

The P U L P

Newsletter of the Hartford User Group Exchange

<http://www.huge.org>

Volume 24 Issue 7



October 18th General Meeting: Topic:

iWork: Pages

East Hartford Public Library
Main St. & Central Ave., East Hartford, CT.

Q&A Session: 6PM–7:15PM
Meeting starts at: 7:15PM



Huge This Month:

- October 18:** General Meeting See above 7:00 P.M.
- November 5:** Deadline for **ALL** Articles. Please upload articles to editors@huge.org, or give them to the Pulp Editor **Midnight**
- November 15:** Q & A Session Starts at 6:00 P.M.
- November 15:** General Meeting Starts at 7:15 P.M.

Contents	Page
From the Editor	2
Stu's Quiz Page	3
Netiquette	4
Indexing the Web	5
Tech News	8

The **PULP** is published monthly by and for members of the Hartford User Group Exchange, Inc. (**HUGE**). **HUGE** is a nonprofit organization whose aim is to provide an exchange of information between users of personal computers. The **PULP** is not in any way affiliated with any computer manufacturer or software company. Original, uncopyrighted articles appearing in the **PULP** may be reproduced without prior permission by other nonprofit groups. Please give credit to the author and the **PULP**, and send a copy to **HUGE**. The opinions and views herein are those of the authors and not necessarily those of **HUGE**. Damages caused by use or abuse of information appearing in the **PULP** are the sole responsibility of the user of the information. We reserve the right to edit or reject any articles submitted for publication in the **PULP**. Trademarks used in this publication belong to the respective owners of those trademarks.

MEETING LOCATIONS

East Hartford Public Library
Main & Central Avenue
in the Lion's Community Room
Downstairs

Wethersfield Public Library
500 Silas Deane Hwy.,
Wethersfield, CT

From The Editor

by Pat Teevan

You've probably gotten bored reading about my [mis]adventures working on the Pulp, so you'll be happy to know that I'm not writing about that this month. Instead, I've got some bits of information to forward that didn't qualify for their own articles.

First, George informs me that nominations for HUGE officer positions will be accepted at the meeting this month and the annual elections will be held at the November meeting. This is your chance to put your talents to work for the club. Please consider running for one of the club offices.

Next, the nice folks at O'Reilly, famous for the "...in a Nutshell" and "Missing Manual" books, send us a newsletter every month as part of their User Group program. They offer to send us a free copy of a book if a review is done and appears in our newsletter. If you'd be interested in writing a review for the Pulp (and getting a free book in the process) let one of the club officers know and they'll get it set up.

Even if you don't want to write a review, O'Reilly offers User Group members 30% off any O'Reilly, No Starch, Paraglyph, PC

Publishing, Pragmatic Bookshelf, SitePoint, or Syngress book you purchase directly from the O'Reilly online store Just use code **DSUG** when ordering online at <http://www.oreilly.com/> or by phone at 800-998-9938.

Finally, as the days shorten and gets gets dark earlier, remember that kids may be out playing in the semi-dark at dusk. Please drive carefully in the evening and especially on Halloween, which will be here before you know it.

...as a matter of fact, it's starting to seem a little creepy around here even now. You know, a sudden chill, an unexplained draft, that whole "Sixth Sense" kind of feeling...

But hey—I'm sure it's just my imagination playing a trick on me. After all there are no such things as ghosts right??

Until next month...

Happy computing!!

The Editor.

Netiquette Equals Friendly Online Neighborhoods

By Gabe Goldberg,

APCUG Advisor; Columnist, AARP Computer & Technology Website, www.aarp.org

Etiquette -- proper behavior in various situations -- has concerned humans through the ages. Google [www.google.com] turns up references to early telephone manners, proper Medieval-era knightly behavior, and how shells and sharp stones became today's knives and forks (and, of course, which hand to hold them in).

Though the Internet became generally accessible about ten years ago, academics and researchers had used it for decades before that. Just as civilization accommodated sudden widespread automobile and telephone usage, developing common practices (stop for red lights, answer telephones with "Hello"), new online technologies created the need for corresponding innovations in manners.

Called Netiquette ('Net + etiquette'), the new discipline really just requires using old manners in a new setting. But since going online is like traveling to another country, a guidebook is useful.

