

Google Search Appliance

Weitere Themen und F&A

März 2014



© 2014 Google

Weitere Themen und F&A

In diesem Dokument werden weitere Themen sowie Fragen und Antworten behandelt, die nicht in anderen Dokumenten der Reihe "GSA – Notes from the Field" ("GSA – Tipps aus der Praxis") erörtert werden.

Über dieses Dokument

Die Empfehlungen und Informationen in diesem Dokument sind das Ergebnis unserer praktischen Arbeit vor Ort mit verschiedenen Kunden und in verschiedenen IT-Umgebungen. Wir danken unseren Kunden und Partnern dafür, dass sie uns ihre Erfahrungen und Erkenntnisse zugänglich gemacht haben.

Inhalt	In diesem Dokument finden Sie eine Reihe von Tipps und Tricks, die vom Bereitstellungsteam der GSA zusammengetragen wurden. Außerdem erhalten Sie Empfehlungen zur Verwendung des öffentlich verfügbaren Admin-Toolkits der GSA.
Hauptzielgruppe	GSA-Administratoren und -entwickler
IT-Umgebung	Für die sichere und öffentliche Suche konfigurierte GSA
Bereitstellungsphasen	Konfiguration der GSA und weitere Schritte nach der Bereitstellung
Weitere Ressourcen	<ul style="list-style-type: none">• Learngsa.com bietet verschiedene Schulungs- und Informationsmaterialien für die GSA.• In der GSA-Produktdokumentation finden Sie umfassende Informationen über die GSA.• Über das Google for Work Support Portal haben Sie Zugriff auf den Google-Support.

Inhalt

[Über dieses Dokument](#)

[Kapitel 1: Apache als Filter-Proxy verwenden](#)

[Überblick](#)

[Apache als Proxy konfigurieren](#)

[GSA für die Verwendung des Proxyservers konfigurieren](#)

[Mehrere Proxy-Konfigurationen verwenden](#)

[Filter erstellen](#)

[Weitere Materialien](#)

[Kapitel 2: Admin-Toolkit der GSA verwenden](#)

[Überblick](#)

[Suchprotokolle mit "searchstats.py" analysieren](#)

[Aufgaben der Admin-Konsole automatisieren](#)

[Dokumente im Index löschen oder neu crawlen](#)

[Eine Kerberos-Konfiguration bestätigen](#)

[Kapitel 3: Fragen und Antworten](#)

[Überblick](#)

[Ergebnisgewichtung](#)

[Metadatensortierung](#)

[GSA überwachen](#)

[Vorschläge für Suchanfragen](#)

[Übereinstimmungen mit unvollständigen Suchanfragen](#)

[Suchanfragenfilter und OneBoxes](#)

Kapitel 1: Apache als Filter-Proxy verwenden

Überblick

Wenn Sie die Google Search Appliance (GSA) konfigurieren, haben Sie nur begrenzte Kontrolle darüber, wie Inhalte gecrawlt oder diese Inhalte für die GSA zur weiteren Verarbeitung zur Verfügung gestellt werden. Wenn Sie jedoch einen Apache-Server als Proxy in der Entwicklungsumgebung einsetzen, können Sie Inhalte beim Crawling so verändern, dass damit mehrere Zwecke erfüllt werden.

Der häufigste Anwendungsfall, bei dem Inhalte angepasst werden, ist das Filtern, z. B. wenn Inhalte beim Crawlen entfernt oder auf Seiten hinzugefügt werden müssen. Mithilfe der Möglichkeit, Inhalte während des Crawlens zu ändern, lässt sich auch beeinflussen, wie sich der Crawler aus der Sicht der Inhaltsquellen darstellt.

Sie können Apache als Filter-Proxy verwenden, indem Sie:

1. [Apache als Proxy konfigurieren](#)
2. [GSA für die Verwendung des Proxyserver konfigurieren](#)

Apache als Proxy konfigurieren

Um Apache als Proxy zu konfigurieren, müssen Sie die folgenden Zeilen zu `httpd.conf` hinzufügen:

```
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>
```

Die ersten zwei Zeilen in dieser Konfiguration laden das Modul [mod_proxy](#) und teilen dem Apache-Server mit, zuerst Port 8080 abzuhören.

Im nächsten Abschnitt wird ein virtueller Host auf Port 8080 definiert und festgelegt, dass Suchanfragen an den Proxy geleitet werden sollen (und nicht wie bei einem normalen Webserver Ergebnisse ausgegeben werden).

