



# Revista de Software Libre de la UCI

**MARZO / 2007**

## Filtrado de paquetes por Contenidos II

### Noticias

La UCI en FLISOL 2007.

NTFS-3G llega a la versión 1.0.  
Ubuntu-cu@list.ubuntu.com

Por qué Microsoft puede perder  
el liderazgo.

Mandriva Linux 2007 Spring-  
Metisse.

Primer Buscador Cubano.

Microsoft y Google a las  
trompadas.

KDE 4 ya tiene fecha de salida.

Disponible Biblia de Ubuntu  
2007.

Los 10 mejores juegos Open  
Source para GNU/Linux del  
2006.

### Migración

¿Qué proyectos usan y desarrollan  
Software Libre en la UCI ?

### Programación

Riesgos de seguridad utilizando  
system() o peopen()

### Humor Libre

### Informática 2007

“Sistema multiagente para  
combinar técnicas de  
aprendizaje automatizado  
sobre plataforma libre.”

“Patentes de Software,  
Monopolios de Ideas.”

flisol2007

Festival Latinoamericano de Instalación  
de Software Libre

El 28 de Abril de este año  
se celebrará el festival de  
instalación de GNU/Linux...



¿Que distribución de GNU/Linux instalo?

El software libre sin duda revolucionó el mundo,  
nos corresponde a nosotros colocarlo en el puesto que merece.



Del colectivo ...

# Editorial

## Equipo Editorial:

### Redacción/Edición:

**Jorge Luis Betancourt G.**  
**Abel García Vitier**

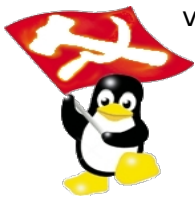
### Diseño Gráfico:

**Jorge Luis Betancourt G.**  
**Angel Alberto Bello**

### Colaboradores:

**Alain Guerrero Enamorado**  
**Abel Meneses Abad**  
**Abdel Rodríguez Abed**  
**Alberto Barrionuevo**

### Patrocinadores:



versión impresa:

**FEU**  
**Facultad X**

versión digital:  
**Proyecto**  
**Unicornios**



Comunidad de  
Software Libre  
en Cuba



Comunidad de  
Software Libre **Cubana**

Cuba  
SOFTWARE LIBRE <http://www.softwarelibre.cu>

### Software utilizado para la realización de esta edición:



Querido lector, ya tiene usted en sus manos (o mejor dicho en su pantalla) una nueva edición de UXI, la Revista de Software Libre y Código Abierto de la Universidad de las Ciencias Informáticas.

En esta edición, hemos querido llevarle a usted novedosas informaciones relacionadas con el mundo del Software Libre y el Código Abierto en nuestra Universidad. Solo esperamos que este número sea de su agrado y que sirva para lograr una mayor difusión del Software Libre en nuestra Universidad.

También es objetivo de esta revista, llevarle a usted el quehacer de nuestra comunidad en la UCI, proyectos que en ella se forman y como el Software Libre ha ido calando entre la comunidad de usuarios de la universidad y por supuesto dentro de la comunidad de desarrolladores fundamentalmente. Mediante diversas secciones esperamos que usted se sienta motivado a migrar hacia el Software Libre si no lo ha hecho aún, y en caso de que ya sea un usuario de Software Libre, pues entonces felicidades, reclínesse en su puesto de trabajo y disfrute del contenido de esta revista con la satisfacción de utilizar Software Libre.

Esperamos que aquellos menos conocedores del tema puedan encontrar en esta Revista una fuente de información, que aunque no pretende ser un manual de usuario, ni una página del man si que lo motive a investigar y a estudiar más sobre los temas tratados a continuación.

Queremos recordarle que el próximo Sábado 28 de Abril se celebrará en nuestra Universidad el FLISOL 2007 (Festival de Instalación de Software Libre) así como en toda Latinoamérica, esperamos su participación y que asista a las Conferencias, Talleres, Charlas, Instalaciones de distribuciones de GNU/Linux y demás actividades que la Comunidad de Software Libre de la UCI ha preparado para usted. Para más información diríjase a la Wiki de Producción (<http://wiki.prod.uci.cu/>).

Equipo Editorial UXI



**Se otorga permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre de GNU, Versión 1.2 o cualquier otra versión posterior publicada por la Free Software Foundation; sin Secciones Invariantes ni Textos de Cubierta Delantera, ni Textos de Cubierta Trasera. Puedes consultar una copia de la licencia en: <http://www.gnu.org/copyleft/fdl.html>**



En esta edición ...

# Contenido



## Sistema Adaptativo de Filtrado de Contenidos (Parte II y final)

El sistema parte del principio de disponer de una Base de Datos (BD) en la cual se encuentran un sinnúmero de URLs preclasificadas en distintas categorías atendiendo al contenido que albergan. ...

Página 1



## Noticias

Disponible primera versión estable de driver libre para lectura/escritura en sistemas de ficheros NTFS. NTFS-3G llega a la versión 1.0, lo que garantiza su estabilidad y fiabilidad. Las versiones de desarrollo previas se incorporaron a más de 60 distribuciones de GNU/Linux y el driver ha sido pasado a otros sistemas operativos como Mac OS X ...

Página 6



## Migración

Para los jóvenes investigadores este artículo intentará mostrar, las escasas, desorganizadas, deficientes y en ocasiones incluso inexistentes fuentes de información a las que deben acudir para experimentar la migración a SWL ...

Página 8



## Programación

Uno de los agujeros de seguridad que todo programador debe tener en mente incluye la utilización de la shell para ejecutar otros programas. Como ejemplo, vamos a considerar un servidor diccionario. Este programa está diseñado para aceptar conexiones ...

Página 10



## Informática 2007

En la actualidad los sistemas multiagente han adquirido un gran auge en diversas áreas. Los problemas abordados con técnicas de aprendizaje automatizado no están exentos de este auge. Se han desarrollado infinidad de sistemas multiagente para combinar sistemas ...