The first and simplest chapter is simply "Follow the golden rule" [en.wikipedia.org/wiki/Ethic_of_Reciprocity]: Treat people online with the same courtesy you expect. Without normal conversational cues such as body language, facial expression, and tone of voice, it's easy to misunderstand someone's meaning. You may remember Emily Litella on classic Saturday Night Live television, mishearing the need to reduce TV violence as "eliminating violins". Even stranger miscommunication occurs online, so give others the benefit of the doubt

Out-of-control ranting is sometimes called "flaming". Before responding angrily, consider how you'd react to receiving the note you're about to

send. Remember that once sent, e-mail and other online communications take on a life of their own, being forever retrievable with your name attached. And asking "Did you mean..." can avoid having to give Emily Litella's trademark "Never mind" if your interpretation was off target. Replying calmly often gets a conversation back on track without anyone suffering a "flame war", an unproductive exchange of angry notes.

**Called Netiquette
(‘Net + etiquette’),
the new discipline really just
requires using old man-
ners in a new setting.**

E-mail is likely the most commonly addressed Netiquette area. Searching Google for email + netiquette provides an encyclopedia of advice, some general, some for personal use, and much for business settings. Common tips are DON'T USE ALL CAPITAL LETTERS (it's hard to read and is considered "shouting"); be brief; use meaningful subject lines; quote just enough from what you're answering to provide useful context; don't forward to everyone you know jokes, rumors, hoaxes, chain letters, charity appeals, and such, even if an arriving note tells you to do so; don't send "Me too" notes to discussion lists; don't send attachments without getting the recipient's permission; send plain text e-mail unless all your addressees prefer HTML-format; etc.

Accept and cheerfully answer questions asked by people newly online -- remember that we all started with basics. Interesting e-mail tips are available at places like Writers Write [www.writerswrite.com/journal/dec99/pirillo1.htm], Emailreplies.com [www.emailreplies.com/], and Yale University Library [www.library.yale.edu/training/netiquette/].

Continued on page 6

Indexing the Web: Spiders, Web Crawlers & Bots

By *Brian K. Lewis, Ph.D.*,

Sarasota PC Monitor, Sarasota FL PC Users Group

www.spcug.org

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

Continued on next page

Netiquette...—continued from page 4**Not all tips are "Don't"!**

If you receive multiple replies to a question you ask online, it's advisable -- this will make people love you -- to summarize answers for everyone else who saw the question. Be careful with humor: what's obvious and sidesplitting to you may baffle or annoy someone else. When appropriate, and sparingly, use "smileys" -- symbols like ;-) -- to show that you're not serious. And if you receive what seems to be an unlikely rumor, do the sender a favor by checking it out at a site like snopes.com [www.snopes.com] and reporting what you find.

The Netiquette Home Page

[www.albion.com/netiquette/] is an entertaining and informative reference. It notes that Netiquette covers both common courtesy online and informal cyberspace "rules of the road". The site lists and explains rules with friendly illustrations. It's interesting that not only do the rules all agree with that Golden Rule mentioned earlier, they're often different ways of giving the same advice. For example, using the same standards of behavior online that you follow in real life is an easy way to make yourself look good online.

[This article originated on AARP's Computers and Technology Web site, www.aarp.org/computers, and is copyrighted by AARP. All rights are reserved; it may be reproduced, downloaded, disseminated, or transferred, for single use, or by nonprofit organizations for educational purposes, with attribution to AARP. It should be unchanged and this paragraph included. Please e-mail Gabe Goldberg at gabe@gabegold.com when you use it, or for permission to excerpt or condense.](http://www.aarp.org/computers)

There is no restriction against any non-profit group using this article as long as it is kept in context with proper credit given the author. The Editorial Committee of the Association of Personal Computer User Groups (APCUG), an international organization of which this group is a member, brings this article to you.

Indexing...—continued from page 5

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

Continued on next page

Indexing...—continued from page 5

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 determine 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 computers for more than thirty years. He can be reached via e-mail: [bwsail at yahoo.com](mailto:bwsail@yahoo.com).

There is no restriction against any non-profit group using this article as long as it is kept in context with proper credit given the author. The Editorial Committee of the Association of Personal Computer User Groups (APCUG), an international organization of which this group is a member, brings this article to you.