Konfiguration sperren

Wenn es sich bei der konfigurierten Maschine um eine öffentliche Maschine handelt, empfehlen wir dringend, die Konfiguration mit einer Sperre zu versehen, um unbefugten Zugriff zu verhindern. Schränken Sie in diesem Fall die erlaubten Suchanfragen an den Proxy mithilfe einfacher IP-Regeln auf den Port 192.168.0.20 (GSA-IP) ein.

Proxyserver testen

Nach dem Start des Servers können Sie ihn folgendermaßen testen:

1. Telnet-Test auf Port 8080
2. Geben Sie "GET http://www.google.com/" ein.

Wenn der Server aktiv ist, gibt er die Quelle von der Google-Startseite zurück. Wenn Sie keine zulässige IP-Adresse haben, wird der Fehler 403 (Unzulässig) angezeigt.

GSA für die Verwendung des Proxyserver konfigurieren

So konfigurieren Sie die GSA für die Verwendung des Proxyserver:

1. Navigieren Sie in der GSA-Admin-Konsole zu **Inhaltsquellen > Web Crawling > Proxyserver** (vor Version 7.2: **Crawl und Index > Proxyserver**).
2. Geben Sie die URL-Muster, die den Proxy verwenden sollen, die IP-Adresse oder den vollständig qualifizierten Domainnamen sowie den Port des konfigurierten Proxyserver ein.
3. Klicken Sie auf **Speichern** (vor Version 7.2: **Crawler-Proxy-Konfiguration speichern**).

In manchen Fällen ist es sinnvoll, den Proxy zu verwenden, um alle Inhalte zu crawlen. Geben Sie dafür nur "/" als URL-Muster ein. In anderen Fällen, z. B. beim Crawlen von Videos oder Bildern, bietet es sich wiederum an, den Proxy nur für diese speziellen Inhalte verwenden. Geben Sie dazu das entsprechende URL-Muster ein.

Mehrere Proxy-Konfigurationen verwenden

Wenn Sie mehrere Proxy-Konfigurationen für Ihre Anwendung benötigen, können Sie mehrere Instanzen von Apache auf verschiedenen Ports ausführen. Alternativ können Sie innerhalb einer Apache-Konfiguration Filter definieren, um Inhalte anhand von URL-Mustern oder anderen Parametern zu verarbeiten.

Filter erstellen

Apache unterstützt zwei Filtertypen:

- Eingabe
- Ausgabe

Bei Proxys ist die Eingabe die Suchanfrage, die die GSA an den Ziel-Webserver sendet. Die Ausgabe sind die Inhalte, die der Webserver zurück an die GSA sendet. Bei den meisten Anwendungen erstellen Sie also einen Ausgabefilter.

Apache verfügt über verschiedene Anweisungen zum Erstellen von Ausgabefiltern, z. B.:

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

Filter werden als Teil des Blocks `Virtueller Proxy-Host` definiert.

Anweisung `SetOutputFilter`

Die Anweisung [SetOutputFilter](#) kann verwendet werden, um einen Filter auf ALLE Inhalte anzuwenden, die durch den Proxy geleitet werden:

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

In diesem Beispiel wird ein externer Filter mit dem Namen "fixrobots" definiert. Dieser Filter passiert "stdin" (das angefragte Dokument) durch "sed" und schneidet die Strings "noarchive", "noindex" und "nofollow" aus.

Dadurch kann die GSA eingebettete Robot-Meta-Tags ignorieren. "sed -r" eignet sich für die schnelle und einfache Bearbeitung von Mustern regulärer Ausdrücke und von einfachen Strings. Die Verwendung eines Pearl-, PHP- oder Shell-Skripts ist jedoch vergleichbar einfach. Apache übergibt die Datei als "stdin" und leitet die Ausgabe des Filters zurück an die GSA.

`AddOutputFilterByType`-Anweisung

Mit der Anweisung [AddOutputFilterByType](#) haben Sie etwas mehr Steuerungsmöglichkeiten, da Sie einen Filter basierend auf dem Typ MIME anwenden können. Diese Anweisung ist nützlich, wenn Sie Inhalte crawlen möchten, die die GSA nicht nativ unterstützt, z. B. Bilder, Videos usw.

```
# 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
```

In diesem Beispiel wird ein externer Filter namens "filtervideo" erstellt. Dieser Filter ruft das externe Skript "mediaFilter.php" auf. In diesem Fall handelt es sich um ein Skript, das Binär-Videodateien als Eingabe akzeptiert und HTML-Dateien (eingebettete Metadaten) und Miniaturansichten ausgibt.

