

Google Search Appliance

Temas adicionales y preguntas
y respuestas

Marzo de 2014



© 2014 Google

Temas adicionales y preguntas y respuestas

En este documento se incluyen temas adicionales, así como preguntas y respuestas que no aparecen en ninguno de los otros documentos de Notas sobre GSA.

Acerca de este documento

La información y recomendaciones que se exponen en este documento se recopilaron gracias al trabajo que realizamos con distintos clientes y entornos del sector. Nos gustaría agradecer a nuestros clientes y partners que hayan compartido sus experiencias y puntos de vista con nosotros.

Temas tratados	En este documento encontrará varios consejos y trucos que ha recabado el equipo de implementación de GSA de Google y sugerencias sobre cómo usar el kit de herramientas de administración de GSA, el cual está disponible de forma pública.
Público objetivo principal	Desarrolladores y administradores de GSA
Entorno de TI	GSA configurado para búsquedas públicas y seguras
Fases de implementación	Configuración y posimplementación de GSA
Otros recursos	<ul style="list-style-type: none">• Learngsa.com proporciona recursos formativos para GSA.• La documentación de producto de GSA proporciona información completa sobre GSA.• El Portal de asistencia de Google for Work proporciona acceso a la asistencia de Google.

Índice

[Acerca de este documento](#)

[Capítulo 1 Usar Apache como servidor proxy de filtrado](#)

[Descripción general](#)

[Configurar Apache como servidor proxy](#)

[Configurar GSA para usar el servidor proxy](#)

[Usar varias configuraciones del servidor proxy](#)

[Crear filtros](#)

[Recursos adicionales](#)

[Capítulo 2 Usar el kit de herramientas de administración de Google Search Appliance](#)

[Descripción general](#)

[Cómo analizar registros de búsqueda mediante searchstats.py](#)

[Cómo automatizar tareas de la Consola del administrador](#)

[Cómo retirar o volver a rastrear documentos en el índice](#)

[Verificar una configuración de Kerberos](#)

[Capítulo 3 Preguntas y respuestas](#)

[Descripción general](#)

[Personalización de resultados](#)

[Clasificación de metadatos](#)

[Supervisar GSA](#)

[Sugerencias de consulta](#)

[Concordancia de consultas parciales](#)

[Filtros de consulta y OneBox](#)

Capítulo 1 Usar Apache como servidor proxy de filtrado

Descripción general

Cuando se configura Google Search Appliance (GSA), el control sobre cómo se rastrea el contenido o cómo se presenta a GSA para procesos posteriores es limitado. Sin embargo, si incorpora Apache como servidor proxy en el entorno de implementación, podrá modificar el contenido a medida que se rastrea para satisfacer varios objetivos.

El caso más habitual para modificar contenido es el filtrado: excluir o añadir contenido a las páginas a medida que se rastrean. La posibilidad de modificar contenido a medida que se rastrea sirve también para cambiar el rastreador ante las fuentes de contenido.

Para usar Apache como servidor proxy de filtrado, debe:

1. [Configurar Apache como servidor proxy](#)
2. [Configurar GSA para usar el servidor proxy](#)

Configurar Apache como servidor proxy

Para configurar Apache como servidor proxy, añade estas líneas a `httpd.conf`:

```
LoadModule proxy_module modules/mod_proxy.so
LoadModule proxy_http_module modules/mod_proxy_http.so
Listen 8080

<VirtualHost *:8080>

ProxyRequests On
<Proxy *>
    Order Deny,Allow
    Deny from all
    Allow from 192.168.0.20
</Proxy>

### Add filters here ###

</VirtualHost>
```

Las dos primeras líneas de esta configuración simplemente cargan el módulo [mod_proxy](#) e indican al servidor Apache que empiece a escuchar en el puerto 8080.

La siguiente sección define un host virtual en el puerto 8080 y le indica que transfiera los resultados al servidor proxy (en lugar de mostrarlos como un servidor web normal).

Bloquear la configuración

Si la máquina en la que se ejecuta esta configuración es pública, debe bloquear esta última aún más para que no se utilice de manera inadecuada. En este caso, utilice reglas de IP simples para permitir solicitudes de servidor proxy únicamente de 192.168.0.20 (el dispositivo GSA).

Probar el servidor proxy

Una vez que haya iniciado el servidor, pruébelo siguiendo estos pasos:

1. Acceda mediante telnet al puerto 8080.
2. Introduzca "GET http://www.google.com/".

Si el servidor funciona, muestra la fuente de la página principal de Google. Pero si su dirección IP no está permitida, aparece el error 403 (prohibido).

Configurar GSA para usar el servidor proxy

Para configurar GSA de forma que utilice el servidor proxy:

1. En la Consola del administrador de GSA, vaya a **Fuentes de contenido > Rastreo web > Servidores proxy** (en versiones anteriores a la 7.2: **Rastrear e indexar > Servidores proxy**).
2. Introduzca los patrones de URL que debería usar el servidor proxy, la dirección IP o el nombre de dominio completo y el puerto del servidor proxy que ha configurado.
3. Haga clic en **Guardar** (en versiones anteriores a la 7.2: **Guardar la configuración de los servidores proxy del rastreador**).