Página 12





Ing. Alain Guerrero Enamorado, Ing. Siovel Rodríguez Morales, Dovier Antonio Ripoll Mendez, Luis Enrique Sanchez Arce, Yordan Suarez Santiago, Jorge Luis Rodríguez Carpio, Jose Ramon Hermosilla Moreno, Kiuver Kaddiel Ibañez Castro, Oscar Andres Casas Machado, Saimel Saez Estrada, Israel Daniel De La Cruz Piñeiro, Ing. Joelsy Porven Rubier.

Universidad de las Ciencias Informáticas. Prof. Instructor [alaing@uci.cu](mailto:alaing@uci.cu)

## PRINCIPIO DE FUNCIONAMIENTO

El sistema parte del principio de disponer de una Base de Datos (BD) en la cual se encuentran un sinnúmero de URLs preclasificadas en distintas categorías atendiendo al contenido que albergan. Las categorías a su vez son clasificadas por el administrador del sistema según el tipo de contenidos al que se estima pertenecen y las características propias del lugar (cultura, normas, regulaciones, leyes vigentes, etc). Así una categoría puede pertenecer a uno de los siguientes tipos: ilícita, nociva o adecuada y a su vez las categorías pudieran ser: pornografía, terrorismo, violencia, juegos, noticias, deportes, salud, otras más y todas las que el administrador estime conveniente adicionar.

Todo comienza cuando un cliente (ver **Figura 1**) le solicita un recurso de Internet al servidor Squid-Proxy, el cual es parte inseparable del sistema, a partir de este momento el Squid le entrega al Subsistema de Filtrado los parámetros de la solicitud: identificador del usuario, dirección IP de origen y URL, comienza así un proceso de interacción con la BD el cual dependiendo de:

características usuario, edad y política de navegación, la existencia o no de la URL en la Base de Datos, las categorías y tipo de contenido al que pertenece la URL,

se toma una decisión (ver **Tabla 1**) sobre si el recurso solicitado se le entrega al usuario o si en su lugar se le entrega una página (ver **Figura 2**) en la que se le explican los motivos por los cuales se le negó el acceso a dicho recurso.

En el punto actual de desarrollo aun no se ha incorporado el Subsistema de Recuperación y Clasificación de Información (RCI, ver **Figura 3**), se prevé que este sistema funcione de manera independiente pero tomando como uno de los posibles puntos de partida la información que le brinde el Subsistema de Filtrado.

Todo se ha diseñado para que pueda funcionar de manera centralizada o distribuida; tal como se muestra en los esquemas de las **Figuras 3 y 4** respectivamente. En la primera variante: el Squid-Proxy, la Base de Datos, el Subsistema de Filtrado y el Subsistema de Recuperación y Clasificación de Información estarán funcionando en el mismo servidor. En la segunda variante: Squid y el Subsistema de Filtrado se encuentran en un primer servidor, la Base de Datos en un segundo servidor y el Subsistema de Recuperación y Clasificación de Información en un tercer servidor independiente. Esta variante es para lugares donde se requiera alto desempeño.

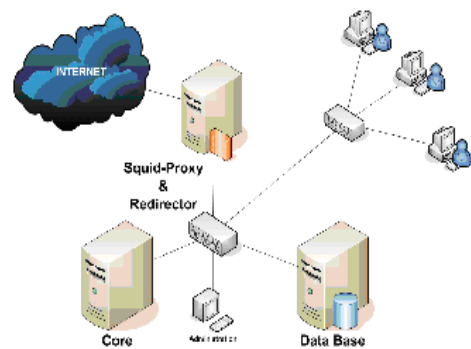


Figura 3. Esquema centralizado del sistema

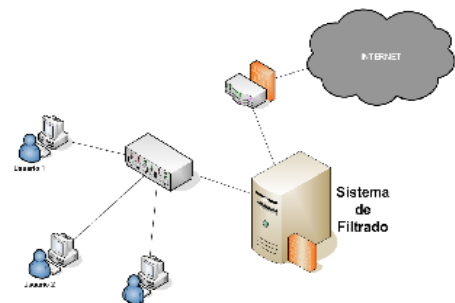


Figura 4. Esquema distribuido del sistema

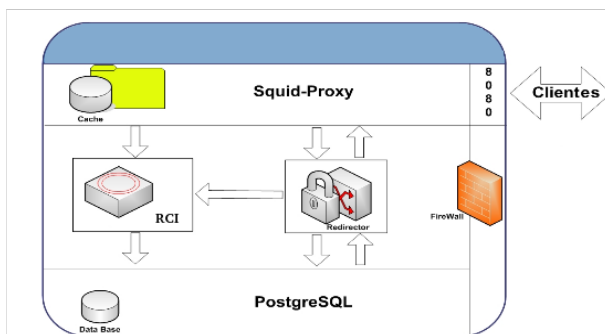


Figura 1. Esquema funcional del sistema

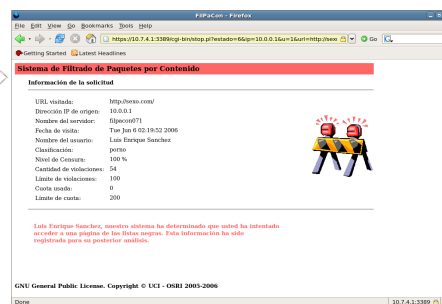


Figura 2. Página de negación de la solicitud

## DETALLES DE LOS COMPONENTES DEL SISTEMA

El sistema esta compuesto por:

- Subsistema de instalación
- Subsistema de filtrado
- Subsistema de administración
- Base de Datos
- Subsistema de recuperación y clasificación de la información

A continuación se explicarán algunas características de los componentes del sistema:

### Subsistema de Instalación

El objetivo principal de este Subsistema es lograr que el proceso de instalación tenga como características las siguientes: instalación desatendida e intuitiva. detección automática y soporte amplio de hardware. instalación y configuración automática del sistema. exigencias de conocimiento, mínimas para el administrador.

