Filtrado de contenidos, control del acceso a sitios web, incluyendo un filtrado de virus(Dansguardian,Squid,SquidGuard,Sarg)

Raúl Casado Sutil

Carlos Hugo Flores Moreno

Department of Computer Science and Automation, University of Salamanca

Plaza de la Merced, s/n, 37008, Salamanca, Spain

Raul.C.S@usal.es

hugoflores@usal.es

Abstract. Responisabilidades, miedo, falta de actitud, seguridad, principales aspectos críticos en el dia a dia de no solo padres y madres, sino tambien de jefes. La duda de que estarán haciendo sus hijos o empleados cuando utilizan un ordenador. Preguntas como ¿que estaran haciendo?, ¿estará viendo páginas para mayores? , ¿de verdad está trabajando?, son muy frecuentes en estos casos. El no cumplir con su trabajo o visitar paginas inadecuadas son de los principales temas a tratar. La solucion a estos y muchos mas problemas.

Keywords: Contenido, seguridad, web, filtrado.

1. Introduction

Con el paso de los años y el estudio de nuevas formas de seguridad, se han desarollado progamas capaces de satisfacer las necesidades que muchos padres y jefes llevan buscando.Casos como por ejemplo, mantener a niños apartados de paginas inapropiadas, trabajadores con el hábito de utilizar los medios para fines lucrativos, son las principales consecuencias. En el primer caso, ningun padre quiere que sus hijos entren en páginas del tipo pornograficas, de apuestas, web que

incitan a la violencia, terrorismo… Con programas como los que vamos a presentar a continuacion evitarian que esto ocurriese.
Por otra parte el uso de Internet entre el ámbito laboral revela que los empleados leen noticias, hacen compras en línea y se comunican con sus amigos y familiares durante las horas de trabajo. Un estudio estadístico revela que el 35% de los empleados usa el Internet en horas de trabajo. El 20% usa el Internet por al menos 2 horas diarias durante su periodo laboral. El 15% Usa el Internet para motivos personales como lo son: compras en línea, leer y contestar correos personales, visitar páginas de entretenimiento, deportes o con contenido sexual. Un empleado pierde al menos 30 minutos en consultar sus correos personales por la mañana y gasta aproximadamente 40 minutos de su tiempo laboral chateando o utilizando mensajeros instantáneos para comunicarse con sus amigos o familiares.

Otro problema son los virus.Hay una enorme cantidad de material dudoso en línea, así como spyware, adware y virus informáticos. Como padres de familia y usuarios de Internet, tenemos que saber cómo mantener alejados los spam perjudiciales sin poner en peligro los beneficios de Internet.Los filtros son elementos imprescindibles para cualquier persona con acceso a Internet, ya que pueden evitar los virus informáticos, el robo de información personal y pueden ayudar a los padres a mantener el contenido inadecuado lejos de los niños.

2. Filtrado de contenidos web.

El filtrado de contenidos web nos permite el bloqueo de paginas web o la habilidad para tener el control de acceso a internet y evitar tiempos muertos.

Imagen tomada al entrar a una página bloqueada por un antivirus.

·Errores de filtrado:

-Exceso de bloqueo: Ocurre cuando se utiliza un filtro de contenidos programado con excesiva severidad, puede acabar bloqueando paginas aceptables por contener parte de una palabra restringida.

-Filtrado insuficiente: Al estar internet siempre activo, se suben contenidos nuevos todos los días y puede que algunos contenidos que deberían ser bloqueados pasen desapercibidos por los administradores si estos no actualizan los filtros temporalmente.

3.Filtrado de virus.

El software antivirus es un tipo de filtro que bloquea específicamente los virus informáticos dañinos, spyware y adware. Los programas antivirus pueden ser filtros activos para prevenir la transmisión de virus a la máquina host a través de la conexión a Internet o pueden ser programas pasivos de "limpieza" que detectan y eliminan los virus después de la infección. Los anti-virus más eficaces son combinaciones de ambos.

4. Control de acceso a contenidos web.

El control de acceso a contenidos web está directamente relacionado con el filtrado. Este te permite evitar que otros usuarios del ordenador puedan tener acceso a contenidos inapropiados mientras navegan por internet. Hay diferentes filtros para ello (filtro infantil, filtro para adolescentes, para empleados.).
Se puede aplicar un filtro diferente a cada usuario del ordenador, o crear nuevos filtros de contenidos adaptados a tus necesidades. Por otra parte, el control de contenidos web te permite tambien determinar un listado de páginas web a las que los usuarios del ordenador podrán acceder (o no), independientemente del filtro de contenidos que se les haya aplicado. Son muchas las aplicaciones que existen hoy en día, pero vamos a centrarnos en DansGuardian.