En algunas situaciones, quizá le interese usar el servidor proxy para rastrear todo el contenido. En ese caso, tan solo debe introducir "/" para el patrón de URL. En otras situaciones, por ejemplo, si rastrea vídeo o imágenes, quizá quiera usar el servidor proxy solo para ese contenido. En ese caso, introduzca un patrón de URL, según corresponda.

Usar varias configuraciones del servidor proxy

Si necesita disponer de varias configuraciones del servidor proxy para la aplicación, puede ejecutar varias instancias de Apache en distintos puertos, o puede definir filtros en una misma configuración de Apache para gestionar el contenido a partir de patrones de URL u otros parámetros.

Crear filtros

Apache admite dos tipos de filtros:

- Input
- Output

En los servidores proxy, los filtros Input (datos) son las solicitudes que GSA envía al servidor web de destino; y los filtros Output (resultado) son los que el servidor web devuelve a GSA. Por lo tanto, para la mayoría de aplicaciones deberá crear filtros Output.

Estas son algunas de las directivas de Apache para crear filtros Output:

- [SetOutputFilter](#)
- [AddOutputFilterByType](#)

Los filtros simplemente están definidos como parte del bloque `Proxy Virtual Host`.

Directiva SetOutputFilter

La directiva [SetOutputFilter](#) se puede usar para aplicar un filtro a todo el contenido que se transfiera a través del servidor proxy:

```
# Filter robots meta tags
ExtFilterDefine fixrobots mode=output intype=text/html \
cmd="/bin/sed -r 's/(noarchive|noindex|nofollow)> //g'"
SetOutputFilter fixrobots
```

En este ejemplo, hemos definido un filtro externo, "fixrobots", que solo transfiere stdin (el documento solicitado) a través de sed, y excluye las cadenas "noarchive", "noindex" y "nofollow".

Esto básicamente indica a GSA que ignore las metaetiquetas robots insertadas. "sed -r" es rápido y fácil para patrones de expresiones regulares y manipulación simple de cadenas. Pero resulta igualmente fácil usar una secuencia de comandos Perl, PHP o shell. Apache simplemente transfiere el archivo como stdin y devuelve el resultado del filtro a GSA.

Directiva AddOutputFilterByType

La directiva [AddOutputFilterByType](#) ofrece un poco más de control, ya que permite aplicar un filtro basado en el tipo MIME. Resulta útil para rastrear contenido que GSA no admite de forma nativa, como imágenes, vídeos, etc.

```
# Filter video files
ExtFilterDefine filtervideo mode=output outtype=text/html \
cmd="/home/ericl/mediaFilter.php"
AddOutputFilterByType filtervideo video/x-msvideo video/mp4 video/ audio/mpeg
audio/ video/quicktime
```

En este ejemplo, hemos creado un filtro externo, "filtervideo", que invoca una secuencia de comandos externa, mediaFilter.php. En este caso es una secuencia de comandos que aceptaría archivos binarios de vídeo como Input, y mostraría HTML (metadatos insertados) y miniaturas como Output.

Dado que solo queremos aplicarlo en determinados tipos de contenido, usamos la directiva **AddOutputFilterByType** para especificar varios formatos multimedia.

Como alternativa, modifique las cabeceras HTTP. Por ejemplo, simplemente sustituya la cadena User-Agent de GSA por otra:

```
# Set User Agent of the Proxy
RequestHeader set User-Agent "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT
6.0; SLCC1; .NET CLR 2.0.50727;)"
```

Con esto no se modifican las cabeceras de GSA porque no se transfieren a través del servidor proxy. Tan solo define la cabecera que utiliza Apache cuando obtiene la página.

Resulta útil para definir una cookie concreta, User-Agent, u otras cabeceras para rastrear el contenido.

Una vez que se haya configurado el servidor proxy y los filtros, puede probarlos enviando sus propias solicitudes GET o a través de estas acciones:

1. Configurar el navegador para que utilice el servidor proxy
2. Solicitar algunas URL
3. Ver la fuente

Cuando esté todo correcto, simplemente añada los patrones del servidor proxy adecuados a GSA e inicie el rastreo. Los documentos almacenados en caché deberían mostrar los resultados filtrados (Output).

Recursos adicionales

- Página del [módulo mod_proxy de Apache](#)
- [Guía del almacenamiento en caché de Apache](#)
- Página del [módulo mod_cache de Apache](#)

Capítulo 2 Usar el kit de herramientas de administración de Google Search Appliance

Descripción general

El kit de herramientas de administración de GSA es una biblioteca de herramientas de código abierto para administradores de GSA. Puede descargar las herramientas por separado en el [kit de herramientas de administración de GSA](#).

En la tabla siguiente se incluyen las herramientas disponibles y una breve descripción de sus funciones.

