged into a standard wall electrical outlet. Any computer in the home, as well as many of the newer entertainment devices, are then connected to any other electrical outlet in the house via an adapter that connects to the computer with a traditional USB or network plug. In terms of security, the broadband signal does not typically pass through the household junction box, making it extremely difficult for neighbors and others to access the connection or network. As with all other forms of networking, a software firewall is still necessary on each computer.

If a user is bearing the cost of broadband, and has more than one computer in the home or office, then one of these networking technologies may be a useful method to connect those computers to the internet.

<div align="center">

◄◄◄◄◊►►►►

</div>