4.1 Dansguardian.

DansGuardian es un software de control de contenidos, diseñado para controlar el acceso a sitios web. Incluye un filtro de virus, importante en sistemas Windows, es

usado principalmente en instituciones de educación, gobierno y empresas. Se caracteriza por su alto grado de flexibilidad y adaptación de la implementación.

DansGuardian se instala en un ordenador (servidor) con el sistema operativo GNU/Linux, y filtrará contenidos de webs solicitadas por el resto de ordenadores (independientemente del sistema operativo que tengan instalado).

La herramienta DansGuardian es código abierto, está desarrollada en C++ y permite una configuración flexible adaptándose a las necesidades del usuario. Al instalar el paquete la configuración por defecto ya limita las visitas a páginas prohibidas para menores, pero dispone de gran cantidad de archivos de configuración para llevar a cabo un ajuste del servicio mas personalizado.
El mecanismo es el siguiente: los clientes mediante sus navegadores web hacen peticiones de páginas que son recibidas por DansGuardian y sólo son redireccionadas al servidor proxy SQUID aquellas que superan la fase de filtrado.
En realidad DansGuardian se ejecuta como un demonio independiente del proxy, acepta peticiones en el puerto 8080 y las redirecciona al proxy SQUID, que escucha en el puerto 3128. Por lo tanto, cuando una petición entra por el puerto 8080, DansGuardian la filtra y la pasa al proxy **SQUID** por el puerto 3128. Es importante, en consecuencia, que ningún otro servicio esté utilizando el puerto 8080.

4.1.1 Instalación y configuración.

Para instalarlo solo hace falta ejecutar la siguiente sentencia en linea de ordenes*:*
Su apt-get install dansguardian
El archivo de configuración de DansGuardian es /etc/dansguardian/dansguardian.conf .
Para poder configurarlo hace falta editar el siguiente archivo:
$sudo gedit /etc/dansguardian/dansguardian.conf
Pasos para la configuración:

1-Establecer la línea que contiene la directiva UNCONFIGURED como un comentario. Para ello añadir al principio de la línea el carácter '#'.
#UNCONFIGURED - Please remove this line after configuration
2-Si no se está trabajando con el antivirus modificar la línea correspondiente desactivando la opción y comentar la indicada:
virusscan = off
#virusengine = 'clamav'
3-En la sección 'Network Settings' comprobar que están las líneas siguientes:
filterport=8080
proxyip=127.0.0.1
proxyport = 3128
4-Esta sería la configuración para una máquina aislada, también llamada standalone o desktop. En el caso de tratarse de un aula con

varias máquinas cliente que salen a Internet a través de un servidor de aula, habría que modificar el valor dado en *proxyip* por la dirección IP de la tarjeta de red del servidor que escucha dentro del aula.
5-Modificar el idioma por defecto. Para ello sustituir el inglés por 'spanish' y dejar las líneas como sigue:

```
languagedir = '/etc/dansguardian/languages'
# language to use from languagedir.
language = 'spanish'
```

6-Salir de gedit salvando los cambios.
7-Reiniciar el servicio dansguardian ejecutando la orden:

```
$sudo /etc/init.d/dansguardian restart
```

4.1.2 Archivos de filtro y excepciones.
Archivos de filtros en /etc/dansguardian :
bannedphraselist: contiene una lista de frases prohibidas. Las frases deben estar entre <>
bannedmimetypelist: contiene una lista de tipos MIME prohibidos. Si una URL devuelve un tipo MIME incluido en la lista, quedará bloqueada.
Bannedextensionlist: contiene una lista de extensiones de archivos no permitidas. Si una URL termina con alguna extensión contenida en esta lista, será bloqueada.
Bannedregexpurllist: contiene una lista de expresiones regulares3 que si se cumplen sobre la URL ésta será bloqueada.
Bannedsitelist: contiene una lista de sitios prohibidos. Si se indica un nombre de dominio todo él será bloqueado
bannedurllist: permite bloquear partes específicas de un sitio web

Archivos de excepciones en /etc/dansguardian :

exceptionsitelist contiene una lista de los nombres de dominio que no serán filtrados Es importante tener en cuenta que el nombre de dominio no debe incluir http:// o www.

