Vol. 15 No. 3 November 2002
$2.50

MUUGlines

The Manitoba UNIX User Group Newsletter

November 12, 2002: Essential Unix for System Administrators

Sean Walberg, who presented GPG in March 2002 and Netfilter in April 2001, will be our guest presenter again this month. This presentation will focus on some of the lower levels of Unix such as files and filesystems, processes, and the boot procedure.

December 10, 2002: Gentoo Linux

Gentoo Linux is a modern distribution that focuses on performance and features. Its packaging system was inspired by the BSD ports tree, but with many enhancements. In this presentation, Shawn Wallbridge will briefly cover the install, how the Portage system works and other neat things about this distribution.

Meetings are held at the IBM offices at 400 Ellice Ave. (between Edmonton and Kennedy). When you arrive, you will have to sign in at the reception desk, and then wait for someone to take you (in groups) to the meeting room. Please try to arrive by about 7:15 PM, so the meeting can start promptly at 7:30 PM. Don't be late, or you may not get in.

Limited parking is available for free on the street, or in a lot across Ellice from IBM, for $1.00 for the evening. Indoor parking is also available nearby, at Portage Place, for $2.00 for the evening.

Xandros User Group Outreach Program

From now till the end of November, Xandros is offering Linux Users Group members worldwide a discount of 45% from the regular list price of US$99 to try Xandros Desktop (previously known as Corel Linux) for themselves. Also until the end of November is free international shipping.

You can learn more about Xandros Desktop at www.xandros.com/desktop.html. For online reviews of Xandros Desktop 1.0, see NewsForge (http://newsforge.com/newsforge/02/10/21/1749230.shtml?tid=23) and ExtremeTech (http://www.extremetech.com/article2/0,3973,648362,00.asp).

Some of the highlights of this distribution:

- Based on Debian GNU/Linux
- Simple, five clicks or less, Express Installation process for novice users and a sophisticated custom setup process for advanced users.
- First and only Linux OS to support non destructive NTFS (default for Microsoft Windows XP installation) partition resizing during installation.
- Automatic hardware detection and configuration including some popular winmodems, automounting of existing Windows and Linux drives and hot plug and play USB device support.
- Easy to use wizards including First Run Wizard, Add Printer Wizard, Connection Wizard, Desktop Shortcut Wizard and many more to simplify your configuration and customization tasks.
- Xandros File Manager drag and drop files from any source to any destination including shared Windows drives and FTP sites.
- Seamless Windows networking integration including domain level authentication.
- Windows application compatibility from CodeWeavers Crossover products allow some popular Windows applications (including MS Office 2000) to be run on Xandros Desktop without Microsoft Windows.
- Control panels for display settings including multi-head monitor support, network configuration and wireless (802.11b) network cards.
- Switch Users allows multiple users to use your system momentarily without forcing you to close down all your work and log out.
- Single mouse click web update service to install feature enhancements and maintenance releases.

POPfile: Naive Bayes (Read: Spam Filtering)

POPfile is an open source POP3 proxy that does email classification and sorting using Naive Bayes. POPfile can be found at http://popfile.sourceforge.net/.

History

The Reverend Thomas Bayes lived in the English town of Tunbridge Wells in the 18th century.

Three years after his death in 1761 the *Philosophical Transactions of the Royal Society of London* published his paper Essay towards solving a problem in the doctrine of chances that laid out the underpinnings of what would become known as Bayesian Statistics.

The most famous result from the paper is *Bayes Theorem* which shows how to calculate the probability of one event given that you know some other event has occurred. Algebraically that is:

P(A|B) = P(A) * P(B|A) / P(B)

Or the probability of A occurring given that B has occurred (P(A|B)) is the probability of A occurring (P(A)) times the probability of B occurring if A has occurred (P(B|A)) divided by the probability of B occurring (P(B)).

No doubt he had time to think this up because he wasn't spending all day sorting 200 emails into appropriate categories and deleting spam.

Meanwhile in the 21st century

Luckily, Bayes' 300 year old idea has a direct application to email sorting and text classification in general.

Imagine that you have three folders you'd like to sort email into: work, personal and spam (POPFile calls these folders 'buckets'). Setting up an email client to know how to sort the mail ranges from hard (in the case of work where you'd have to tell it about everyone in your company) and impossible (spammers keep changing their emails to evade filtering).