### Subsistema de Filtrado

El objetivo principal de este Subsistema es tomar una decisión sobre que contenido será bloqueado y cual no, siempre teniendo en cuenta la información que le facilita la Base de Datos. La **tabla 1** expresa la relación que existe entre las características de usuario: edad y política de navegación, y el tipo de contenido al que pertenece la categoría de la URL que se está procesando. Con este esquema se logra una gran flexibilidad del sistema, permitiendo que pueda adaptarse a disímiles situaciones.

### Subsistema de administración

El objetivo de este subsistema es brindar a los administradores una GUI intuitiva que permita realizar las tareas de configuración y ajustes necesarios para el correcto funcionamiento y adecuación a las características propias del lugar. Esta interfaz no necesita de la instalación de ningún software adicional para su explotación, es una aplicación Web que funciona con cualquier navegador (Internet Explorer, NetScape, Mozilla, etc) disponible.

**Tabla 1. Política de navegación del sistema.**

Edad	Política de Navegación	Decisión a tomar	Consulta en la Base de Datos
Menor/Mayor	Ninguna	No tiene acceso a Internet	No se hace nada
Menor/Mayor	Básica	Sólo tiene acceso a las direcciones definidas como básicas	La tabla de dominios básicos define las direcciones que pertenecen a esta política de navegación
Menor	Restringida	Permite solamente las categorías adecuadas	Si la dirección no existe en la base de datos se deniega por defecto
Mayor	Restringida	Permite solamente las categorías adecuadas y nocivas	Si la dirección no existe en la base de datos se deniega por defecto
Menor	Permisiva	Deniega solamente las categorías ilícitas y nocivas	Si la dirección no existe en la base de datos se deniega por defecto
Mayor	Permisiva	Deniega solamente las categorías ilícitas	Si la dirección no existe en la base de datos se deniega por defecto
Menor/Mayor	Plena	Acceso Total	No importa

Esta interfaz le brinda al administrador un conjunto de funcionalidades imprescindibles para garantizar: gestión de los usuarios, gestión de las categorías existentes en la BD, gestión de las URL's, la configuración del sistema y la generación de reportes sobre el uso de Internet.

# Sistema Adaptativo de Filtrado por Contenidos (Parte II)

## Módulos y Funcionalidades que brinda la interfaz de administración.

### Módulos:

Gestión de Usuarios  
 Gestión de categorías  
 Gestión de URL's  
 Configuración del sistema.  
 Reportes

### Funcionalidades:

- ◆ Insertar usuario.
- ◆ Eliminar usuario.
- ◆ Actualizar los datos del usuario.
- ◆ Cambiar contraseña.
- ◆ Denegar la navegación.
- ◆ Habilitar la navegación.
- ◆ Insertar nueva categoría.
- ◆ Eliminar categoría existente.
- ◆ Seleccionar tipo de contenido para cada categoría.
- ◆ Insertar categoría a una URL.
- ◆ Actualizar categoría de una URL.
- ◆ Comprobar categorías a la que pertenece una URL.

- ◆ Eliminar una categoría de una URL.
- ◆ Eliminar URL.
- ◆ Configurar sección del sistema.
- ◆ Configurar sección de la Base de Datos.
- ◆ Configurar sección de la Interfaz de Administración.
- ◆ Restablecer configuración.
- ◆ Establecer los dominios de navegación básica.
- ◆ Incidencias por usuarios.
- ◆ Ranking de las URL's denegadas.
- ◆ Ranking de usuarios por incidencias.
- ◆ Ranking de días de la semana por incidencias. (escoger rango de fecha).
- ◆ Cantidad de incidencias por tipo de usuarios. (escoger rango de fecha).
- ◆ Reporte de la cantidad de sitios por categoría.
- ◆ Reporte de posibles regulaciones incorrectas.

### Base de datos

El objetivo que se persigue con este medio es almacenar las listas de URL's y la información relacionada con los usuarios y sus actividades. En la Tabla 2 se muestran las categorías, una descripción y la cantidad de URL que están clasificadas hasta la fecha.

**Tabla 2. Categorías con que cuenta el sistema**

Categoría	Descripción	Cantidad
Pornografía	Sitios con contenidos pornográficos.	830999
Regiones	Sitios relacionados con algunas regiones del planeta.	670954
Mundo	Sitios de contenidos variados relacionados con países.	503416
Artes	Sitios relacionados con las artes.	289227
Sociedad	Sitios relacionados con diferentes temas de la sociedad (religión, política, organizaciones no gubernamentales, etc).	262964
Negocios	Sitios para negocios.	251491
Noticias	Agencias de noticias on-line.	229342
Computadoras	Sitios de negocios de computadoras.	138149
Recreación	Sitios de contenido recreativo.	118605
Tiendas	Tiendas virtuales.	116699
Deportes	Sitios relacionados con el deporte.	107539
Ciencias	Sitios relacionados con las ciencias.	103590
Referencia	Sitios de referencia bibliográfica.	64745
Salud	Sitios que promueven la salud.	64415
Juegos	Sitios dedicados a dar publicidad a los juegos.	61772
Adolescentes	Sitios para niños y adolescentes.	38997
Hogar	Sitios relacionados con artículos para el hogar.	33227
Spyware	Sitios que capturen información de los usuarios.	10453
Drogas	Sitios donde se promueve el uso de drogas.	2029
Lotería	Sitios que promueven los juegos de azar.	1223
Publicidad	Sitios publicitarios.	591
Hackeo	Sitios que publican herramientas o contenidos para ataques informáticos.	373
Audio_video	Sitios relacionados con Audio o Vídeo.	359
Redirector	Sitios que burlan los sistemas de filtrado.	329
Violencia	Sitios con contenido agresivo.	319
Warez	Sitios relacionados a la piratería informática.	87
Webmail	Sitios que brindan servicios de correo on-line.	47
Foros	Foros de discusión en Internet.	40
Radio	Emisoras de radio en Internet.	25
Materiales_peligrosos	Fabricación de explosivos, venenos, etc.	18
Educacion_sexual	Sitios de educación sexual.	3
Total	Total de categorías clasificadas.	3902027