Exceptioniplist: contiene una lista de las direcciones IP de los clientes a los que se permite el acceso sin restricciones. este sería el caso de la dirección IP del administrator.

Exceptionuserlist: lista de los nombres de usuarios que no serán filtrados en el caso de utilizar control de acceso por usuario. Requiere autenticación básica o "ident".

Exceptionphraselist: lista de las frases que, si aparecen en una página web, pasará el filtro.

5. Squid.

Squid. Es un popular programa de software libre que implementa un servidor proxy y un dominio para caché de páginas web, publicado bajo licencia GPL(General Public Licence). Tiene una amplia variedad de utilidades, desde acelerar un servidor web, guardando en caché peticiones repetidas a DNS(Servidor de nombre de dominio) y otras búsquedas para un grupo de gente que comparte recursos de la red, hasta caché de Web, además de añadir seguridad filtrando el tráfico. Está especialmente diseñado para ejecutarse bajo entornos tipo Unix

5.1 Instalacion y configuracion.

Para descargarlo ejecutamos la orden: sudo apt-get install squid
Para iniciar el servidor: sudo service squid3 start
Normalmente, tras la instalación se incluye un fichero de configuración que se encuentra en /etc/squid3/squid.conf y es este el que debemos editar con un editor como nano o gedit.
Otro detalle es que en este fichero se especifica el puerto por defecto que usa el servicio Squid, que por defecto es 3128 (ver la línea o directiva "http_port 3128" y quítale el # para activarla).
Y otra cosa necesaria es configurar el hostname, busca el comentario "TAG: Visible_hostname" y verás una línea "visible_hostname" donde debes poner tu hostname.

Para saber tu hostname, puedes teclear en el terminal: hostname.

Y el nombre que te aparezca lo añades a la línea que no debe estar precedida por # para que no sea ignorada como un comentario.

Hay que tener en cuenta que cuando quites un # para activar una regla, asegurarse de no dejar espacios al inicio de la línea.

Un ejemplo bastante práctico, imagina que tu equipo lo usan niños menores de 18 años y quieres restrigir el acceso a ciertos sitios de contenido adulto. Lo primero es crear un fichero llamado /etc/squid3/lista con el contenido (se puede poner ips, dominios..):

sex

adult

porno

Y para que esto funcione habra que modificar el fichero squid.conf de la siguiente forma:

acl denegados url_regex "/etc/squid3/lista"
http_access deny denegados

6. SquidGuard*:*

SquidGuard es un plug-in para Squid , es decir, es necesario tener Squid instalado para que funcione y al igual que DansGuardian, es necesario que esté instalado en un ordenador con sistema GNU/Linux pero puede actuar en cualquier ordenador conectado a la red independientemente del sistema operativo que tenga.

SquidGuard hace las funciones de redireccionamiento web, al igual que DansGuardian usa un sistema de listas negras para saber que URLs o dominios tiene que bloquear, esto quiere decir que la efectividad va a estar condicionada por la calidad de la lista negra que utilicemos, en internet podemos encontrar listas predefinidas para ahorrarnos el trabajo de configurarlo todo manualmente por ejemplo en la pagina

http://dsi.ut-capitole.fr/blacklists/index_en.php

podemos encontrar una infinidad de listas negras filtradas por categorías para elegir la que más nos convenga según nuestras necesidades.

SquidGuard también permite elegir espacios de tiempo en los que hacer efectivos los bloqueos, de esta manera si lo queremos usar en una oficina podemos restringir el acceso solo en horas laborales.

También permite crear grupos jerárquicos, de esta manera podemos hacer más leves los bloqueos según ascendamos en esta jerarquía.

Otra de las funciones interesantes de SquidGuard es que no solo bloquea URLs que puedan ser molestas, sino que nos permite redireccionar al usuario según nuestras necesidades, por ejemplo, redirigir automáticamente a un usuario que no esta registrado a la página de registro, o redirigir las páginas bloqueadas a una pagina de información sobre los bloqueos establecidos.

6.1 Instalacion:

Para instalar SquidGuard lo mas fácil es ir a la página oficial http://www.squidguard.org/Doc/install.html

donde nos dan todos los pasos necesarios para instalarlo, simplemente tenemos que descargar un archivo zip desde la página oficial descomprimirlo compilarlo y ejecutarlo.