Da dies nur bei bestimmten Inhaltstypen durchgeführt werden soll, wird die Anweisung **AddOutputFilterByType** verwendet, um verschiedene Multimedia-Formate anzugeben.

Sie können zu demselben Zweck aber auch die HTTP-Header bearbeiten. Ein einfaches Beispiel hierfür wäre es, den User-Agent-String der GSA durch einen anderen zu ersetzen:

```
# 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;)"
```

Dadurch werden nicht die von der GSA stammenden Header verändert, da diese nicht vom Proxy weitergeleitet werden. Sondern es werden nur die Header festgelegt, die Apache beim Abrufen einer Seite verwendet.

Diese Anweisung ist hilfreich, wenn Sie ein bestimmtes Cookie, einen User-Agent oder einen anderen Header festlegen müssen, um Ihre Inhalte zu crawlen.

Nach Konfiguration des Proxys und der Filter können Sie beides testen. Senden Sie dazu Ihre eigenen GET-Suchanfragen oder führen Sie folgende Schritte aus:

1. Konfigurieren Sie Ihren Browser für die Verwendung mit einem Proxy.
2. Fordern Sie einige URLs an.
3. Sehen Sie sich die Quelle an.

Wenn alles ordnungsgemäß aussieht, fügen Sie einfach die entsprechenden Proxy-Muster der GSA hinzu und starten Sie das Crawling. Die Dokumente in Ihrem Cache müssten die gefilterte Ausgabe anzeigen.

Weitere Materialien

- [Seite Apache-Module "mod_proxy"](#)
- [Apache-Caching-Handbuch](#)
- [Seite Apache-Modul "mod_cache"](#)

Kapitel 2: Admin-Toolkit der GSA verwenden

Überblick

Das Admin-Toolkit der GSA ist eine Bibliothek von Open-Source-Tools für GSA-Administratoren. Sie können die einzelnen Tools im [GSA-Admin-Toolkit](#) herunterladen.

In der folgenden Tabelle sind die verfügbaren Tools mit kurzen Funktionsbeschreibungen aufgeführt.

Tool	Beschreibung
monitor.sh	Überwachungsskript, das das Serving auf der GSA prüft
load.py	Führt Lasttests mit der GSA durch
authn.py	Webserver zum Testen der Authentifizierungs-SPI
authz.py	Webserver zum Testen der Autorisierungs-SPI
Connect.java	Java-Klasse zum Testen der JDBC-Verbindung zur Datenbank
cached_copy_checker.py	Überwachungsskript, das prüft, ob Crawling, Indexierung und das Serving auf der GSA funktionieren
sso.py	Webserver zum Testen von Cookie-Websites. Dieser kann so konfiguriert werden, dass er Oblix imitiert.
searchstats.py	Analysiert die Such-Protokolldatei (Fehlerraten, Suchanfragen pro Sekunde, durchschnittliche Antwortzeit)
smbcrawler.py	Skript, das imitiert, wie SMB von der GSA gecrawlt wird. Nützliches Tool zur Behebung von SMB-Crawlingproblemen, bei denen die Fehlermeldung der GSA keine brauchbaren Informationen liefert.
reverse_proxy.py	Reverseproxy, der verwendet werden kann, um Suchanfragen an die GSA zur Warteschlange hinzuzufügen, um die Anzahl der gleichzeitigen Verbindungen zu begrenzen. Wurde als Machbarkeitsstudie geschrieben und noch nicht in einer Produktionsumgebung getestet.
gsa_admin.py	Python-Skript zur Automatisierung von Aufgaben der Admin-Konsole. Dieses Skript wird verwendet, wenn die GData API der Admin-Konsole hierfür nicht verwendbar ist, z. B. in Softwareversionen vor Version 6.0 oder wenn eine Funktion in der API fehlt.
remove-or-recrawl-urls.html	Hilfeseite zu HTML+JavaScript, mit der Feeds entwickelt werden können, um URLs zu entfernen oder erneut zu crawlen
urlstats.py	Python-Skript, das Berichte über URLs in der GSA erstellt
ssoproxy.py	Konfigurierbares Python-Skript, das SSO-System-Anmelderegeln an den Proxy übermittelt, um der GSA SSO-Cookies für das Crawling und Serving zur Verfügung zu stellen.