Herramienta	Descripción
monitor.sh	Supervisa secuencias de comandos que verifican la publicación de resultados en GSA
load.py	Ejecuta pruebas de carga en GSA
authn.py	Servidor web para realizar pruebas de SPI de autenticación
authz.py	Servidor web para realizar pruebas de SPI de autorización
Connect.java	Clase Java para probar la conexión JDBC a la base de datos
cached_copy_checker.py	Supervisar secuencias de comandos que verifican que el rastreo, la indexación y la publicación de resultados funciona en GSA
sso.py	Servidor web para probar sitios de cookies; se puede configurar para que imite a Oblix
searchstats.py	Análisis del archivo de registro de búsqueda (porcentaje de errores, consultas por segundo, tiempo medio de respuesta)
smbcrawler.py	Secuencia de comandos que imita la manera en que GSA rastrea SMB; sirve para solucionar problemas de rastreo con SMB cuando el mensaje de error en GSA no resulta de ayuda
reverse_proxy.py	Servidor proxy inverso que puede servir para poner en cola las solicitudes dirigidas a GSA y limitar el número de conexiones simultáneas; se escribió para demostrar los principios en los que se basa y no se ha probado en un entorno de producción
gsa_admin.py	Secuencia de comandos Python para automatizar las tareas de la Consola del administrador; se utiliza en casos en los que el API GData de la Consola del administrador no funciona (p. ej., versión de software anterior a la 6.0 o si falta una función en el API)
remove-or-recrawl-urls.html	Página de métodos auxiliares HTML+JavaScript que crea feeds para retirar o volver a rastrear URL
urlstats.py	Secuencia de comandos Python que genera informes sobre URLs en GSA

ssoproxy.py	Secuencia de comandos Python configurable que transfiere reglas de inicio de sesión de sistemas de inicio de sesión único a través del servidor proxy para proporcionar a GSA cookies de inicio de sesión único de rastreo/publicación de resultados
CrawlReport.java	Clase Java para obtener, a través de la nueva API de administración (versión de software 6.0.0), el número de URLs que se han rastreado desde el día anterior
connectormanager.py	Gestor de conectores simple y conectores de ejemplo con documentación sobre cómo escribir un conector nuevo
Herramienta para validar Kerberos	Aplicación HTML para validar la configuración de Kerberos/Autenticación de Windows integrada (keytab/Active Directory, etc.)
interactive-feed-client.html	Página HTML/JavaScript que genera feeds XML de datos introducidos tales como la URL y el nombre de host para luego enviarlos a GSA
search_report_xhtml.xsl	Hoja de estilo XSLT para transformar XMLs de informes de búsqueda exportados en XHTML que pueda leer una persona
search_results_analyzer.py	Herramienta para analizar resultados de búsqueda y comparar resultados entre dos GSA
convert_cached_copy_to_feed.py	Herramienta para convertir versiones en caché en feeds de contenido para migrarlos
fetch_secure.py	Obtener resultados de búsqueda seguros de GSA siguiendo todos los redireccionamientos de inicio de sesión universal
GSA-GA.zip	Recursos de integración de Google Analytics
SHT	Herramienta de autoayuda

Cuatro de nuestras herramientas más usadas de esta lista se explican en las secciones siguientes:

- [Cómo analizar registros de búsqueda mediante searchstats.py](#)
- [Cómo automatizar tareas de la Consola del administrador](#)
- [Cómo retirar o volver a rastrear documentos en el índice](#)
- [Verificar la configuración de Kerberos](#)

Cómo analizar registros de búsqueda mediante searchstats.py

searchstats.py es una herramienta de análisis de registros de búsqueda. En la captura de pantalla siguiente se incluye el análisis de un archivo de registro de búsqueda, 2011-03-14-web_log.log, a intervalos de una hora. El registro de búsqueda se descarga de la Consola del administrador de GSA, en **Informes > Registros de búsqueda** (en versiones anteriores a la 7.2: **Estado e informes > Registros de búsqueda**).

```

-bash-3.1$ python searchstats.py 2011-03-14-web_log.log 1h
Summary for 14/Mar/2011: total searches: 59
-----
time      200 non-200 %err total %tot qps av. response
00:00:00-00:59:59 0 0 0.00 0 0.00 0.00 0.000
01:00:00-01:59:59 0 0 0.00 0 0.00 0.00 0.000
02:00:00-02:59:59 1 0 0.00 1 1.69 0.00 0.033
03:00:00-03:59:59 0 0 0.00 0 0.00 0.00 0.000
04:00:00-04:59:59 48 0 0.00 48 77.97 0.01 0.022
05:00:00-05:59:59 9 0 0.00 9 15.25 0.00 0.014
06:00:00-06:59:59 3 0 0.00 3 5.08 0.00 0.018
07:00:00-07:59:59 0 0 0.00 0 0.00 0.00 0.000
08:00:00-08:59:59 0 0 0.00 0 0.00 0.00 0.000
09:00:00-09:59:59 0 0 0.00 0 0.00 0.00 0.000
10:00:00-10:59:59 0 0 0.00 0 0.00 0.00 0.000
11:00:00-11:59:59 0 0 0.00 0 0.00 0.00 0.000
12:00:00-12:59:59 0 0 0.00 0 0.00 0.00 0.000
13:00:00-13:59:59 0 0 0.00 0 0.00 0.00 0.000
14:00:00-14:59:59 0 0 0.00 0 0.00 0.00 0.000
15:00:00-15:59:59 0 0 0.00 0 0.00 0.00 0.000
16:00:00-16:59:59 0 0 0.00 0 0.00 0.00 0.000
17:00:00-17:59:59 0 0 0.00 0 0.00 0.00 0.000
18:00:00-18:59:59 0 0 0.00 0 0.00 0.00 0.000
19:00:00-19:59:59 0 0 0.00 0 0.00 0.00 0.000
20:00:00-20:59:59 0 0 0.00 0 0.00 0.00 0.000
21:00:00-21:59:59 0 0 0.00 0 0.00 0.00 0.000
22:00:00-22:59:59 0 0 0.00 0 0.00 0.00 0.000
23:00:00-23:59:59 0 0 0.00 0 0.00 0.00 0.000
-----
Total entries: 59