6.2 Bloquear páginas:

Voy a poner una captura de la página oficial de SquidGuard porque es más fácil de entender de esta manera

```
#
# CONFIG FILE FOR SQUIDGUARD
#

dbhome /usr/local/squidGuard/db
logdir /usr/local/squidGuard/logs

dest porn {
        domainlist porn/domains
        urllist porn/urls
        }

acl {
        default {
                pass !porn all
                redirect http://localhost/block.html
        }
 }
```

Esta es la sintaxis básica para hacer un bloqueo usando las listas negras de SquidGuard,las dos primeras líneas dbome y logdir indican la rutas de las listas negras y de los archivos log.

Dest lo que hace es definir una categoría en este caso porn que es la que vamos a bloquear, dentro de dest indicamos las urls y dominios que estarán dentro de esta categoría, y dentro de acl indicamos la página a la que queremos que redirija todas estas páginas.

7. Sarg*:*

Sarg es una herramienta que usada junto a Squid y DansGuardian nos permitirá tener el máximo control de los usuarios conectados a nuestra red, al igual que los otros dos Sarg tiene que estar instalado en un ordenador con GNU/Linux y configurado de manera adecuada con Squid para que pueda funcionar.

La función de Sarg es interpretar los archivos log de Squid para generar reportes de actividad en formato HTML para que sean fáciles de interpretar.

7.1 Instalación:

Simplemente tenemos que usar la orden apt-get install sarg y ya lo tendríamos instalado, una vez hecho esto ejecutando la orden sarg -x generaría un reporte del

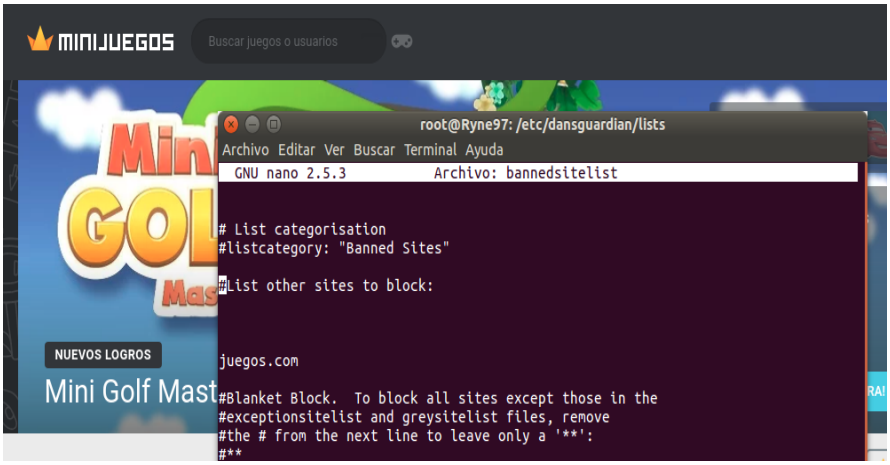
tráfico generado desde la instalación hasta el momento en que ejecutemos el comando.

Podemos cambiar la carpeta en la que se guardan los reportes desde el archivo de configuración de SARG

```
# TAG:  output_dir
#       The reports will be saved in that directory
#       sarg -o dir
#
output_dir /var/www/html/squid-reports
output_dir /var/lib/sarg
```

8. Resultados y conclusiones.

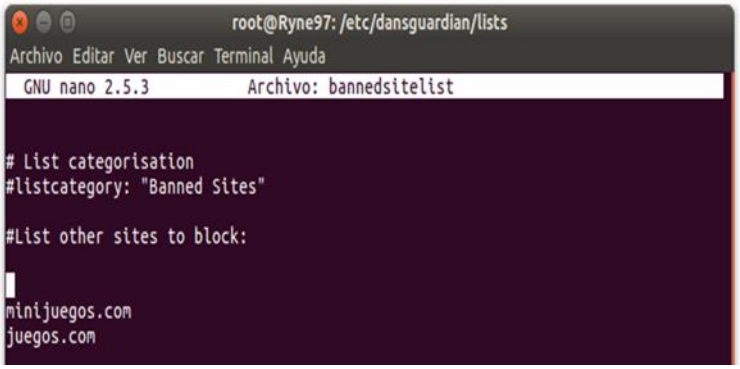
Ejecución de dansguardian**:**



No se puede acceder a este sitio web

Es posible que la página web **https://www.minijuegos.com/** esté temporalmente inactiva o que se haya trasladado definitivamente a otra dirección.