CrawlReport.java	Java-Klasse, um über die neue Admin API (Softwareversion 6.0.0) die Anzahl der URLs abzurufen, die seit dem Vortag gecrawlt wurden
connectormanager.py	Einfacher Connector-Manager und Beispiel-Connectors mit Dokumentation zum Schreiben eines neuen Connectors
Kerberos-Validierungs-Tool	HTML-Anwendung, um die Einrichtung für Kerberos/IWA (Keytab/AD usw.) zu bestätigen
interactive-feed-client.html	HTML-/JavaScript-Seite, die XML-Feeds von Eingaben erstellt, z. B. URL und Hostname, und diese dann an eine GSA übermittelt
search_report_xhtml.xsl	XSLT-Stylesheet, um exportierte Suchbericht-XMLs in das von Menschen lesbare XHTML-Format umzuwandeln
search_results_analyzer.py	Tool zur Analyse von Suchergebnissen und zum Vergleich der Ergebnisse zweier GSAs
convert_cached_copy_to_feed.py	Tool zum Konvertieren von im Cache gespeicherten Versionen in Inhaltsfeeds zu Migrationszwecken
fetch_secure.py	Tool, das sichere Suchergebnisse von der GSA abrufen, indem es allen Weiterleitungen für die universelle Anmeldung folgt
GSA-GA.zip	Integrationsressourcen zu Google Analytics
SHT	Selbsthilfe-Tool

In den folgenden Abschnitten werden vier der am häufigsten benötigten Tools aus der obigen Tabelle beschrieben:

- [Suchprotokolle mit "searchstats.py" analysieren](#)
- [Aufgaben der Admin-Konsole automatisieren](#)
- [Dokumente im Index löschen oder neu crawlen](#)
- [Kerberos-Konfiguration bestätigen](#)

Suchprotokolle mit "searchstats.py" analysieren

"Searchstats.py" ist ein Analysetool für Suchprotokolle. Im folgenden Screenshot ist die Analyse einer Suchprotokolldatei mit dem Namen "2011-03-14-web_log.log" in einstündigen Intervallen dargestellt. Das Suchprotokoll wurde in der GSA-Admin-Konsole unter **Berichte > Suchprotokolle** (vor Version 7.2: **Status und Berichte > Suchprotokolle**) heruntergeladen.

```

-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	46	0	0.00	46	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

```

Beachten Sie, dass "searchstats.py" in der Programmierskriptsprache Python geschrieben ist. Um dieses Skript auszuführen, benötigen Sie die Python-Laufzeitumgebung. Sie können diese Laufzeitumgebung auf der [Downloadseite von Python](#) herunterladen.

Wenn Sie die Laufzeitumgebung installiert haben, können Sie das Skript mit einer Syntax ausführen, die der im oben angezeigten Beispiel ähnelt.

Aufgaben der Admin-Konsole automatisieren

Die [Administrative API der GSA](#) ist die bevorzugte Schnittstelle, um Aufgaben in der Admin-Konsole zu automatisieren. Sie deckt jedoch nicht alle Aufgaben ab, die in der Admin-Konsole verfügbar sind. Beispielsweise wird die Datenbanksynchronisierung nicht unterstützt. Außerdem ist es nicht möglich, Datenbanken mit der Admin API ferngesteuert zu synchronisieren. Für diese nicht unterstützten Anwendungsfälle verwenden Sie "gsa_admin.py" – ein Skript, das sich in der Admin-Konsole anmelden und programmatisch auf die erforderlichen Menüpunkte klicken kann.

Die folgenden Screenshots illustrieren:

- Syntax von "gsa_admin.py"
- Beispielbefehl zur Synchronisation von zwei Datenbanken

```

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

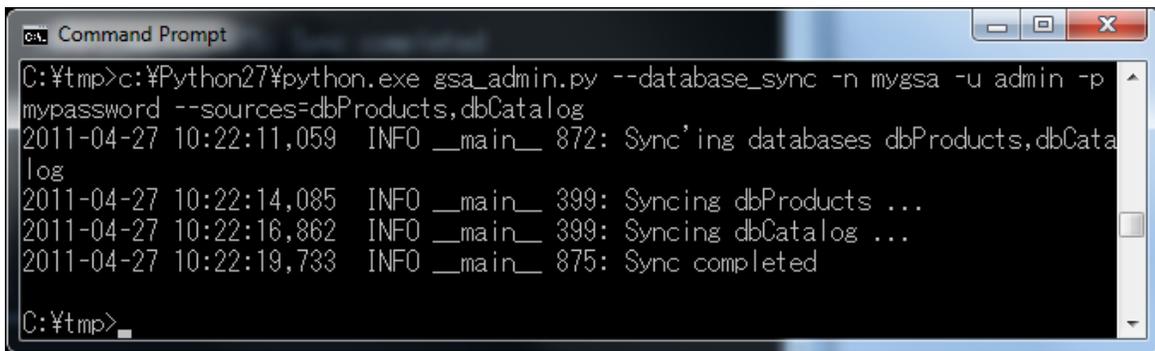
Options:
  -h, --help            show this help message and exit
  -f FILE, --input-file=FILE
                        Input XML file
  -o FILE, --output=FILE
                        Output file name
  -g SIGNPASSWORD, --sign-password=SIGNPASSWORD
                        Sign password for signing/import/export
  --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