```

Tenga en cuenta que searchstats.py está escrito en el lenguaje de secuencia de comandos de programación Python. Para ejecutar la secuencia de comandos, se necesita el entorno de tiempo de ejecución Python. Puede descargar el entorno en la [página de descargas de Python](#).

Tras la instalación, ejecute la secuencia de comandos usando una sintaxis similar a la que se muestra en las capturas de pantalla anteriores.

Cómo automatizar tareas de la Consola del administrador

El [API de administración de GSA](#) es el método preferido para automatizar tareas en la Consola del administrador. Sin embargo, no cubre todas las tareas disponibles en dicha consola. Por ejemplo, no se admite la sincronización de bases de datos y tampoco se pueden sincronizar de forma remota mediante el API de administración. Para estos casos, utilice gsa_admin.py, una secuencia de comandos que mediante programación puede iniciar sesión en la Consola del administrador y hacer clic en los elementos de menú necesarios.

En las capturas de pantalla siguientes se muestra:

- gsa_admin.py syntax
- Comando de muestra para sincronizar dos bases de datos

```

C:\tmp>c:\Python27\python.exe gsa_admin.py --help
Usage: gsa_admin.py [options]

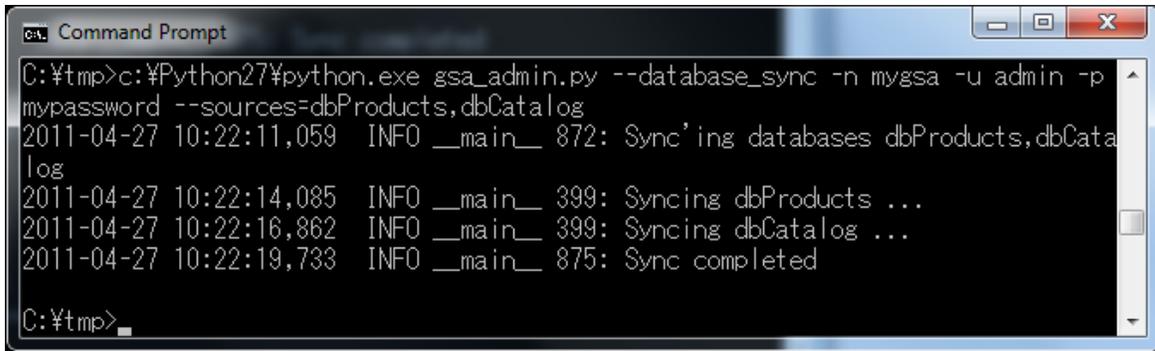
Options:
  -h, --help                show this help message and exit
  -f FILE, --input-file=FILE
                           Input XML file
  -o FILE, --output=FILE   Output file name
  -g SIGNPASSWORD, --sign-password=SIGNPASSWORD
                           Sign password for signing/import/export
  --cache-timeout=CACHETIMEOUT
                           Value for Authorization Cache Timeout
  --max-hostload=MAXHOSTLOAD
                           Value for max number of concurrent authz requests per
                           server
  --sources=SOURCES        List of databases to sync
                           (database1,database2,database3)
  --frontend=FRONTEND      Frontend used to export keymatches or related queries
  -v, --verbose             Specify multiple times to increase verbosity

Actions::
  -i, --import              Import config file to GSA
  -e, --export              Export GSA config file from GSA
  -s, --sign                Sign input XML file
  -r, --verify              Verify signature/HMAC in XML Config
  -a, --set                 Set Access Control settings
  -l, --all_urls            Export all URLs from GSA
  -d, --database_sync      Sync databases
  -k, --keymatches_export  Export All Keymatches
  -y, --synonyms_export    Export All Related Queries
  -z, --get-status         Get GSA Status

GSA info:
  -n GSAHOSTNAME, --hostname=GSAHOSTNAME
                           GSA hostname
  --port=PORT              Upload port. Defaults to 8000
  -u GSAUSERNAME, --username=GSAUSERNAME
                           Username to login GSA
  -p GSAPASSWORD, --password=GSAPASSWORD
                           Password for GSA user