Tech News

By Sue Crane, Vice President & Editor, Big Bear Computer Club, CA
sue.crane@charter.net

Nothing is Ever Really FREE

Linus Torvalds, creator of the Linux OS, has begun protecting his trademark for the term Linux. He's doing so to protect users from unauthorized and confusing use of the name. Torvalds first trademarked Linux several years ago but has rarely defended its use. The licensing fees for Linux are modest. LMI is charging \$200 to \$5000 for each license; most sublicensees end up paying \$200 to \$500. Most Linux companies have enthusiastically agreed to the licensing terms.

AMD Asks Intel for a duel

Major US newspapers in September showed an AMD processor standing in a boxing ring waiting for Intel to appear for a duel later this year so consumers can see who has the best dual core processor. Dual-core chips have two processor cores on a single piece of silicon, boosting performance and reducing power consumption compared to two single-core processors. AMD released its first dual-core Opteron processors earlier this year. Intel said that it will release its first dual-core server chips later this year, ahead of its original schedule. But regarding the duel, Intel was not available for comment.

PC Has Water-cooled Radiator

NEC took the wraps off its Valuestar G Type C, a new desktop PC that has a radiator embedded in the water cooler unit built into the back of the chassis. The company said the added refrigeration will let consumers overwork the included Intel Celeron processor but keep the PC running at 30 decibels, which is whisper voice. PC noise is increasingly a concern as more powerful computers require stronger and often louder cooling systems.

New IE 7.0 is for Windows Only!

Sources at Microsoft have announced that IE 7.0 will henceforth be referred to as Windows IE 7.0, highlighting the fact that the browser is integrated with Windows and isn't a standalone product. According to a source at Microsoft, the change signifies that IE will no longer be available for other platforms. Users should simply consider IE 7.0 to be part of Windows.

No More Textbooks?

Students at Empire High School in Vail, AZ started class this year with no textbooks. Instead, the school issued laptop computers to each of its 340 students, becoming one of the first U.S. public schools to turn away from printed textbooks. Empire High, which opened for the first time this year, was designed specifically to have a textbook-free environment.

TMPs Make E-Commerce Safer

Trusted Platform Modules, if you've never heard of them, are chips that store cryptographic information needed to unlock hard drives, authenticate network log-ons and perform similar tasks. What is exciting about TPMs is their role in electronic transactions because they make sure the money and the product get to their proper destinations, via "anonymous certificates." Hacking such a certificate would not give criminals access to personal information such as credit card numbers, since the transaction is done on your side of the network, not on the server-side.

eBay Rethinks Recycling

Faster, more advanced computers, cell phones and electronics offer endless possibilities for enrichment, learning and entertainment. But as consumers and businesses keep pace with the introduction of exciting new products, we are faced with a mounting challenge: what to do with the products we're upgrading from. The Rethink Initiative brings together industry, government and environmental organizations to offer a fresh perspective and new answers to the challenge of e-waste. On their website you can find information, tools and solutions that make it easy – and even profitable – to find new users for idle computers and electronics, and responsibly recycle unwanted products. <http://rethink.ebay.com/>

There is no restriction against any non-profit group using this article as long as it is kept in context with proper credit given the author. The Editorial Committee of the Association of Personal Computer User Groups (APCUG), an international organization of which this group is a member, brings this article to you.

PULP Staff

Interim Editor Pat Teevan
Distribution George Carbonell

Membership: Anyone may become a member. Dues are \$12 per year and include a one-year subscription to The Pulp as well as access to the HUGE Public Domain disk libraries and BBS. Meeting topics, times and places can be found on page 1 of this issue.

Officers & SIG Leaders

Leader: _____ George Carbonell 568-0492 george.carbonell@comcast.net
Secretary: _____ Charles Gagliardi
Treasurer: _____ Charles Gagliardi 233-0370 epencil@worldnet.att.net
Director at Large: ___ Phil Manaker 659-4584 amanaker@earthlink.net
Director at Large: ___ Ted Bade 643-0430 tbade@cox.net
Director at Large: ___ Stuart Rabinowitz 633-9038 s.e.rabinowitz@worldnet.att.net
Web Manager: _____ Gail Smith
Membership: _____ Richard Sztaba richer1@aol.com
Integrated SIG: _____ Stuart Rabinowitz

October 2005

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18 General Meeting	19	20	21	22
23	24	25	26	27	28	29
30	31 Halloween					