SOCIETY OF ACTUARIES

Article from:

CompAct

July 2009 – Issue 32

Cool Tech

By Matthew Wilson

For this edition of Cool Tech I'm going to show you more than 200 free applications that you can download and try out. They're all free but not necessarily open source.

I start off with OpenDisc, which contains 40 free programs. You download an ISO image then burn a CD from the image. Making a CD from an ISO image is slightly different than copying a file to CD.

Next we go to PC Magazine's list of 173 free programs for you to try out. The programs range from application launchers, audio and backup all the way to networking and video applications.

Then there is BitNami. BitNami, or more correctly BitNami stacks, makes "it incredibly easy to install your favorite open source software," according to their Web site.

About a month ago I tried an experiment to see if I could improve traffic to my blog. I show you what I did and my results.

Finally, I briefly go over a Google custom search engine, a mail list manager, Feedblitz and a couple of Perl programs for scraping news articles off of your favorite news Web site.

MY VIEW

There are so many free applications available these days. It really pays to do a little investigating before purchasing a program or writing it yourself. Many Web hosts install open source applications for free.

Here are a few Web sites worth checking out:

www.Koders.com—Code snippets.
www.Hotscripts.com, *www.Sourceforge.net*—Open source software.
www.Bitnami.com, *www.Apachefriends.org/en/xampp. html*—Application packages.
www.Fantastico.com, *www.Installatron.com*—Typical Web host packages installed for free.
www.Elance.com—Hire help.

Recently, I was playing around with StumbleUpon *(www.stumbleupon.com)* and it served up a game that sucked me in. StumbleUpon serves up pages that might interest you based on your viewing history.

The game, Hell of Sand, has several streams of sand dropping into an area where you can draw walls to affect the flow. I probably spent 20 minutes drawing walls changing the sand flow. Ok, I know it sounds stupid. I got sucked in.

Hell of Sand:

http://www.andyslife.org/games/sand.php

If you're interested in games, then the *BUSINESS WEEK* Arcade has a cool set of independently produced games on the Web. And the games are free.

http://tinyurl.com/dhlszw

Did you notice that I'm using "tinyurl.com" for the URL. Check out http://tinyurl.com as a way to take those long URLs and shorten them for easy manual input.

It turns out that tinyurl.com is on *TIME* magazine's list of top 50 Web sites for 2008: http://tinyurl.com/cc45dl.

TIME magazine doesn't present the list in a convenient format, so I created my own list by using the advanced search features from Google:

http://www.google.com/advanced_search?hl=en

I searched the *TIME* Web site (Time.Com) for "50 Best Websites 2008." I included omitted results in order to get 144 links. I grabbed the HTML code and cleaned it up in a text editor. Later I show you how to clean up Google HTML using Regular Expressions (RegEx). My results are here: *http://tinyurl.com/cc45dl.*

Here's the search text: "50 Best Websites 2008," site time.com

Matthew Wilson, ASA, MAAA, can be reached at matt_wilson@ farmersinsurance.com

Stratfor.com is an intelligence Web site that charges about $300 per year for a subscription. Since I don't particularly want to spend that much for a subscription, I do the next best thing. I use the advanced search features in Google to see what I can grab for free.

Search the stratfor.com Web site and target filetype PDF. site: stratfor.com, filetype.pdf

The search results turn up over 50 reports. Here is their annual intelligence forecast for 2009: *http://tinyurl.com/c75u49.*

Now you know how to use Google to grab interesting material from a Web site. What about developing some spiders to grab more interesting items? At the end of this article I share a couple of Perl programs on spidering. You can get more information on this topic by checking out these books:

Spidering Hacks—*http://tinyurl.com/c2qvn3*

Webbots, Spiders and Screen Scrapers: A Guide to Developing Internet Agents with PHP/CURL—*http://tinyurl.com/dn4pck*

Spidering Hacks will introduce you to Perl and Regular Expressions for scraping purposes.

OpenDisc
OpenDisc is literally a disk or ISO image that you can download containing a lot of open source programs that work on Windows. I've summarized the list of programs below. You can get a better description of the programs on this page:

http://www.theopendisc.com/programs/

Design (Seven Programs)
Blender (3D graphics), Dia (Flow Charts), GIMP (Image Manipulation), Inkscape (Vector Graphics Editor), Nvu (Web page Editor), Scribus (Desktop Publishing) and Tux Paint (Children's Paint Program).

Games (Four Programs)
The Battle for Wesnoth, Enigma, Neverball and Neverputt.