ERR_TUNNEL_CONNECTION_FAILED

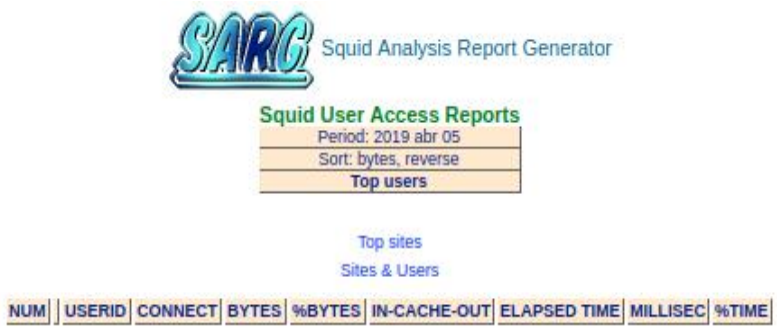


Sarg:

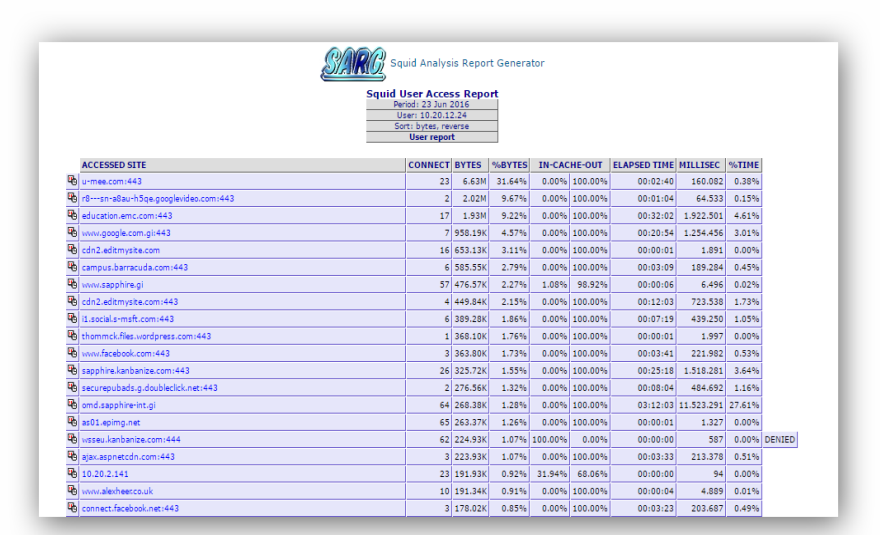
Como ya hemos dicho lo que hace SARG es interpretar los datos recopilados por Squid y convertirlos a HTML para que podemos leerlos de una manera sencilla y podamos configurarlos como nosotros queramos. Básicamente el archivo de configuración de SARG es HTML así que si manejamos este lenguaje podremos configurarlo como nosotros queramos.

```
# sarg.conf
#
# TAG:  access_log file
#       Where is the access.log file
#       sarg -l file
#
access_log /var/log/squid/access.log

# TAG: graphs yes|no
#       Use graphics where is possible.
#           graph_days_bytes_bar_color blue|green|yellow|orange|brown|red
#
#graphs yes
#graph_days_bytes_bar_color orange

# TAG:  graph_font
#       The full path to the TTF font file to use to create the graphs. It is required
#       if graphs is set to yes.
#
#graph_font /usr/share/fonts/truetype/ttf-dejavu/DejaVuSans.ttf

# TAG:  title
#       Especify the title for html page.
#
title "Squid User Access Reports"

# TAG:  font_face
#       Especify the font for html page.
#
font_face Tahoma,Verdana,Arial

# TAG:  header_color
#       Especify the header color
#
header_color darkblue

# TAG:  header_bgcolor
#       Especify the header bgcolor
#
header_bgcolor blanchedalmond

# TAG:  font_size
#       Especify the text font size
#
font_size 9px

# TAG:  header_font_size
#       Especify the header font size
#
#header_font_size 9px

# TAG:  title_font_size
#       Especify the title font size
#
```

Un reporte vacío de SARG tiene este aspecto.