```

Befehlssyntax



```
Command Prompt
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,dbCata
log
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>
```

Beispielsyntax zur Synchronisation von zwei Datenbanken ("dbProducts" und "dbCatalog")

Dokumente im Index löschen oder neu crawlen

Wenn Sie bestimmte Dokumente im Index entfernen oder neu crawlen müssen, ist das Tool "remove-or-recrawl-urls.html" sehr gut für diese Zwecke geeignet. Es ist eine in JavaScript-eingebettete HTML-Seite, die eine Liste von URLs als Eingabe akzeptiert und Feeddateien für die GSA erstellt und übermittelt.

Anhand der Informationen in den URL-Listen entfernt die GSA die angegebenen URLs oder crawlt sie erneut. Das Skript kann mit jedem Browser geöffnet und ausgeführt werden, der JavaScript unterstützt.

Zurzeit werden die drei folgenden Eingabeparameter unterstützt:

- GSA-Hostname oder IP-Adresse
- Liste von URLs, die entfernt oder neu gecrawlt werden sollen
- Option zum Entfernen oder erneuten Crawlen

Beachten Sie jedoch, dass Sie die IP-Adresse des Browsers auf der weißen Liste für IPs registrieren müssen, bevor Sie einen Feed an die GSA übermitteln. Sie können die weiße Liste auf der Seite **Inhaltsquellen > Feeds** (vor Version 7.2: **Crawl und Index > Feeds**) in der Admin-Konsole ansehen und bearbeiten.

Der folgende Screenshot zeigt die Seite "remove-or-recrawl-urls.html" dargestellt.

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:

Eine Kerberos-Konfiguration bestätigen

Zur Konfiguration der GSA für die stille Kerberos-Authentifizierung sind zahlreiche heterogene Komponenten erforderlich. Dazu zählen unter anderem Komponenten wie Active Directory, Keytab-Dateien, DNS, Internet Explorer-Sicherheitszone, Datenverschlüsselungsmethoden usw. Wenn bei der Konfiguration der Kerberos-Authentifizierung Probleme auftreten, können Sie das Kerberos-Validierungstool verwenden, um schnell die einzelnen konfigurierbaren Komponenten zu überprüfen.

Der folgende Screenshot zeigt das Kerberos-Validierungstool und eine Liste von zu überprüfenden konfigurierbaren Komponenten.