Internet (10 Programs)
Azureus (bitTorrent Client), FileZilla (File Transfer), Firefox (Browser), HTTrack (Offline Browsing), Pigdin (Instant Messaging), RSSOwl (RSS Reader), SeaMonkey (Internet Suite), Thunderbird (E-Mail), TightVNC (PC Remote Control) and WinSCP (Secure File Transfer).

Multimedia (Six Programs)
Audacity (Audio), Celestia (Universe Simulation), Really Slick Screensavers, Stellarium (Planitarium), Sumatra PDF (PDF Reader) and VLC (Media Player).

Productivity (Five Programs)
GnuCash (Accounting), MoinMoinWiki (Wiki), Notepad2 (Replace Notepad), OpenOffice (Replace Microsoft Office) and PDFCreator (Create PDF Files).

Utilities (Eight Programs)
7-Zip (Replace WinZip), Abakt (Backup Tool), Clamwin (Anti-Virus), GTK+ (Create Graphical User Interfaces), HealthMonitor (Windows Monitoring Tool), TrueCrypt, Encryption and Workrave (Prevent Repetitive Strain Injury).

Burn a CD from an ISO image
If you want to burn a CD for the OpenDisc programs, then check out the following link for software capable of burning a CD from an ISO image:

http://en.wikipedia.org/wiki/List_of_ISO_image_software

The Best Free Software of 2009
PC Magazine's mother load of free software list:

http://www.pcmag.com/article2/0,2817,2338803,00.asp

Here's the print version of the article with everything on one page:

http://www.pcmag.com/print_article2/0,1217,a%253D235942,00.asp

There are 173 software packages listed below. The article contains links to each application. Also, some applications are classified as in the "Hall of Fame." The application categories:

App Launchers (four), Audio/Music (11), Backup (six), Blogs (eight), Browsers (nine), Calendar/PIMs (10), Communication/E-Mail (10), Conferencing (four), File Transfer/Download (eight), File Viewers/Converters (eight), Finance (four), Fun/Home (seven), Graphics (17), Ims (four), Interface Enhancers (13), Local Search (three), Office (19), Operating Systems (three), Networking (nine), RSS Readers (six), Synchronization (five) and Video (five).

BitNami

BitNami stacks make it "easy to install your favorite open source software." A stack includes a primary application plus all the dependencies necessary to run it. The installation wizard makes it very easy to complete the installation process quickly.

If you've ever tried to resolve dependency issues then you know that they can be a real black hole at times. They can suck up huge amounts of time if you let them. Here's the link:

http://bitnami.org

Try to find the general category of software that interests you. Then go to the Bitnami Web site to find out a little more about a specific application:
- **Infrastructure**
  - DjangoStack, JRubyStack, LAMPStack, LAPPStack, MAMPStack, MAPPStack, RubyStack, SAMPStack, WAMPStack and WAPPStack
- **Blog**
  - Roller and WordPress
- **Bug-Tracking**
  - Mantis, Redmine and Trac
- **Business Intelligence**
  - JasperServer
- **Content Management System (CMS)**
  - Alfresco, Drupal, Enano CMS, eZ Publish, Joomla and KnowledgeTree
- **Client Relations Management (CRM)**
  - SugarCRM
- **ECM**
  - Alfresco and KnowledgeTree
- **Forum**
  - phpBB
- **Photo Sharing**
  - Coppermine Photo Gallery
- **Planning**
  - Tracks
- **Poll Management**
  - Opina
- **Portal Server**
  - JasperServer and Liferay
- **Version Control**
  - Subversion
- **Wiki**
  - DokuWiki and MediaWiki
- **eLearning**
  - Moodle

HOW TO INCREASE YOUR BLOG TRAFFIC BY 30 PERCENT IN ONE MONTH

I've been running a blog since 2006. I have now posted thousands of articles. Recently, I started examining my Google Analytics which show how many hits specific pages got in the last month. I noticed that some older articles are still continuing to get a decent amount of traffic. I wondered if it was possible to improve these pages in order to get even more hits.

Google Analytics allows you to drill into the statistics for a specific page. For example, you can find out which keywords people used to find that page.

Once I knew what people were looking for, I set out to improve specific pages. I added more content specific to the top keywords for that page. For content I added excerpts from more news articles, content from Wikipedia and targeted videos from YouTube.

One article that I posted awhile back is called "What is Suter?" Suter is the name of a system that Israel used to hack into Syria's attack air defense system as their jets were flying toward a secret Syrian nuclear facility. I guess that's literally hacking on-the-fly. I thought

people would find this page by searching for Suter. It turns out that most people were finding the page while searching for "Big Safari." Suter is a part of the Big Safari project run by the U.S. Air Force. I had only briefly mentioned the Big Safari project in the original article.

Once I learned Big Safari was rather important, I added a paragraph explaining more about it.