## Subsistema de Recuperación y Clasificación de la Información

Dado el dinamismo de Internet las listas blancas y negras constantemente estarán desactualizadas. Surge así la necesidad de contar con un mecanismo que permita realizar el análisis de los contenidos presentes en las páginas Web para apoyar la toma de decisiones y hacer frente al continuo crecimiento y expansión de la red de redes. Para ello se está desarrollando un Subsistema que será el encargado de recuperar y clasificar la información. Este Subsistema está compuesto por un módulo de recuperación de información, un clasificador de textos, un clasificador de imágenes y un módulo decisor; este último entrega un resultado de categorización combinando varios factores, entre ellos el resultado de ambos clasificadores. Aunque actualmente no está integrado al sistema y se encuentra en franco proceso de desarrollo, se le han realizado pruebas y se continúa perfeccionando.

### Módulo de Recuperación de Información

Para poder llevar a cabo las acciones descritas anteriormente es necesario contar con un mecanismo que permita recuperar el contenido de las páginas Web. Para esto se desarrolló un robot (spider) que navega de manera automática por la Web a partir de una serie de sitios que se le especifican en un fichero o que simplemente recupera la información de una URL en particular para la cual sea invocado.

### Módulo de Clasificación Automática de Texto

Las numerosas desventajas de la búsqueda de palabras claves conduce irremediablemente a la necesidad de encontrar un algoritmo más eficiente para clasificar los contenidos semánticos de las páginas Web. Para ello se empleó la categorización automática de textos lo que puede contemplarse como un proceso de aprendizaje, durante el cual un programa capta las características que distinguen una clase o categoría de las demás. Antes de comenzar la clasificación se verifica que el idioma esté soportado por el clasificador. Si el idioma identificado es válido se realiza un proceso de Indexación automática que persigue lograr una representación del documento que pueda ser entendible por el ordenador. Esta representación se hace en forma de vectores de pesos de términos que son usados en la clasificación, la cual se materializa hallando la medida de similitud entre estos vectores y los vectores representativos de cada categoría. Esto produce como salida la probabilidad que tiene el documento en cuestión de pertenecer a cada una de las categorías, siendo clasificado en la que mayor probabilidad tenga. Actualmente están soportados los idiomas español e inglés.

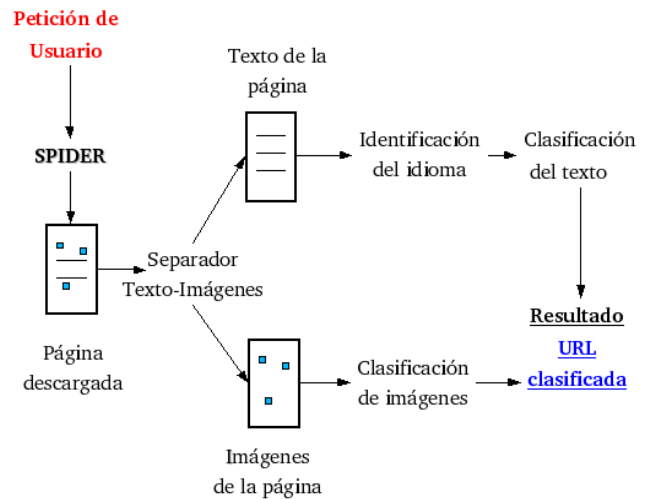


Figura. Subsistema de Recuperación y Clasificación de Información.

### Módulo de clasificación Automática de Imágenes Pornográficas

Aunque el análisis del texto que está contenido en las páginas Web con técnicas de inteligencia artificial es una solución que elimina los problemas asociados a la búsqueda de palabras claves aún persiste uno de ellos y es la imposibilidad de clasificar la información contenida en las imágenes presentes en el hipertexto.

El objetivo de este módulo es el de apoyar la toma de decisiones mediante el análisis de las imágenes contenidas en las páginas Web para determinar si son pornográficas o no. Este módulo está desarrollado siguiendo la estructura básica de las aplicaciones de reconocimientos de patrones: adquisición, segmentación, extracción de características y entrenamiento-clasificación. En este proceso se toma el color de la piel como característica primaria. Hasta ahora se ha utilizado el método de los n-vecinos más cercanos de conjunto con redes neuronales; obteniéndose en las pruebas realizadas una efectividad de un 60%.

## RESULTADOS

Este sistema aunque se encuentra en desarrollo posee una serie de funcionalidades básicas que han permitido obtener los siguientes resultados experimentales:

Se ha evitado el acceso a sitios con contenidos inadecuados pre-clasificados.

El tiempo que tarda hacer la consulta en la base de datos es despreciable, en el orden de los 20 milisegundos.

Algunas URL's han sido denegadas incorrectamente debido a que el dominio al que pertenecen alberga algunas páginas con contenidos inadecuados y por tanto se ha bloqueado el dominio completo; en algunos casos se ha cambiado la clasificación del dominio para permitir la navegación por este.



El 10 de febrero del 2006 se realizó una prueba en los laboratorios docentes de la Facultad 10 de la Universidad de las Ciencias Informáticas, el objetivo fundamental de la misma era comprobar el funcionamiento adecuado bajo condiciones de carga media en el servidor experimental donde está instalado el sistema; esta se realizó con 120 estudiantes distribuidos en 6 laboratorios docentes con 20 PC. Los resultados de esta prueba se recogen en la Tabla 3. Las características del servidor:

*CPU: Intel(R) Pentium(R) 4 CPU 2.40GHz, 8KB L1 Cache, 512 KB L2 Cache.  
MM : 1GB Memoria Principal  
Red : Ethernet interface 100MB/s  
HDD: 74 GB*

**Tabla 3. Resultados de la prueba de carga**

Proceso	CPU%	Memoria%
Squid Proxy	15.703	5.7752
Subsistema de filtrado	10.080	1.5670
Gestor Base de Datos	3.0562	0.5014
Servidor Web	1.0066	0.3499

Al finalizar la prueba se realizó una encuesta a los 120 estudiantes; donde se anularon 3 por respuestas incongruentes.