Validate GSA Kerberos Configurator

- 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
```

Weitere Informationen finden Sie in der folgenden Online-Dokumentation:

- [Validierungs-Dienstprogramm für die Kerberos-Einrichtung verwenden](#). Achtung: Das Programm hat sehr strenge Systemanforderungen: 32-Bit-Version von Windows XP, Vista, Windows 7.
- [Fehlerbehebung bei der Kerberos-Einrichtung und der sicheren Suche](#)

Kapitel 3: Fragen und Antworten

Überblick

In diesem Abschnitt werden einige allgemeine Fragen zur GSA behandelt, die mehrere Themen abdecken.

Ergebnisgewichtung

Frage: Warum erscheint auf meiner Suchergebnisseite kein Dokument, obwohl ich stark nach Inhaltstyp gewichte?

Antwort: Einer der Gründe, warum nicht das erwartete Ergebnis angezeigt wird, kann damit zusammenhängen, wie die GSA relevante Ergebnisse für Ihren Suchbegriff abrufen. Wenn Sie beispielsweise nach "Schmerz" suchen, führt die GSA folgende Schritte aus:

1. Die GSA findet zuerst die 1.000 (nach PageRank) ranghöchsten Dokumente im Index, die den Begriff enthalten. Hinweis: Dieser Schritt hat nichts zu tun mit der Begriffshäufigkeit oder einer Gewichtungsrichtlinie, die eventuell für das Frontend konfiguriert wurde.
2. Diese 1.000 Ergebnisse durchlaufen dann verschiedene Algorithmen, die Prozesse ausführen, z. B. einer Sortierung anhand der Häufigkeit des Begriffs in Dokumenten, anhand der Gewichtungsrelevanz und einer Reihe anderer Faktoren.

Wenn die von Ihnen gesuchte URL einen niedrigeren PageRank als die ersten 1.000 Dokumente in Schritt 1 hat, wird sie nicht von weiteren Algorithmen bearbeitet, die Gewichtungsrichtlinien berücksichtigen.

Auch wenn eine URL zu den ersten 1.000 Dokumenten gehört, die in Schritt 1 beschrieben sind, bedeutet sogar die Verwendung einer "ausgeprägten" Gewichtungsrichtlinie nicht zwangsläufig, dass die URL am Anfang der Suchergebnisse erscheint. Gewichtung ist nur einer der Faktoren, die in Schritt 2 des Prozesses betrachtet werden. Zahlreiche weitere Faktoren beeinflussen das Endergebnis.

Metadatensortierung

Frage: Wie kommt es, dass ein bestimmtes Dokument, von dem ich weiß, dass es existiert und gültige Metadaten besitzt, bei der Metadatensortierung nicht berücksichtigt wird?

Antwort: Die Metadatensortierung wird nur bei den 1.000 (nach PageRank) ranghöchsten Dokumenten angewendet. Weitere Informationen finden Sie unter [Ergebnisgewichtung](#).

GSA überwachen

Frage: Wie kann ich den Status der GSA und anderer eingesetzter Architekturkomponenten überwachen? Ist es unverhältnismäßiger Aufwand, wenn eine Skript-Suchanfrage am Index der GSA ausgeführt wird?

Antwort: Es gibt mehrere Möglichkeiten, den Heartbeat zu überwachen. Viele unserer Kunden überprüfen einfach, ob Port 80 reagiert. Andere Kunden machen auch noch den nächsten Schritt und überwachen Serving, Crawling und Indexierung. Informationen zu Überwachungsstrategien finden Sie unter *Überwachung einrichten* im Abschnitt [Konzipieren einer Suchlösung](#).

Über welchen Zeitraum Sie die Lösung überwachen möchten, hängt davon ab, wie komplex und gründlich die einzusetzende Überwachungsstrategie ist. Sie können zwischen verschiedenen Optionen mit jeweils eigenen Vor- und Nachteilen wählen. Im Folgenden einige dieser Optionen:

- SNMP verwenden – siehe [SNMP-Objekte](#) auf der Hilfeseite zu **Administration > SNMP-Konfiguration**
- Ein [Überwachungsskript](#) ausführen
- [Cached-Copy-Checker](#) ausführen
- Eine Suchanfrage nach einer bekannten URL im Index ausführen und sicherstellen, dass das Ergebnis in XSLT ausgegeben wird
- Eine Suchanfrage ausführen und überprüfen, ob Status 200 ausgegeben wird
- Das Frontend aufrufen, ohne eine Suchanfrage auszuführen, um sicherzugehen, dass Status 200 ausgegeben wird

Eine Suchanfrage nach einem Heartbeat auszuführen, ist kein unverhältnismäßiger Aufwand. Wenn Sie diese Lösung aktuell umsetzen und sie funktioniert, ist sie wahrscheinlich zum Erkennen eines Problems und Bereitstellen des Failovers genau ausreichend. Sie können diese Lösung möglicherweise optimieren, indem Sie ein Frontend erstellen, das nur über die minimalen Funktionen verfügt, d. h. keine KeyMatches, keine Suchanfragenvorschläge usw. Sie können auch eine Abfrage nach einer bekannten URL mit dem Syntaxelement "info:" in der Suchanfrage durchführen.

Weitere verfügbare Tools sind z. B.:

- [monitor.sh](#) – Überwachungsskript, das das Serving auf der GSA prüft
- [cached_copy_checker.py](#) – Überwachungsskript zur Überprüfung von Crawling, Indexierung und Serving auf der GSA
- [searchstats.py](#) – Analyse der Such-Protokolldatei (Fehlerraten, Suchanfragen pro Sekunde, durchschnittliche Antwortzeit)
- [urlstats.py](#) – Python-Skript, das Berichte über URLs in der GSA erstellt
- [search_report_xhtml.xsl](#) – XSLT-Stylesheet zur Umwandlung von exportierten XML-Suchberichten in das von Menschen lesbare XHTML-Format

Einen Connector können Sie mit dem Verbindungstester-Servlet unter "http://[cm-address-and-port]/connector-manager/testConnectivity" überwachen.

Mit dem Abruf der oben stehenden URL sollten Sie eine XML-Antwort mit einem HTTP-200-Status erhalten. Eine <StatusId>0</StatusId> in der XML-Datei zeigt an, dass alles ordnungsgemäß funktioniert.

Das Format der eingehenden XML-Datei sieht folgendermaßen aus:

```
<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>
```

Denken Sie daran, den Zugriff über das System zu erlauben, von dem aus Sie die Überwachung ausführen.

Es empfiehlt sich auch, den Java-Prozess zu überwachen, der die Connector-Tomcat-Instanz ausführt, auf seine CPU- und Speichernutzung zu überwachen.

Vorschläge für Suchanfragen

Frage: Die Funktion für die Vorschläge für Suchanfragen gibt unangemessene oder explizite Begriffe zurück. Wie kann ich die Rückgabe dieser Begriffe in Suchanfragenvorschlägen einschränken?

Antwort: Grundsätzlich kann es immer vorkommen, dass Nutzer auf der GSA Suchanfragen nach Begriffen ausführen, die in Ihrer Organisation als unangemessen gelten. Es gibt mehrere Möglichkeiten zu verhindern, dass diese Begriffe von der Funktion für Suchanfragenvorschläge angezeigt werden.

Die wirksamste Methode besteht darin, mithilfe der GDATA API eine schwarze Liste von Begriffen hochzuladen. Die Funktion unterstützt reguläre Ausdrücke. Sie können damit also einen Ausdruck definieren, der einen exakten oder teilweisen Treffer ausgibt. Hinweis: Wenn auf Ihrer GSA eine Version vor Version 6.14 ausgeführt wird, können Sie über die Benutzeroberfläche der Admin-Konsole keine Einträge zur schwarzen Liste hinzufügen oder ändern. Dies müssen Sie über die APIs durchführen. Weitere Informationen hierzu finden Sie unter dem Thema "Schwarze Liste für Suchanfragenvorschläge" in der entsprechenden GSA-API-Dokumentation:

- [Administrative API Developer's Guide: Protocol \("Administrative API – Entwicklerhandbuch: Protokoll"\)](#)
- [Administrative API Developer's Guide: Java \("Administrative API – Entwicklerhandbuch: Java"\)](#)
- [Administrative API Developer's Guide: .NET \("Administrative API – Entwicklerhandbuch: .NET"\)](#)

Es gibt ein Open-Source-Java-Dienstprogramm zum Hochladen von schwarzen Listen mit der GDATA API:

<http://code.google.com/p/gsa-admin-reports/>

Weitere Details zur Verwendung finden Sie unter diesem Link:

<http://code.google.com/p/gsa-admin-reports/wiki/BlacklistEditor>

Alternativ können Sie, um Begriffe für Suchanfragenvorschläge auf die schwarze Liste zu setzen, den Frontend-Code bearbeiten, der Vorschläge anzeigt. Sie können z. B. folgendes Codesnippet verwenden:

```
<!--
*****
Rechtschreibvorschläge auf Ergebnisseite (nicht anpassen)
*****
-->
<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'>&amp;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'>&amp;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] /@qe } &amp; spell=1 &amp; { $base_url } { $apps_param }">
        <xsl:value-of disable-output-escaping="yes" select="/GSP/Spelling/Suggestion[1]"/>
      </a>
    </p>
  </xsl:if>
</xsl:template>
```

Sie müssen die `if`-Bedingung aktualisieren, damit diese eine schwarze Liste von Begriffen enthält und die Vorschläge nicht im Frontend angezeigt werden.

Beispiel für einen exakten Treffer:

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

Beispiel für eine Übereinstimmung, die den Suchbegriff enthält:

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

Frage: Wie kann ich die Inhalte der Vorschläge für Suchanfragen komplett zurücksetzen?

Antwort: Ab Version 7.2 können Sie über die Seite **Suche > Suchfunktionen > Vorschläge** alle Vorschläge für Suchanfragen löschen. Klicken Sie dann auf **Zurücksetzen** für die Option **Vorschläge zurücksetzen** . Vor Version 7.2 müssen Sie ein Ticket für den Google-Support öffnen.

Übereinstimmungen mit unvollständigen Suchanfragen

Frage: Ich verwende die GSA für die Suche nach Personen. Meine Nutzer beschwerten sich, dass die Suche nach unvollständigen Namen keine Ergebnisse liefert. Was kann ich tun, um das Suchergebnis bei der Personensuche zu verbessern, wenn Übereinstimmungen mit Suchanfragen unter Verwendung unvollständiger Vor- und Nachnamen gefunden werden sollen?

Antwort: Die GSA unterstützt bislang keinen Abgleich mit Platzhaltern für unvollständige Begriffe. Nutzer sind es gewohnt, Kollegen zu finden, auch wenn Sie nur einen Teil des Vor- oder Zunamens eingeben. In dieser Hinsicht besteht bei der GSA noch Verbesserungsbedarf.

Die erste Empfehlung lautet, den Personendatensätzen in der GSA Metadaten hinzuzufügen. Diese Metadaten enthalten Tags für jede Kombination der ersten sechs Buchstaben des Vor- und Nachnamens einer Person. Im folgenden Beispiel mit "Jennifer Johnson" würden die folgenden 12 Meta-Tags zum Datensatz dieser Person als zusätzliche Metadaten hinzugefügt:

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

Je nachdem, wie Personeninhalte in Ihrer Umgebung abgerufen werden, kann über ein Skript entweder definiert werden, dass diese auf einer ausgegebenen, von der GSA gecrawlten HTML-Seite angezeigt oder dass sie über einen Contentfeed in der GSA bereitgestellt werden.

Eine weitere Möglichkeit ist es, eine benutzerdefinierte Liste von Synonymen von Namenstypen in die Suchanfragenerweiterungs-Funktion der GSA aufzunehmen, um bestimmte Nicknames zu vollständigen Namen zu erweitern.

An bestimmten Arbeitsplätzen ist es üblich, für die Mitglieder in einer Organisation Nicknames zu verwenden. Wenn diese Nicknames nicht direkt mit den Personendatensätzen in Form von Metadaten oder Inhalten verknüpft sind, kann die Suchanfragenerweiterung bei der Verwendung von Nicknames Vorteile bieten. Besonderes Augenmerk kann z. B. auf Verknüpfungen ausländischer und lokaler Nicknames gelegt werden. Beispielsweise kann der Name "Mateusz" so eingerichtet werden, dass er auch zu "Mat", "Matt" oder "Matthew" erweitert wird.

Eine weitere Möglichkeit ist es, die Funktion für die Suchanfragenvorschläge zu verwenden, um die Nutzerzufriedenheit bei der Personensuche zu steigern. Das Lab "Suche während der Eingabe" kann so implementiert werden, dass eine Datenbank mit benutzerdefinierten Namensvorschlägen Nutzern hilft, Suchanfragen nach Kollegen zu formulieren.

Suchanfragenfilter und OneBoxes

Frage: Die OneBox zur Personensuche verschwindet, wenn ich meiner Suchanfrage einen Filter wie "Zeitraum" hinzufüge. Was kann ich tun, damit die OneBox zur Personensuche erhalten bleibt, nachdem ich meiner Suchanfrage einen Filter hinzugefügt habe?

Antwort: Die Ergebnisse der OneBox-Personensuche werden bei Übereinstimmungen mit Suchbegriffen im Index ausgegeben. Wenn Sie versuchen, organische Suchergebnisse mithilfe eines Filters wie "Zeitraum" zu filtern, möchten Sie möglicherweise nicht, dass dieser Filter auf die Ergebnisse der Personensuche angewendet wird. Da die Ergebnisse der Personensuche auch aus Treffern im Index einer Personensammlung stammen, entspricht dieses Ergebnis der normalen Funktionsweise der GSA.

Beispiel: Wenn die der Sammlung zugeordneten Dokumente, die in den Personen-Suchergebnissen ausgegeben werden, keine Metadaten für das Datum enthalten, werden keine Suchergebnisse für Personen ausgegeben, wenn ein Datumsfilter in der Suchanfrage enthalten war.

Um hier die Sucherfahrung zu verbessern, können Sie sicherstellen, dass die für das Filtern erforderlichen Metadaten auch in den Datensätzen enthalten sind, auf denen die OneBox-Suchergebnisse beruhen.

Im Fall von "Zeitraum" verhält es sich wie folgt: Wenn ein Datum für Personendatensätze nicht ermittelt werden kann, sollten Sie einen statischen Datumswert verwenden. Tragen Sie ihn in der Suchanfrage im Backend mit einer "OR-Bedingung" ein.

Beispiel: Fügen Sie ein Datum "1901-01-01" als verändertes Datum für alle Personendatensätze hinzu. Wenn Sie dann einen Filter auf das Frontend anwenden, stellen Sie sicher, dass ein "OR Zeitraum:...1900-01-02" an die Suchanfrage angehängt wird, um den statischen Teil des Datums zu erfassen, der den Datensätzen hinzugefügt wurde.