Un ejemplo de un reporte de SARG completo es este:


Squid Analysis Report Generator

Squid User Access Report

Period: 23 Jun 2016
User: 10.20.12.24
Sort: bytes, reverse
User report

ACCESSED SITE	CONNECT	BYTES	%BYTES	IN-CACHE-OUT	ELAPSED TIME	MILLISEC	%TIME	
u-mee.com:443	23	6.63M	31.64%	0.00%	100.00%	00:02:40	160.082	0.38%	
r8---sn-a8au-h5qe.googlevideo.com:443	2	2.02M	9.67%	0.00%	100.00%	00:01:04	64.533	0.15%	
education.emc.com:443	17	1.93M	9.22%	0.00%	100.00%	00:32:02	1.922.501	4.61%	
www.google.com.gi:443	7	958.19K	4.57%	0.00%	100.00%	00:20:54	1.254.456	3.01%	
cdn2.editmysite.com	16	653.13K	3.11%	0.00%	100.00%	00:00:01	1.891	0.00%	
campus.barracuda.com:443	6	585.55K	2.79%	0.00%	100.00%	00:03:09	189.284	0.45%	
www.sapphire.gi	57	476.57K	2.27%	1.08%	98.92%	00:00:06	6.496	0.02%	
cdn2.editmysite.com:443	4	449.84K	2.15%	0.00%	100.00%	00:12:03	723.538	1.73%	
i1.social.s-msft.com:443	6	389.28K	1.86%	0.00%	100.00%	00:07:19	439.250	1.05%	
thommck.files.wordpress.com:443	1	368.10K	1.76%	0.00%	100.00%	00:00:01	1.997	0.00%	
www.facebook.com:443	3	363.80K	1.73%	0.00%	100.00%	00:03:41	221.982	0.53%	
sapphire.kanbanize.com:443	26	325.72K	1.55%	0.00%	100.00%	00:25:18	1.518.281	3.64%	
securepubads.g.doubleclick.net:443	2	276.56K	1.32%	0.00%	100.00%	00:08:04	484.692	1.16%	
omd.sapphire-int.gi	64	268.38K	1.28%	0.00%	100.00%	03:12:03	11.523.291	27.61%	
as01.epimg.net	65	263.37K	1.26%	0.00%	100.00%	00:00:01	1.327	0.00%	
wsseu.kanbanize.com:444	62	224.93K	1.07%	100.00%	0.00%	00:00:00	587	0.00%	DENIED
ajax.aspnetcdn.com:443	3	223.93K	1.07%	0.00%	100.00%	00:03:33	213.378	0.51%	
10.20.2.141	23	191.93K	0.92%	31.94%	68.06%	00:00:00	94	0.00%	
www.alexheerco.uk	10	191.34K	0.91%	0.00%	100.00%	00:00:04	4.889	0.01%	
connect.facebook.net:443	3	178.02K	0.85%	0.00%	100.00%	00:03:23	203.687	0.49%	

He usado capturas de internet porque erá más fácil que simular el tráfico de un red en casa.

En este caso podemos ver el tráfico de un solo usuario, SARG nos aporta datos sobre los sitios web visitados y el tiempo que se ha estado en ellos, también nos muestra si se ha visitado alguna página cuyo acceso se haya denegado.

Todos los programas son de gran utilidad para la seguridad de cualquier persona, no solo por su simplicidad sino también por su eficiencia. Es por ello que debería estar más presente en nuestro día a día, que como en un principio se explicó, una gran parte de la sociedad utiliza los ordenadores de una forma incorrecta. Son aplicaciones muy útiles y que la mayoría desconoce.

References

1. https://www.salixnetworks.com/filtrado_web.html
2. https://es.wikipedia.org/wiki/Filtro_de_contenido
3. https://www.pandasecurity.com/homeusers/downloads/docs/product/help/is/2009/sp/84.htm
4. https://www.ecured.cu/DansGuardian
5. https://www.ecured.cu/Squid
6. http://recursostic.educacion.es/observatorio/web/en/software/software-general/524-dansguardian-filtro-de-contenidos
7. https://www.linuxadictos.com/introduccion-a-squid-configuracion-paso-a-paso.html
8. https://techlandia.com/filtro-computadora-sobre_96865/
9. http://www.squidguard.org/about.html
10. https://blogdesistemas.com/instalar-sarg-squid-analysis-report-generator/