Los resultados fueron los siguientes:

- ◆ 89% respondió que tenía experiencia anterior de navegación por Internet, el 11% que nunca lo había hecho.
- ◆ 23% respondió que se le denegó alguna página por error, 71% que no, 6% Sin respuesta.
- ◆ Velocidad de navegación: muy rápida 13%, rápida 59%, normal 25%, lenta 2%, muy lenta 1%. Comparado con el servidor de internet de la Universidad.
- ◆ El promedio de uso del procesador fue de un 30%.

### Requerimientos del sistema

El sistema está diseñado para funcionar sobre la distribución Debian de linux. Los requerimientos se muestran en la Tabla 4.

**Tabla 4. Requerimientos del sistema**

Cantidad de usuarios	CPU	Memoria	HDD
0 - 50	450 Mhz	64 MB	2GB
50-100	1.7 Ghz	256 MB	20 GB
+100	3.2 Ghz	1 GB	80 GB

### CONCLUSIONES

A pesar de que el sistema se encuentra aun en fase de desarrollo si permite palpar resultados parciales en las pruebas que se le han hecho. Los tiempos de respuesta, de la Base de Datos aun teniendo millones de campos son mínimos, sin embargo aun faltan por hacer pruebas más profundas en este sentido. El objetivo fundamental de describir el Sistema de Filtrado Adaptativo que se desarrolla en la Universidad de las Ciencias Informáticas, al mismo tiempo que se brinda una panorámica sobre la problemática que existe a nivel mundial sobre la regulación de los contenidos ilícitos y nocivos ha sido cumplido con la redacción de este trabajo. Sin embargo aun queda por recorrer el camino que convertiría a esta solución en una realidad palpable en toda su dimensión: los algoritmos de IA que garanticen una Base de Datos lo suficientemente actualizada para garantizar una navegación confiable para los usuarios.



Grupo de Investigación y Soluciones para Internet





## **NTFS-3G**

Disponible primera versión estable de driver libre para lectura/escritura en sistemas de ficheros NTFS. NTFS-3G llega a la versión 1.0, lo que garantiza su estabilidad y fiabilidad. Las versiones de desarrollo previas se incorporaron a más de 60 distribuciones de GNU/Linux y el driver ha sido pasado a otros sistemas operativos como Mac OS X, por lo que se le prevé un futuro brillante.

## **Ubuntu-cu@list.ubuntu.com**

Ya esta creada la lista de distribución: [ubuntu-cu@list.ubuntu.com](mailto:ubuntu-cu@list.ubuntu.com). Desde ya puede suscribirse entrando a: <https://lists.ubuntu.com/mailman/listinfo/ubuntu-cu> o enviando un correo a: [Ubuntu-cu-request@lists.ubuntu.com](mailto:Ubuntu-cu-request@lists.ubuntu.com) y ponen *subscribe* en el asunto.

## **Media Player mejor reproductor del año**

Según la encuesta que se realiza anualmente por el sitio [linuxquestions](http://linuxquestions.com), los miembros de este sitio escogieron el reproductor MPlayer, en lo alto del podio, siendo este catalogado como la mejor aplicación de reproducción de videos del año.

## **Por qué Microsoft puede perder el liderazgo**

El propio Bill Gates fue el encargado de escribir los motivos en un memo secreto dirigido a los principales directivos de la empresa. Gates alertó en un informe interno dirigido a los principales directivos de su compañía que la "próxima ola de servicios será muy perjudicial (...) Tenemos competidores que se aprovecharán de esas tendencias y nos retarán". En el mismo mensaje instó a todos a reaccionar de manera inmediata para evitar que ello suceda, aseguran varios medios internacionales citando al diario británico Financial Times y al norteamericano Wall Street Journal. Y es terminante al señalar: "Está claro que si fracasamos en ello, nuestro negocio, tal como lo conocemos, está en riesgo (...) debemos responder rápidamente y de forma decisiva".

## **Mandriva Linux 2007 Spring - Metisse**

Mandriva Linux 2007 Spring - Metisse, es la más nueva LiveCd de los chicos de Mandriva, esta nueva distro LiveCd cuenta con escritorio 3D clasico de cubo, además cuenta con Virtual desktops donde

podrás cambiar de aplicaciones de una manera novedosa.

## **Primer Buscador Cubano**

En el marco de la XII Convención y Expo Informática 2007, celebrada en La Habana, Cuba, se presentó el primer buscador cubano en Internet para los sitios con dominio ".cu". Se llama 2x3 y ya se encuentra disponible la versión beta con más de 150.000 páginas indexadas.

## **Microsoft y Google a las trompadas**

Una pugna se ha desatado entre el gigante informático Microsoft y el motor de búsqueda de internet Google cuando un alto funcionario de Microsoft, Tom Rubin, en un discurso ante empresarios de la industria editorial, acusó a Google de sistemáticamente violar los derechos de autor con su ambicioso proyecto de digitalizar millones de libros y ofrecerlos, libre de costo, en internet.

## **Ubuntu Ultimate Gamers Version**

Ultimate Ubuntu es una versión adaptada de Ubuntu Edgy Eft, que incorpora todos los extras preinstalados, tales como Beagle, QEMU, Inkscape, Gdesklets, Wine, aMule, Gambas, etc... Tan sólo falta por instalar los plugins de Flash y Java por problemas de licencia. Destaca de la versión "Gamers" la inclusión de múltiples y numerosos juegos, listos para ejecutar.

## **KDE 4 ya tiene fecha de salida**

La cuarta revisión de este popular entorno de escritorio aparecerá en su formato final el 23 de octubre de este año. Pero antes de eso tendremos oportunidad de ir comprobando sus posibilidades gracias a varias betas y Release Candidates.