```

Sintaxis del comando



```
C:\tmp>c:\Python27\python.exe gsa_admin.py --database_sync -n mygsa -u admin -p  
mypassword --sources=dbProducts,dbCatalog  
2011-04-27 10:22:11,059 INFO __main__ 872: Sync'ing databases dbProducts,dbCatalog  
2011-04-27 10:22:14,085 INFO __main__ 399: Syncing dbProducts ...  
2011-04-27 10:22:16,862 INFO __main__ 399: Syncing dbCatalog ...  
2011-04-27 10:22:19,733 INFO __main__ 875: Sync completed  
C:\tmp>
```

Sintaxis de muestra para sincronizar dos bases de datos (dbProducts y dbCatalog)

Cómo retirar o volver a rastrear documentos en el índice

Cuando deba retirar o volver a rastrear ciertos documentos en el índice, la herramienta `remove-or-recrawl-urls.html` le será de gran ayuda. Es una página HTML con JavaScript insertado que acepta una lista de URLs como datos de entrada y crea y envía archivos de feeds a GSA. Según la información de la lista de URLs, GSA retira o vuelve a rastrear las URL especificadas. Esta secuencia de comandos puede abrirse y ejecutarse mediante cualquier navegador web que admita JavaScript. En estos momentos admite los siguientes tres parámetros de entrada de datos:

- Nombre de host de GSA o dirección IP
- Lista de URLs que se deben retirar o volver a rastrear
- URL concretas que se deben retirar o volver a rastrear

Sin embargo, tenga en cuenta que antes de enviar un feed al dispositivo de búsqueda, debe registrar la dirección IP de su navegador web en la lista blanca de IPs. Para ver o modificar la lista blanca, vaya a la página **Fuentes de contenido > Feeds** (en versiones anteriores a la 7.2: **Rastrear e indexar > Feeds**) de la Consola del administrador.

En la captura de pantalla siguiente se muestra la página `remove-or-recrawl-urls.html`.

Remove or recrawl URLs by sending an XML feed

This form generates an XML feed for the Google Search Appliance, which will tell the appliance to remove or recrawl the provided URLs.

1. Enter the **Appliance hostname or IP address**, no port is necessary --it always POSTs to port 19900.
2. Enter URLs into the **URLs to remove/recrawl** text box, one URL per line, no comments are allowed.
3. Press the appropriate button:
 - **Remove them** will submit an incremental content feed (`datasource remove-urls`) with one record per URL, all in a group with `action="delete"` to remove them from the index.
 - **Recrawl them** will submit a web feed (`datasource web`) with one record per URL, with no meta data to inject them in the crawl queue.

Read the [Feeds Protocol Developer's Guide](#) to learn more about how this works.

Appliance hostname or IP address:

URLs to remove/recrawl:

Verificar una configuración de Kerberos

Configurar el dispositivo de búsqueda para que realice autenticaciones silenciosas de Kerberos implica componentes muy diversos. Por ejemplo, Active Directory, archivos keytab, DNS, área de seguridad de Internet Explorer, métodos de encriptación de datos, etc. Si se produce algún problema al configurar la autenticación de Kerberos, utilice la herramienta para validar Kerberos para realizar una comprobación rápida de cada uno de los componentes de la configuración.

En la captura de pantalla siguiente se muestra la herramienta de validación de Kerberos y una lista de los elementos configurables que comprueba.

Validate GSA Kerberos Configuraton

- FullyQualified DNS AName for GSA:
- GSA username (not DOMAIN\username):
- Keytab file:

Running Diagnostics for:
CN=Administrator,CN=Users,DC=esodomain,DC=com
Forest DNS Name: esodomain.com
DNSServer: ESOTEST.esodomain.com
Root Domain: DC=esodomain,DC=com
Domain Controller Functionality Level: WIN 2008
Forest Functionality Level: WIN 2008