Bayes Theorem gives POPFile a way to calculate the probability that an email is work, personal or spam by calculating P(work|E), P(personal|E), and P(spam|E) where E is the new email and P(work|E) is the probability of email E being a 'work' email and so on. By picking the largest probability of the three POPFile can automatically pick the appropriate folder. POPFile calculates these probabilities by looking at the frequency with which words occur in each folder and applying Bayes Theorem.

A complete description of how POPFile calculates the probability for each email can be found at the project's SourceForge home.

Once POPFile has determined a folder for an email it modified the Subject: line of the email to include the folder name. For example a mail with the subject Subject: hello john that belonged in the personal folder would have the subject Subject: [personal] hello john. POPFile also adds a new mail header called X-Text-Classification containing the folder name as well.

Either header can be used to set up simple filters in almost any mail client (it's been tested with Eudora, Outlook, Outlook Express, Mozilla, Netscape, ...). Instead of tens of complex half working filters you add one filter per folder and point your mail client at POPFile and away it goes.

Getting POPFile

POPFile is an open source project written in the Perl programming language and can be downloaded from http://sourceforge.net/projects/popfile/. Comments are very welcome in the Forums, bug reports should be placed in the Bug Database. To install POPFile follow the instructions in the latest README file or read the manual.

POPFile is intended to be cross-platform (at least Windows, Unix and Macintosh), it is totally free, and works with any mail client that uses the POP3 protocol.

Xandros Licences CodeWeavers' CrossOver

CodeWeavers, Inc announced that it has licensed its award-winning CrossOver software technology to Xandros® for use in the upcoming Xandros Desktop operating system. The technology will enable Xandros users — many for the first time — to run the most popular Windows applications, including Microsoft® Office, Lotus Notes®, Intuit Quicken, and many more, dependably in the Linux environment.

Xandros, a leading developer of a complete Linux computing environment, will launch its premiere Xandros Desktop product later this month.

Among the product's key features, Xandros Desktop is an easy-to-install, user-friendly and cost effective Linux operating environment appropriate for individual and enterprise use. Fully documented and backed by Xandros technical support, Xandros Desktop is expected to be a highly attractive alternative to the Windows OS for all PC users.

Via its licensing of CodeWeavers' CrossOver technology, Xandros will offer "CrossOver for Xandros" as a leading feature of the new Xandros Desktop product. "CrossOver for Xandros" will combine functionality from both CrossOver Office™ and CrossOver Plugin™, CodeWeavers' two best-known software applications; the resulting capability will enable users to run or open the following without the need for the Windows operating system:

- Microsoft Office (including Microsoft Word, Excel, PowerPoint®, and Outlook®)
- Microsoft Visio® Drawing and Diagramming Software
- Microsoft Internet Explorer
- Microsoft Windows Media™ Player plugin
- IBM Lotus Notes
- Intuit Quicken
- Apple QuickTime® plugin
- Macromedia Shockwave Director® plugin
- Macromedia Authorware® plugin
- RealPlayer® plugin
- iPIX® plugin
- Trillian, a convergent instant messaging plugin that supports AOL Instant Messenger, MSN Messenger and Yahoo Messenger; and
- Chime, a scientific plugin used in the chemical industry as well as at many colleges and universities.

Other popular Windows programs are known to run on CrossOver technology, most notably Intuit's QuickBooks® small business finance package. These applications are not currently supported officially, but are expected to be supported in the near future.

"CodeWeavers' CrossOver is a powerful enabling technology that is playing a major role in the Linux movement," said Mike Bego, Vice President, Xandros. "By incorporating CrossOver technology into the Xandros Desktop, we will be able to give users of many Windows applications the assurance they need to stride confidently into the more dependable Linux environment."

"CodeWeavers and Xandros have taken a solid step toward the Holy Grail of popular computing — a competitive, Windows-compatible operating system," said Jeremy White, founder and chief executive officer of CodeWeavers, Inc. "In a very real sense our two companies have created the competitive situation the market has been seeking. We believe our combined technologies will be the first to successfully tap into the incredible latent demand for viable Windows alternatives."

CrossOver Office, the first Windows-to-Linux product from CodeWeavers, debuted in early 2002. The product has won widespread acceptance among consumers and business users alike, winning industry acclaim in the process. The application recently was honored at LinuxWorld 2002 Conference & Expo with the 'Best Desktop Office Solution' award. In receiving the award, CodeWeavers joined such world-class companies as Computer Associates, HP, and IBM, all of whom were honored at the presentation.

For more information about CodeWeavers and its CrossOver product line, visit http://www.codeweavers.com.