## **Biblia de Ubuntu 2007**

Disponible para toda la comunidad en el Servidor de Documentación la "Biblia de Ubuntu 2007" en idioma inglés y con una capacidad de 44,5MB. La misma la puede encontrar en la siguiente dirección:

[http://10.128.50.121/Documentacion/Distribuciones/Debian\\_Ubuntu\\_Knoppix/Biblia\\_de\\_Ubuntu/](http://10.128.50.121/Documentacion/Distribuciones/Debian_Ubuntu_Knoppix/Biblia_de_Ubuntu/)

más noticias...



## Los 10 mejores juegos Open Source para GNU/Linux del 2006.

Según votación publicada, aquí están los 10 mejores juegos seleccionados por los usuarios de GNU/Linux.

- 1- Battle for Wesnoth
- 2- Frets On Fire
- 3- Glest
- 4- Frozen Bubble
- 5- Super Mario War
- 6- Tremulous
- 7- SuperTux
- 8- Nexuiz
- 9- ManiaDrive
- 10-Torcs

## El 30% de los usuarios usa Firefox en Europa

La compañía alemana de publicidad en línea Adtech publicó sus nuevas estadísticas sobre el mercado europeo de los navegadores. Las cifras se basan en varios miles de millones de anuncios distribuidos en 4.000 sitios web durante el período noviembre de 2006 a febrero de 2007.

En el marco de un estudio similar publicado en 206, Adtech afirmó que el crecimiento de Firefox se había detenido y que Internet Explorer comenzaba a recuperar terreno. La tendencia parece haberse revertido nuevamente, a pesar de que Microsoft ha lanzado una nueva versión de Internet Explorer.

## Kaspersky: "Vista es menos seguro que XP"

Natalya Kaspersky, CEO de la empresa, reveló que la característica UAC de Windows Vista es tan molesta que los usuarios la desactivarán, y eso acabará provocando que Vista sea un sistema menos seguro que Windows XP.

## FLISOL 2007

Con el objetivo de promover y difundir el uso de Software Libre, el próximo 28 de Abril se llevará a cabo el Festival Latinoamericano de Instalación de Software Libre en Cuba. Debido a la cantidad de usuarios de Software Libre que existen en nuestra Universidad, también fuimos incluidos como un grupo de usuarios en el evento. FLISOL 2007 está dirigido a cualquier tipo de persona, de cualquier índole, y de cualquier grado escolar; como bien se expone en sus bases organizativas. Persiguiendo estos objetivos tenemos como propósito adaptarlo a nuestra Universidad y celebrar el evento, dirigido a cualquier personal que trabaje o estudie en la misma. Para conocer más sobre este evento y la participación de nuestra comunidad, visite la página en <http://wiki.prod.uci.cu>.

## Reunión de la Comunidad

El 23 de marzo pasado se realizó la reunión de la Comunidad donde se presentaron los siguientes temas:

1. Presentación oficial del Concurso de Software Libre de la Universidad de las Ciencias Informáticas. Antes de exponer el nuestro se hizo un breve análisis del resultado del Concurso de SWL de España.
2. Exposición de las experiencias de la Facultad # 4, la cual migrará a Software Libre paulatinamente.
3. Conferencia Magistral sobre Migración, la cual fue impartida por el Ingeniero Abel Meneses Abad, haciendo énfasis en los siguientes temas:

- Aplicaciones Libres para la Docencia.
- Perspectivas del Software Libre en la UCI.
- Proyectos que usan Software Libre.

## Repositorio de Ubuntu Feisty

Ya se encuentra disponible en nuestra universidad el repositorio de Ubuntu 7.04 Feisty para todos aquellos que quieran instalarlo, también se encuentran las imágenes alfa de los CD's de instalación de esta versión de Ubuntu. El lanzamiento oficial de esta versión será el próximo mes, se prevé que cuenta con algunas mejoras en cuanto a detección de hardware como dispositivos Plug&Play.

## Nuevo Repositorio de Debian

Hace algunos días se descargó un nuevo repositorio de Debian para todos aquellos que usan esta distribución, estamos hablando de Debian Multimedia, un repositorio que cuenta con todos los codecs de multimedia y facilidades para trabajar videos. Está disponible en el repositorio de debian de producción y en el de la facultad 10.



# Migración

¿Qué proyectos usan y desarrollan Software Libre en la UCI?

Ing. Abel Meneses Abad

## ¿Qué proyectos usan y desarrollan SWL en la UCI?

Con frecuencia se habla ya del software libre en la UCI, y muchos se pregunta: **¿Qué se está haciendo? ¿Qué proyectos existen?**

Para los jóvenes investigadores este artículo intentará mostrar, las escasas, desorganizadas, deficientes y en ocasiones incluso inexistentes fuentes de información a las que deben acudir para experimentar la migración a SWL.

## ¿Dónde saber en que proyectos, facultades, etc se está usando SWL?

Wikiprod (<http://wiki.prod.uci.cu>)

El primero de los lugares a donde nos podemos dirigir para conseguir información sobre los proyectos es la wiki de la UCI, allí se encuentra, en dependencia de la cultura de los líderes de los proyectos, gran parte de la información. Ejemplo: en la portada inicial podemos encontrar una categoría llamada Producción UCI. Existe en esta categoría un enlace al "portal de producción", pero desafortunadamente la información disponible es "0", los enlaces están abandonados.

Al realizar una investigación tratamos de acceder a todos los puntos importantes que nos lleven a cumplir el objetivo final, si intentamos encontrar las fuentes de información es por ello importante hacer uso de cada link. Así el ejemplo descrito anteriormente nos llevó a la dirección: <http://dirproduccion2.uci.cu/>; deduciendo, probamos también los enlaces: <http://dirproduccion1.uci.cu/> (NO EXISTE); <http://dirproduccion3.uci.cu/> (Si existe); también se intentó con los números 4 y 5 pero no existen. Volveremos entonces a esta fuente de información.

Regresemos a la información disponible en la wikiprod. Allí encontrará usted los eventos, el grupo de trabajo de Producción de la FEU-UJC, la categoría Producción UCI que nos lleva a más información aún.