This utility validates various GSA-related kerberos infrastructure and configuration options for Secure Serving. Please run this script on any windows system that has the [MIT Kerberos client](#) installed. It is assumed the user ran through the [GSA Kerberos Setup](#). This script also tests the IE version and Zone settings for the current user.
Additional Kerberos Troubleshooting [link](#) on the GSA:

```
Starting Diagnostics
1 Attempting verify DNS ANAME
  --->DNS Entry for the GSA is an A-Name
2 Attempting to parse keytab
  --->Found KVNO: 3
  --->Found Principal: HTTP/gsademo4.esodomain.com@ESODOMAIN.COM
  --->Found Encryption is DES: (DES cbc mode with RSA-MD5)
  --->Principal in keytab is in correct case format: HTTP/gsa.yourdomain.com@YOURDOMAIN.COM
3 Attempting to contact Active Directory
  ---> Found User account [gsademo4]
  ---> AD UserPrincipalName: HTTP/gsademo4.esodomain.com@ESODOMAIN.COM
  ---> AD KVNO: 3
  ---> KVNOs in AD and keytab matched
  ---> AD Useraccount is unlocked
  ---> AD Account Trusted for Delegation
  ---> AD DES Encryption Enabled
  ---> AD Password has not Expired
  ---> AD SPN Found: HTTP/gsademo4.esodomain.com
  ---> Comparing SPN from AD against keytab...
  ---> SPN/Principal in Keytab and AD Matched: HTTP/gsademo4.esodomain.com
  ---> DNS ANAME matches the host portion of the SPN
  ---> Principal in Keytab matches UPN in AD: HTTP/gsademo4.esodomain.com@ESODOMAIN.COM
4 Attempting to verify password in keytab
  ---> Password in Keytab Verified in AD
5 Attempting to verify SPN uniqueness in AD
  ---> SPN is unique: HTTP/gsademo4.esodomain.com--->[CN=gsademo4,CN=Users,DC=esodomain,DC=com]
6 Attempting to verify IE Zone settings
  ---> Internet Explorer Version: 8.0.6001.19048
  ---> GSA ANAME Found in IE Local Intranet Zone
```

Si desea obtener más información al respecto, consulte la siguiente documentación online:

- [Usar la herramienta para validar la configuración de Kerberos](#); los requisitos del sistema son muy estrictos: Windows XP, Windows Vista o Windows 7 de 32 bits
- [Solucionar problemas con la configuración de Kerberos y búsquedas seguras](#)

Capítulo 3 Preguntas y respuestas

Descripción general

En esta sección se incluyen algunas preguntas y respuestas generales de GSA que tratan varios temas.

Personalización de resultados

Pregunta: ¿Por qué no aparece un documento en la página de resultados de búsqueda cuando lo he personalizado específicamente por tipo de contenido?

Respuesta: Uno de los motivos por los que no ve el resultado que espera puede estar relacionado con la manera en que GSA obtiene los resultados más relevantes a partir del término de búsqueda. Si busca, por ejemplo, "dolor", GSA sigue estos pasos:

1. GSA localiza primero los 1.000 documentos con mayor ranking (mediante PageRank) del índice que contienen el término. Este paso no está relacionado con la frecuencia de términos ni con ninguna política de personalización configurada para la interfaz.
2. Esos 1.000 resultados se ejecutan con varios algoritmos que realizan procesos, como la clasificación basada en la frecuencia del término en los documentos, resultado de la personalización y muchos otros factores.

Si la URL que busca tiene un PageRank menor que los primeros 1.000 documentos del paso 1, ni siquiera se incluirá en los demás algoritmos que tienen en cuenta las políticas de personalización.

Incluso cuando una URL se incluye en los primeros 1.000 documentos del paso 1, el hecho de tener la política más estricta no significa necesariamente que la URL se sitúe entre los primeros resultados. La personalización es solo uno de los factores que se contemplan en el paso 2 del proceso. Hay muchos otros factores que influyen en el resultado final.

Clasificación de metadatos

Pregunta: ¿Por qué un determinado documento que sé que existe y tiene metadatos válidos no se incluye cuando clasifico por metadatos?

Respuesta: La clasificación por metadatos afecta solamente a los 1.000 documentos con mayor ranking (mediante PageRank). Consulte también [Personalización de resultados](#).

Supervisar GSA

Pregunta: ¿Qué debo hacer para supervisar el estado de GSA y de otros componentes de la arquitectura que se han implementado? ¿Es una exageración ejecutar una consulta con secuencia de comandos en el índice de GSA?

Respuesta: Tiene varias opciones para supervisar el estado de la solución. Muchos de nuestros clientes simplemente comprueban si el puerto 80 responde. Otros, además, supervisan la publicación de contenido, el rastreo y la indexación. Si desea obtener información sobre las estrategias de supervisión, consulte [Configurar la supervisión](#) en *Diseñar una solución de búsqueda*.

El tiempo durante el cual supervise la solución depende de lo compleja y profunda que quiera que sea la estrategia de supervisión. Tiene varias opciones, cada una con sus ventajas y desventajas. Estas son algunas de ellas:

Usar SNMP: consulte [Objetos SNMP](#) en la página de ayuda de **Administración > Configuración SNMP**

- Ejecutar [Monitor](#) con secuencia de comandos
- Ejecutar [Cached-Copy-Checker](#) con secuencia de comandos.
- Ejecutar una consulta para una URL conocida del índice y asegurarse de que se devuelve un resultado usando XSLT
- Ejecutar una consulta y asegurarse de que se devuelve el código de estado 200
- Mostrar la interfaz sin ejecutar ninguna consulta para asegurarse de que se devuelve el código de estado 200

Ejecutar una consulta para controlar el estado de la solución no es una exageración. Si ya tiene implementada la solución y funciona, probablemente sea suficiente para detectar un problema y solucionarlo. Es posible que pueda ajustar esto creando una interfaz con los mínimos elementos (es decir, sin KeyMatch, Sugerencias de consulta, etc.) o enviando una consulta de una URL conocida incluyendo la cláusula "info:" en la consulta.