On2 And Xiph Release Theora Alpha

On2 Technologies Inc., The Duck Corporation (Amex: ONT), the industry leader in compression technology and Xiph.org Foundation, developers of market-leading open-source multimedia solutions, announced the completion and immediate availability of the initial Alpha code release of Theora.

Theora is the project name for the combination of VP3, Vorbis Audio, and Ogg media framework that represents a breakthrough in license- and royalty-free open-source solutions for multimedia developers and users. The release and associated documentation are available now for download at http://www.theora.org.

"This preliminary code release represents the first time developers will have access to a completely license- and royalty-free system that includes world-class video and audio codecs in an integrated, streaming-friendly format, with all the source code and intellectual property open, customizable, and available for immediate, anonymous download," said Dan Miller, CTO and Founder of On2 Technologies.

"We are very, very happy with Theora Alpha One, and we've already got some fantastic plans for Theora Alpha Two. This is an engineering release, so it's definitely for the wireheads, but it is definitely a milestone on the way to Theora 1.0 next June. The code wizards of the known universe are cordially invited to start hacking," said Emmett Plant, CEO of Xiph.org Foundation. Theora 1.0 is scheduled for release in June of 2003.

On2 Technologies is a video compression software company. On2 licenses its high quality video codecs for use in set-top boxes, consumer electronics devices and wireless applications. In addition, On2 offers a suite of products and services, including high-level video encoding, customized technical support, and consulting services.

Xiph.org Foundation is a non-profit company dedicated to protecting the foundations of Internet multimedia from control by private interests by creating, producing and maintaining open multimedia standards. The Xiph.org Foundation can be found at http://www.xiph.org.

HP Itanium 2 Server Sets Records

HP announced further record-breaking performance benchmark results across multiple workloads on the rack-optimized HP Server rx5670 featuring the Intel® Itanium® 2 processor.

Close design and development collaboration between HP and its partners has produced industry-leading performance benchmarks that satisfy customers' most complex and data-intensive workloads. The HP Server rx5670 has achieved the first certified Windows-based benchmark result for the SAP® Advanced Planner and Optimizer-Demand Planning (SAP APO-DP) solution running on Microsoft Windows, as well as record-setting benchmark results for SPECweb99_SSL running Linux and for SPECjbb2000 running HP-UX.

"Itanium 2's processor performance numbers are impressive, matching or besting the leading RISC processor results," said Jonathan Eunice, principal analyst and IT advisor, Illuminata, Inc. "More conclusive from a user perspective, however, are the stream of excellent results coming out of recent database, Web and application benchmarks. They confirm the suitability of the platform for commercial computing."

The HP Server rx5670, with four 1-GHz Itanium 2 processors, holds the distinction of being the first Windows-based server to gain a certified SAP Standard Application APO-DP Benchmark. The rx5670 result beat the previous industry-leading four-processor result by 15 percent, achieving a result of 157,555/hour planned characteristic combinations on an aggregate level running Microsoft Windows Advanced Limited Edition, version 1.2, and Microsoft SQL Server 2000 (64-bit) beta version database. The result showcases to customers the server's leading performance for running supply chain demand planning models.

HP, in cooperation with Zeus Technology, Ltd., a leading provider of Web server infrastructure, also has produced the world's top four-processor system performance result on the industry standard SPECweb99_SSL benchmark running Zeus Web Server for Linux. The HP Server rx5670 easily outperformed the competition with a score of 1,498 simultaneous connections. The result is more than 50 percent higher than IBM's p630 result using the 1-GHz Power4 processor and is more than double Sun's V480 result using the 900-MHz UltraSPARC III processor. The SPECwebb99 SSL benchmark measures a server's ability to effectively handle secure encrypted Web transactions and is a critical benchmark for companies executing high levels of secure transactions over the Internet.

Sending Us E-Mail?

Due to the amount of e-mail MUUG receives, we've set up an auto-reply to give you immediate feedback, and redirect some of the e-mail to the appropriate places. Why not look at http://www.muug.mb.ca/about.html#contacts first?

Share Your Thoughts

E-mail us with your comments on the newsletter, whether it's criticisms or commendations, and continue to send in articles or ideas for same. If you have a How-To or other idea, and aren't ready to give a presentation at MUUG, an article is a great alternative! If you can write better than the editor, that's terrific; if you can't, submit it anyway and we'll get it into shape for publication. We know that many of you have some great ideas and lots of knowledge. Why not share? Mailto: editor@muug.mb.ca.