## ¿A quiénes preguntar por los proyectos libres?

El redactor considera que la disponibilidad en la web de información a veces no confiable hace necesario para la credibilidad de los resultados de esta investigación poner a disposición de todos: las personas, a las cuales se les envió el siguiente correo:

*"En estos momentos realizo una investigación sobre que proyectos pueden estar trabajando en SWL. ¿Conoce usted los lugares en la Web UCI, Líderes de proyecto, o los nombres de proyectos con los cuales conformar mi investigación?"*

*David Leyva Leyva, Vicedecano Producción Ftad X  
Loan Joa Matos, Vicedecano Producción Ftad 4  
Pedro Y. Piñeiro Pérez, Vicedecano Producción Ftad 3  
Aurelio Antelo Collado, Vicedecano Producción Ftad 6  
Grupo de Producción FEU – UCI.*

*A pesar de la simplicidad de la información a enviar que se respondía en el caso más organizado, con un link hasta un lugar donde los investigadores puedan continuar el proceso de búsqueda, solo dos compañeros de la FEU y un vicedecano de producción respondieron. Lo que indica la primera dificultad de los proyectos libres, poca promoción y divulgación.*

*Los compañeros escogidos en esta lista no están de manera casual, las cuatro facultades mencionadas son las que mayor desempeño han tenido hasta el momento, a consideración del autor, en desarrollos sobre plataformas libres (eclipse, Zope/Plone, Perl, etc.)*

## ¿Qué proyectos de SWL existen hoy en la UCI?

*Para decir si efectivamente cada proyecto de los que encontramos en la web de la UCI existe aún, dado que las actualizaciones no son muy constantes es necesario enviar correos a sus líderes y cuestionar el seguimiento al proyecto. Ponemos entonces a disposición de la comunidad universitaria la información que motivara esta investigación:*

**Nota:** *Algunos proyectos más pueden ser encontrados en la página de la Facultad 3, [facultad3.uci.cu](http://facultad3.uci.cu), sobre todo relacionado con desarrollo para sistemas de gestión basados en tecnologías Zope/Plone.*

Nombre del Proyecto	Descripción	@ Líder	Tecnologías
Nova – GNU/Lnx	Desarrollo de una distribución libre para la República de Cuba.	agoni@est...	Gentoo, C++, gtk
Filpacon	Desarrollo de un filtro de internet para Cuba, y empresas.	alaing@uci	Debian, Perl, PHP, MySQL
RINDE	Red de Integración y Desarrollo, una red para el trabajo, la promoción del SWL en Vzla.	jpernia@uci	gforge, wiki
Biblioteca Nacional	Desarrollo de un sistema de gestión documental.	lgongora@est...	Zope/Plone
Portales	Desarrollo de portales Web para la intranet nacional.	maikelm@uci leansy@uci	Drupal/PHP & Zope/Plone/Python
Proyecto UNICORNIOS	Soporte, y desarrollo de aplicaciones, para la migración a SWL.	abelma@uci yvillazon@est..	C++, PHP, Python, Perl, Bash
SCADA (Ftad 5)	Desarrollo en JAVA de un simulador para PDVSA	rene@uci	Java
Intranet PDVSA	Desarrollado de una intranet corporativa	sergioenrique@	PHP
Software de Gestión para la Oficina Nacional de Recursos Mineros (Ftad 9)	Software para controlar lo referido a yacimientos minerales y su explotación en el país.	vito@uci.cu	J2EE(JBOSS), Eclipse
Automatización de las refinerías de PDVSA (Ftad 9).	Simulación de una refinería. Automatización y control. Software de Gestión.	lperez@uci.cu	PostgreSQL, PHP, Python+GTK
Aduana General de la Rep. (Ftad 4)	Proyecto integral de desarrollo y estrategia de migración de la aduana general.		

### Conclusiones:

El auge de los proyectos que desarrollan en software libre es cada día mayor, podemos encontrar hoy muchos desarrollos basados en PHP, Java y Python. No hemos de dudar que en breve tiempo estos se harán sobre sistemas operativos libres. Dejamos a los lectores la posibilidad de solicitar una nueva revisión de esta investigación, que en pocos meses podría arrojar nuevos proyectos libres dentro de la UCI.





# Programación

Riesgos de seguridad utilizando `system()` o `peopen()`

Jorge Luis Betancourt González  
E-Mail: [jlbetancourt@estudiantes.uci.cu](mailto:jlbetancourt@estudiantes.uci.cu)  
Universidad de las Ciencias Informáticas

Uno de los agujeros de seguridad que todo programador debe tener en mente incluye la utilización de la shell para ejecutar otros programas. Como ejemplo, vamos a considerar un servidor diccionario. Este programa está diseñado para aceptar conexiones vía Internet, cada cliente envía una palabra, y el servidor dice si es o no una palabra válida del idioma Inglés. Debido a que cada sistema GNU/Linux viene con una lista de aproximadamente 45 000 palabras en inglés en `/usr/share/dict/words`, una forma inteligente de resolver este problema es invocando al programa `grep`, como por ejemplo:

```
% grep -x word /usr/share/dict/words
```

En este caso `word` es la palabra que el usuario está buscando. El código de salida de `grep` dirá si esa palabra aparece o no en `/usr/dict/words`.

```
/*
 * grep_dictionary.c
 */

#include <stdio.h>
#include <stdlib.h>
#include <string.h>

/*
 * Returns a nonzero value if and only if
 * WORD appears in /usr/dict/words
 */

int grep_for_word(const char* word)
{
    size_t length;
    char *buffer;
    int exit_code;

    /*
     * Build up the string 'grep -x WORD
     * /usr/dict/words'.
     * Allocate the string dynamically to
     * avoid buffer overruns.
     */
    length = strlen("grep -x") +
             strlen(word) +
```

```
strlen("/usr/share/dict/words") + 1;
    buffer = (char*)malloc(length);
    sprintf(buffer, "grep -x %s
/usr/dicts/words", word);

    /*
     * Run the command, using the
     * system() function.
     */

    exit_code = system(buffer);
    /*
     * Free the buffer
     */
    free(buffer);
    /*
     * If grep returned 0, then the word
     * was present in the dictionary,
     * otherwise not.
     */
    return (exit_code == 0);
}
```

Nótese que calculando el número de caracteres que necesitamos y entonces reservando el buffer dinámicamente, nos aseguramos que nunca exista una sobrecarga del buffer.