Otras herramientas disponibles:

- [monitor.sh](#): supervisa secuencias de comandos que verifican la publicación de resultados en GSA
- [cached_copy_checker.py](#): supervisar secuencias de comandos que verifican que el rastreo, la indexación y la publicación de resultados funcionan en el dispositivo de búsqueda
- [searchstats.py](#): análisis del archivo de registro de búsqueda (porcentaje de errores, consultas por segundo, tiempo medio de respuesta)
- [urlstats.py](#): secuencia de comandos Python que genera informes sobre URLs en GSA
- [search_report_xhtml.xsl](#): hoja de estilo XSLT para transformar XMLs de informes de búsqueda exportados en XHTML que pueda leer una persona

Supervise un conector comprobando el servlet de conectividad de pruebas: `http://[cm-address-and-port]/connector-manager/testConnectivity`.

Tras obtener la URL anterior, debería recibir la respuesta XML con el código de estado HTTP 200. Un `<StatusId>0</StatusId>` en XML indica que todo es correcto.

Este es el formato del XML entrante:

```
<CmResponse>
  <Info>Google Search Appliance Connector Manager 3.0.8 (build 3222 3.0.8-RC3
  May 24 2013); Sun Microsystems Inc. Java HotSpot(TM) 64-Bit Server VM 1.6.0_33;
  Windows Server 2008 R2 6.1 (amd64)</Info>
  <StatusCode>5501</StatusCode>
  <StatusId>0</StatusId>
</CmResponse>
```

Debe permitir el acceso de la máquina desde la cual realiza la supervisión.

También se recomienda supervisar el proceso Java que ejecuta la instancia del conector Tomcat para CPU/uso de la memoria.

Sugerencias de consulta

Pregunta: La función Sugerencias de consulta muestra términos inapropiados o explícitos. ¿Qué debo hacer para que estos términos no se muestren en las sugerencias de consulta?

Respuesta: Siempre cabe la posibilidad de que los usuarios realicen búsquedas en GSA de términos que son inadecuados para su empresa. Hay dos formas de evitar que este tipo de términos aparezcan mediante la función Sugerencias de consulta.

La forma más eficaz es que el API GDATA suba una lista negra de términos. La función admite expresiones regulares, así que puede crear expresiones de concordancia exacta o parcial. Si GSA ejecuta una versión anterior a la 6.14, no podrá añadir ni modificar la lista negra desde la interfaz de la Consola del administrador. Deberá hacerlo a través de APIs. Si desea obtener más información sobre este procedimiento, consulte el apartado sobre listas negras para Sugerencias de consulta en la documentación del API de GSA correspondiente:

- [Guía para desarrolladores del API de administración: protocolo](#)
- [Guía para desarrolladores del API de administración: Java](#)
- [Guía para desarrolladores del API de administración: .NET](#)

Existe una herramienta Java de código abierto para subir listas negras mediante el API GDATA: <http://code.google.com/p/gsa-admin-reports/>.

Si desea obtener información detallada sobre el uso, consulte este documento de la wiki: <http://code.google.com/p/gsa-admin-reports/wiki/BlacklistEditor>.

El otro método para incluir en una lista negra términos de sugerencias de consulta es modificar el código de la interfaz donde se muestran las sugerencias. Por ejemplo, observe el siguiente fragmento de código:

```
<!--
*****
Spelling suggestions in result page (do not customize)
*****
-->
<xsl:template name="spelling">
  <xsl:if test="/GSP/Spelling/Suggestion">
    <p>
      <span class="p">
        <font color="{ $spelling_text_color }">
          <xsl:value-of select="$spelling_text"/>
          <xsl:call-template name="nbsp"/>
        </font>
      </span>
      <xsl:variable name="apps_param">
        <xsl:choose>
          <xsl:when test="/GSP/PARAM[@name='exclude_apps']">
            <xsl:text disable-output-escaping='yes'>&exclude_apps=</xsl:text>
            <xsl:value-of select="/GSP/PARAM[@name='exclude_apps']/@original_value" />
          </xsl:when>
          <xsl:when test="/GSP/PARAM[@name='only_apps']">
            <xsl:text disable-output-escaping='yes'>&only_apps=</xsl:text>
            <xsl:value-of select="/GSP/PARAM[@name='only_apps']/@original_value" />
          </xsl:when>
        </xsl:choose>
      </xsl:variable>
      <a ctype="spell"
href="search?q={/GSP/Spelling/Suggestion[1]/@q}&spell=1&{$base_url}{$apps_param}">
        <xsl:value-of disable-output-escaping="yes" select="/GSP/Spelling/Suggestion[1]"/>
      </a>
    </p>
  </xsl:if>
</xsl:template>
```

Deberá modificar la condición `if` para que incluya una lista negra de términos y las sugerencias no se muestren en la interfaz.