I found out which tags are most popular. For example, the tag "global-financial-crisis" is more popular than "financial-crisis." Unfortunately, I had been mostly using the tag financial-crisis. So I went back to many articles in the financial-crisis category and added the tag global-financial-crisis.

I found a few popular pages that surprised me. I posted one page on Google Earth secrets that was very popular. So I went back to this page and beefed up the content even more.

I found a few new categories that people were interested in. One category is on the powerful global elite. For example, the 50 most powerful people in the world or the world's top billionaires. Another category is on Mexico, but this pertains more to the crisis in Mexico. These types of pages were very popular, so I created entirely new categories just for these topics.

I recently started something new as well. I started using Twitter (*http://www.twitter.com*). Of course, with a name like that it took someone hitting me over the head before I would give it a try. Now I'm hooked.

Think of twitter as a mini-blogging system. You have a maximum of 140 characters in each post. So why bother? I use it to post article titles and links. You can direct people back to your blog or to new articles on a topic.

With long URLs in use today, it can be difficult to squeeze an article title and link into 140 characters. Fortunately, you can use *www.tinyurl.com* to shorten the length of your links.

Sometimes I use content from Google in order to give my readers more article choices. Unfortunately, I hate the HTML code that Goggle uses. So I need to spend a few minutes cleaning up the code. Here's how I do it:

1. I use the ConText text editor which provides RegEx in its Find/Replace section.
2. I use Regular Expressions (RegEx) code for the clean-up.
   a. Replace <a .*?href with <a href
   b. Replace <div.*?> with blank
   c. Replace <td.*?> with blank
3. ".*?" means any character 0 or more times. "." means any character. "*" means 0 or more times. The "?" in this context means stop as soon as possible.
4. You can clean up Wikipedia pages using \[.*?\] which removes those pesky edit and reference marks.
5. You can find a tutorial on RegEx here: http://www.1913intel.com/demo/RegEx_Tutorial.zip

MAIL LIST MANAGER

Here is a free e-mail list manager. Now you can manage your e-mail lists yourself. You will need to set up a MySQL database to get this working, but that's usually pretty easy. Ask your host if you don't know how to create a database.

http://www.phplist.com
http://en.wikipedia.org/wiki/phplist

I used this package for about a year with good results.

I now use Feedblitz (www.feedblitz.com) for convenience. Feedblitz automatically scans my Web site each night and sends out an e-mail with new articles to my list. It's free, and it's easy for people to sign up or remove one's self from the list.

I also use Feedblitz for a newsletter where the content comes from Google. I feed Google news into Yahoo! Pipes *(http://pipes.yahoo.com/pipes/)*. Feedblitz automatically checks my link *(http://tinyurl.com/ctbfn9)* and sends out an e-mail to the subscribers. There are about 200 subscribers to this newsletter.

SOME CUSTOM SEARCH ENGINES

Google has a custom search engine where you select the sites to be searched. Just enter the URLs that you want to be searched when configuring the search engine.

http://www.google.com/coop/cse/

Check out my instance of a Google custom search engine:

http://www.1913intel.com/custom-search/

Try entering "Russia" in the search box.

Since Google came up with this custom search engine, the following section is not really needed. However, the scrapers/spiders are still pretty cool to play around with.

DOWNLOAD THREE PERL PROGRAMS

The first program, xoopSpider4.cgi, reads several news Web sites and grabs the links to news articles. The results are printed to the screen. The program is capable of writing out results to a file, but I've commented out that part of the code. Go into subroutine writeFILES and uncomment the code in order to write to files.

Now you can quickly scan articles from your favorite news sites.

Don't forget to change the properties of xoopSpider4. cgi so you can execute it. You'll need to do this to all CGI programs.

This program reads one Web site at a time, so you'll want to keep the number of Web Sites down to a manageable level.

newsScaper4.cgi is the hardcore version of xoopSpider4.cgi. It uses parallel processing, so you can load up on the Web sites. It only writes to files. However, you'll need to set up the ParallelUserAgent folder using the ParallelUserAgent-2.57.zip file.

In the past I developed newsScaper4.cgi for a custom search engine. I placed the results from scraping into a searchable MySQL database. Users simply entered keywords and pressed a button to get relevant articles for that day.

The second program, parallelSpider.cgi, grabs several Web pages in parallel and prints them to the screen.

http://www.1913intel.com/demo/scrapers.zip
http://www.1913intel.com/demo/ParallelUser Agent-2.57.zip

You will need to unzip ParalletUserAgent-2.57.zip. Place the unzipped folder on your Web site and point your program to the library.

This is a basic program showing how to do parallel processing. ■