Desafortunadamente el uso de la función `system()` es inseguro. Esta función invoca al shell estándar del sistema para ejecutar el comando y entonces devolver el valor del código salida. ¿Pero que sucedería si un usuario malicioso enviara una "palabra" como la que sigue a continuación:

```
% foo /dev/null; rm -rf /
```

En este caso el servidor ejecutaría el siguiente comando:

```
%grep -x foo /dev/null; rm -rf / /usr/dict/words
```

Ahora el problema es obvio. El usuario ha transformado un comando, la invocación a `grep`, en dos comandos porque la shell trata al punto y coma (;) como un separador de comandos. El primer comando sigue siendo una inofensiva invocación a `grep`, pero el segundo eliminará todos los ficheros del sistema completo. Incluso si la aplicación no está ejecutándose como `root`, todos los ficheros que puedan eliminarse por el usuario que está ejecutando el servidor serán eliminados.

más...

El mismo problema puede presentarse con la función `popen()`, la cual crea una tubería entre el proceso padre y el proceso hijo pero sigue utilizando la shell para ejecutar el comando.

Existen dos formas de prevenir este problema: Uno es utilizando la familia de las funciones `exec()` en vez de `system()` y `popen()`. Esta solución evita los problemas porque los caracteres que el shell trata especialmente (como el punto y coma en el ejemplo anterior) no son tratados especialmente cuando aparecen en la lista de argumentos de una llamada a `exec()`. Por supuesto, usted puede ver las ventajas sobre `system()` y `popen()`.

La otra alternativa es validar la cadena y asegurarse que es benigna. En el ejemplo del servidor diccionario, usted debe asegurarse que la palabra proveída contiene solamente caracteres alfabéticos, utilizando la función `isalpha()`. Si no contiene otro carácter entonces no hay forma de engañar a la shell para ejecutar un segundo comando. No se recomienda implementar el chequeo de forma tal que busque los caracteres peligrosos o inesperados, siempre es más seguro chequear de forma explícita los caracteres que usted conoce que son seguros que tratar de anticipar todos los caracteres que pueden causar problemas.



## Humor Libre



COPYRIGHT (c) TIRA ECOL - Javier Malonda



(versión española): tira.escomposlinux.org



(english version): comic.escomposlinux.org



Javi  
Ecol (4)





# Informática 2007

## SISTEMA MULTIAGENTE PARA COMBINAR TÉCNICAS DE APRENDIZAJE AUTOMATIZADO SOBRE PLATAFORMA LIBRE

Abdel Rodríguez Abed  
abdelr@uclv.edu.cu  
Universidad Central de Las Villas, Cuba

En la actualidad los sistemas multiagente han adquirido un gran auge en diversas áreas. Los problemas abordados con técnicas de aprendizaje automatizado no están exentos de este auge. Se han desarrollado ininidad de sistemas multiagente para combinar sistemas basados en aprendizaje automatizado. Por otra parte ha habido una gran tendencia, por parte de los investigadores, a desarrollar y probar sus sistemas en herramientas desarrolladas a código abierto. En el caso del aprendizaje automatizado, la Universidad de Weikato, Nueva Zelanda, produjo el software libre Weka (Weikato Environment for Knowledge Analysis), donde se implementan las diversas técnicas de clasificación, asociación, agrupamiento, y predicción existentes en la actualidad. Pero esta herramienta no cuenta con facilidades para implementar sistemas multiagente para combinar estas técnicas. Es por ello que se ha decidido añadir un paquete para el desarrollo de estos sistemas para poder lograr mayor certeza en las soluciones. Al publicar este paquete buscamos el objetivo de facilitar a los investigadores una plataforma en la cual puedan poner a prueba los diseños ideados para situaciones dadas o incluso de técnicas novedosas basadas en este esquema, reutilizando lo que se ha desarrollado e implementando únicamente los cambios que desee hacer sobre los esquemas estándares.



## PATENTES DE SOFTWARE, MONOPOLIOS DE IDEAS

Alberto Barrionuevo  
abarrio@ffii.org  
Foundation for a Free Information Infrastructure  
España

Todo el exitoso desarrollo de la ciencia y economía del software ha estado basado en cuatro pilares legales: derechos de autor (copyright), secreto industrial y comercial, y, ligado al derecho de autor, los acuerdos de licencia de usuario final (EULAs). En EEUU desde principios de los 80 y por medio de la jurisprudencia, se fue añadiendo paulatinamente una controvertida quinta figura legal, que es la patente (de software), y que nunca ha conseguido probar su interés económico para la sociedad y la innovación en software. Sin embargo, no ha sido hasta el estallido del '201cpunto COM'201d, cuando las patentes de software han jugado un papel significativo en la economía del software. En Europa, con legislación distinta a la estadounidense, las patentes en general se regulan por el Convenio de la Patente Europea de Munich del 1973 (CPE). En él se cita explícitamente a los programas de ordenador como un sujeto no patentable. Sin embargo, la Oficina Europea de Patentes (OEP) hace imaginativas interpretaciones favorables a la masiva patentabilidad de software, constituyendo ésta su práctica actual. Los pocos juicios de relevancia que se están produciendo en la Unión Europea están decantándose mayormente por una interpretación más restrictiva a la patentabilidad del software y más cercana a la postura contraria a la misma del Parlamento Europeo en la recientemente rechazada directiva de patentes de software. En España se puede concluir que existe una ley de las más restrictivas de Europa en cuanto a la patentabilidad del software, aunque esto no evita que también exista cierta controversia legal generada por la malabarística interpretación del CPE realizada por la OEP.