Por ejemplo, en el caso de una concordancia exacta:

```
<xsl:if test="/GSP/Spelling/Suggestion and
GSP/Spelling/Suggestion[1]/@q!='offending_term'">
```

Para el parámetro "contains":

```
<xsl:if test="/GSP/ Spelling/Suggestion and
not(contains(/GSP/Spelling/Suggestion[1]/@q,'offending_term'))">
```

Pregunta: ¿Qué debo hacer para restablecer por completo el contenido de Sugerencias de consulta?

Respuesta: A partir de la versión 7.2, se puede eliminar todo el contenido de Sugerencias de consulta. Para ello, vaya a la página **Búsqueda > Funciones de búsqueda > Sugerencias** y haga clic en **Restablecer** en la opción **Restablecer sugerencias** (en versiones anteriores a la 7.2 deberá abrir una incidencia con el servicio de asistencia de Google).

Concordancia de consultas parciales

Pregunta: Estoy usando GSA para un caso práctico de búsqueda de personas. Los usuarios se quejan de que no se muestran resultados para las búsquedas que hacen de nombres parciales. ¿Qué puedo hacer para que la búsqueda de personas admita concordancias de consultas parciales del nombre y del apellido y mejorar así la experiencia de búsqueda de los usuarios?

Respuesta: En estos momentos GSA no admite la concordancia comodín de términos parciales. Dado que los usuarios están habituados a encontrar personas con solo introducir parte del nombre o del apellido, tiene varias opciones para optimizar GSA en este sentido.

En primer lugar, puede añadir metadatos a los registros de las personas en GSA que incluyan etiquetas para cada combinación de las 6 primeras letras del nombre y apellido de una persona. En el ejemplo de "Jennifer Johnson", se añadirían las siguientes 12 metaetiquetas a su registro como metadatos adicionales:

j	j
je	jo
jen	joh
jenn	john
jenni	johns
jennif	johnso

Según la manera en que el contenido de personas se adquiera en el entorno, se puede transformar en secuencia de comandos para mostrarse en una página HTML publicada, que GSA rastreará, o en GSA a través de un feed de contenido.

Otra sugerencia es incluir una lista personalizada de sinónimos name-type en la función Ampliación de consultas de GSA para ampliar ciertos apodos a nombres completos.

Es posible que en el trabajo se usen apodos para algunas personas. Si estos apodos no se adjuntan directamente a los registros de las personas como metadatos o contenido, la función Ampliación de consultas puede ayudar a solucionar el uso de apodos. Hay que prestar especial atención a las asociaciones de apodos extranjeros con apodos locales, por ejemplo, "Mateusz" se puede ampliar a "Mat", "Matt" o "Matthew".

Otra posibilidad es usar la función Sugerencias de consulta para mejorar la experiencia de usuario con el buscador de personas. Se puede implementar el laboratorio "Busca mientras escribes" para proporcionar una base de datos de sugerencias de nombres personalizados que ayude a los usuarios con la consulta sobre sus colegas.

Filtros de consulta y OneBox

Pregunta: El cuadro OneBox de la Búsqueda de perfiles desaparece cuando añado un filtro como "daterange" a la consulta. ¿Qué debo hacer para que no desaparezca tras añadir un filtro a la consulta?

Respuesta: En OneBox de la Búsqueda de perfiles se muestran los resultados de las coincidencias de los términos de consulta del índice. Cuando intenta filtrar los resultados de búsqueda orgánica mediante un filtro, por ejemplo, "daterange", quizá no lo aplica a los resultados de la Búsqueda de perfiles. Dado que los resultados de la Búsqueda de perfiles también provienen de coincidencias en el índice de una colección de personas, este resultado obedece al funcionamiento normal de GSA.

Por ejemplo, si los documentos que se incluyen en la colección y que se muestran en los resultados de la Búsqueda de perfiles no contienen valores de metadatos de fecha, no se mostrarán resultados de la Búsqueda de perfiles si se incluye un filtro de ese tipo en la consulta de búsqueda.

Una forma de mejorar esta experiencia es asegurarse de que los metadatos necesarios para el filtrado también se incluyen en los registros que activan el cuadro OneBox de la Búsqueda de perfiles. En el caso de "daterange", si no se puede determinar una fecha para los registros de los perfiles, podría usar un valor de fecha estático que puede rellenar en la consulta de búsqueda del servidor mediante la condición OR.

Por ejemplo, adjunte la fecha "1901-01-01" como fecha modificada para todos los registros de perfiles. Luego, cuando aplique un filtro en la interfaz, deberá añadirse "OR daterange:..1900-01-02" a la consulta para poder captar la parte estática de los datos que se han adjuntado a los registros.